Dallas N. Little
David H. Allen
Amit Bhasin

Modeling and Design of Flexible Pavements and Materials

Springer

Modeling and Design of Flexible Pavements and Materials

Dallas N. Little · David H. Allen
Amit Bhasin

Modeling and Design of Flexible Pavements and Materials

Springer

Dallas N. Little
Texas A&M University
College Station, TX
USA

David H. Allen
Texas A&M University
College Station, TX
USA

Amit Bhasin
The University of Texas at Austin
Austin, TX
USA

ISBN 978-3-319-58441-6 ISBN 978-3-319-58443-0 (eBook)
https://doi.org/10.1007/978-3-319-58443-0

Library of Congress Control Number: 2017939620

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

There is no denying the profound changes that our planet has seen since the advent of the twentieth century. Chief among these are the dramatic developments in all things derived from electromagnetism—television, computers, cell phones, just to name a few. But where materials and mechanics are concerned, the developments are less obvious. Though aircraft, automobiles, buildings, and bridges have also undergone development over the past half-century and more, these changes are less apparent to the consumer. And so it is with roadways—we seem to be traveling on the same types of surfaces as our grandparents did. Some might even say we are still copying the roadway designs developed by the Romans. Unfortunately, roadway traffic worldwide has skyrocketed within the last century, thereby leading to rapid decay of these transportation facilities. Indeed, within the USA alone it has become an annual struggle by local, state, and federal governments to allocate sufficient resources to fund roadway construction and maintenance.

There are examples within the field of transportation of cost containment and even cost decreases per passenger mile traveled, including the costs of both automobiles and air transport. However, so long as the antiquated design procedures of the past continue to be utilized to design, build, and maintain roadways, the cost of these facilities will likely continue to increase. Given that the cost of other transportation facilities is becoming more cost-effective, it would seem that such cost containments should be attainable with roadways as well.

The authors have undertaken this text with the intent of demonstrating that while developments in roadways are not as apparent as they may be in other fields, there has nevertheless been substantial headway made in roadway design since the days of the construction of the US Interstate Highway System. Indeed, one may even say that these innovations are revolutionary. Hopefully, this text will serve our purpose—to demonstrate these new developments for the practicing engineer in a clear, concise, and useful way. If so, then we would hope that the enormous cost to societies worldwide of roadway design, construction, and maintenance should decrease substantially as a result of the technologies elucidated within this text. And if that is indeed the result of our labor, we will have accomplished our intent.

Chapters 4 and 5 of this text were written by the first author. In addition, Chaps. 1 and 9–15 were written by the second author. Finally, Chaps. 2, 3, and 6–8 were written by the third author. And while we have each written on disparate topics, we have taken every effort to ensure that the subject of roadway design is treated herein in a contiguous manner.

And so, we who have spent our lives studying the subject within these pages, wish you a happy and enlightening read.

College Station, USA — Dallas N. Little
College Station, USA — David H. Allen
Austin, USA — Amit Bhasin

Acknowledgements

We wish to thank Michael Luby, our Springer Editor, for pursuing this challenging undertaking with us. We also wish to thank Brian Halm of Springer who worked tirelessly to help bring the text to publication. In addition, we want to thank the reviewers of the manuscript who provided valuable advice as to content and format. Finally, we wish to thank our respective family members who patiently supported us during the long process it took to bring this text to completion.

The authors wish to gratefully acknowledge the following contributors to this book: Dr. Arno Hefer and Dr. Didier Lesueur for their willingness to share from their graduate and postdoctoral work, respectively, at Texas A&M University and from subsequent publications, partially with the first author, to the text in Chaps. 4 and 5; Ph.D. candidates at Texas A&M University Pavan Akula, Narain Hariharan, Javier Grajales-Saavedra, and Atish Nadkarni for their help in preparing key figures and in review of the text.

Contents

About the Authors

Dallas N. Little is the E. B. Snead Chair Professor and Regents Professor in the Zachry Department of Civil Engineering at Texas A&M University, where he has been a faculty member since completing his Ph.D. in 1979. Dr. Little holds an M.S. from the University of Illinois at Urbana-Champaign (1973) and a B.S.C.E. from the United States Air Force Academy in 1970. He has been a registered professional engineer since 1976.

David H. Allen is currently Director of the Center for Railway Research within the Texas A&M Transportation Institute. Prior to that, he was a faculty member at Virginia Tech (1980–81), Texas A&M University (1981–2002), The University of Nebraska-Lincoln (2002–10), and The University of Texas-Pan American (2010–13). He obtained his B.S., M. Eng., and Ph.D. degrees from Texas A&M University.

Amit Bhasin is a faculty member in the Department of Civil, Architectural, and Environmental Engineering at The University of Texas at Austin. He has been in this position since 2008. He received his B. Tech. in Civil Engineering from IIT Varanasi, India, and his M.S. and Ph.D. from Texas A&M University (2003, 2006).

D.H. Allen's Other Publications

Introduction to the Mechanics of Deformable Solids—Springer (2013)
How Mechanics Shaped the Modern World—Springer (2014)
Introduction to Aerospace Structural Analysis (with W.E. Haisler)—Wiley (1985).

Chapter 1
Introduction

1.1 Historical Introduction

Evidence suggests that humans were traveling along identifiable paths and trails, many previously carved out by animals, long before the first roads were built. For example, at Langebaan, South Africa, human footprints found within a trackway have been dated at approximately 117,000 years old. But of course, these cannot be accurately characterized as roadways.

In fact, at the time of this writing, it is not clear when or where the first road was built on Earth. This is at least in part due to the fact that newly discovered ancient roadways are not uncommon. Combine that with oftentimes imperfect means of determining the ages of these roadways, as well as the problem of defining exactly what one means by "building" a roadway, and the controversy emerges quite naturally. Perhaps the best candidate for the longest ancient road is the Silk Road, which ran approximately 5,000 miles, connecting Central China and Europe. But was this really a road, or was it a series of tacks that were at interval partially paved with rubble, gravel, even refuse, or perhaps logs?

Among the prominent candidates are the mysterious Misrah Ghar il-Kbir (also called Clapham Junction) Cart Tracks located on the island of Malta. Although most archeologists date the site to about 2000 BCE, some believe that they might in fact be much older. Still, these mysterious tracks should not accurately be termed "roadways," as they appear to have been gouged naturally from the soft limestone, rather than via any observable construction project.

Another equally enticing possibility is termed a timber trackway, a number of which dot the landscape across Northern Europe. These include the corduroy trackways (ca. 3000 BCE), the Lindholme Trackway (ca. 2900 BCE), and, most recently, the Belmarsh Trackway (ca. 3,900 BCE) (Keel 2009). But perhaps the most impressive candidate for "world's oldest roadway" to date is the 8-mile-long roadway discovered in 1994 within the Faiyum Depression 45 miles southwest of

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_1

Cairo, which was apparently used to transport basalt carvings from a quarry to the Nile River. Geologists have dated this roadway to approximately 4,600 BCE (Maugh 1994).

What is known is that roads were being built several thousand years ago. In ancient times, the Greeks are known to have constructed the Via Pythia as a route to Delphi, where stood the Omphalos stone, considered at that time the center of the universe. Indeed, it is now well known that the Egyptians, Carthaginians, Greeks, Turks, and Etruscans all built roads. But for shear enormity of roadway networks in antiquity, surely the Romans must be regarded as the forerunners of modern road construction. Livy mentions in his *Ab Urbe Condita Libri (Books from the Foundation of the City)* (Livy 1994) that the Romans were building roads as far back as 500 BCE.

As shown in Fig. 1.1, by the time of Julius Caesar, the Romans had constructed a vast "interstate" highway system.

The following was written in the *Itinerary of Antoninus* (Antoninus 2010):

> With the exception of some outlying portions, such as Britain north of the Wall, Dacia, and certain provinces east of the Euphrates, the whole Empire was penetrated by these *itinera* (plural of *iter*). There is hardly a district to which we might expect a Roman official to be sent, on service either civil or military, where we do not find roads. They reach the Wall in Britain; run along the Rhine, the Danube, and the Euphrates; and cover, as with a network, the interior provinces of the Empire.

The Roman roads were marvels of mechanics, as depicted in Fig. 1.2. They were composed of several layers, the top layer being of large paving stones. Where there was a supply of iron nearby, small iron filings were spread over the surface, whereupon they rusted and bonded together, forming a fine patina on the surface. The roads were normally built either by military engineers or by slaves. In either case, the paving materials were located as near to the construction site as possible.

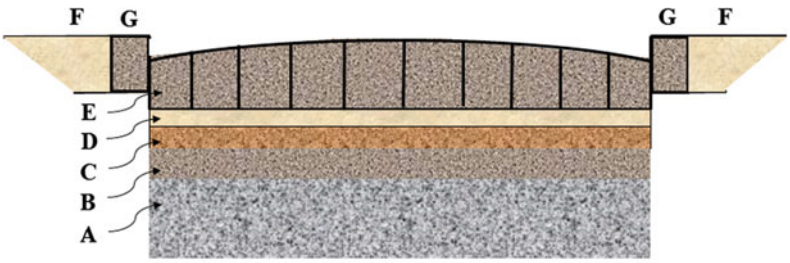
In all, the Romans built more than 80,000 km of roadways, and it is a testament to the Romans that some of these roads still exist today. Perhaps the most famous is the Via Appia (begun in 312 BCE), which stretches eastward from Rome, connecting it to Brindisi (the port of embarkation to Greece in ancient times), on the Southeast coast of Italy (Fig. 1.3). It has recently been discovered that the part of the Appian Way that lies directly east of Rome is built on top of an old lava flow field, thus both providing further reinforcement for the roadway and underscoring the ingenuity of Rome's ancient engineers.

Although Portland cement concrete was not used for roadways until modern times, during Hadrian's reign (during the first century A.D.), the Pantheon was rebuilt (it had been destroyed by fire), as shown in Fig. 1.4. And in what is perhaps the most amazing feat of all of the mechanics employed by the Romans, the roof of the Pantheon was rebuilt using concrete (apparently by an engineer named Neri). This was the first structure of its type on Earth. In the center of the dome is the oculus, as shown in Fig. 1.5, and this is today the last completely intact Roman structure on Earth, at least in part due to the fact that an ancient form of hydraulic cement was utilized to build the dome.

Green: Via Aurelia -> Via Aumelia Scaura -> Via Julia Augusta
Bright green: Via Clodia
Yellow: Via Cassia
Blue: Via Flaminia
Grey: Via Salaria
Dark purple -> purple: Via Tiburtina -> Via Valeria
Carrot red: Via Latina
Bright red: Via Appia
Dark blue: Via Ostiensis

Additional major roads specified on the map:
Via Popilia, going to southern Italy (Rhegium) from Via Latina
and Via Appia crossroads. (cyan)
Via Traiana, branching from Via Appia and also going to Brundisium. (dark red)
Via Aemilia, starting from the end of Via Flaminia at Ariminum and going to Placentia. (orange)
Via Postumia, going from Genua (Genoa) on the northwest of Italy to Aquileia on the northeast. (dark green)

Fig. 1.1 Depiction of the Roman Roads in Italy by Agamemnus



(A). Native earth, levelled and, if necessary, rammed tight.
(B). Statumen: stones of a size to fill the hand.
(C). Audits: rubble or concrete of broken stones and lime.
(D). Nucleus: kernel or bedding of fine cement made of pounded potshards and lime.
(E). Dorsum or agger viae: the elliptical surface or crown of the road (media stratae eminentia) made of polygonal blocks of silex (basaltic lava) or rectangular blocks of saxum qitadratum (travertine, peperino, or other stone of the country). The upper surface was designed to cast off rain or water like the shell of a tortoise. The lower surfaces of the separate stones, here shown as flat, were sometimes cut to a point or edge in order to grasp the nucleus, or next layer, more firmly.
(F). Crepido, margo or semita: raised footway, or sidewalk, on each side of the via.
(G). Umbones or edge-stones.

Fig. 1.2 Depiction of a Roman Road

Fig. 1.3 Via Appia today (note the crowning of the roadway)

No one knows how the Romans accomplished this amazing feat, but one thing that is known is that concrete was not used again in modern times until the eighteenth century (Englishman John Smeaton made the first modern concrete in 1756 by mixing aggregate with cement. The first reinforced concrete was developed by

Fig. 1.4 The Pantheon (note the Egyptian obelisk in front)

Fig. 1.5 The Oculus in the Dome of the Pantheon (note the inlaid pattern in the concrete ceiling)

Frenchman Joseph Monier, who was granted a patent in 1867). The intricate inlaid pattern on the interior surface of the dome suggests that a massive wooden false-work structure must have been constructed to span from opposite edges to the center of the dome, much like an arch is constructed. The workers would then have raised the cement in powder form, mixed the concrete at elevation, and then poured the wet concrete for two opposing lobes of the dome at a time (there are a total of

thirty lobes within the dome). The falsework would then have been partially disassembled and shifted to the next set of lobes, and in so doing slowly completing the circle of the dome.

Visitors to the Pantheon can also see perhaps the most massive set of twenty granite columns from antiquity (as well as a granite Egyptian obelisk in the square in front). These columns were quarried in Egypt and shipped across the Mediterranean to their present location in the portico of the Pantheon. And surely the fact that these columns still stand today is a testament to those who believe that granite cannot be quarried for use as pavement aggregate.

The Romans utilized the arch in order to provide larger spans for stone structures, and this invention allowed the Romans to create many of the most famous structures still standing today from that time period. Figure 1.6 shows a row of arches within the hypogeum of the Colosseum of Rome, built beneath the stadium

Fig. 1.6 Ancient Roman Arches in the Hypogeum of the Colosseum

shortly after the colosseum was completed in 80 AD. A telling example of an arch is the portico from the Canopus at Hadrian's villa, as shown in Fig. 1.7. The portico has both flat (beams) and curved (arches) stone members between the columns, and the discerning reader will recognize that the span between the arches is slightly larger than that between the beams, attesting to the fact that arches can span larger dimensions than beams made of stone because they carry loads strictly in compression, whereas beams necessarily undergo tensile loading on the bottom edge due to their own weight, a circumstance that precludes the use of stone beams for large spans. This is an example of the mechanics of deformable bodies, suggesting that the Romans determined rules for limiting the length of stone beams via careful experimentation.

Despite their proven ability to construct both massive and impressive structures, ancient engineers did not possess or follow rigorous design methodologies. Theirs was an experimental and therefore necessarily expensive discipline. For example, it is known that the Pont du Gard (Fig. 1.8), built in the first century AD, was constructed at a cost that would have bankrupted a small nation today. This absolutely massive aqueduct that spans the Gardon River in Southern France is approximately 49 m in height and 275 m in length. It is part of a 50-km-long aqueduct system that was built to transport water to the city of Nemausus (modern-day Nimes).

Amazingly, the total elevation change from the spring to the city is only 17 m! This is indeed a tribute to the Romans' ingenuity with mechanics. Roman engineers possessed a number of leveling tools that worked on the same principle as the modern water bubble level. It is noteworthy that this massive aqueduct in the South

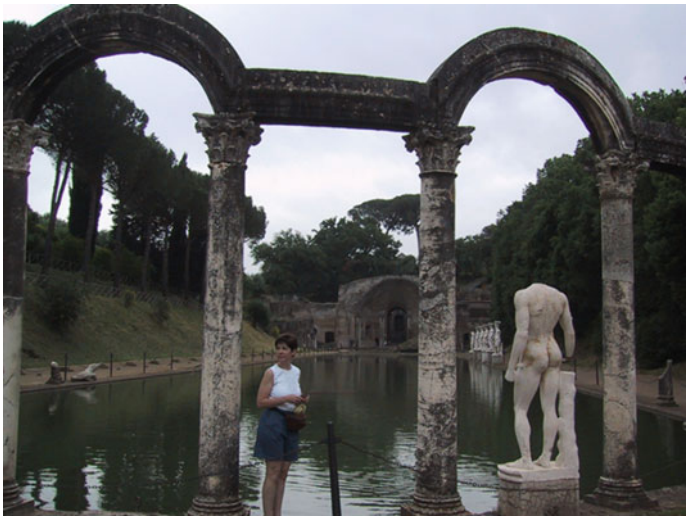


Fig. 1.7 Beams and Arches connecting columns adjacent to the Canopus at Hadrian's Villa

Fig. 1.8 Photograph of the Pont du Gard

of France still stands today, so that the cost may not sound so astronomical if amortized over two millennia.

After the fall of the Western Roman Empire in 476 AD, the art of building roads seems to have reached a stagnation point, one that gripped the world for nearly a thousand years. The construction of Christian cathedrals, commencing with the Gothic period in the mid-twelfth century, seems to have fueled a rebirth in road construction, and this appears to have been as a direct consequence of cathedral construction. In fact, quite a few incomprehensibly massive Gothic cathedrals collapsed shortly after (or even during) construction. In many cases, it was painfully evident that structural failure was due at least in part to erosion of the surface around the cathedral walls. As a means of correcting this shortcoming, it became standard practice to build stone coverings surrounding the cathedrals; these stones were "cobbled" from the leftover remnants of the cathedral construction. Thus was born the cobblestone roadway and, given the obvious superiority of the surface coated with cobblestones, cobblestoning entire townships became commonplace across Europe.

Nonetheless, paved roads between townships was nearly nonexistent (with the obvious exception of the old Roman roads) until the dawn of the industrial revolution in the early nineteenth century. This unfortunate circumstance can perhaps be better understood when one considers that prior to the introduction of the steam engine, essentially all work on Earth was performed by human or animal power. Even then, it was not actually roads that were constructed. Instead, mass transport shifted from almost entirely waterborne to transport by rail, as the steam engine was adapted for use on rails, thereby inventing the railroad. The first steam locomotive was built in 1804, and shortly thereafter, the Puffing Billy became the first working locomotive, as shown in Fig. 1.9. By the middle of the nineteenth century, rail transport had connected large portions of Western Europe, and it was soon to race across the USA.

Fig. 1.9 Photograph of the Puffing Billy—an early steam locomotive

As a by-product of the development of railroads, it was necessary to design and build a new generation of bridges on hitherto unimaginable scales. Perhaps the first of these was the Menai Suspension Bridge, designed by Thomas Telford (1757–1834), and completed in 1826 (Fig. 1.10). This graceful suspension bridge connects the island of Anglesey to the mainland of Wales.

Who could have known what such a structure would portend, but within a century the world had changed remarkably. The Brooklyn bridge, opened on May 24, 1883, was perhaps the first truly modern massive structure built on Earth (Fig. 1.11).

But for a truly amazing stretch of the imagination, consider a structure that is not even a bridge—the Eiffel Tower, as shown in Fig. 1.12. The tower, conceived by Maurice Koechlin and Émile Nouguier, was built by French bridge engineer Gustav Eiffel (1832–1923), for whom it is named. When construction was completed in March 1889, it was the tallest structure on Earth (320 m), thereby creating what is arguably the most recognizable structure on Earth.

Fig. 1.10 Photograph of the Menai Suspension Bridge

Fig. 1.11 Photograph of the Brooklyn Bridge

Fig. 1.12 Photographs of the Eiffel during construction

Unfortunately, rail transport suffered from a significant shortcoming—it was not readily accessible for the purpose of door-to-door movement. Instead, one rode the train to the township of choice and, subsequently disembarking, was transported by horse-drawn carriage to their destination. What was needed was a sort of "personal" steam engine, one that was not limited to following a preconstructed track. Unfortunately, the steam engine proved unable to provide a solution to this problem, as steam engines could not be designed that were both small and safe for personal use.

Enter the reciprocating engine, invented in the third quarter of the nineteenth century. This type of engine was not only more compact than the steam engine, and it was far safer. By the 1870s, prototypical predecessors of the modern-day

automobile were springing up everywhere. Henry Ford (1863–1947) was instrumental in developing mass production techniques that made the automobile affordable by the masses, as shown in Fig. 1.13.

Within a few short years, the age of powered personal travel had commenced in earnest. Unfortunately, the heretofore cobbled streets of yore proved to be much too dangerous for lengthy travel by automobile. No matter what was attempted (such as soft tires and compliant springs), riding in an automobile over cobblestones proved to be far too uncomfortable, not to mention dangerous, for the pilots of these new-fangled vehicles. This shortcoming was magnified as for economic reasons the automobile was pressed to move passengers at ever-increasing speeds.

For the first time in nearly a thousand years, there arose a need for a new type of roadway. Certainly Napoleon's troops would have preferred a surfaced roadway to the mud-filled pathways they endured (or in most cases died within) during the long retreat from Russia in the winter of 1812–1813. And rest assured, the Civil War would have been dramatically altered had there been paved roadways across the USA in the 1860s.

As early as 1717, John Metcalf was building roadways from stone in England (NAPA 2015), but these roadways quickly succumbed to the effects of weather. In the years 1803–1821, the same Thomas Telford mentioned above paved nine hundred miles of Scottish roads with broken stones, creating the first of what we would nowadays call "gravel" roads (NAPA 2015). Much later, it was recognized in England that hot tar could be used to bond the broken stones together, and thus was born the term "tarmac."

Nonetheless, there remained few if any paved roads anywhere outside the cobblestoned cities of the world, right up to the outset of the twentieth century. But the need for safe and efficient roadways for transporting automobiles was all that was needed to ensure that this circumstance would change dramatically. As mentioned above, hydraulic cement concrete was invented in the late eighteenth century and reinforced concrete in the nineteenth. It soon became apparent that this type of

Fig. 1.13 Photograph of Mr. and Mrs. Henry Ford in his first automobile

concrete was not only a cost-effective alternative to laying stone, and if poured and smoothed properly, it offered a safe and comfortable roadway surface. Unfortunately, it soon became apparent that hydraulic cement concrete cracked unmercifully (even under its own weight) unless poured in significant thickness, thereby leading to monumentally large amounts of concrete required to build a roadway. This was compounded by the fact that concrete is not easily mixed on site. Accordingly, there sprung up an entirely new industry—the concrete mixing and transport business—in the first quarter of the twentieth century. Before long, hydraulic cement concrete roads were being poured in a host of urban areas across Europe and the USA. The first street in the USA to be paved with hydraulic cement concrete was Court Avenue in Bellefontaine, Ohio, in 1891 (Lee and Lee 2004). The first mile of hydraulic cement concrete pavement in the USA was on Woodward Avenue in Detroit, Michigan, in 1909 (Kulsea et al. 1980).

But this "concrete rush," like the gold rushes of the nineteenth century, proved to be somewhat short-lived, as countries, states, counties, provinces, and townships realized that pouring hydraulic cement concrete wherever a roadway was needed proved to be fiscally impossible.

Enter another fortuitous development. It did not take a genius to realize that what was needed was a very cheap mastic, far cheaper than concrete, and the only thing as cheap as concrete must necessarily also come out of the ground. By the late 1860s, bituminous asphalt was being used for sidewalks and roads in the USA (NAPA 2015).

Laura Ingalls Wilder, author of the beloved *Little House on the Prairie*, tells of her first encounter with an asphalt pavement. She was on a wagon journey with her parents in 1894 that took them through Topeka.

> In the very midst of the city, the ground was covered by some dark stuff that silenced all the wheels and muffled the sound of hoofs. It was like tar, but Papa was sure it was not tar, and it was something like rubber, but it could not be rubber because rubber cost too much. We saw ladies all in silks and carrying ruffled parasols, walking with their escorts across the street. Their heels dented the street, and while we watched, these dents slowly filled up and smoothed themselves out. It was as if that stuff were alive. It was like magic (NAPA 2015).

We now know that asphalt roads are much older than anyone could have imagined. The inscription on the fired clay tablet shown in Fig. 1.14 records the receipt of bitumen circa 2250 BCE (NAPA 2015). In fact, the first recorded use of asphalt in a roadway occurred in Babylon around 625 BCE, in the reign of Nabopolassar (Gillespie 1992). And although there is no evidence that the Ancient Greeks used asphalt in roadways, it is known that they used them to seal baths, reservoirs, and aqueducts (NAPA 2015).

The first asphalt pavement laid in the USA was constructed by Belgian Edmund J. DeSmedt in 1870 in front of City Hall in Newark, New Jersey (NAPA 2015). DeSmedt later paved Pennsylvania Avenue in Washington, DC, the first large-scale asphalt paving construction in the USA. But these roadways were paved with expensive asphalt that came from Lakes Trinidad and Bermudez in Venezuela.

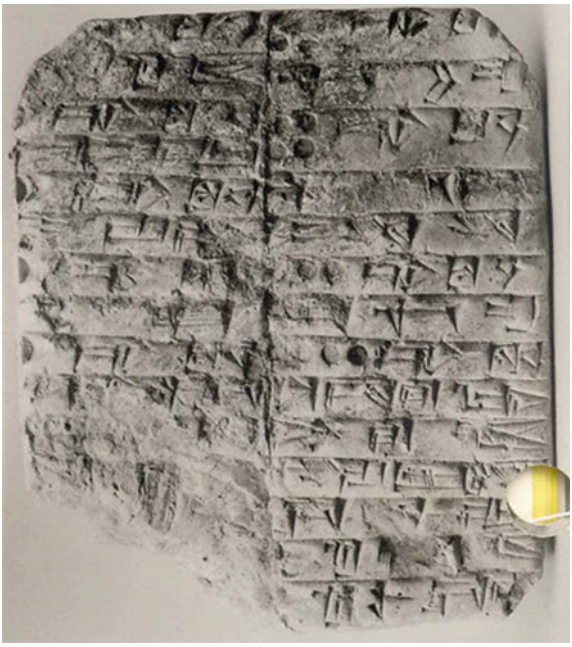


Fig. 1.14 Babylonian clay tablet describing the use of asphalt in an ancient roadway

As it so happened, the fuel needs of the automobile produced asphalt as a sort of serendipitous by-product. The internal combustion engine required a cheap fuel, and it quickly developed that the most accessible fuel was gasoline, distilled from oil that was easily mined nearly everywhere that mining engineers went in search of it. But to their dismay, much of the oil that came out of the ground was plagued with this nasty semisolid material called bitumen, one whose viscosity was far too high, and energy content far too low, for use as an energy source by the automobile. So here lay this nasty by-product wherever there was an oilfield—and what to do with it? Of course, those few who had already tried asphalt as a binder in roadways eventually realized that this new refined source of asphalt might be the solution to what was a mind-boggling estimate of the required amount of asphalt to pave roadways. And somewhat surprisingly, refined asphalt turned out to be just as useful as a binder as was naturally occurring asphalt. And by 1907, refined asphalt had become the binder of choice in asphalt concrete roadways (NAPA 2015).

Today, asphalt concrete (a mixture of bitumen or asphalt binder with mineral aggregates) covers 94% of the paved roads in the USA (NAPA 2015). It is estimated that approximately 1.6 billion metric tons of asphalt concrete produced worldwide annually (2007) and 85% of bitumen or asphalt binder produced is used within roadways (NAPA-EAPA 2011).

1.2 Technical Introduction

The process of designing pavement structures is complex. First, because voluminous amounts of materials are required to construct the roadway and pavement structure, it is imperative that the materials deployed are as inexpensive as possible. Because the availability of such materials varies dramatically from point to point, it is often not possible to utilize the same materials or even at different locations. This inconsistency of availability alone impacts materials and construction cost.

Second, the cost of roadway materials is highly sensitive to the binder used to improve properties and consistency. Often these additives or binders are required to bring the basic materials within specification tolerances. Whether the binder is cementitious, pozzolanic, or bituminous, it is relatively much more expensive than the fill or subbase material, base aggregate or the aggregate used in base and surface asphalt layers. In fact the cost ratio of binder to aggregate typically varies within a range of 3:1–50:1. The cost differential requires careful binder selection, evaluation of compatibility between aggregate and binder, and efficacious mixture design in order to optimize performance. Furthermore, in order to introduce material properties into the design, analysis, and performance prediction process, appropriate material characterization is mandatory. This includes assessment of the impact of external variables such as temperature, moisture ingress, and oxidative aging of the asphalt binder. This book focuses on the impact of material variations on performance of the principal structural layer of the pavement, the asphalt layer, at multiple scales.

Third, roadways can fail to perform their intended purpose for a wide range of reasons, including (Distress Guide 2015):

- Fatigue cracking, also called alligator cracking (Fig. 1.15), normally caused by excessive loading of the wheel path and/or weakness of sublayers

Fig. 1.15 Photograph of fatigue cracking. (*Source* Road Science, a Division of ArrMaz)

- Surface aging (Fig. 1.16)
- Bleeding (Fig. 1.17), normally caused by migration of liquid asphalt to the surface
- Longitudinal cracking (Fig. 1.18), normally caused by excessive wheel path loading, weak sublayers or environmental conditions and volumetrically unstable subgrade soils
- Pushing, also called shoving (Fig. 1.19), oftentimes accompanied by surface delamination
- Reflective cracking (Fig. 1.20), caused by previously occurring cracks in an underlying layer due to volume change, differential thermal movements, etc.

Fig. 1.16 Photograph of surface aging. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.17 Photograph of bleeding. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.18 Longitudinal cracking. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.19 Photograph of pushing accompanied by delamination. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.20 Photograph of reflective cracking. (*Source* Road Science, a Division of ArrMaz)

- Thermal cracking (Fig. 1.21), normally caused by excessive temperature variations
- Raveling (Fig. 1.22), whereby aggregate spalls from the pavement, typically near the edge and often due to mixture design inadequacies, oftentimes resulting in depressions that pool moisture
- Stripping (Fig. 1.23), probably caused by replacement of moisture for asphalt bond at the aggregate surface due to surface chemistry effects, deleterious fines content or a combination of several factors
- Potholing (Fig. 1.24), caused by a variety of factors including moisture damage, high dynamic loads, pore water pressure, freeze and freeze–thaw damage, etc.
- Rutting (Fig. 1.25), typically due to excessive plastic deformation in the upper hot mix asphalt layers related to mixture design and/or pavement materials selection inadequacies, high tire pressures, and general inadequacies of design.

Fig. 1.21 Photograph of thermal cracking. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.22 Photograph of raveling. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.23 Photograph of stripping. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.24 Photograph of potholing. (*Source* Road Science, a Division of ArrMaz)

Rutting can be near surface and asphalt concrete related or deep layer rutting related to weakness in sublayers

- Surface wear (Fig. 1.26), oftentimes resulting in surface polishing

Each of these failure mechanisms is difficult in its own right to predict. Accordingly, developing a model that is capable of predicting all of these failure modes is daunting to say the least. Nonetheless, the development of a robust cost model demands that the life of the roadway be known a priori, something that is rarely actually possible.

Fig. 1.25 Photograph of rutting accompanied by alligator cracking. (*Source* Road Science, a Division of ArrMaz)

Fig. 1.26 Photograph of surface wear resulting in aggregate polishing. (*Source* Road Science, a Division of ArrMaz)

Fourth, the typical roadway may perform admirably over significant portions of its length, while failing rapidly over short stretches such that the entire roadway must be replaced. This issue is related to the fact that loads, environmental effects, and material properties vary spatially within roadways over length scales that are oftentimes sufficiently small that there is little possibility of accounting for these variations within any cogent predictive methodology.

Fifth, when roadways fail, and eventually all roadways will, there is the issue of what is to be done to restore the roadway to its intended purpose. In the case

wherein an overlay is chosen for reasons of cost over the complete reconstruction of the roadway, much less is known as to what constitutes the optimum treatment.

Sixth, because climatic conditions vary widely across our planet, each newly designed pavement is unique in that it must meet specific environmental constraints, including rainfall, snow and ice, solar radiation, and temperature extremes.

Finally, both new roadway materials and application methods are being developed all the time, and oftentimes little is known about the long-term performance of these.

The result of all of the complicating factors described above is a design process that is oftentimes termed "open-ended," in the sense that there are so many design variables to choose from that the designer finds it difficult if not impossible to determine the optimum choice of design variables for the roadway/pavement in question. As a result, it has become customary to design pavements across our planet in an essentially forensic way, implying that future pavements are designed based on experimental observation of the performance of previous pavements. Unfortunately, just as in the case of other forensically based methodologies (take forensic medicine, for instance), there is oftentimes little scientific support for the resulting methodology.

Still worse, forensically based models do not normally address the underlying causes for failure, so that they cannot be utilized in any scientific way to predict improvements or remedial corrections in roadways.

What is needed is a scientifically based approach to pavement design, one in which the effects of changes in design constraints can be accurately predicted as a means of determining the most efficient and cost-effective roadway design. Based on the observations of failure mechanisms in asphalt materials and pavements listed above, it should be readily apparent to the reader that this is a complex problem!

In seeking out a methodology for solving such a complicated problem, perhaps it is well to recall Ockham's Razor (William of Ockham, c. 1287–1347), which may be paraphrased as follows: *the best solution is the simplest one that works for the problem at hand*. Because the roadway problem is such a complex one, Ockham's razor is especially significant, as a model that is overly complicated is unlikely to be practical for the purpose of roadway design. Accordingly, consider a set of the minimum physical phenomena that should be included in any cogent model of roadway performance:

- Permanent deformations, implying inelastic material models
- Cracking, implying models capable of predicting evolutionary fracture
- Temperature dependence, implying adherence to thermodynamic principles
- Effects of loading variations, implying adherence to principles of mechanics
- Aging, implying adherence to chemistry and conservation of mass

As a result, finding a balance between the actual complexity of this problem and the need for simplicity in constructive models is a daunting challenge. As an initial step toward this end, there have been a number of attempts to approach pavement design in a more scientifically rigorous way, thereby resulting in what is now termed the mechanistic-empirical approach. And while these efforts have produced marked advances in both our understanding and our ability to model pavement, we submit that there is nonetheless still room for improvement. Because roadways are both expensive and absolutely essential to our way of life on this planet, the potential rewards from advances in these models are enormous. We the authors have written this text with the intent of stimulating ideas for continued development of these more scientifically based models that are based in mechanics, chemistry, and thermodynamics.

The development of a physically and scientifically based predictive model for pavement materials and pavements would therefore appear to be a quite complex and daunting task, one that has not to our knowledge yet been solved. However, armed with a cadre of recent developments, we submit that the solution to this problem is now within the reach of the scientific community. While it would be presumptuous to say that this textbook accomplishes that challenge, it is nonetheless posited by the authors that the concepts contained herein will address all of the issues listed above, thereby pointing the way toward a new generation of roadway design models and procedures that are dramatically more accurate than those previously employed on our planet.

1.3 A Road map for Using This Text

The text is composed of two parts: materials science of asphalt concrete, normally the primary structural component of flexible pavements, and mechanics of pavement structures and paving materials with a major emphasis on the asphalt concrete structural layer. We have attempted to cover these significantly different topics in a single text for the simple reason that both are necessary for a more complete understanding of modern flexible pavements. Several textbooks and reference books are available that provide reviews, summaries, and/or commentaries on basic design approaches from the most empirical to the most advanced and most

mechanistic (Yoder and Witczak 1975; Roberts et al. 1996; The Aggregates Handbook 2013; Wang 2011; Huang 2004; Papagiannakis and Masad 2008; Kim 2009; MEPDG 2008). Such references are excellent for design and for some degree of analysis. This book is not intended to compete with texts or reference documents that focus on design or even analysis. Rather this book is intended to serve as a unique bridge between materials science and mechanics and toward advanced analysis of the pavement structure. This book may well serve as a complementary text in a design course for upper level undergraduate or graduate students. When used to support of complement other texts or resources in a pavement design course, it is intended to improve understanding of the performance of infrastructure paving materials as components of the structural pavement analysis. Toward that end, both students and practicing engineers will find Part I especially useful. Part II offers the fundamental mechanics required to apply the materials science described in Part I and to provide a glimpse into the future of pavement design, analysis, and performance modeling methodologies. As such, it is perhaps not at the stage where designers can utilize the computational tools described herein. However, it is anticipated that these will become ubiquitous in the not too distant future at the current rate of development of computational power. In the case wherein this text is to be used as an academic teaching tool, we have attempted to intimately relate Parts I and II.

In fact, we envision that Part I may be taught as an upper-level undergraduate or entry-level graduate course, whereas Part II would constitute a second advanced level graduate course requiring Part I as a prerequisite. For this reason, there are some concepts that are introduced in Part I and then revisited in more detail in Part II. This redundancy is by design and is intended to help the reader develop a better appreciation for the importance of mechanics (Part II) in the context of pavement materials (Part I) and vice versa. Accordingly, we have devised a road map for use of this textbook, as shown in Fig. 1.27. We, the authors, have taught collectively for more than seventy-five years at several major universities across the USA and abroad. During that span of time, we have taught the materials herein (as they have become available) within this two course sequence. It is our hope that this text will become an essential tool for both academics and roadway engineers. If we have hit both of these targets, then we have struck the mark intended.

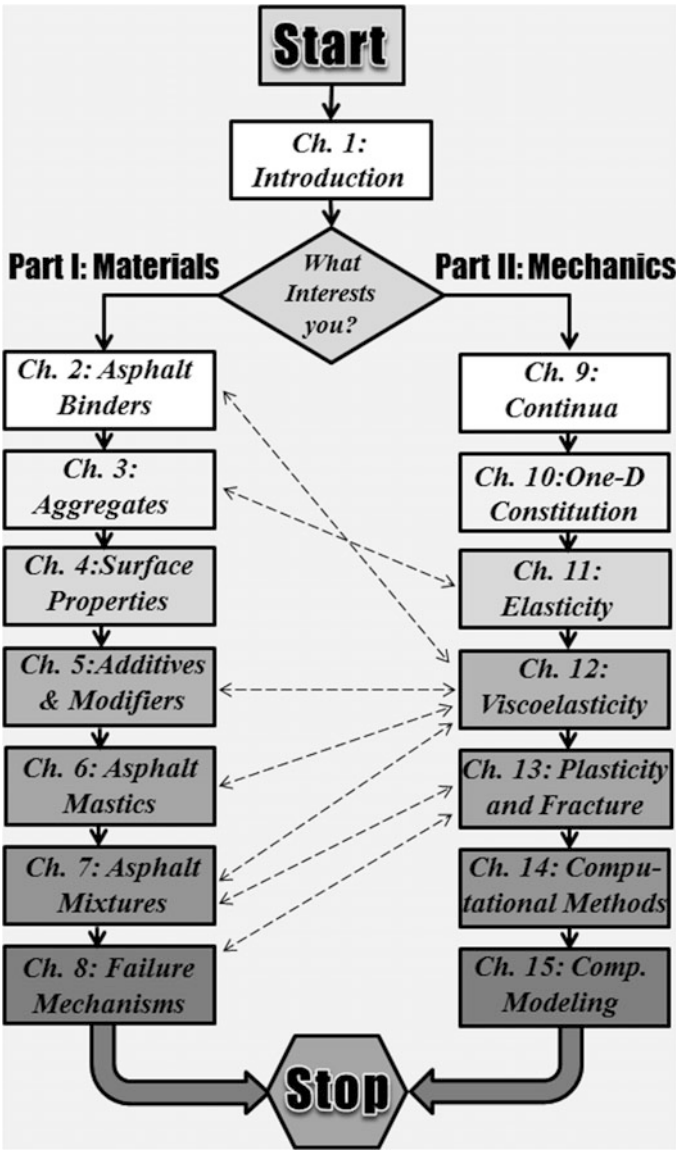


Fig. 1.27 A road map for using this textbook

References

Antoninus of Piacenza. (2010). *Itinerary of Antoninus*. Encyclopedia Britannica.
Distress Guide. (2015). Road Science. http://www.roadscience.net/services/distress-guide.
Gillespie, H. (1992). *A century of progress: The history of hot asphalt mix*. National Asphalt Pavement Association.
Huang, Y. (2004). *Pavement analysis and design* (2nd ed.). Prentice Hall.
Keel, D. (2009). BELMARSH: Timber trackway 500 years older than Stonehenge found by archaeologists. News.
Kim, R. (2009). *Modeling of asphalt concrete*. McGraw-Hill.
Kulsea, B., Shawver, T., & Kach, C. (1980). *Making Michigan move: A history of Michigan highways and the Michigan department of transportation*. Michigan Department of Transportation.
Lee, B., & Lee, H. (2004). Position-invariant neural network for digital pavement crack analysis. *Computer-Aided Civil and Infrastructure Engineering, 19,* 105.
Livy. (1994). *Ab Urbe Condita Libri*. Cambridge University Press.
Maugh, T. (1994). Archeologists find world's oldest paved road in Egypt. Los Angeles Times.
MEPDG. (2008). *Mechanistic-empirical pavement design guide*. American Association of State Highway and Transportation Officials.
NAPA. (2015). National Asphalt Paving Association. http://www.asphaltpavement.org/index.php?option=com_content&task=view&id=21&Itemid=41.
NAPA-EAPA. (2011). *The asphalt paving industry—A global perspective* (2nd ed.). National Asphalt Pavement Association (NAPA) and European Asphalt Pavement Association (EAPA) joint report.
Papagiannakis, A., & Masad, E. (2008). *Pavement design and materials*. Wiley.
Roberts, F., Kandhal, P., Brwon, E., Lee, D., & Kennedy, T. (1996). *Hot mix asphalt materials, mixture design, and construction* (2nd ed.). National Asphalt Pavement Association Research and Education Foundation.
The National Stone, Sand and Gravel Association. (2013). *The aggregates handbook* (2nd ed.). The National Stone, Sand and Gravel Association.
Wang, L. (2011). *Mechanics of asphalt microstructure and micromechanics*. McGraw-Hill.
Yoder, E., & Witczak, M. (1975). *Principles of pavement design* (2nd ed.). Wiley-Interscience.

Part I
Materials

Chapter 2
Asphalt Binders

2.1 Introduction

Asphalt mixtures are a composite comprising asphalt binder and mineral aggregates. Asphalt binder, also referred to as bitumen in Europe and other parts of the world, can be regarded as glue (hence the term binder) that holds the aggregate particles together in the mixture. Asphalt binders make up approximately 4–8% of the asphalt mixture by weight. From a volumetric standpoint, typical asphalt mixtures contain 9–18% of asphalt binder, 87–78% of mineral aggregates, and 3–6% of air voids. The size and distribution of air voids within the asphalt mixture is very important in order to understand mechanisms related to oxidation and moisture-induced damage and will be discussed in more detail in later chapters. Despite its low mass or volume fraction, from a cost standpoint as of this writing, asphalt binders make up approximately 40–50% of the cost of materials used to produce an asphalt mixture. Consequently, the cost of an asphalt mixture is very sensitive to the price of asphalt, which is in turn related to the price of crude oil in most cases.

The mechanical properties of an asphalt binder are time, temperature, and age dependent. Consequently, asphalt mixtures inherit these characteristics (time, temperature, and age dependency) from the asphalt binder. It is impossible to fully understand and appreciate the behavior and performance of asphalt mixtures without having some knowledge about the performance and behavior of asphalt binders. For example, consider the typical distresses in a flexible pavement: rutting (or plastic or permanent deformation), fatigue cracking, thermal cracking, and moisture-induced damage.

In terms of rutting, the resistance to deformation is provided by both the aggregate structure and the asphalt binder that holds the aggregate structure in place. The contribution of the binder to resist rutting is significant and cannot be ignored in understanding or predicting the rutting of asphalt mixtures.

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_2

One of the mechanisms by which asphalt mixtures fail is the action of moisture referred to as stripping (other mechanisms will be discussed in more detail in Chap. 8). Stripping occurs when water displaces the asphalt binder from the surface of the aggregate particle. The ability of the asphalt binder-aggregate interface to resist this form of moisture-induced damage is dictated by the physicochemical surface properties of both the binder and the aggregate.

The relationship between the properties of the binder and distresses such as fatigue and thermal cracking is even more direct. The ability of an asphalt mixture to resist tensile stresses can almost exclusively be attributed to the asphalt binder. Although the aggregate size and distribution (gradation) may play a role in defining distribution of internal stresses within the binder, without the asphalt binder the aggregate particles would remain unbound and unable to resist any tensile stresses. Similarly, one of the key characteristics of an asphalt pavement is its ability to resist cracking due to thermal shrinkage when the temperature drops. As the air and consequently pavement temperature drops, tensile stresses develop within the asphalt mixture. Due to the inherent viscoelastic nature of asphalt binders, it is possible for the binder to relax and partially relieve these tensile stresses over time such that the cumulative tensile stresses do not exceed the strength of the binder.

The aforementioned examples demonstrate the importance of understanding the behavior and performance of asphalt binders as a first step to better understand and predict the performance of asphalt mixtures and pavements. However, one may raise the question that since asphalt binders are used exclusively with aggregates, it would be more meaningful to evaluate and understand the behavior of asphalt mixtures as a whole as opposed to evaluating one of the ingredients within the asphalt mixture. There is a twofold answer to this question.

First, since the performance of an asphalt mixture is dictated by the properties of the asphalt binder, a poor quality binder will not result in a durable asphalt mixture. In other words, it is difficult to produce a durable mixture if one of the critical ingredients of the mixture is of poor quality. Selecting and using an asphalt binder that is appropriate for a given climatic and traffic condition is necessary (but not sufficient) to ensure that the asphalt mixture does not fail during its service life.

Second, significant improvements in the durability and life-cycle cost of asphalt mixtures and pavements can be made by engineering and producing binders with superior performing characteristics. In short, when designing asphalt mixtures, it is critical to screen and use asphalt binders with superior mechanical and engineering properties. As of this writing, the practice of designing asphalt mixtures and pavements requires that both the asphalt binder and mixture meet certain performance requirements before being approved for use in construction.

This chapter is organized into four sections. The first section briefly describes the origins and production of asphalt binders. The remaining three sections go over the chemical, mechanical, and surface properties of asphalt binders. Each of these sections addresses two important questions: (1) Why the property is important in terms of mixture or pavement performance; and (2) how the property or its attributes are measured and used in engineering practice or research? A significant amount of information presented in this chapter is based on the research conducted

during and following the Strategic Highway Research Program (SHRP pronounced sharp) during the late 1980s and early 1990s. Information presented in this chapter must be considered as a starting point for individuals new to the asphalt industry, graduate and undergraduate students, and engineers in the asphalt industry who need a refresher on the basics of asphalt binder properties. For those interested in a more in-depth study of one or more of the topics listed in this chapter, a list of references for additional reading has been provided at the end of this chapter. Finally, by design, this chapter does not focus on details of standard tests or methods to characterize asphalt binders. Instead, the goal of this chapter is to explain to the readers the rationale behind the development of such methods. Detailed descriptions of the methods can be found in the list presented in the additional reading material.

2.2 Production of Asphalt Binders

Most of the asphalt binder used in commercial paving applications is produced by refining crude oil. Asphalt binder is also naturally found impregnated in rocks and minerals referred to as rock asphalt, mixed with sand and other fine minerals referred to as oil sands, or in bulk as in the case of Lake Trinidad. In the case of natural asphalts, the "refining" is mostly done by nature over a much longer period of time. Although most of this chapter is based on asphalt binder produced from the distillation of crude oil, concepts related to chemical makeup and mechanical properties can also be extended to natural asphalts.

Crude oil is formed by the decomposition of organic materials at high temperatures and pressures over millions of years. The nature of the organic material and the conditions under which the organic material decomposes dictates the chemical makeup of crude oil. In other words, the chemistry of crude oil will vary significantly depending on its source. Crude oils from different sources are generally described as heavy or light depending on the percentage of asphalt binder or heavy residue that is left behind after the distillation of lighter fractions such as gasoline and kerosene. Crude oils may also be described as sweet or sour depending on the percentage of sulfur (sweet refers to lower percentage of sulfur). Figure 2.1 illustrates typical compositions of crude oil from two very different geographical sources.

Crude oil is distilled to produce familiar products such as gasoline, kerosene, diesel, and lubricating oil. These products are also referred to as high-value refining products. The residue left behind after the distillation of the high-value products is typically referred to as short residue or in most cases straight-run asphalt binder. The chemical and mechanical properties of a straight-run binder can vary significantly depending on the source and batch of the crude oil as well as the processes used during the distillation of the crude oil. Note that the properties of the high-value products (e.g., gasoline) produced during distillation are relatively well

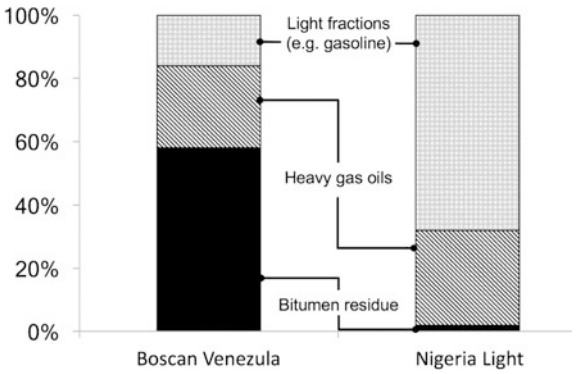


Fig. 2.1 A comparison of two different crude oil sources; the Boscan crude is regarded as sour with approximately 6% sulfur compared to the Nigerian light that has 0.2% sulfur

controlled. Consequently, the differences in the chemical makeup of different crude oil sources are amplified in the short residue or straight-run binder that is left behind after distillation.

In many cases, the properties of the straight-run asphalt binder are not suitable for paving application. This may prompt the binder producers to further process the straight-run binder to produce an end product that meets the requirements of end users for pavement applications. There are three main variables that are manipulated to produce an asphalt binder with specific properties: (i) proportioning of different straight-run binders or feedstocks, (ii) the selection of the techniques and chemical catalysts to process the straight-run asphalt binder (e.g., air blowing or half air blowing), and (iii) selection of an additive or combination of additives to modify the binder. Refineries that produce asphalt binders optimize the above variables to produce an asphalt binder that meets the requirements of the consumer with the lowest production cost. The desired properties of the asphalt binders vary with geographic region and type of application and will be discussed later.

2.3 Chemical Properties

2.3.1 The Need to Understand Binder Chemistry

This section addresses the question of why it is important for engineers or researchers working with asphalt materials to be knowledgeable, at least to some extent, about the chemistry of asphalt binders. The importance of understanding the chemistry of asphalt binders can be exemplified during the three stages of the binder's life: When the binder is produced in a refinery, when it is put to use for construction of an asphalt pavement, and when it is reclaimed after the end of the service life of an asphalt pavement.

First, consider the production of an asphalt binder. It is no surprise that the engineering properties of a material can be significantly altered by subtle changes in

the chemical composition of the material. The simplest example of this is steel. Addition of a small percentage of carbon to iron and changes in the heat treatment during production can result in significant improvements in the engineering or mechanical properties of the resulting steel. By understanding the relationship between the elemental composition of steel to its crystal structure and concomitant engineering properties, it is possible to design grades of steel that meet specific requirements for engineering applications. Also, different grades of steel can be consistently produced in different parts of the world irrespective of the source of the iron ore.

While the relationship between the chemical makeup and engineering properties is just as important for asphalt binders as it is for steel, the ability to produce different grades of asphalt binders with consistent engineering properties poses a unique set of challenges. For example, the chemical makeup of the raw material used to produce asphalt binders (crude oil) is highly dependent on the source and batch. As mentioned before, these differences are amplified in the bottom residue after the distillation of other high-value products. In addition, the chemical structure of the organic molecules that constitute asphalt binders is extremely diverse and complex. It is imperative that the personnel working with the production of asphalt binders have a thorough knowledge of the different chemical species that make up crude oils and asphalt binders as well as their relationship to the binders' engineering properties.

This section will help the reader develop a basic understanding of the chemistry of asphalt binders, and references are provided at the end as additional reading material for those interested. This knowledge will be useful for practitioners and researchers to better appreciate the differences in asphalt binders procured from different refineries or sources, even when the binders have similar mechanical and physical properties. In other words, *just because two asphalt binders have been classified to have the same "grade" based on their mechanical properties, there is no reason to expect that the two binders would have the same chemistry or would behave similarly under conditions where the chemistry of the material is important.* This knowledge is also important to appreciate the long-term implications, advantages, and limitations associated with using asphalt binders that have been modified or extended using any one of the several different environment friendly products (e.g., used oil by products, bio-binders).

After being produced at a refinery, the second stage in the life of an asphalt binder commences when it is used to produce an asphalt mixture and subsequently spans its service life in a flexible pavement. A basic understanding of asphalt chemistry is again important for researchers and practitioners alike to better appreciate:

1. the mechanisms that allow the use of chemical additives to facilitate mixing and compaction at reduced temperatures, referred to as warm mix asphalt,
2. the mechanisms that allow certain liquid additives or chemically active fillers such as hydrated lime to act as anti-stripping agents,

3. oxidative aging that occurs in the asphalt binder during production and service life of the pavement, changes in the binder chemistry due to oxidative aging and its impact on the rheological properties of the binder and distresses such as low-temperature cracking, and
4. the production and use of emulsions for application in surface treatments such as chip seals and cold asphalt mixes.

Finally, at the third stage, i.e., at the end of life stage, asphalt pavements are frequently demolished and pulverized or milled to produce fractionated black rock or recycled asphalt pavement (RAP). Since the asphalt binder in the recycled material is highly oxidized, the mechanical properties of the recycled binder must be counterbalanced by adding a binder that is less oxidized (soft binder) or by using a rejuvenator. With the recent emphasis on using higher percentages of recycled asphalt in the construction of new asphalt pavements, the knowledge of asphalt chemistry is crucial to maximize the use of recycled materials with or without rejuvenators while avoiding any short- or long-term detrimental impact. Having discussed the question of why understanding the chemistry of asphalt binders is important, the next subsection discusses the attributes that are used to describe the chemical makeup of asphalt binders.

2.3.2 Attributes of Chemical Properties and Methods of Measurement

Asphalt binders comprise a very diverse and complex range of hydrocarbons. The molecular structure of these hydrocarbons varies with the source and batch of the crude oil used to produce the asphalt binder. Owing to this diversity and complexity, it is neither feasible nor useful to identify the exact molecular structure of any given asphalt binder. Instead, a more practical approach to understand the chemical makeup of asphalt binders is to classify the molecules in the asphalt binder based on different attributes. Such classifications can then be used to discover relationships between binder chemistry and engineering properties. Four different attributes are commonly used to classify the chemical nature of asphalt binders. These attributes are:

1. elemental composition of the asphalt binder,
2. size of the molecules,
3. ionic character of the molecules, and
4. polarity of the molecules.

This section will discuss chemistry of the asphalt binder based on the above attributes as well as its implications in terms of the properties of the binder in the context of asphalt mixtures and pavements.

2.3.2.1 Elemental Composition

A typical asphalt binder comprises 90–98% by weight of hydrocarbons. The hydrocarbons exist in the form of saturated branched or unbranched alkanes and cyclic alkanes as well as unsaturated aromatics. Just as a refresher, recall that saturated hydrocarbons are molecules that do not have a carbon atom with a double or triple bond. The saturated alkanes and cyclic alkanes are also referred to as paraffins and naphthenes, respectively, in the petrochemical industry. The hydrocarbons typically have a carbon to hydrogen ratio of 1.5. This is somewhere in between the carbon to hydrogen ratio of saturated alkanes (approaching 2.0) and that of an aromatic ring structure such as benzene (1.0).

In addition to carbon and hydrogen, asphalt binders also contain other heteroatoms such as oxygen, sulfur, and nitrogen and trace metals. Although the heteroatoms constitute a small percentage by weight of the asphalt binder, they have a significant influence on the overall properties of the binder. For example, the functional groups such as pyrroles and carboxylic acids have been shown to be responsible for promoting adhesion although with different levels of resistance to debonding (resulting in stripping) in the presence of water. Oxygen and sulfur impart polarity to the molecules and are responsible for generating strong intermolecular interactions, which in turn contribute to the stiffness and strength of the binder. The oxygen content in an asphalt binder does not remain unchanged during its service life. Oxidation of the asphalt binder during production and service life results in the production of functional groups such as ketones, sulfoxides, and carboxylic acid. The typical heteroatoms and functional groups found in an asphalt binder are perhaps best summarized by Petersen (1984) and some of which are illustrated in Fig. 2.2.

Elemental composition provides the basic building blocks for any given asphalt binder. However, due to the diversity and complexity of the molecular structures, it is difficult to relate the elemental composition of the binder to it's mechanical properties. One notable exception to this is the amount of oxygen present in an asphalt binder that can be used as an indicator of the extent of oxidation in the binder. Oxidation results in the formation of polar functional groups as shown in Fig. 2.2 and reduce the ductility of the asphalt binder as well as its ability to relax. The subject of oxidation and aging in general will be revisited in the last section at the end of this chapter.

In the context of evaluating asphalt binders based on their elemental composition, the amount of oxygen in the binder can be used to track the extent of aging that the binder has experienced during production or its service life. This information can also be used to determine the rate and extent of oxidation in any given asphalt binder. The extent of oxidation in an asphalt binder is strongly related to its mechanical and rheological properties; however, such relationships may be different for each asphalt binder and dictated by its overall chemical makeup. An understanding of the oxidation process and concomitant changes in the binder properties is very important in the context of asphalt mixtures, and this topic will be revisited several more times in this and subsequent chapters.

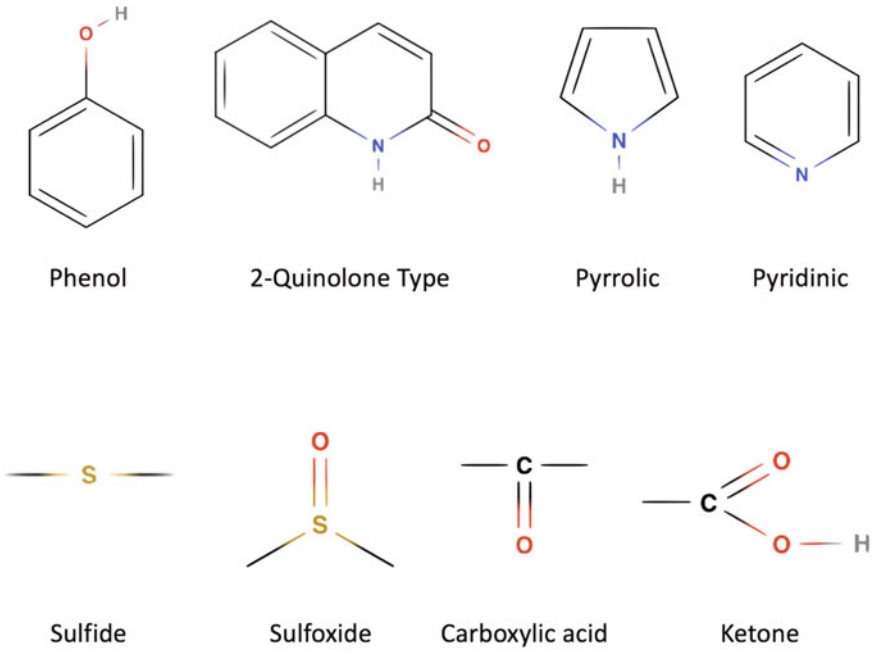


Fig. 2.2 Typical functional groups in asphalt binder (Recreated based on Petersen 1984)

On the subject of functional groups and oxidation, it is important to briefly mention one of the most commonly used techniques to quantify the concentration of different functional groups in asphalt binder, i.e., Fourier transform infrared spectroscopy or FTIR. When infrared light, typically comprising wave numbers that range from 4000 to 600 cm^{-1}, is incident on a sample of the material, molecular bonds corresponding to different functional groups absorb energy corresponding to specific wave numbers. An absorption spectrum is obtained by comparing the incident and reflected (or transmitted) infrared spectra. The absorption spectrum provides a fingerprint of the different functional groups that are present in the sample. The magnitude of the absorption peak at a given wave number is related to the concentration of the functional group that absorbs energy at that wave number in the sample. For example, the carbon-oxygen bond in the carbonyl (C = O) functional group stretches by absorbing energy corresponding to wave numbers from 1820 to 1670 cm^{-1}. Therefore, an absorption peak between these wave numbers indicates the possible presence of the carbonyl functional group; the size and area under the absorption peak is an indirect indication of the concentration of the carbonyl functional group.

The FTIR absorption spectra for a material can be obtained in either (i) transmission mode or (ii) reflectance mode. In the transmission mode, the infrared beam

passes through the sample to the detector. Liquid or gaseous samples in infrared transparent cells are typically used in the transmission mode. A very thin solid film or small amount of solid mixed with a powdered infrared transparent material (like KBr) to form a pellet can also be used in the transmission mode. If a solvent is used to carry the sample (e.g., binder in a solution form), then absorption by the pure solvent must be subtracted from the absorption spectra. In the reflectance mode, the infrared beam passes through one end of an infrared transparent crystal at a grazing angle.

The beam then reflects multiple times from the interface of the sample (solid or liquid) with the crystal surface. This type of test is referred to as attenuated total reflectance (ATR) spectroscopy (Fig. 2.3). Most FTIR instruments support both types of test modes. Although the FTIR is a powerful and easy to use technique, the results must be interpreted with caution because each functional group may have more than one characteristic absorption peak (e.g., corresponding to bond stretching

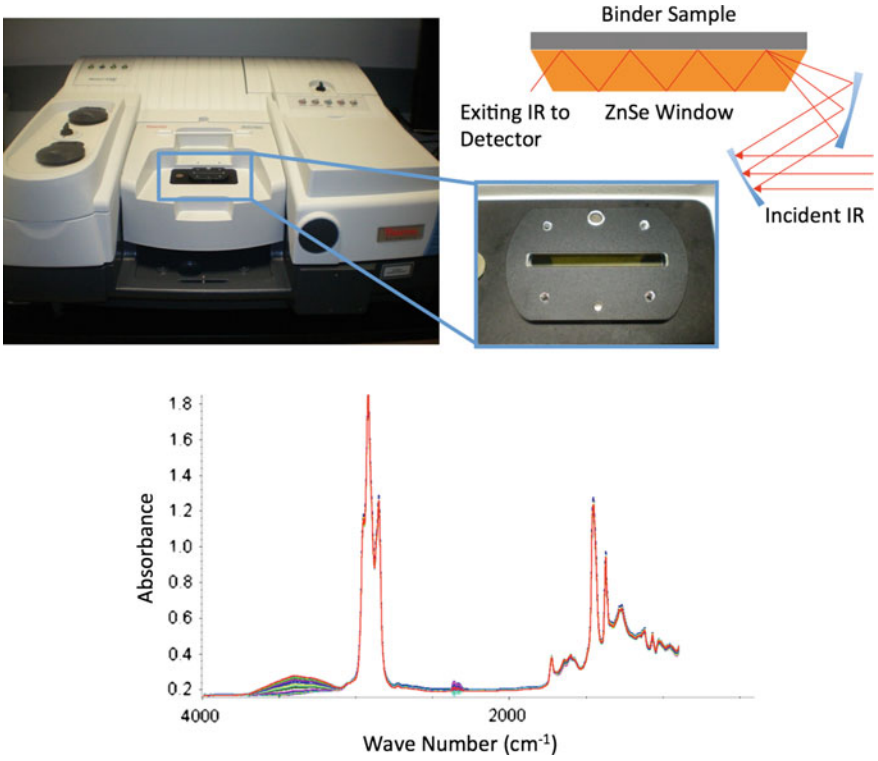


Fig. 2.3 Typical FTIR setup (*top* and *top right*) being used in the multi-bounce ATR mode; typical stack of spectra (*bottom*) for an asphalt binder exposed to water on the surface, the two highest peaks correspond to the methyl and ethyl functional groups that are predominant in asphalt binders, the several peaks with increasing intensity in the 3100 to 3700 cm^{-1} correspond to water due to diffusion in water over a period of time; peak around 1700 cm^{-1} is often used to characterize extent of oxidation (with permission from Vasconcelos 2010)

and rotating), and different functional groups may have peaks that overlap. Semiquantitative estimates must also be made carefully to account for sample-to-sample variability. As of this writing, FTIR is extensively used for research purposes (e.g., for oxidative aging kinetics). Although FTIR is not used a common tool by transportation agencies to specify or verify the chemical makeup of asphalt binders, there has been some work in this direction to promote the use of FTIR as a tool to detect and verify the polymer content in modified asphalt binders or for QA/QC to verify the use of liquid anti-stripping agent or hydrated lime in asphalt binders during mixture production.

In summary, by comparing the elemental composition (e.g., concentration of heteroatoms), it is possible to get some idea about the relative differences in functional groups between asphalt binders and expected resistance to shear, ability to bond with aggregate surfaces, and stripping in the presence of water. For any given binder, by tracking the amount of oxygen it is possible to estimate the extent to which a binder oxidizes as a function of time and temperature, also referred to as oxidation kinetics. Extent of oxidation is strongly related to the mechanical and rheological properties of the asphalt binder, although such relationships are unique for each asphalt binder and dictated by their overall chemical makeup.

2.3.2.2 Molecular Size Distribution

Attempts to relate the mechanical properties of the asphalt binder to its molecular size distribution were made during the 1960s and 1970s. The inspiration for such attempts was the observation that at the macroscopic level, the viscoelastic behavior of asphalt binders is similar to that of polymers. In the case of polymers, the mechanical properties can be modified by adjusting their molecular composition and size distribution. Therefore, it is reasonable to assume that, to some degree, the properties of the asphalt binder can also be related to molecular size distribution.

The molecular size distribution of asphalt binders is typically obtained using a method referred to as size exclusion chromatography (SEC) or gel permeation chromatography (GPC). This method entails dissolving the asphalt binder in a solvent and then pumping the solvent through a column packed with porous particles or beads. The beads are typically made of polystyrene cross-linked with divinyl benzene and are saturated in the solvent prior to use. Saturating the beads causes them to expand and form a pore structure with a broad size distribution. The asphalt binder solution is then pumped through the gel-packed column. Smaller molecules will tend to migrate into the smaller pores within the gel whereas the larger particles, owing to their size, would tend to travel in the larger spaces between the beads and within the gel. Consequently, the smaller molecules will take the longest time to travel to the outlet of the gel-packed column and vice versa. By tracking the concentration of hydrocarbons at the outlet of the GPC column over time, it is possible to obtain a distribution of molecular sizes in the asphalt binder. Figure 2.4 illustrates typical SEC results for two different asphalt binders.

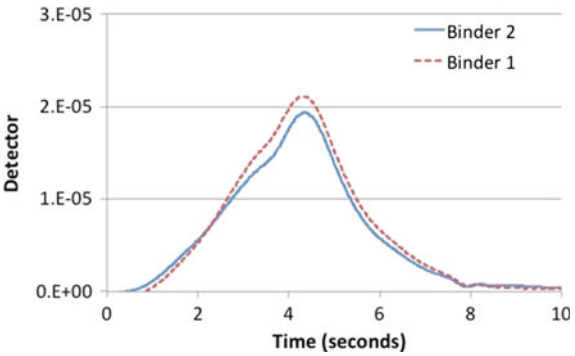


Fig. 2.4 Typical results for molecular size distribution for two binders using the SEC; B2 was oxidized version of Binder B1 and is considered to have slightly more polar and associated molecules

Careful consideration is required in comparing the results from SEC for different asphalt binders. First, results from the SEC are extremely sensitive to the particulars of the method used. Several factors influence the results from SEC. For example, results from SEC are sensitive to the choice of gel material, pore sizes in the gel material, concentration of the binder in the carrier solvent, temperature of the column, dimensions of the column, and method used to pump the solution through the column. Therefore, SEC results for different binders from different labs may not be comparable.

The second consideration is regarding the definition of molecular size itself. Molecules with high polarity have intermolecular forces that keep such molecules together. Molecules bonded by such forces are referred to as associated molecules. Although intermolecular forces are much weaker compared to intramolecular forces (e.g., a covalent bond between a hydrogen and a carbon atom), the association between molecules may be strong enough to resist separation in the solvent used as a carrier for the SEC. As a result, several smaller associated molecules may appear as a single large molecule when separated using the SEC. One may argue that since associated molecules are likely to remain as such in an asphalt binder, the SEC must be carried out using a solvent that preserves these associations.

In a study following SHRP, researchers from the Western Research Institute compared the molecular size distribution of different asphalt binders to their rheological properties. They used the SEC to separate the asphalt binder into two fractions: SEC Fraction-I and SEC Fraction-II. The demarcation between the two fractions was based on fluorescence under ultraviolet (UV) radiation at 350 nm wavelength and reported to be very distinct. The material extracted from SEC Fraction-I is often in the form of dark friable solid particles and represents the strongly associated molecules in the asphalt binder. In contrast, the material extracted from SEC Fraction-II is a dark viscous liquid comprising molecules that are not strongly associated and hence fluorescent to UV light. In terms of relationship to the mechanical properties of the asphalt binder, the phase angle δ obtained by measuring the phase shift between sinusoidal stress and strain response correlates with the concentration of the SEC Fraction-I (more on phase angle later in this chapter). In summary, the molecular size distributions from SEC can be used to compare the relative molecular associations and size distribution of different binders provided the exact same testing conditions are used during the SEC separation.

Concentration of the SEC fractions can also provide some indication of the relative elastic to viscous nature of the asphalt binder. However, the success in relating molecular weight distribution to mechanical properties is limited because although asphalt binders have similar mechanical or viscoelastic behavior as polymers, the molecular origins of such behaviors are vastly different.

2.3.2.3 Classification Based on Ionic Character

The third category to relate the chemical makeup of asphalt binders to their mechanical properties is based on the ionic character of the molecules. Ion exclusion chromatography (IEC) is used to separate the acid, base, and neutral fractions in any given asphalt binder. This method involves dissolving the asphalt binder in a solvent and pumping it through columns containing activated anion or cation resins. Molecules with a basic functional group are adsorbed on the cation resins, and molecules with an acidic functional group are adsorbed on the anion resins. Neutral fractions are eluted through the column without being absorbed by these resins. A stronger solvent is then used to desorb the basic and acidic fractions from the cationic and anionic resins. Results from the IEC can vary depending on the conditions used during chromatography, including but not limited to the type of solvent being used and the rate at which the solvent is pumped through the columns. Therefore, care must be exercised in interpreting results or comparing the results from IEC on different binders from different testing sources. Also, there are several amphoteric fractions in the asphalt binder, i.e., fractions that are both acidic and basic in nature. In sequential chromatographic columns (e.g., cation resin column followed by anion resin column), the amphoteric fractions can be adsorbed in the first column that they come into contact with.

One of the studies from SHRP (Report A-368) evaluated seven different asphalt binders based on their IEC fractions. The asphalt binders were separated into five different fractions by using solvents and adsorption resins with different strengths: weak acids, strong acids, weak bases, strong bases, and neutrals. Typical results from this study indicate that the neutral fractions make up approximately 50% of the binder by mass. Approximately 23–36% of mass fraction of the binders was weakly or strongly acidic, and approximately 13–20% of the mass fraction of the binders was weakly or strongly basic. In terms of functional groups, sulfoxides and pyrrolic functional groups were found in all polar fractions, ketones were mostly concentrated in weak acids and bases, carboxylic acids were mostly found in strong acids, quinolones were mostly found in strong acids, and phenolic functional groups were found in strong and weak acids. One of the goals of this exercise was to relate binder chemistry to ion exclusion chromatography. To this end, findings from the SHRP studies indicate that the molecular weight of the IEC neutral fraction positively correlated well with the viscosity of the binders at 25 °C. Also, the molecular weight of the strong acid fraction negatively correlated with $\tan \delta$ of the asphalt binders, where δ is the viscoelastic phase angle. However, the aforementioned correlations do not necessarily explain the mechanisms or causes that result in the viscosity building of a binder.

2.3.2.4 Classification Based on Polarity

The last form of classifying the different molecular species in an asphalt binder, and perhaps the one that has received the most attention, is based on the polarity of the molecules. Asphalt binders comprise molecules that exhibit a range of polarity from nonpolar to highly polar. As discussed previously, highly polar molecules also tend to be the most associated. Recall the aphorism from basic chemistry that "like dissolves like". In this context, if a nonpolar solvent is used with the asphalt binder, it is likely to dissolve the nonpolar molecules as well as molecules that are at least partially nonpolar leaving behind the highly polar molecules as a precipitate. In some of the earliest works related to separating the asphalt binder based on polarity, a nonpolar solvent such as n-hexane was used to separate the highly polar fractions from the less polar or nonpolar fractions. The precipitate obtained after dissolving an asphalt binder in a nonpolar solvent was referred to as asphaltene, whereas the fraction that remained dissolved in the binder was referred to as maltene. Asphaltenes are dark friable solid particles, whereas maltenes are a thick viscous fluid. Referring back to the SEC, the SEC Fraction-I typically corresponds to asphaltenes, whereas SEC Fraction-II corresponds to maltenes. Figure 2.5 shows the asphaltene and maltene fractions extracted from a typical asphalt binder.

During the late 1940s, Hubbard and Stanfield (1948) further separated the maltenes into resins and oils using other slightly polar solvents, with resins being

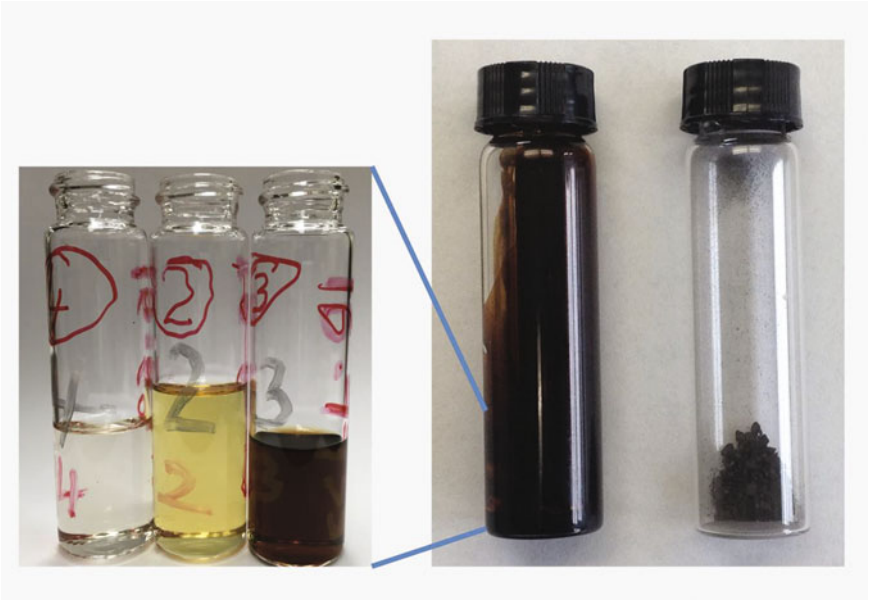


Fig. 2.5 Asphaltene (*right*) and maltene (*middle*) fractions from a typical asphalt binder separated using n-heptane; maltene further separated into Saturates, Aromatics and Resins (inset—*left* to *right*) (Photograph by A Bhasin)

slightly more polar than oils but much less polar than the asphaltenes. Finally, in the 1970s Corbett (1969) used liquid chromatography and solvents with different levels of polarity to separate maltenes into three fractions. The four fractions thus obtained are referred to in increasing order of polarity as: (S) saturates, naphthene aromatics or simply (A) aromatics, polar aromatics or (R) resins, and (A) asphaltenes. These fractions, also referred to by their acronym SARA fractions, are still being used to understand the relationship between the chemical makeup of asphalt binders and their mechanical properties. Saturates (typically 5–15% by weight) are the least polar fraction of all and have a glass transition temperature of approximately −70 °C that is approximately 40 °C below the glass transition temperature of the parent binder.

Saturates also have the lowest density of the four fractions with typical values in the range of 0.9 g/cm^3. On the other extreme, asphaltenes (typically 5–20% by weight) are highly polar dark colored solid particles at room temperatures. Asphaltenes have the highest density of the four fractions with typical values in the range of 1.15 g/cm^3. Asphaltene molecules typically have several fused aromatic rings in a plate-like structure; these molecules have been shown to stack together as micelles both experimentally and using computational modeling tools such as molecular dynamics. The aromatics and resins have polarity and properties that are in between these two extremes (saturates and asphaltenes). These four molecular species collectively work together to impart the time-dependent viscoelastic properties of the asphalt binder.

The solubility of different chemical species is typically measured using Hansen solubility parameters with units of MPa$^{0.5}$. Molecules with similar Hansen solubility parameters are more likely to remain dissolved. Saturates, aromatics, resins, and asphaltenes have a solubility range of approximately 15–17 MPa$^{0.5}$, 17–18.5 MPa$^{0.5}$, 18.5–20 MPa$^{0.5}$, and 17.6–21.7 MPa$^{0.5}$, respectively. The monotonically increasing range of solubility is also a reflection of the methodology used to separate these four fractions. The aforementioned ranges must not be considered as a definitive cutoff for these fractions but more of an approximate quantitative indicator for the polarity of the SARA fractions. As with other techniques discussed earlier, the SARA fractions are a qualitative indicator and do not have a universal definition. For any given binder, the quantitative estimates of these fractions depend on the solvents and adsorption medium used during liquid chromatography. Also, the solubility parameter, especially for asphaltenes, can be influenced by the aggregation of the different molecules and must be interpreted carefully. Having considered the asphalt binder as an ensemble of organic molecules with different structures and polarities, it is important to address the spatial assembly of such molecules.

2.3.3 Microstructure of Asphalt Binders

Figure 2.6 illustrates SARA fractions of some of the core SHRP binders. As mentioned earlier, Saturates normally compose between about 5–15 weight percent

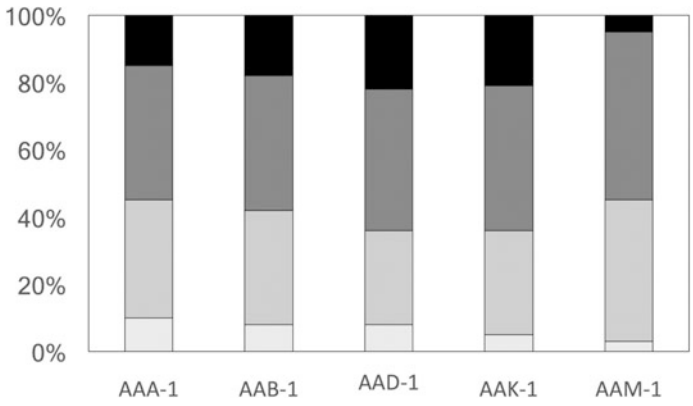


Fig. 2.6 SARA fractions for some of the binders from the core Strategic Highway Research Program, SHRP (Recreated using data from SHRP Report A-646 by Mortazavi and Moulthrop 1993)

of the paving grade bitumen (Corbett 1969). The structures are mainly aliphatic with some branching of the chains and with a solubility parameter (related to the energy required to completely remove a unit volume of molecules from their neighbors to infinite separation) between 15 and 17 $MPa^{0.5}$ (Speight 1999). The aromatics compose between about 30–45 weight percent of the bitumen and have a carbon skeleton that is slightly aliphatic with lightly condensed aromatic rings and with a notably higher molecular weight than the saturate fraction (Corbett 1969). The solubility parameter is between about 17 and 18.5 $MPa^{0.5}$ (Speight 1999).

The resins, also referred to as polar aromatics, compose between about 30 to 45 wt% of the bitumen and form a black solid at room temperature (Corbett 1969). They play a key role in the stability of the bitumen as they stabilize the asphaltene fraction. Their stability parameter lies between about 18.5 and 20 $MPa^{0.5}$ (Speight 1999). Asphaltenes generally represent between about 5 and 20 wt% of the bitumen of paving grade (Corbett 1969). Figure 2.7 presents average molecular structures of asphaltenes in oil sands. Several characteristics of the asphaltene structure may be gleaned from this figure. They contain condensed aromatic rings and more polar groups and fused aromatic structures. They also form planar molecules that can associate through π-π bonding to form staked-type structures, Fig. 2.8. The stacking phenomenon of asphaltenes has been captured by X-ray diffraction of "pure" solid asphaltenes from which the Yen model was derived as illustrated in Fig. 2.9.

Researchers have investigated the spatial organization of different molecular species within the asphalt binder since the 1940s. Knowing that the molecules that make up the asphalt binder are diverse in terms of their elemental composition, size,

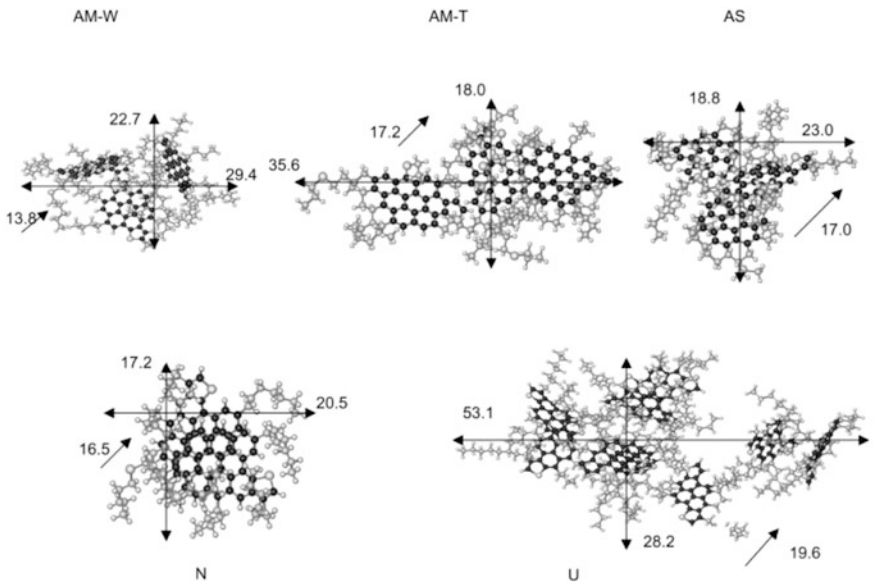


Fig. 2.7 Average molecular structures for asphaltenes from oil sands: Athabasca (Canada) asphaltenes recovered using distinct procedures (AM-W, AM-T and AS), Dahomey (Nigeria) asphlatenes (N) and Sunnydale (Utah, USA) asphaltenes. The numbers are in Angstroms (Reprinted from Fuel, 80/13, Zhao et al., Solids contents, properties and molecular structures of asphaltenes from different oilsands, pp 1907–1914, Copyright (2001), with permission from Elsevier.)

and polarity, one can imagine two possibilities for their spatial organization, henceforth referred to as the microstructure of asphalt binders. The first possibility is that different molecular species agglomerate to form well-defined structural entities within the binder. Such agglomerates may further be a function of the thermal history of the asphalt binder. The second and contrasting possibility is that the different molecular species are homogenously dispersed within the asphalt binder. The first possibility is referred to as the colloidal model, and the latter is referred to as the dispersed fluid model.

The idea that molecular species with different polarities coexist in the asphalt binder in the form of a colloid dates back to the late 1940s and 1950s from the work of Traxler and Romberg (1952). These early models proposed that asphaltenes floated in the form of micelles within a matrix of maltenes. Later studies further separated maltenes into saturates, aromatics, and resins and also demonstrated that aromatics and resins were critical in ensuring the stability of the binder. In other words, without a certain minimum percentage of aromatics and resins, the asphaltenes and saturates would simply separate. Earlier studies also proposed that these fractions created macrostructures, referred to as the sol-gel model (Pfeiffer

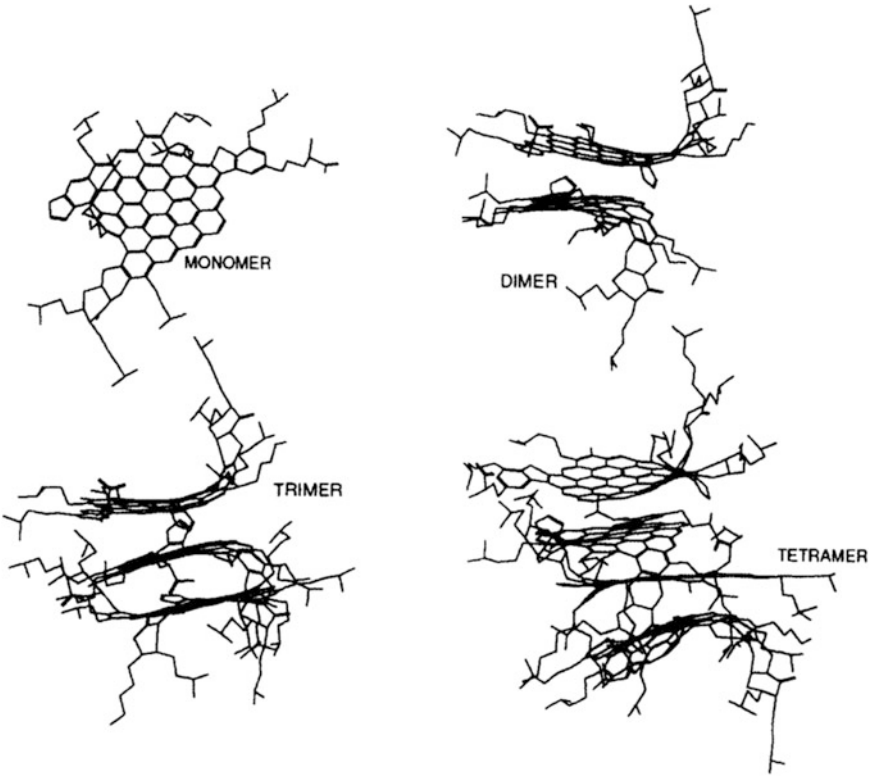


Fig. 2.8 Formation of dimer, trimer and tetramer for as Venezuelan molecule (Reprinted from Colloids and Surfaces A: Physicochemical and Engineering Aspects, 104/1, Rogel. E., Studies on asphaltene aggregation via computational chemistry, pp 85–93, Copyright (1995), with permission from Elsevier)

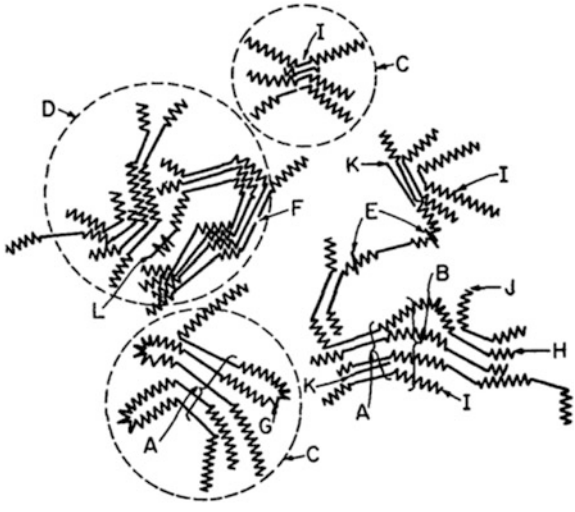


Fig. 2.9 The Yen model of "pure" solid asphaltenes (Reprinted with permission from Analytical Chemistry, Dickie, J.P. and Yen, T.F., Macrostructures of the asphaltic fractions by various instrumental methods, Vol. 39, No. 14, pp 1847–1852, Copyright (1967), American Chemical Society)

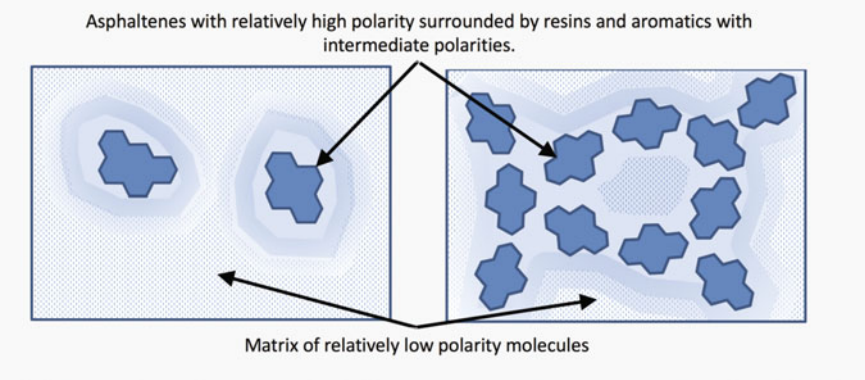


Fig. 2.10 Original colloidal model: sol and gel bitumen (Recreated based on concept from Read and Whiteoak 2003)

et al. 1940), that could be used to explain differences in the rheological behavior of bitumen from different sources. Figure 2.10 presents a schematic of the original colloidal model for sol and gel bitumens. During the most recent decade, significant experimental evidence has emerged to support the colloidal model. For example, results based on small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) indicate that asphaltenes tend to agglomerate and form micelles in asphalt binders as well as other organic solvents.

A contrasting theory to the colloidal model is the dispersed polar fluid model, which, as the name suggests, proposes that the polar molecular fractions are more uniformly dispersed within the asphalt binder. As of this writing, there seems to be increasing consensus toward the colloidal model. Readers interested in learning more about the differences between these two models for asphalt binders are encouraged to review papers by Lesueur (2009a, b) and Redelius (2006). These two papers present very thorough, and often contradicting, synthesis of observed experimental results and their interpretations supporting one or the other of the two microstructure models discussed above. For example, Lesueur (2009a, b) uses results from SAXS and SANS to support the colloidal model, whereas Redelius (2006) suggests that the SAXS and SANS cannot be used to establish whether microstructures exist, rather results from these techniques can be interpreted only if it is known a priori' that such structures exist.

A more recent body of experimental work that is being used to investigate the microstructure of asphalt binders is based on the use of atomic force microscopy (AFM). Unlike an optical (or electron) microscope, AFM scans the surface of a test specimen by "feeling" the surface of a test specimen using a very sharp conical or spherical tip attached to the end of a cantilever. The end of the tip is only a few nanometers in size and the cantilever is no more than 1 or 2 mm in length. By moving the tip over the test surface, it is possible to obtain a profile of the surface. By measuring the friction and forces of attraction between the tip and the surface, it is possible to identify whether the material underneath the tip has changed. The

latter method produces phase images highlighting the heterogeneities on the surface of the test material. Typical AFM images have a spatial resolution of a few nanometers over an image size of approximately 25 × 25 to 50 × 50 μm.

Loeber et al. (1996) were perhaps the first to identify the existence of two or more phases in the asphalt binder using AFM. They labeled the discrete phases as "bee structures" based on their peculiar shape. In more general terms, typical phases observed using AFM in an asphalt binder may be regarded as the matrix, inclusion, and matrix-inclusion interface (Fig. 2.11). Several independent research groups have confirmed the presence of these phases (e.g., Loeber et al. 1996). However, as of this writing there is a lack of consensus on the relationship between specific phases and molecular species in the asphalt binder. Readers must also note that the term "phase" in this context is not employed in the strictest sense. In physics and chemistry, the term phase is used to describe one of many states or configurations for a material with the same chemical composition. For example, water, ice, and steam are different phases of the same material. In the context of asphalt binders and literature relevant to AFM, the term phase may refer to a microstructural entity with entirely different chemistry. The term domain or microdomain, although not commonly used in the asphalt literature, may be more appropriate.

As mentioned before, as of this writing there is no conclusive evidence in the literature to fully explain the chemical makeup of the different domains that appear in the AFM images (Fig. 2.11). For example, Masson and coworkers suggest that the inclusions represented asphaltenes in the asphalt binders (2007). Kringos et al. (2009) and Schmets et al. (2010, 2009) suggest that the asphalt binder comprises two different phases: a wax rich phase and a continuous matrix surrounding it. They attributed the inclusions (aka "bee structures") to waxes in the asphalt binder as opposed to the asphaltenes as suggested by Masson et al. (2007, 2006). However, the wax content in most asphalt binders is typically between 0 and 7%, which is not consistent with the relative percentage of the inclusions in the AFM images.

Other researchers conducted a systematic evaluation of the different domains observed using the AFM. In one study, the microrheology of the different domains observed using the AFM before and after oxidative aging of different binders was evaluated and compared (Allen et al. 2012). Results showed that (1) oxidative aging increased the size and frequency of the inclusions, and (2) the inclusions and inclusion-matrix interface were much stiffer than the paraphase. Other studies have established that oxidative aging increases the percentage of asphaltenes through the oxidation and concomitant decrease in the concentration of resins and aromatics (Lin et al. 1995; Petersen et al. 1993; Petersen 1984). This information in conjunction with the AFM results suggests that the inclusions may be related to the presence of asphaltenes. Also, since the stiffness of asphaltenes is typically an order of magnitude higher than the aromatics or saturates, it is unlikely that the inclusions or inclusion-matrix interface are made up entirely of asphaltenes but are more likely to contain a higher concentration of asphaltenes and resins. Other studies related to the use of the AFM were conducted by different derivative asphalt binders that were produced by doping the original binder with each one of the four chemical fractions (SARA) (Allen et al. 2013). These studies also showed that the difference between

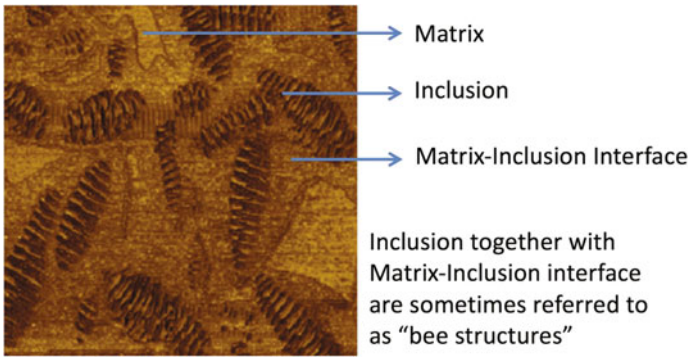


Fig. 2.11 Typical phase image obtained for an asphalt binder using AFM; the image shows different phases in the asphalt binder

the percentage of asphaltenes and saturates (the two extreme polars in the SARA fraction) was related to the size and dispersion of the catanaphase and periphase. Finally, more recent studies have shown that these structures are much smaller within the bulk of the binder compared to the surface and highly dependent on the composition and thermal history of the binder (Ramm et al. 2016).

It is possible that the inclusions and the inclusion-matrix interface may not exclusively correspond to asphaltenes or waxes as suggested by other researchers. A more realistic explanation for these phases, which is also consistent with the reported experimental evidence, is that these domains have higher concentrations of asphaltenes and resins, while the matrix has a higher concentration of saturates and aromatics. This proposed model of the asphalt microstructure reconciles the differences reported by the proponents of the colloidal model and the dispersed fluid model.

One of the major drawbacks of using the AFM is that this technique does not investigate the bulk of the asphalt binder and evaluates a test material only at its surface. Also, unless a series of functionalized tips are used with AFM imaging, the phase images cannot be used to identify the chemical composition of the asphalt binder.

2.3.4 Relationship Between Microstructure and Engineering Properties of Asphalt Binder

The ultimate goal of understanding the relationship between chemical composition, microstructure, and rheological properties of an asphalt binder is that such knowledge can eventually be used to design newer asphalt binders, binder modifiers, or eco-friendly substitutes with superior mechanical properties and performance characteristics.

Thus far, we have discussed the possible relationship between the chemical composition and microstructure of the asphalt binders but we have not fully addressed a very relevant question: Can the properties of the constituent chemical fractions be used to predict the properties of the binder without knowing the microstructural arrangement of these fractions within the binder? In other words, Why is understanding the chemical makeup and microstructure of the asphalt binder important?

To answer this question, let us consider the problem of predicting the mechanical or rheological properties of a composite (in this case asphalt binder) of different constituent materials (in this case SARA fractions). The simplest approach to solve such a problem would be to use the rule of mixtures or something similar to Einstein's model. However, studies also show that the shape, size distribution, and interfacial properties of the inclusions in a matrix significantly influence the properties of the composite (Chow 1980; Demjén et al. 1998; Verghese and Lesko 1999). One of the models that has been used with some success to predict the rheological properties of a composite of two different viscoelastic materials was proposed by Palierne (1990). Other studies have shown that even the Palierne model cannot accurately predict the composite properties, when there are more than two components in the composite (Elias et al. 2007). To summarize, in addition to the volume concentration, the microstructure of a multicomponent or multiphase material must also be known to predict the properties of the composite.

The fact that mixture models cannot be used to predict the properties of a composite binder has been well known within the industry for many years. For example, Siegmann (1950) demonstrated that the mixture of two crude stock binders from two different sources with very high and similar stiffness resulted in a composite that had a stiffness that was lower than the stiffness of the constituent binders. The only explanation for the results observed by Siegmann is that combining the crude stocks from these two chemically diverse sources resulted in a microstructure that reduced the overall stiffness of the composite binder. Other researchers (Kringos et al. 2009) have also hypothesized that the microstructure within the asphalt and the difference in the stiffness of the microstructural entities is responsible for nucleating fatigue cracks in the asphalt binder.

To further demonstrate the significance of microstructure on the properties of the composite, consider a simple 2D finite element analysis of an asphalt binder based on the geometry and microrheology obtained using AFM. Figure 2.12 illustrates the creep compliance of the different phases within the binder as well as the creep compliance of the composite. The colored overlay on the AFM images shows regions of stress concentration, where the local stress intensity is two to five times the far field stress.

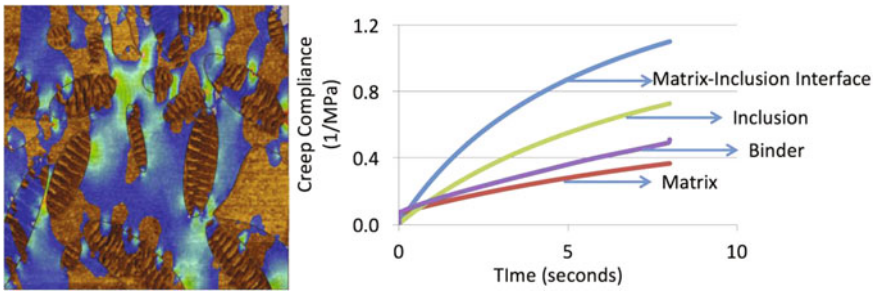


Fig. 2.12 Results from a simple 2D finite element analysis that demonstrate the influence of microstructure on the overall properties of the binder

2.3.5 Concluding Thoughts on the Chemical Properties of Asphalt Binders

In summary, the chemical makeup of asphalt binders and their microstructure (defined as the spatial arrangement of different chemical species within the binder) significantly influence the engineering properties of the binder. Although chemical properties of asphalt binders discussed in the preceding sections are typically not used for purchase specifications or as an input in the design of asphalt binders, mixtures, or pavements, this information is extremely useful to better understand several aspects related to binder selection and mixture design. For example, the knowledge of binder chemistry is important to understand the effect of binder oxidation on mixture performance, role of chemical and other additives to enhance binder and mixture properties, and handling of recycled asphalt materials in design of new asphalt mixtures. These and other aspects will be discussed in the following sections and chapters. Engineers and researchers involved in the production and modification of asphalt binders should treat the aforementioned section as a simplified starting point for further reading.

2.4 Aging in Asphalt Binders

Understanding aging in asphalt binders is very important to appreciate the significant changes that occur in the mechanical response of the asphalt binder during the various stages of its life starting from asphalt mixture production in a hot-mix plant to end of service life in an asphalt pavement, and beyond as the binder may be used in the form of recycled asphalt material incorporated in new asphalt mixtures. Typically the term “aging” of asphalt binders generally refers to oxidative aging of binders. However, in a more general sense the term aging may sometimes also be used to describe the hardening of asphalt binders due to steric hardening (that is

reversible), oxidative aging, loss of volatiles from the binder, or any combination of these factors. This section briefly describes these processes. The importance of aging in characterizing the properties of asphalt binders and composites cannot be over emphasized. Although the subject of aging is introduced here only briefly, it will be a recurring theme throughout this book as we investigate its influence on physical and mechanical properties of asphalt binders and composites.

2.4.1 Steric Hardening

Steric hardening is also referred to as reversible aging and can occur at intermediate to low temperatures. At intermediate temperatures (typically around 5–25 °C), steric hardening is typically driven by the polar nature of the molecules. Recall from the previous section that asphalt binders comprise molecules with varying degrees of polarity. Note that polar molecules also tend to have opposite charges at different ends. For example, consider the pyridine molecule as shown in Fig. 2.13. The nitrogen atom imparts polarity to this molecule, with one end of the molecule being positively charged and the other being negatively charged.

Molecules above absolute zero possess some amount of energy that keeps them in constant motion. As a result, over time, polar molecules within the material tend to rearrange themselves such that the opposite charges between two adjacent polar molecules are aligned next to each other. This results in the formation of secondary bonds or associations between the polar molecules and a concomitant increase in the capacity of the material to resist deformation; engineers recognize this as an increase in stiffness. As a corollary, the breakdown of these secondary bonds or associations causes the material to loose its stiffness or become soft. Secondary bonds are weak bonds and require very little energy to be disrupted. This energy can be provided in the form of thermal energy by heating the binder or in the form of external work by applying stress. Steric hardening is also referred to as reversible aging, since the formation of secondary bonds between molecules and concomitant increase in stiffness can be reversed by heat or work.

Steric hardening or reversible aging also takes place at lower temperatures (typically below 0 °C), albeit via a different mechanism. At lower temperatures, neutral or apolar molecules (and typically molecules with few or no branches) tend to reorient themselves into a well-organized structure. This organization typically tends to increase the overall stiffness of the asphalt binder. This process is typically referred to as crystallization of naturally occurring “waxes” in the asphalt binder. The resulting increase in stiffness is also referred to as physical hardening in some literature. In fact, a temperature sweep conducted using a differential scanning calorimeter can be used to detect the energy released during the crystallization of the these molecules. Similar to steric hardening at intermediate temperatures, physical hardening at low temperatures is also easily reversed with an increase in temperature. The existence of this form of hardening must be recognized while measuring engineering properties of the material. For example, certain studies have

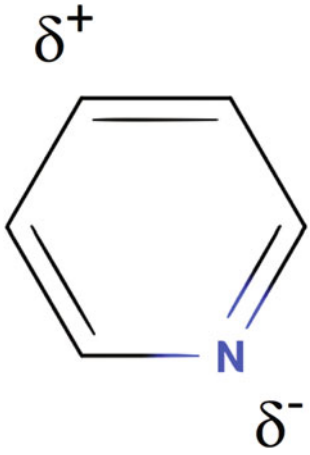


Fig. 2.13 Structure of polar pyridine molecule showing regions of positive and negative charge

shown that asphalt binders conditioned at low temperatures (approximately −5 to −10 °C) continue to show an increase in stiffness with an increase in the conditioning time up to a certain limit.

2.4.2 Volatilization and Oxidative Aging

Asphalt binder experiences oxidation during two stages of its life: (1) during mixture production and placement and (2) during its service life. During production and placement, a process that can last typically for a few hours, hot or warm mix asphalt binder is maintained at very high temperatures (typically above 100 °C) in the form of a thin film coating aggregate particles in a loose mix. These conditions (high surface area, high temperature, and easy access to oxygen from air in a loose mix) result in significant oxidation of the asphalt binder by the time it is placed and compacted in the field. Also during this time, some of the lighter and more volatile fractions escape from the asphalt binder. Oxidation of the binder continues during the service life of the asphalt pavement albeit at a much slower rate, since the temperatures are relatively lower and access to oxygen is limited.

As the name suggests, oxidative aging occurs due to the oxidation of unsaturated bonds typically in the aromatics and resins fraction of the binder. Unlike steric hardening, oxidation results in a change in the molecular structure that cannot be easily reversed. The addition of an oxygen atom to a molecule also typically results in an increase in its polarity. As a result, over time the aromatic and resin molecules are consumed to form more asphaltene like polar molecules. This in turn leads to the formation of larger and more associated groups of molecules that increase the stiffness and decrease the ductility of the binder. Since oxidative aging and concomitant changes in the properties of the asphalt binder and mixture continue to occur over the service life of the pavement, it is imperative that any approach to

predict the performance of a mixture or pavement structure take into consideration this continually changing nature of the asphalt binder.

2.4.3 Simulating Aging in Asphalt Binders

Asphalt binders are typically evaluated for their mechanical properties at different levels of aging. For practical purposes, three different points or levels of aging are defined for asphalt binders: (1) unaged, (2) short-term aged, and (3) long-term aged. As the name suggests, unaged binder is the binder as received from the refinery by the contractor responsible for producing the hot mix asphalt. The viscosity of the unaged binder is of interest, because it dictates the workability of the asphalt mixture during production at the hot mix plant. The short-term aged binder refers to the binder immediately after placement in the field. In this sense, short-term aging encompasses aging that occurs during the few hours starting from mixture production at the hot mix plant, transport to the construction site, until placement and compaction. Recall that although the short-term aging occurs only for a duration of a few hours, the loss of volatiles and extent of oxidation is significantly higher on account of: (1) the high mixing temperatures and (2) excessive exposed surface area of the binder-coated aggregates in the loose mix. Finally, long-term aged binder reflects the state of the binder in the mixture at the end of the expected service life of the pavement. The duration of time between the short-term aged state and the long-term aged state is several years, but in this case the oxidation occurs at typical service temperatures of the pavement.

Several methods have been developed over the years to simulate short-term and long-term aging in asphalt binders in a laboratory environment in a short accelerated time frame. Two of the most common methods are the rolling thin film oven (RTFO) and pressure aging vessel (PAV) to simulate short-term and long-term aging, respectively. The RTFO is a device (typical device shown in Fig. 2.14) that injects a stream of hot air (typically at 163 °C) into an open mouthed bottle filled with asphalt binder. The bottle or bottles filled with the asphalt binder are placed on a rack that constantly rotates. The rotation helps the binder film flow over itself, thus avoiding a scenario in which only the surface of the binder is oxidized. The aging is carried out typically for 85 min, and the binder sample is weighed before and after aging to estimate the mass percent of volatiles that escape from the binder. At the end of the test, the binder sample is treated as a short-term aged binder. A very important consideration here is that the binder industry typically uses standard temperature and duration for producing these so-called short-term aged binders and for subsequent grading of the binder. Given the fact that the binders used in the field may be produced using a hot mix asphalt or warm mix asphalt technology (at 20–30 °C lower than hot mix asphalt), transported over varying distances for construction at elevated temperatures, and possibly even stored in a heated silo for extended durations, it is unlikely that a universal standard time and temperature for RTFO aging would result in an accurate representation of the binder

aged in the field. On the other hand, when the binder is being produced in a refinery, there is no way for the producer to ascertain the circumstances or technology that will be used to produce and place the asphalt mixture in the field. As a result, despite this limitation the RTFO is still used as a standard benchmark to physically produce a short-term aged sample of the asphalt binder.

The sample aged using the RTFO is further aged with the pressure aging vessel or PAV to simulate long-term aging (typical device shown in Fig. 2.15). The RTFO binder sample is placed as a thin film on a metal tray in the PAV. The sample is then subjected typically to a temperature of 100 °C at 300 psi (2.07 MPa) pressure for 20 h in the presence of air which naturally contains oxygen. The aging temperature may vary based on climatic conditions. The high temperature and pressure accelerate oxidation in the binder to simulate several years of aging. Similar to the RTFO, the conditions used for PAV are standardized, which helps with the grading and specification of the asphalt binder but is not always the most accurate representation of what occurs in the field. Also, there has been some criticism that the PAV aging does not really simulate the claimed 7–10 years of aging. For example, some recent studies have aged the asphalt binder under atmospheric conditions at realistic high pavement temperatures (e.g., 60 °C) and compared the binder to the PAV-aged binder to suggest that the PAV does not age the binder as severely as claimed. Another criticism about the PAV aging method is that the high temperature and pressures may trigger reactions that would normally not occur in the pavement at the typical service temperatures (recall from basic chemistry that certain reactions have an energy barrier or threshold that needs to be crossed for the reaction to occur).

2.5 Mechanical Properties

2.5.1 Scope

In the previous sections, we have emphasized the need to be knowledgeable of binder chemistry (at least to some extent) in order to better understand and appreciate the mechanical behavior of asphalt binders and consequently asphalt mixtures. This section will first discuss the challenges and differences in characterizing the properties of an asphalt binder, which are also applicable for asphalt mixtures. This will be followed by an overview of the methods typically used to characterize the time, temperature, and age dependent properties of an asphalt binder. This section is a starting point for readers who are being introduced for the first time to asphalt binders or for practicing engineers who are interested in refreshing their knowledge on the basics of rheology that are the basis for current standards. For readers interested in details of the current industry standards and protocols used to grade asphalt binders, a list of easily downloadable references is provided at the end of this chapter. Readers interested in a more in-depth

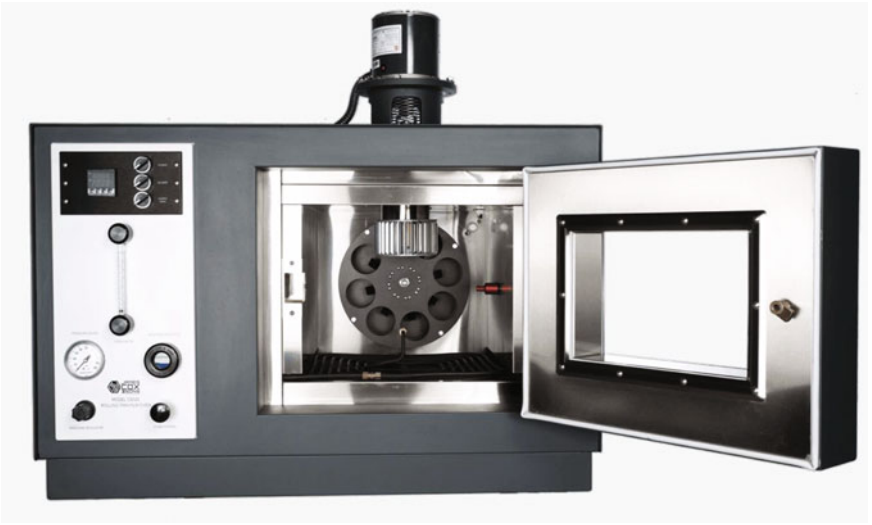


Fig. 2.14 Image of a typical RTFO device for short-term aging (Image courtesy of James Cox and Sons, USA)

mathematical analysis of the methods described in this chapter can refer to Part II of this book. *Note that the information presented in this section and chapter is for a reader interested in having the basic working knowledge of the methods used to characterize asphalt binders and model its response. A more detailed and mathematically rigorous account of these models is revisited again in Part II and Chap. 13 of this book. A slight overlap of information and repetition is by design to enable the readers to appreciate the application of the mathematical methods in the context of the material (asphalt binder or mixture) as well as in the context of the general principles of mechanics.*

2.5.2 *Significance of Mechanical Properties of Binder and Challenges*

Characterizing and reporting binder properties is far more complex compared to other construction materials. To illustrate this, let us once again use the example of steel and contrast the design of an asphalt pavement to the design of a steel structure in three areas: material characterization, structural design, and failure.

First, in terms of material characterization, steel used in construction can typically be considered to be linear elastic within its service range. In other words, steel has a well-defined modulus and tensile strength that does not vary with rate of loading, typical service temperatures, or age of the structure (exceptions to this may be

Fig. 2.15 Image of a typical PAV device for long-term aging (Image courtesy of Gilson Company Inc., USA)

designing structures that resist high temperatures in the event of a fire). In contrast, the mechanical properties of asphalt binders and concomitantly those of asphalt mixtures are highly dependent on time histories or rate of loading (e.g., traffic speeds or rate of cooling), temperature of the environment, and age of the material.

Second, in terms of structural design, a steel structure can take any simple or complex geometric form. The job of a design engineer is to calculate stresses at different locations within the structure and to determine the cross sections of steel members. In contrast, an asphalt pavement always has a geometry of a slab; the only design variable that can be manipulated by design engineer is the layer

thickness and in some cases the use of an interface layer with special properties between these layers (e.g., geosynthetics or other membranes).

Third, in terms of failure criteria, steel structures oftentimes fail due to lack of serviceability (e.g., unacceptable deflection) much before they fail due to lack of strength (e.g., fracture or plastic failure). In contrast, failure in asphalt pavements due to lack of serviceability occurs much after failure due to lack of strength. For example, microcracks and concomitant weakening of the pavement or small amounts of plastic deformation are allowed to a certain extent as long as it does not severely deteriorate the pavement structure and ride quality.

In summary, the most challenging aspect of designing a pavement structure is not the complexity of the structure but rather the complexity of the stress-strain behavior that is dependent on the time or rate of loading, the temperature of the material, and the age of the binder. More importantly, flexible pavement structures are allowed to accrue small amounts of damage (e.g., plastic deformation or rutting and microcracking) before they become unserviceable. Therefore, characterization of the stress-strain response of the material in an undamaged state as well as the characterization of damage evolution (as a function of time or rate of loading, temperature, and age) in asphalt materials is critically important and perhaps the most challenging aspect of material and pavement design.

2.5.3 Time Dependency

Asphalt binders are observed to be time dependent or viscoelastic in nature. The term viscoelastic indicates that the material response is somewhere in between that of a viscous fluid and an elastic solid. The stress-strain behavior of an isotropic linear elastic solid in one dimension is typically given as $E = \sigma/\in$, where E is the modulus of elasticity, σ is the stress, and $\in$ is the strain. Also, recall the behavior of a simple Newtonian fluid, wherein the shear stress is directly proportional to the strain rate. This relationship is typically denoted as $\eta = \tau/\dot{\gamma}$, where η is the shear viscosity, τ is the shear stress, and $\dot{\gamma}$ is the shear strain rate. Asphalt binders are typically shown to have a stress-strain response that is a combination of an elastic solid and a viscous fluid (although not necessarily a Newtonian fluid). As a result, any model that describes the stress-strain relationship of an asphalt binder will inevitably include time or rate of loading. The influence of time decreases with a decrease in temperature. This will be discussed in more detail in the next subsection.

The time or rate dependency of asphalt binders is typically modeled and measured either in time domain or frequency domain. The two most common tests used in the time domain are the creep test and relaxation test. These two tests yield the creep compliance and relaxation modulus of the material, respectively. *The test used in the frequency domain is the frequency sweep test that yields the complex or dynamic modulus of the material.* Note that the tests discussed here are typically considered to represent the linear viscoelastic response of the material without any damage. In other words, *these tests are typically conducted by applying a stress or*

strain of magnitude that is not large enough to induce significant permanent damage to the material. A more detailed mathematical description of these behaviors and test methods is discussed in Part II of this book (Chaps. 11 and 12).

The creep test involves applying an instantaneous constant load or strictly speaking stress to the test specimen and measuring its deformation over time. The material property measured using such a test is referred to as the creep compliance, $D(t) = \varepsilon(t)/\sigma_0$, where σ_0 represents the constant stress applied during the creep test, and $\varepsilon(t)$ denotes the measured strain that changes with time. The creep compliance $D(t)$ can be represented using any mathematical function that appropriately describes the observed material behavior. For example, mechanical analog models are sometimes used to model the measured response. Such models comprise one or more elastic springs and viscous dashpots configured in series or parallel or a combination of the two such the overall mechanical model can mathematically simulate the measured response of the material. Another approach is to use simple phenomenological models to model the measured creep behavior. For example, the power law model $D(t) = D_0 + D_1 t^m$ or the Prony series of the form $D(t) = D_0 - \sum_{i=1}^{n} D_i e^{t/\tau_i}$ are often used to model the measured creep compliance of the material as a function of time. The terms D_0, D_i, m, and τ_i are constants that are unique to the material and obtained by fitting the experimentally measured response to the model. It is important to emphasize during this discussion that one should not decide a model a priori and then use it to represent the material behavior. Ideally, it is important to first observe the material response under well-controlled laboratory conditions and then identify the best (and simplest) model that can be used to represent the material response (revisited in more detail in Chap. 13).

A variation of the creep test is the creep-recovery test. In this case, the constant load is instantaneously removed after a specific duration of loading time and the strain recovery of the material is recorded. The observed strain from such a test is often decomposed into components such as elastic or instantaneous strain typically denoted as ε_e, time dependent or viscoelastic strain typically denoted as ε_{ve}, and permanent or irrecoverable viscoplastic strain typically denoted as ε_{vp}. As a side note, in some literature time dependent elastic and viscoelastic mean different things; the former refers to the time dependent elastic behavior with complete recovery of strain after removal of all loads at some point in time whereas the latter suggests that there is some extent of permanent strain after the application of loads. The existence or relative magnitude of these components depends on the material being tested as well as the test conditions. In the context of asphalt materials, the irrecoverable or viscoplastic strain is regarded as the source for distresses such as rutting. The constitutive models for materials that are intended to incorporate damage must include terms to account for such strains. Figure 2.16 presents a schematic of the creep compliance test with typical response for a viscoelastic material and an elastic material.

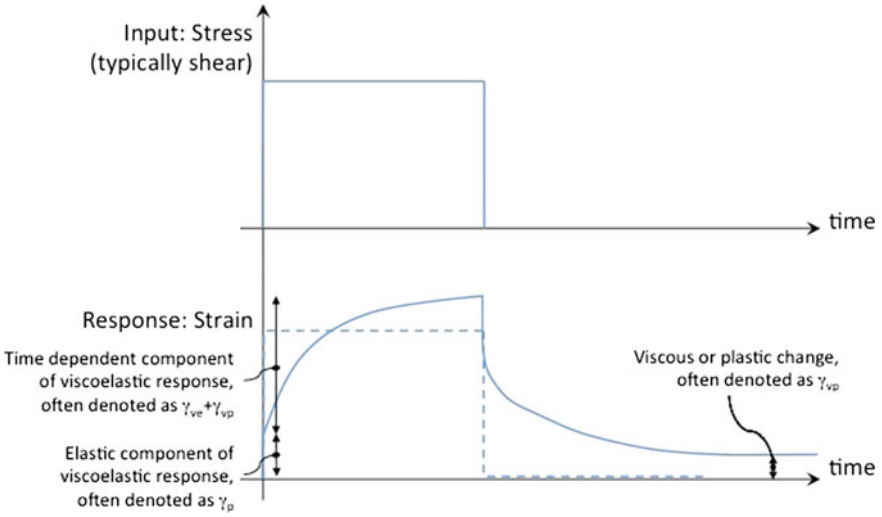


Fig. 2.16 Typical response from a creep-recovery test

The relaxation test involves applying an instantaneous constant deformation or strictly speaking strain to the test specimen and measuring the reaction stress over time. The material property measured using such a test is referred to as the relaxation modulus, $E(t)$, given by $E(t) = \sigma(t)/\in_0$, where $\in_0$ represents the constant strain applied during the relaxation test, and $\sigma(t)$ denotes the measured stress that relaxes with time. Similar to the creep compliance, the relaxation modulus $E(t)$ can also be represented using mechanical analog models or phenomenological models such as the power law or the Prony series. Note that the power law or Prony series will have a different sign for some of the coefficients because of the decreasing nature of the relaxation modulus with time. Figure 2.17 presents a schematic of the relaxation test with typical response for an elastic and viscoelastic material.

The frequency sweep test involves applying several cycles with either constant stress amplitude or constant strain amplitude in a sinusoidal waveform. A constant stress amplitude is applied and the resulting strain amplitude is measured and recorded over time or vice versa. When subject to stress (or strain) following a sinusoidal waveform, the strain (or stress) response of a viscoelastic material reaches a steady state after a few load cycles. It is important to remember that this property is measured only when the material has reached a steady state response. During this steady state, a constant time lag is observed between the stress and strain waves. The time lag, Δt, or phase shift is expressed in the form of a phase angle, δ, in degrees given as $360^\circ f\Delta t$ or $360^\circ\omega\Delta t/2\pi$, where f is the frequency of the sinusoidal wave in Hz or ω is the angular frequency in radians/second. Also, once the material response has reached steady state, the magnitude of complex modulus, $|E^*|$, is defined as $\sigma_0/\in_0$, where σ_0 and $\in_0$ are the stress and strain amplitudes, respectively. Note that the complex modulus is also sometimes referred

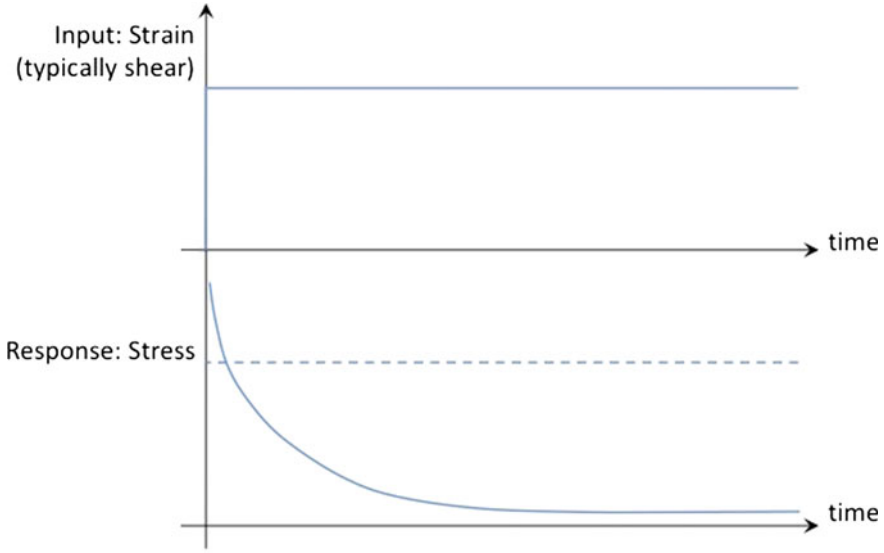


Fig. 2.17 Typical response from a relaxation test

to as dynamic modulus in the common asphalt literature, however, strictly speaking from a mechanics point of view the term dynamic modulus is not appropriate. The notation G^* may be used in lieu of E^*, when the test is conducted in a cyclic shear mode as opposed to a cyclic tension or compression mode. Figure 2.18 presents a schematic of the dynamic modulus test.

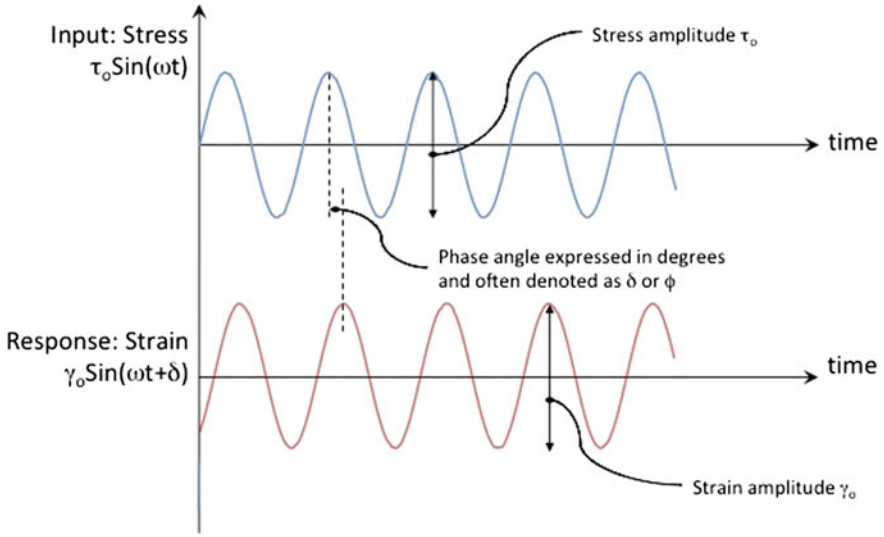


Fig. 2.18 Typical input and response at one frequency for a dynamic modulus test

The complex modulus is also sometimes expressed using complex notation as $E^* = E' + iE''$, where the tangent of the phase angle $\tan\delta = E''/E'$ and the magnitude of the complex modulus E^* is given as $|E^*| = \sqrt{E'^2 + E''^2}$. The term E' denotes the storage or elastic part of the modulus that is in complete sync with the applied load. The term E'' denotes the loss modulus or fluid part of the modulus, that is, 90° out of phase with the applied load. A more detailed mathematical description of these terms can be found in Chap. 13. Also, note that the complex modulus and phase angle are functions of the frequency f or angular frequency ω. Therefore, strictly speaking these terms are measured at multiple frequencies and must be represented as $E^*(\omega)$ and $\delta(\omega)$. Tests conducted at multiple frequencies are referred as frequency sweep tests. The typical protocol for such tests involves subjecting the test specimen to several cycles at a given frequency of sinusoidal loading with a small stress or strain amplitude. The material response from the last few cycles in steady state is then used to determine the modulus and phase angle at that frequency. The process is repeated at several different frequencies. The applied stress or strain amplitudes are usually very small to ensure that the specimen does not experience any permanent change or damage during the process. In addition to complex modulus, the term complex compliance denoted by $|D^*|(\omega)$ or $|J^*|(\omega)$ is also sometimes used. Complex compliance is simply the ratio of the strain amplitude to the stress amplitude, and its magnitude is the inverse of the complex modulus.

It is important to briefly describe the interrelationship between creep compliance, relaxation modulus, complex modulus, and complex compliance. These four metrics are material properties, and therefore it is expected that these be related to each other for a given material. This is true as long as the material is evaluated within its linear viscoelastic limits (i.e., limits within which the stress and strain response are proportional to each other for a given time history). In other words, if any one of the four properties $E(t)$, $D(t)$,$|E^*|(\omega)$, or $|D^*|(\omega)$ is known, then the other three properties can be determined. For example, the magnitude $|E^*|(\omega)$ and $|D^*|(\omega)$ are inverses of each other at any given frequency and one can easily be obtained from the other. Along the same lines, although the properties $E(t)$ and $D(t)$ are not numerical inverses of each other at any given time, one can be obtained from the other by using appropriate mathematical transformations. Also, properties in the frequency domain such as complex modulus or compliance can be obtained using hereditary integrals from properties measured in the time domain such as creep compliance or relaxation modulus. In fact, hereditary integrals can also be used to obtain the response for any arbitrary loading history if either the relaxation modulus or creep compliance is known. For example, asphalt mixtures (although not asphalt binders) are traditionally evaluated using a resilient modulus test, wherein a haversine pulse with 0.1 s loading time is applied followed by a rest or dwell time of 0.9 s. The response of the material under such a loading history can be calculated (assuming the material is linear viscoelastic) if any of the four basic material properties are known. *In fact, uniaxial forms of the hereditary integrals (Eqs. 2.1 and 2.2) are the most basic constitutive models to obtain stress response over time*

given a strain history or vice versa. For a linear elastic material subjected to uniaxial stresses, the stress is computed as the product of the elastic modulus and the given strain or the strain is computed simply as the ratio of the given stress to the elastic modulus. Equations 2.1 and 2.2 are analogous to this relationship for a time-dependent material. A more detailed description of the hereditary integral and interconversion between these four metrics can be found in Chap. 13.

$$\sigma(t) = \int_0^t E(t-\tau)\frac{\partial \varepsilon(\tau)}{\partial \tau}d\tau \tag{2.1}$$

$$\varepsilon(t) = \int_0^t D(t-\tau)\frac{\partial \sigma(\tau)}{\partial \tau}d\tau \tag{2.2}$$

Finally, readers must note that while this information is being presented in the context of mechanical properties of an asphalt binder, it is also applicable to characterize the basic undamaged properties of asphalt mixtures, which will be discussed in subsequent chapters.

2.5.4 Temperature Dependency

In the previous subsection, we reviewed the methods and metrics that are typically used to characterize the time-dependent stress-strain relationship of asphalt binders. In this section, we will review the temperature dependency of asphalt binders with some introduction of the mathematical tools and methods that are used to model such dependencies.

The creep compliance, relaxation modulus, and complex modulus described in the previous section are very sensitive to the temperature of the binder. For example, it is not unusual for the complex modulus to increase by four or five orders of magnitude as the temperature is gradually decreased from a typical maximum in-service temperature to a typical minimum in-service temperature. Perhaps the most straightforward way to handle the temperature sensitivity of the asphalt binder is to define the basic material properties such as relaxation, creep compliance, and complex modulus as functions of both time (or frequency) and temperature denoted by T, e.g., $E(t,T)$ or $D(t,T)$ or $|E^*|(\omega,T)$. However, several materials including asphalt binders exhibit a behavior referred to as being thermorheologically simple. This provides a more elegant approach to quantify and work with the temperature sensitivity of asphalt binders.

Consider, for example, Fig. 2.19. The figure shows complex modulus, $|G^*|(\omega)$, collected from a frequency sweep test on a binder at five different temperatures. As expected, at any given frequency of loading, an increase in temperature tends to soften the material and consequently reduce the complex modulus. Further observation of the data reveals that there is a smooth overlap of data shifted from different temperatures. If the data at different temperatures were to be horizontally

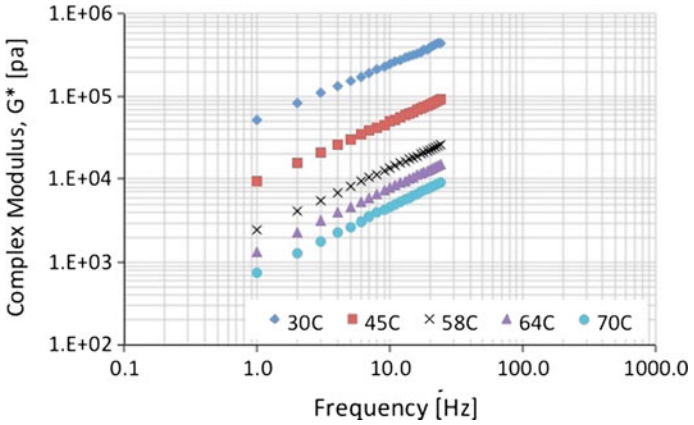


Fig. 2.19 Typical complex modulus of an asphalt binder at five different temperatures plotted against frequency

shifted on the log frequency axis, it is observed that all data eventually fall on a single smooth curve. In this example, let us define 58 °C as our *reference temperature*. Now leaving the data from 58 °C as it is, let us *horizontally* shift data from the relatively higher temperatures to the right and data from the relatively lower temperatures to the left. This horizontal shifting is carried out such that all data fall on a continuous smooth curve as shown in Fig. 2.20. This curve is referred to as the master curve. Note that all data from a given temperature are subjected to the same horizontal shift. Also individual data points from any given temperature

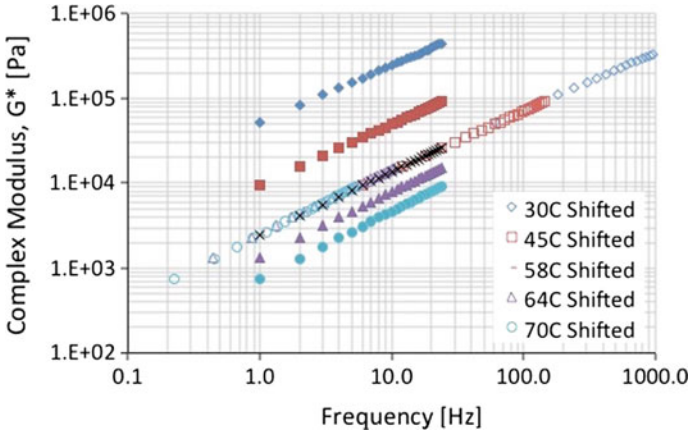


Fig. 2.20 Typical master curve obtained by horizontal shifting of the data from different temperatures

do not necessarily coincide with specific data points at the reference temperature, but rather these points overlap to form a smooth curve constructed using the data at the reference temperature as the basis. Materials that demonstrate this characteristic, wherein the data from different temperatures can be horizontally shifted onto a reference temperature to generate a single continuous smooth curve, are referred to as *thermorheologically simple materials*.

The magnitude of horizontal shift for data from any given temperature, T, is referred to as the *shift factor* and is typically denoted by the term a_T. Based on the method used to shift the graph, it is also easy to see that the horizontal shift $\log(a_T) = \log(f_r) - \log(f_T)$, where f_T is the frequency of any given data point at the test temperature T and f_r is the frequency at which the data point lands on, after being shifted to the reference temperature T_r. The term f_r is referred to as the *reduced frequency*. Finally, the shift factors for different temperatures are used to construct a shift factor versus temperature relationship. Figure 2.21 illustrates this relationship for the data shown in Figs. 2.19 and 2.20. Also note that by definition, at the reference temperature this line passes through an ordinate of zero since there is no shifting of the data involved at the reference temperature.

The master curve obtained based on the process described above can be fitted to an appropriate mathematical form that best represents the shape of the smooth curve obtained after shifting. A common mathematical form used for this purpose is the sigmoidal function shown in Eq. 2.3. The terms α through δ in Eq. 2.3 are shape parameters, and ω_r is the reduced angular frequency. The sigmoidal function has an upper and lower asymptote and is more appropriate for asphalt mixtures than binders. This is because asphalt mixtures tend to have a lower asymptote for the modulus even at very low frequencies (or high temperatures) due to the interaction between aggregate particles. A model that is more common in the area of asphalt binders is the Christensen-Anderson model or a generalized form

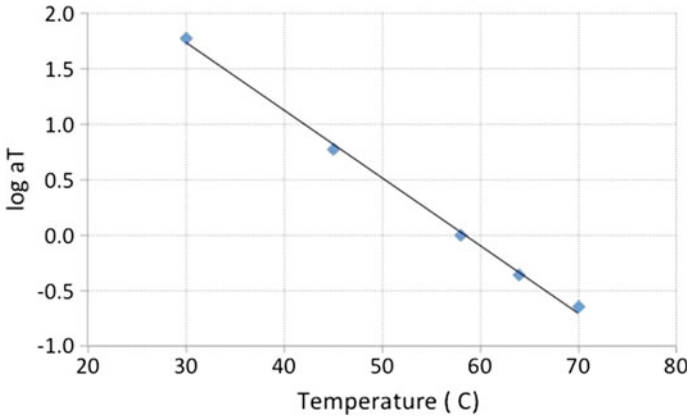


Fig. 2.21 Shift factor for time–temperature superposition derived by shifting data shown in Fig. 2.21

Christensen-Anderson-Marasteanu (CAM) model shown in Eq. 2.4. The parameters G_g is related to the glassy modulus of the binder, ω_c is the cross-over frequency (the frequency at which the loss and storage modulus are the same), and w and v are fitting parameters.

$$\log|G^*|(\omega) = \delta + \frac{\alpha}{1 + e^{\beta + \gamma \log \omega_r}} \tag{2.3}$$

$$|G^*|(\omega) = G_g \left[1 + \left(\frac{\omega_C}{\omega_r}\right)^v\right]^{-w/v} \tag{2.4}$$

Similar to the master curve, the shift factor versus temperature relationship can be fitted to any appropriate mathematical function that best describes the data. For example, researchers in the area of asphalt binders and mixtures have typically used polynomial models. However, it is important to briefly discuss a couple of forms of this relationship that have received a lot of attention in the literature (on asphalt and polymeric materials). First is the WLF equation named after its discoverers (Williams, Landel, and Ferry) and shown in Eq. 2.5. The terms C_1 and C_2 are fitting or shape parameters. The term T_g is the glass transition temperature, which is also used as the reference temperature. Broadly speaking, the glass transition temperature is the temperature below which the material behaves as a solid elastic material (the term glassy behavior is used to represent a solid amorphous material) and can be measured using techniques such as dilatometry, differential scanning calorimetry, or by conducting a temperature sweep using a dynamic shear rheometer. The WLF equation was extensively used by polymer researchers, because the parameters C_1 and C_2 were almost similar for several different polymers with values of approximately 17.4 and 51.6, respectively. Aklonis briefly describes the rationale for the generality of this model based on the concept of free volume. Another model that is also used to describe the shift factor versus temperature relationship is the Arrhenius model shown in Eq. 2.6. In this equation, the parameter k is obtained from fitting the data. The Arrhenius model has been shown to work better than the WLF model for temperatures below the glass transition temperature.

$$\log a_T = \frac{-C_1(T - T_g)}{C_2 + T - T_g} \tag{2.5}$$

$$\ln a_T = k\left(\frac{1}{T} - \frac{1}{T_r}\right) \tag{2.6}$$

The master curve along with the shift factor versus temperature relationship can be used to determine the modulus of the material for any given combination of frequency and temperature. As an example, consider a scenario where one is interested in the complex modulus at 6 Hz and 40 °C for the binder shown in Figs. 2.19 and 2.20. In this case, the first step would be use the model from

Fig. 2.21 to determine the shift factor $\log(a_T)$ at 40 °C, which is approximately 1.2 in this case. This shift factor would then be used to obtain the reduced frequency f_{Tr} based on the frequency of interest f_T using the expression described above; in this case, the reduced frequency would be 95 Hz. Phenomenologically, this means that the complex modulus of the binder at 95 Hz and reference temperature (58 °C in this case) would be the same at the complex modulus at 6 Hz and temperature of interest (40 °C in this case). Note that since the reference temperature is higher than the temperature of interest, the reduced frequency will have to be increased to compensate for the reduced stiffness associated with the higher temperature. The reduced frequency can now be used with the master curve (Fig. 2.20) to obtain the complex modulus of approximately 7×10^4 Pa. The convenience of using this approach is that once the mathematical forms for the functions shown in Figs. 2.20 and 2.21 are known, and the complex modulus at any given temperature and frequency can easily be computed using a simple expression.

Despite, which does not work well at temperatures substantially below the glass transition temperature, it is instructive to understand the derivation of the WLF equation. Aklonis et al., 1972, take us through the derivation. To begin with, the Doolittle equation for viscosity of a liquid (Doolittle 1951) is used to give an expression of viscosity based on two constants, A and B. In the Doolittle equation (shown below as Eq. 2.7), η refers to viscosity, V is the total volume of the system, and V_f is the free volume of the system.

$$\ln \eta = \ln A + B\left(\frac{V - V_f}{V_f}\right) \tag{2.7}$$

This equation demonstrates that viscosity is intimately connected to mobility and mobility to free volume; as free volume increases, viscosity rapidly decreases. Rearrangement of Eq. 2.7 gives:

$$\ln \eta = \ln A + b\left(\frac{1}{f} - 1\right) \tag{2.8}$$

where f is the fractional free volume, V_f/V. With the assumption that above the glass transition (normally used as the reference temperature), fractional free volume increases linearly, one can state:

$$f = f_g + \alpha_f\left(T - T_g\right) \tag{2.9}$$

where f is the fractional free volume at T, any temperature above T_g, f_g is the fractional free volume at T_g, and α_f is the thermal coefficient of change of the fractional free volume above T_g. If we then rewrite the Doolittle equation in terms of Eq. 2.9, we see that with some algebraic manipulation:

$$\log \frac{n(T)}{\eta(T_g)} = \log a_T = -\frac{B}{2.303 f_g}\left(\frac{T - T_g}{\left(\frac{f_g}{\alpha_f}\right) + T - T_g}\right) \tag{2.10}$$

The concept of increased mobility as a function of temperature and free volume is clear and evident, and one can now understand the form of the celebrated WLF equation. In Eq. 2.10, $C_1 = \frac{B}{2{,}303 f_f}$ and $C_2 = \left(\frac{f_g}{\alpha_f}\right)$.

2.5.5 Age Dependency

In addition to the time and temperature dependency, properties of the asphalt binder are also dependent on the age of the binder. Typically, the term aging in the context of asphalt binder during its service life refers to oxidative aging that was discussed in the previous subsections. Oxidative aging results in a change in the chemical composition of the asphalt binder, which in turn results in a change in its rheological properties. In general, oxidation results in the production of more polar molecules, which tend to associate more than the unaged binder. As a result, the stiffness of the binder increases with oxidative aging. This aspect must be carefully considered while using models and approaches to characterize the performance of asphalt binders and mixtures over the long term. In general, properties of the aged binder are measured at discrete aging stages (unaged, short-term aged, or long-term aged). There has been some work, although mostly from a research point of view at the time of this writing, to extend to time-temperature master curve concept to incorporate aging with shift-factors for aging.

2.5.6 Typical Measurement Techniques

In this section, we will discuss the instruments and hardware typically (but not exclusively) used to measure the mechanical properties of the asphalt binder. The dynamic shear rheometer (DSR) and the bending beam rheometer (BBR) are often used to characterize the mechanical properties of asphalt binders.

The DSR can be used to obtain both time- and frequency-dependent properties at different temperatures, i.e., creep compliance, relaxation modulus, and complex modulus. There are two different geometric configurations that are typically used to test asphalt binders with a DSR. The first and the most common is a cylindrical specimen, approximately 1–2 mm thick, between two parallel plates with the top plate being subjected to torsion and the bottom plate being fixed. The choice of the diameter of the plate and specimen is dictated by the torsion capacity of the instrument as well as the test temperature and aging condition of the binder. Long-term aged or significantly oxidized binders at intermediate or low temperatures (approximately 10–20 °C) are considerably stiffer than unaged or short-term aged binders tested at high temperatures (approximately 50 °C or above).

Consequently, a smaller diameter plate (8 mm) and thicker sample is more appropriate to test the former. Researchers have recently experimented using plates and specimen as small as 4 mm in diameter to test asphalt binders at much lower temperatures. For a parallel plate geometry, the shear strain rate at any point is given as $\dot{\gamma} = r\Omega/h$, where r is the radius from the axis of rotation, h is the height of the specimen, and Ω is angular speed of the top plate. Since the shear rate depends on the distance from the axis of rotation, the specimen is subjected to different strain rates and stresses at different radial positions within the specimen. As a result, it is important that this method be used to characterize a material within its linear viscoelastic limit (more on this discussed later). In some cases, it may be possible to use the method in nonlinear range, but the discussion of such conditions is beyond the scope of this chapter. For materials that exhibit significant nonlinear response, cone, and plate geometry is recommended. Figure 2.22 illustrates the parallel plate and (less commonly used) cone and plate geometries.

Consider a binder specimen that is to be evaluated after long-term aging and/or at low temperatures (typically below 0 °C). Under these conditions, the stiffness of the binder is typically much higher compared to the stiffness of the short-term aged binder or binder at elevated temperatures. Also, consider the fact that typical DSRs that are commercially available have a limited torque capacity. Consequently, if one were to use the typical 8 or 25 mm diameter and 1or 2-mm-thick sample that is long-term aged particularly at low temperatures with the DSR, the instrument will not be able to induce adequate deformation to the specimen required to make repeatable measurements. For such scenarios, some researchers have tried using a very small diameter specimen (e.g., 4 mm) with the DSR. While theoretically the use of a smaller diameter reduces the torque required to achieve a certain level of deformation, other factors must be carefully considered. For example, if the compliance of the test specimen is in the same order of magnitude as the compliance of the testing frame, then depending on the location of the displacement transducers, the instrument may record cumulative deformations that occur in the loading frame and the test specimen. One must always be on the lookout for such artifacts while making such measurements. Owing to the above limitations, the binder industry typically employs a different device, the bending beam rheometer or BBR, to make measurements for long-term aged binder at low temperature.

The bending beam rheometer or BBR is comprised of a loading frame immersed in a temperature-controlled bath of typically ethylene glycol mixed with water

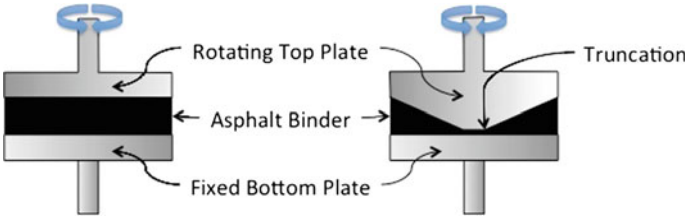


Fig. 2.22 Schematic of the parallel plate and cone and plate geometry for testing with a DSR

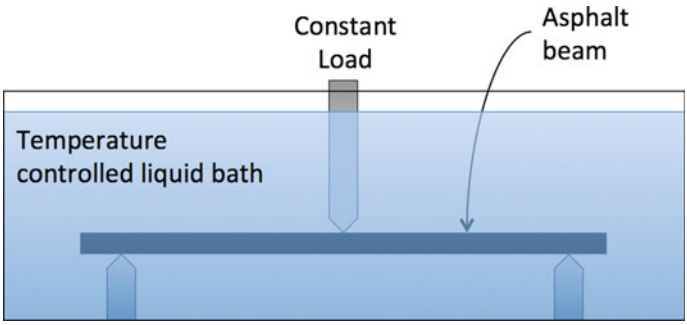


Fig. 2.23 Schematic of the bending beam rhoemeter

(Fig. 2.23). The load frame is designed to apply a constant load at the center of a beam specimen with rectangular cross section that is simply supported at its two ends. The loading axis also measures the deformation at the center of the specimen over time. Recall the relationship between load P, deformation δ, and stiffness S, for a simply supported rectangular beam with width b, depth d, and span l (Eq. 2.11). Notice in Eq. 2.9 that the stiffness and deflection are expressed as a function of time as $S(t)$ and $\delta(t)$, respectively. This is because of the time-dependent nature of the material. By estimating the change in stiffness of the material over time, it is possible to get some estimate of the ability of the material to relax. A couple of things must be noted with the use of BBR. First, the information collected in the manner as shown in Eq. 2.9. is an indicator of material behavior but not a true fundamental material property (such as creep compliance, relaxation modulus, or dynamic modulus). However, with the appropriate use of boundary conditions and mathematical analysis the data from the BBR can be used to estimate the creep compliance or relaxation modulus of the material. Second, the most common types of BBR currently available are not capable of performing tests in the frequency domain. However, it is possible to replicate this arrangement in a universal testing machine to conduct such and other tests (e.g. Four point bending, notched fracture etc.).

$$S(t) = \frac{Pl^3}{4bd^3\delta(t)} \tag{2.11}$$

At this point, we have discussed some of the basic material properties that are used to characterize a viscoelastic material such as the asphalt binder, i.e., creep compliance, relaxation modulus, and dynamic or complex modulus. We have also discussed the influence of temperature and age on these properties as well as some of the common laboratory methods used to measure these properties (DSR and BBR). However, not all of these properties are measured all the time to characterize asphalt binder for routine use. In the following section, we will briefly discuss the metrics based on these fundamental measures of material behavior that are currently being used to grade asphalt binders.

2.5.7 Desired Binder Properties to Produce Durable Asphalt Mixtures and PG System

The optimal rheological properties of an asphalt binder that will result in a durable mix are dictated by the geographic location and the type of mix. Simply put, asphalt binders cannot be produced using a one-size fits all approach. The most recently developed industry standard, referred to as the Superpave performance grade or PG specifications, requires that the rheological properties of an asphalt binder (with appropriate aging) must meet certain criteria at the highest, intermediate, and lowest service temperatures to prevent premature failure due to rutting, fatigue cracking, or low-temperature cracking, respectively. In summary, an asphalt binder rated as PGXX-YY is capable of being used in asphalt pavements with a maximum pavement temperature of XX and a minimum pavement temperature of -YY. The methodology is briefly described here, and the readers are encouraged to try the exercise questions at the end of this chapter to get a better understanding of the process.

The high-grade temperature XX is determined as the maximum temperature at which the binder fails a high-temperature criterion. In fact, there are two criteria, one each for unaged and laboratory short-term aged (or RTFO aged) binder and the lower of the two is used after rounding to the nearest lower six-degree increment (the increments are in steps of 6° starting from 46 °C). The current criterion for high temperature is a minimum value of $|G^*|/\sin\delta$ to mitigate rutting at the maximum pavement temperature measured on unaged and short-term aged (partially oxidized) binder at 10 rads/second. Here, $|G^*|$ is the magnitude of the complex shear modulus, and δ is the phase angle measured using a DSR (recall discussion related to Fig. 2.18). As of this writing, the industry is also considering to replace the above criterion with a maximum requirement of the nonrecoverable compliance J_{nr} measured at high temperatures as an indicator of the binder's susceptibility to permanent deformation. The nonrecoverable compliance is measured using a creep-recovery test with a DSR (Fig. 2.22) on short-term aged asphalt binders by applying cycles of constant shear stress for 1 s followed by zero shear stress for 9 s. The nonrecoverable compliance is defined as the ratio of the permanent or nonrecoverable strain (difference between residual strain at the end of the recovery period and the strain at the beginning of the creep period for any given cycle) to the applied stress (see Fig. 2.16).

The low-grade temperature YY is determined as the minimum temperature at which the binder fails a low-temperature criterion. In fact, there are two criteria, one based on stiffness and one based on rate of relaxation and the temperature corresponding to the higher of the two is used after rounding to the next highest six-degree increment. The creep compliance of long-term aged (significantly oxidized) asphalt binder at low temperatures is related to the low-temperature cracking resistance of the asphalt mixture. Binders with lower stiffness and higher rate of relaxation will tend to develop lower tensile stresses as the pavement cools. Consequently, an aged binder must meet a maximum stiffness and minimum rate of

relaxation criteria at the lowest expected service temperature. The stiffness and rate of relaxation are measured using the BBR (recall discussion around Fig. 2.23).

Although the grade is set based on the high and low temperatures, the binder must also meet an intermediate temperature requirement. The intermediate temperature is typically defined as 4 + the average of the sum of high and low temperatures in Celsius, although agencies have adopted other variations of this definition. The current requirement in the PG specification for intermediate temperature is a maximum value of $|G^*| \sin \delta$ at the intermediate service temperature to mitigate fatigue cracks, where $|G^*|$ is the complex shear modulus and δ is the phase angle measured on long-term aged (oxidized) binder at 10 rads/second measured using a DSR.

Note that the above tests and criteria are typically used to determine and assign a performance grade or PG to a binder produced in a refinery. Since refineries produce several different PGs, the user is then required to select the appropriate PG binder for their specific project. This is typically done by first estimating the maximum and minimum pavement temperature based on historical air temperature data. In some cases, this may involve "grade bumping" if the traffic speed and/or volume is higher than the typical speeds and volumes used as a basis for the high-temperature grade. Simply put, if the traffic speed is slower than typical (e.g., weigh stations) or traffic volume is higher (e.g., busy interstate highways), then one or two higher grades than the high-temperature grade based on pavement temperature alone are recommended for use.

Readers are also encouraged to work through the examples provided at the end of this chapter to get a better understanding of how the PG is determined for a binder and the selection of the appropriate PG binder for a specific construction project. Readers are also referred to additional reading material for a more detailed description of the standard test methods and criteria used in the PG specification.

2.5.8 Limitations of the PG System

Although the PG specification was a significant improvement over the previous characterization methods such as the penetration grade, there are two major shortcomings of this approach. First, recall that in the subsection on time dependency earlier we had mentioned that the "tests are typically conducted by applying a stress or strain magnitude that is not large enough to induce significant permanent damage to the material". A common criticism of the PG specification is that it is based on the stiffness of the binder at different critical conditions measured using smaller magnitudes of stresses or strains and not strength. While stiffness (even time dependent stiffness) characterizes the load deformation or stress-strain response of the material, it does not characterize the failure limit or capacity of the material to deform prior to failure initiation and propagation. Several studies have shown binders with very similar performance grade often have failure characteristics that are very different from each other. Improvement of the aforementioned tests and

criteria for specification is a subject of several ongoing studies. For example, the nonrecoverable compliance, J_{nr}, discussed in the previous paragraph as a replacement for $|G^*|/\sin\delta$ is a step in this direction, i.e., to evaluate the material for its ability to resist permanent deformation at high stresses.

Second, and the more challenging criticism to overcome, is that rheological properties measured at specific temperature and aging conditions do not fully reflect the performance of the binder in the mixture. This is particularly important for aspects related to binder performance that have to do with its physicochemical interactions with the aggregate or other chemical additives included in the mix. For example, binders with the same PG can have very different chemical compositions and may react differently with different mineral aggregates or additives such as liquid anti-strip agents resulting in very different performance characteristics. These issues will be discussed in more detail in subsequent chapters.

2.6 Properties of Liquid Asphalt Binder

In the previous section, we discussed properties of asphalt binder that are relevant during its service life in an asphalt pavement. In this last short section, we will briefly discuss properties of asphalt binder that are relevant, while the asphalt binder is mixed with aggregates in an asphalt mixture production plant and being transported to the construction site.

The two most important properties of an asphalt binder that dictate its ability to coat aggregate particles are its viscosity and surface tension. Perhaps a simplified albeit clear explanation of the interplay between these two properties on wetting and coating is given by Wulf et al. (2000): "without going into details surface tension can be regarded as the driving force and viscosity as the resistance of wetting". In the asphalt industry, viscosity of the asphalt binder is routinely used as a measure of its workability during asphalt mixture production and placement. In most cases, standards and specifications dictating the temperatures for mixing and compaction of asphalt mixtures are based on the premise that achieving a certain binder viscosity will ensure that the mixture remains workable. For example, according to the current Superpave specifications mixture production must be carried out when the binder viscosity is within 0.15–0.19 Pa-sec as measured using a rotational viscometer with a specific spindle diameter and rate of rotation (20 rpm). Figure 2.24 illustrates a schematic of the rotational viscometer along with a photograph of a typical device. The working assumption in this case is that at this viscosity the binder is fluid enough to adequate coat the surface of the aggregate particles. It is also recommended that the mixture be placed and compacted when the binder viscosity is between 0.25 and 0.31 Pa-sec measured as before. The working assumption in this case is that in this viscosity range the mixture is workable to be placed, spread, and compacted to achieve target densities, while at the same time it is not prone to segregation.

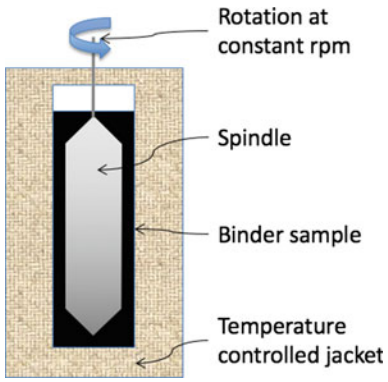


Fig. 2.24 Schematic and photograph of a rotational viscometer

Since each binder has a different viscosity-temperature relationship, in practice, the recommended viscosities need to be translated into production and compaction temperatures. Typically, a rotational viscometer is used to develop a viscosity-temperature relationship for the asphalt binder by measuring the binder viscosity at multiple temperatures. Figure 2.25 shows such a relationship and its use to determine the mixing and compaction temperatures for a typical asphalt binder. However, before we conclude this discussion, it is important to point out a couple of important considerations. First, the shear rate of 20 rpm used in typical viscometer measurements is not representative of typical conditions encountered by the mix during production. Second, mechanisms other than viscosity (e.g., surface tension) also dictate the ability of the binder to wet and coat the aggregate surfaces during production. These two considerations have important implications

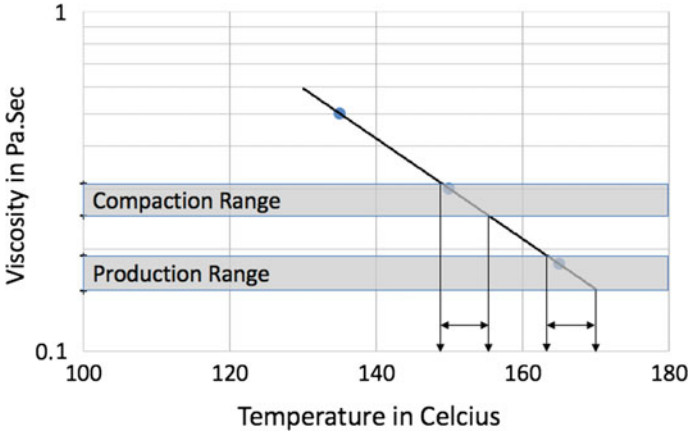


Fig. 2.25 Typical temperature viscosity relationship for an asphalt binder and determination of mixing and compaction temperatures

particularly in the context of the polymer-modified binders and warm mix asphalt, as discussed below.

In the case of polymer-modified binders, the viscosity of the binder is not only temperature dependent but also shear rate dependent. This is referred to as nonNewtonian fluid behavior or sometimes as shear-thinning or shear-thickening behavior depending on whether the increase in shear rate reduces or increases the viscosity, respectively. This behavior is routinely encountered even in daily life. For example, corn starch in water is an excellent example of a shear-thickening liquid. In this context, asphalt binders, particularly polymer-modified binders are highly rate dependent. Therefore, the use of 20 rpm in a viscometer can sometimes result in viscosities and temperatures that are not realistic representations of what happens to the binder in the production facility. In this case, either higher shear rates during viscosity-temperature relationships or experiential data must be used to determine the optimal mixing and compaction temperatures. Readers interested in additional information can refer to the findings from a study focused on this subject (West et al. 2010).

There are several technologies that are used to produce warm mix asphalt or WMA that is produced and compacted at temperatures that are about 20–40 °C below the conventional HMA temperatures. In some cases, it has been observed that the WMA technology does not facilitate mixing by reducing binder viscosity at the lower WMA production temperatures. Instead, other mechanisms such as lubricity (Baumgardner et al. 2012) and the reduction of surface tension (Osmari et al. 2015) allow mixture production. As of this writing, these dimensions of improving workability are still a subject of research, while viscosity of the binder continues to serve a benchmark for its workability.

2.7 Exercises

2.1. You are given two different asphalt binders. Binder A has very high percentage of asphaltenes and Binder B has very low percentage of asphaltenes.

   a. Which of these binders is likely to have a lower stiffness and why?
   b. Two mixtures are prepared using the exact same aggregate gradation and binder content using Binders A and B. Identify with reasoning, the problems that are likely to occur with each of these two mixtures. Also, explain when these problems would become more likely in terms of the pavement age.
   c. If you were told that both binders are from the same source, except that one of them is oxidized, which one of these two binders would you consider to be oxidized and why?

2.2. In static creep experiment, the normal strain on a specimen (material A) subjected to 100 Pa stress was measured over time. The resulting data is

provided in an accompanying data file. Calculate the parameters for the power law model for creep compliance, $D(t) = D_0 + D_1 t^n$. Plot the data and your predicted compliance over time to make sure the two matches reasonably. Do not forget to add the appropriate dimensions for compliance.

2.3. Assume you now have three new time dependent materials, B, C, and D. The compliance of all of these materials follows power law. The properties of these materials are as follows: (8 points)

Material	D_0	D_1	n
A	x	y	z
B	$0.5x$	y	$2z$
C	$2x$	y	$0.5z$
D	x	y	0

Note You will find out the values of x, y, and z from solving the previous question.

Plot the strain for materials A, B, C, and D on the same graph for time t = 0–3 s. Use time steps for 0.1 s for these plots. Answer the following questions:

a. If a person not familiar with time-dependent or viscoelastic materials were to ask, which of the four materials has the highest stiffness, what would be the answer? (Hint: the answer will change with time)
b. Which of the four materials is relatively more time dependent than others?
c. Which of the four materials can be classified as an elastic solid?
d. Does higher "n" value indicate that the material is more time dependent or less time dependent?

2.4. If you are conducting a dynamic test on the materials discussed above and you are also recording the phase angle. Rank the phase angles for the material and test conditions below and provide reasoning for your ranking.

Test number	Material	Frequency (Hz)
1	A	5
2	A	1
3	D	0.1
4	D	25

(*Note* You should be able to estimate the exact phase angle for material D)

2.5. Consider a material that follows power law and the creep compliance is given as $D(t) = 15t^{0.45}$ 1/MPa. A monotonically increasing stress is applied at the rate of 0.02 MPa/sec (i.e., $\sigma(t) = 0.02t$). What is the strain after 2 s? (Hint: Recall the discussion on hereditary integrals and also at some point to make things easier to solve you may have to substitute a variable of integration).

2.6. This exercise is to determine the PG or performance grade of a random binder based on laboratory tests.

Based on a specification system, the high-temperature grade of the binder is defined as follows: The lower of (i) the maximum temperature above which the $|G^*|/\sin\delta$ of the unaged asphalt binder falls below 1.00 kPa and (ii) the maximum temperature above which the $|G^*|/\sin\delta$ of the short-term aged or RTFO-aged asphalt binder falls below 2.2 kPa. Note that, the final temperature must be rounded to the next lower temperature step in six-degree increments starting from 46 C; i.e., the final high temperature will be rounded down and noted as one of 46, 52, 58, 64, 70, 76, or 82.

The low-temperature grade of the binder is defined based on a creep test conducted using the long-term aged or PAV-aged binder as follows: 10 °C below the higher of (i) the minimum temperature below which the stiffness of the binder is higher than 300 MPa and (ii) the minimum temperature below which the rate of relaxation at 60 s is lower than 0.3. Note that, the final temperature must be rounded to the next higher temperature step in six-degree increments starting from −10 C; i.e., the final low temperature will be rounded up and noted as one of −10, −16, −22, −28, or −34 C.

Using the above two criteria, determine the high and low PG grade of the binder if the following measurements were made and provided to you.

DSR with unaged binder			DSR with RTFO-aged binder			BBR with PAV-aged binder		
Temp. in C	|G*| in kPa	Phase angle in degrees	Temp. in C	|G*| in kPa	Phase angle in degrees	Temp. in C	Stiffness in MPa	m-value at 60 s.
64	3.7	79.4	64	6.1	75.9	−24	817	0.214
70	2.0	80.8	70	2.5	78.2	−18	453	0.442
76	1.1	82.2	76	1.1	80.4	−12	251	0.670

For the above exercise, use an interpolation scheme that best fits the data provided as needed. Typically, $|G^*|/\sin\delta$ and temperature follow a straight line on a log-linear scale; similarly, stiffness from BBR and temperature also follow straight line on a log-linear scale. The rate of relaxation or *m*-value is typically found to follow a linear relationship with temperature. Note that in practice, most binder grading specifications also require that a criterion be met at the intermediate temperature for the PAV-aged binder as well as other requirements related to safety and mass loss.

2.7 You need to select a PG binder for the construction of a flexible pavement. The following is some of the information about the construction site.

Longitude 99.43		
Latitude 31.83		
Elevation 528 m		
Low air temperature	Mean = −12 °C	Std. Dev. = 4 °C
High air temperature	Mean = +38 °C	Std. Dev. = 2 °C

Consider that the maximum pavement temperature is at 20 mm below the surface of the pavement. Also, the relationship between the maximum air temperature and pavement temperature is given as follows:

$$T_{pavement} = 54.32 + 0.78T_{air} - 0.0025\,Lat^2 - 15.14\log_{10}(H + 25) + z(9 + 0.61\sigma_{air}^2)^{1/2}$$

The above equation is an empirical model based on data collected from a long-term pavement performance (LTPP) study supported by the Federal Highway Administration (FHWA) of the USA. The term *Lat* refers to the latitude, H is the depth from surface of the pavement in mm, and σ_{air} is the standard deviation in the 7-day mean air temperature in centrigrate. Also for the purpose of this exercise, consider that the minimum pavement temperature is the air temperature at the surface of the pavement. The high-temperature binder grades start at 46 °C in increments of 6 °C and low-temperature grades start at −10 °C decreasing in increments of 6 °C.

a. What the PG binder grade you would recommend if you wanted to achieve 50% reliability in your design? (Hint: recall that z is zero for 50% reliability for high temperature and for low temperature this indicates using the mean value of the temperature).
b. What the PG binder grade you would recommend if you wanted to achieve 98% reliability in your design? (Hint: recall that z is 2.055 for 98% reliability for a normal distribution; for low temperature use, the mean and standard deviation and assuming a normal distribution find the temperature such that the air temperature is above this temperature 98% of the time).
c. The traffic scenario for the project is such that the traffic is slow and higher in volume compared to the typical scenario that was used to develop PG specifications. One high-temperature "grade bump" is recommended for each of these scenarios. What is the final grade of the binder you would recommend?

2.8 You have two binders for which you have obtained a temperature-viscosity relationship. For Binder 1 the mixing temperature range is 150–160 C and for Binder 2 the mixing temperature range is 180–190 C. Based on what you have read, do you think any of these two binders are polymer modified? If so, which one of the binders do you think is polymer modified and what is your recommendation to select a mixing temperature for this binder.

2.9. What are risks of mixing at temperatures higher and lower than the recommended mixing temperature?

Additional Reading

Robertson, R. E. (2000). *Chemical properties of asphalts and their effects on pavement performance*.

Redelius, P. G. (2006a). The structure of asphaltenes in bitumen. *Road Materials and Pavement Design, 7*(sup1), 143–162.
Lesueur, D. (2009a). Evidence of colloidal structure of bitumen. *Advances in Colloid and Interface Science, 145*(1–2), 42–82.

References

Aklonis, J. J., MacKnight, W. J., & Shen, M. (1972). *Introduction to polymer viscoelasticity* (pp. 52–53). Wiley Interscience.
Allen, R. G., et al. (2013). Identification of the composite relaxation modulus of asphalt binder using AFM nanoindentation. *Journal of Materials in Civil Engineering (ASCE), 25*(4), 530–539.
Allen, R. G., Little, D. N., & Bhasin, A. (2012). Structural characterization of micromechanical properties in asphalt using atomic force microscopy. *Journal of Materials in Civil Engineering, 24*(10), 1317–1327.
Baumgardner, G. L., Reinke, G. R., & Brown, J. (2012). *Lubricity properties of asphalt binders used in hot-mix and warm-mix asphalt pavements*. Turkey: In Eurasphalt and Eurobitume Congress.
Chow, T. S. (1980). The effect of particle shape on the mechanical properties of filled polymers. *Journal of Materials Science*, *15*(8), 1873–1888. Retrieved from http://link.springer.com/article/10.1007/BF00550613.
Corbett, L. W. (1969). Composition of asphalt based on generic fractionation, using solvent deasphaltening, elution-adsorption chromatography, and densimetric characterization. *Analytical Chemistry, 41*(4), 576–579.
Demjén, Z., Pukánszky, B. & Nagy, J. (1998). Evaluation of interfacial interaction in polypropylene/surface treated $CaCO_3$ composites. *Composites Part A: Applied Science and Manufacturing*, *29*(3), 323–329. Retrieved from http://www.sciencedirect.com/science/article/pii/S1359835X97000328.
Dickie, J. P., & Yen, T. F. (1967). Marcrostructuare of the asphaltic fractions by various instrumental methods. *Analytical Chemistry, 39,* 1847–1852.
Doolittle, A. K. (1951). Studies in Newtonian flow. I. The dependence of the viscosity of liquids on temperature. *Journal of Applied Physics, 22*(8), 1031–1035.
Elias, L. et al., 2007. Morphology and rheology of immiscible polymer blends filled with silica nanoparticles. *Polymer*, *48*(20), pp. 6029–6040. Retrieved from http://www.sciencedirect.com/science/article/pii/S0032386107007665.
Hubbard, R. L., & Stanfield, K. E. (1948). Determinatio of asphaltenes, oils, and resins in asphalt. *Analytical Chemistry, 20,* 460–465.
Kringos, N. et al. (2009). A thermodynamical approach to healing in bitumen. In A. Loizos et al. (Eds), *Advanced testing and characterisation of bituminous materials* (pp. 123–128). Rhodes, Greece: Taylor & Francis Group, Boca Raton, Florida.
Koots, J. A., & Speight, J. G. (1975). Relation of petroleum resins to asphaltenes. *Fuel, 54*(3), 179–184.
Lesueur, D. (2009b). Evidence of colloidal structure of bitumen. *Advances in Colloid and Interface Science, 145*(1–2), 42–82.
Lesueur, D. (2009c). The colloidal structure of bitumen: Consequences of the rheology and on the mechanisms of bitumen modification. *Advances in Colloid and Interface Science, 145*(1–2), 42–82.
Lin, M. S., et al. (1995). The effects of asphaltenes on asphalt recycling and aging. *Transportation Research Record, 1507,* 86–95.
Loeber, L., et al. (1996). New direct observations of asphalts and asphalt binders by scanning electron microscopy and atomic force microscopy. *Journal of Microscopy, 182*(1), 32–39.
Masson, J.-F., Leblond, V., & Margeson, J. (2006). Bitumen morphologies by phase-detection atomic force microscopy. *Journal of Microscopy, 221*(1), 17–29.

Masson, J. F., et al. (2007). Low-temperature bitumen stiffness and viscous paraffinic nano-and micro-domains by cryogenic AFM and PDM. *Journal of Microscopy, 227*(3), 191–202.

Mortazavi, M., & Moulthrop, J. S. (1993). *The SHRP materials reference library*. SHRP Report A-646. National Research Council. Washington, D.C.

Osmari, P. H., Arega, Z. A., & Bhasin, A. (2015). Wetting characteristics of asphalt binders at mixing temperatures. *Transportation Research Record: Journal of the Transportation Research Board*, *In Press*.

Palierne, J. F. (1990). Linear rheology of viscoelastic emulsions with interfacial tension. *Rheologica Acta*, *29*(3), 204–214. Retrieved from http://link.springer.com/article/10.1007/BF01331356.

Petersen, J. C. (1984). Chemical composition of asphalt as related to asphalt durability: State of the art. *Transportation Research Record: Journal of the Transportation Research Board, 999,* 13–30.

Petersen, J. C., et al. (1993). Effects of physiochemical factors on asphalt oxidation kinetics. *Transportation Research Record, 1391,* 1.

Pfeiffer, J. P., & Saal, R. N. J. (1940). Asphaltic bitumen as colloid system. *The Journal of Physical Chemistry, 44*(2), 139–149.

Ramm, A., Sakib, N., Bhasin, A., & Downer, M. C. (2016). Optical characterization of temperature-and composition-dependent microstructure in asphalt binders. *Journal of Microscopy*, *262*(3), 216–225.

Read, J., & Whiteoak, D. (2003). *The shell bitumen handbook* (5th ed.). London: Thomas Telford Publishing.

Redelius, P. G. (2006b). The structure of asphaltenes in bitumen. *Road Materials and Pavement Design, 7*(sup1), 143–162.

Rogel, E. (1995). Studies of asphaltene aggregation via computational chemistry. *Colloids and Surfaces A, 104,* 85–93.

Schmets, A., et al. (2009). First-principles investigation of the multiple phases in bituminous materials: the case of asphaltene stacking. In A. Loizos et al. (Eds.), *Advanced testing and characterisation of bituminous materials* (pp. 143–150). Boca Raton, Florida: Taylor & Francis Group.

Schmets, A., et al. (2010). On the existence of wax-induced phase separation in bitumen. *International Journal of Pavement Engineering, 11*(6), 555–563.

Siegmann, M. (1950). Manufacture of asphaltic bitumen. In J. P. Pfeiffer (Ed.) *The properties of asphaltic bitumen* (pp. 121–154). Amsterdam: Elsevier.

Speight, J. G. (1999). *The chemistry and technology of petroleum*. Springer.

Traxler, R. N., & Romberg, J. W. (1952). Asphalt, a colloidal material. *Industrial & Engineering Chemistry*, *44*(1), 155–158. Retrieved from http://dx.doi.org/10.1021/ie50505a045.

Vasconcelos, K. L. (2010). *Moisture Diffusion in Asphalt Binders and Fine Aggregate Mixtures. Doctoral dissertation*, Texas A&M University

Verghese, N. E., & Lesko, J. J. (1999). *Fatigue performance: The role of the Interphase* (pp. 336–348). New York, NY: CRC Press.

West, R. C., Watson, D. E., Turner, P. A., & Casola, J. R. (2010). *Mixing and compaction temperatures of asphalt binders in hot-mix asphalt*. Washington D.C.

Wulf, M., Uhlmann, P., Michel, S., & Grundke, K. (2000). Surface tension studies of levelling additives in powder coatings. *Progress in Organic Coatings, 38*(2), 59–66.

Zhao, S., et al. (2001). Solids contents, properties and molecular structures for various asphaltenes from different oilsands. *Fuel, 80,* 1907–1914.

Chapter 3
Aggregates

3.1 Introduction

As we discussed in the previous chapter, asphalt mixture is a composite that comprises asphalt binder and mineral aggregates. We also discussed the different attributes of the asphalt binder that are important from the point of view of performance and durability of asphalt mixtures and pavements. This chapter will discuss the different properties of mineral aggregates that are important in the context of asphalt mixture and pavement design.

Mineral aggregates (or simply aggregates for brevity) make up approximately 80% of an asphalt mixture by volume and 95% by weight. Despite the large volume or mass proportion, mineral aggregates typically make up 50% or less of the total cost of the materials used to produce an asphalt mixture. The cost of mineral aggregates includes the cost of producing aggregates of different sizes, typically from an opencast mine, and the cost of transporting the aggregates to the hot mix asphalt production site. Owing to the relatively high specific gravity of aggregates, the cost of transporting aggregates from the mine to the asphalt mixture production site can increase significantly with the haul distance. Consequently, for economic reasons, aggregates are almost always used from the nearest available location. This creates an interesting problem that couples engineering and economics because in most cases engineers will have to develop a mixture and pavement design that can accommodate the characteristics of locally available aggregates. For example, in some cases locally available aggregates may not have high durability. In this scenario, the design engineer may want to avoid the use of a gradation that increases direct point-to-point contact of large aggregate particles and results in crushing of the aggregate. Another example is that of locally available aggregates having a surface chemistry and microstructure that makes it difficult to bond with certain types of asphalt binders. In such cases, the design engineer may have to resort to

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_3

using asphalt binders refined using specific crude oil sources and/or the use of anti-strip agents.

This chapter presents an overview of several different attributes of aggregates that are important from the point of view of designing a durable asphalt mixture as well as a brief description of the methods that can be used to quantify each attribute. As in the case of the previous chapter, a significant amount of information covered in this chapter is based on the research conducted during and following the Superpave Highway Research Program (SHRP). Note that more emphasis is placed on the properties and the relevance of these properties in the performance of asphalt mixtures; details of the test methods to measure various properties are intentionally avoided in order to focus more on the intent and implications. Adequate references are provided throughout the text for a more detailed description of specific test methods.

3.2 Sources of Mineral Aggregates

Depending on the source, aggregates can be classified as natural or artificial. Artificial aggregates typically refer to some kind of an industrial waste product such as slag or in some cases crushed Portland cement concrete. Most aggregates used in asphalt pavement construction are natural aggregates that are crushed and/or processed by aggregate producers into different sizes. Natural aggregates are a non-renewable resource, and owing to the expensive hauling costs, aggregates are often mined and produced in the vicinity of urban areas. Therefore, any effort towards designing and constructing asphalt pavements in an environment friendly manner must consider making the most efficient use of aggregates. One such strategy is to use reclaimed asphalt pavements (RAP), which also serve as a source of mineral aggregates for the production of asphalt mixtures. The subject of RAP in asphalt mixtures is more involved and will be discussed in a subsequent chapter.

Most aggregates are produced by blasting large pieces of rock in an opencast mine and crushing the rocks to different sizes or by collecting and crushing unconsolidated rock fragments from large natural deposits. The latter form of aggregate is generally referred to as gravel. The process of crushing typically reduces the aggregates to different size fractions that are separated by sieving and stacked in what is referred to as stockpiles. Each stockpile typically has aggregates predominantly from one range of sieve size mixed with a small percentage of aggregates from smaller sieves.

Engineers dealing with the production of asphalt mixtures typically tend to refer to aggregates based on their mineral composition. The terms granite, basalt, limestone, and gravel are commonly used. However, this general and overly simplistic vernacular can oftentimes mask important differences in aggregate properties

that are very relevant to estimate the performance of asphalt mixtures. The following paragraphs try to highlight this issue while establishing a more formal understanding of the different aggregate types used in asphalt mixture production.

Most of us are familiar with the three common types of rocks: igneous, sedimentary, and metamorphic. Igneous rocks are typically formed from the cooling and solidification of molten magma (e.g., granite), whereas sedimentary rocks are formed from the deposition of mineral or other particles (e.g., limestone). Metamorphic rocks are rocks that have been transformed under the action of high temperature and pressure. Intuitively, it is apparent that igneous rocks are more likely to be harder and tougher compared to sedimentary rocks. Other factors related to the origin of a rock also have a lot to do with the way it tends to ultimately perform when used in an asphalt mixture. For example, the rate at which an igneous rock cooled during its formation can significantly affect its ability to adhere to the asphalt binder or provide adequate friction resistance, which will be discussed later.

In the context of description of aggregate types, it must be noted that the terms granite, limestone, etc. are very generic lithological descriptions of the aggregate type. Aggregates having the same lithology can have very different mineral composition and performance. For example, the term "gravel" can refer to aggregates that may have very different geologic origins and mineral composition. As another example, consider two different aggregates, both referred to as granite from specific locations in two different states of Georgia and California in the USA. Although both aggregates are referred to as granite, they have different mineral compositions and physical structures, and as a result they can behave very differently when used with the same asphalt binder. In this example, one of the "granites" is predominantly quartz and potassium feldspar, whereas the other "granite" is predominantly quartz and plagioclase. Also the former has a specific surface area (measured using a method called BET adsorption) of approximately 0.2 m^2/g, whereas the other has a specific surface area of approximately 1.6 m^2/g. Specific surface area is nothing but the total surface area of 1 g of the aggregate of a given size fraction. In other words, one of these granites has about 8 times the surface area of the other. In the context of mixture performance, the aggregate with the higher surface area will have much better mechanical adhesion and bonding with the asphalt binder. In fact, the granite with the smaller surface area is known to have problems related to moisture damage and poor binder adhesion. This example just demonstrates that generalizations, if at all, regarding the expected behavior of a given aggregate based on its lithological description must be made with caution. Figure 3.1 illustrates a photomicrograph of three different aggregate surfaces. The figure clearly shows the diverse range of mineral crystals on the surface of a granite particle with a relatively smooth texture, the rough surface texture of a basalt particle with a localized intrusion of clay, and the surface of a limestone particle with intermediate roughness and visible traces of the organic material that led to the formation of the mineral.

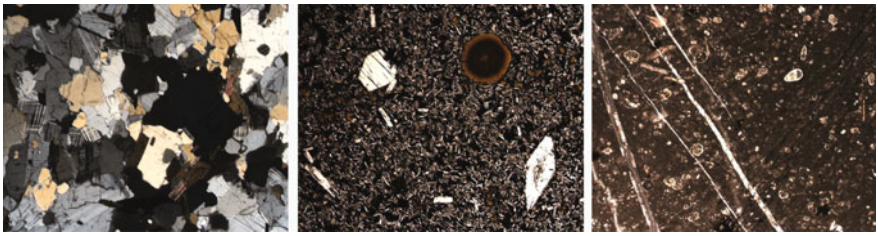


Fig. 3.1 Photo micrographs (2.5X) of three different aggregate surfaces: granite, basalt, and limestone (*left to right*); granite is under cross polarized light to highlight mineral structures (With permission from TRB through the National Academy of Sciences, Little and Bhasin 2006)

3.3 Physical Attributes of Mineral Aggregates

3.3.1 Size and Gradation

Owing to the weight and percentage of aggregates used in typical asphalt mixtures, it is cost prohibitive to mine and haul aggregates from long distances. As a result, engineers have to deal with the good or poor inherent qualities of locally available aggregates. One aspect of aggregate selection and mixture design that engineers have control over is the selection of aggregate size and gradation. Note that aggregate size and gradation in turn dictates the aggregate structure within an asphalt mixture and the ability of the aggregate structure to distribute stresses internally and transfer loads. Typically, the process of selecting an aggregate size and gradation begins with the pavement application. For the construction of asphalt pavement layers, the maximum aggregate size selected is typically one-third the thickness of the pavement layer. The term "maximum aggregate size" is somewhat ambiguous. For example, some states in the USA define maximum aggregate size as the smallest sieve size through which 100% of the aggregates pass and nominal maximum aggregate size as the next smaller sieve size. The Superpave mix design method developed by the SHRP refers to the nominal maximum aggregate size as one sieve size larger than the first sieve to retain more than 10% of the aggregates. In any case, readers are encouraged to make themselves familiar with the definitions that are relevant for their condition. A more interesting attribute is the gradation that defines the internal aggregate structure of the asphalt mixture.

Aggregates of different sizes can typically be graded in four different ways to achieve different internal structures for specific applications:

1. Dense graded.
2. Gap graded or stone matrix asphalt (SMA).
3. Open graded or permeable friction course (PFC).
4. Single size or uniform.

Figure 3.2 shows a picture of a limestone aggregate in different size fractions, and Fig. 3.3 compares the four different gradations that will be further discussed in the following section.

Fig. 3.2 Photograph showing typical aggregate particles of different size fractions that are combined to achieve different gradations; picture below is referred to as fines or filler and corresponds to particles finer than 75 micrometer or passing number 200 sieve (all photographs are at the same scale) (Photograph by A Bhasin)

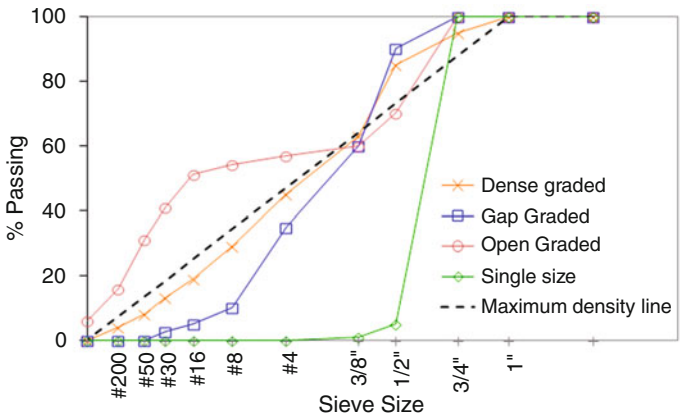


Fig. 3.3 Typical aggregate gradations shown along with the maximum density line for 1-inch. maximum aggregate size

3.3.1.1 Dense Graded Aggregate Structure

Dense graded aggregate structures are perhaps the most commonly used gradation to construct the structural or the primary load transferring layer in a pavement structure. As the name suggests, the objective in this case is to design an aggregate structure that achieves the maximum aggregate density possible. The rationale for using such mixtures is simple; since aggregates are generally orders of magnitude stiffer than the typical asphalt binder used to hold these aggregates together, by designing an aggregate mixture with the densest possible particle packing it should be possible to transfer loads more efficiently through the pavement structure. A secondary reason why a dense gradation would be preferred is an economic one. Dense particle packing would imply very little void space, which would in turn imply reduced amounts of the relatively much more expensive asphalt binder that

would be required to fill this space. The next question then is how does one achieve the maximum packing of aggregate particles of different sizes in a given volume?

One of the earliest reported studies to investigate packing of aggregate particles was by Fuller and Thomson (1907). Fuller and Thompson conducted a very thorough and systematic study on the factors that affect the strength of cement concrete and reported their findings in a paper entitled "The laws of proportioning concrete." One of the findings from their paper was that for a given maximum aggregate size, maximum strength was achieved when the particle gradation resulted in the maximum density. They stated that this was possible when the curve for percent passing versus sieve size resembled "*a parabola, which is a combination of a curve approaching an ellipse for the sand portion and a tangent straight line for the stone portion. The ellipse runs to a diameter of one-tenth of the diameter of the maximum size of stone, and the stone from this point is uniformly graded.*" This finding was later reduced to its mathematical form commonly referred to as the 0.45 power line or the maximum density line:

$$CPP = \left(\frac{d}{D}\right)^{n} 100 \tag{3.1}$$

where, CPP is the cumulative percentage passing, d is the particle size, D is the maximum particle size, and n is 0.45. Interestingly, although several studies and publications attribute Fuller and Thompson for this formulation with $n = 0.5$, their original paper only treated the maximum density line to be a combination of an ellipse and a tangent as mentioned above. In any case, the 0.45 power curve has been in use as the basis to design a dense graded asphalt mixture for many decades now. Figure 3.4 illustrates the 0.45 power line or maximum density line. Note that the maximum density line in the illustration is a straight line on account of the fact

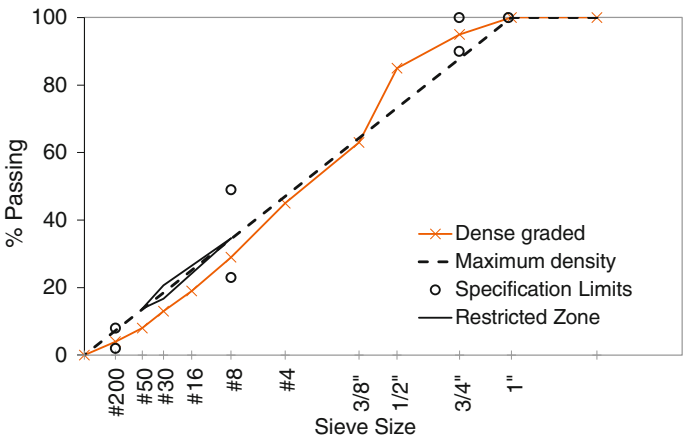


Fig. 3.4 Typical aggregate gradations shown along with the maximum density line for 1-inch. maximum aggregate size

that the x-axis is not linear but rather on 0.45 power scale. The figure also shows the aggregate gradation for a typical dense graded mixture.

Two considerations must be made while selecting an aggregate gradation for a dense graded mix. First, the aggregate gradation must not exactly follow the maximum density line. In other words, for an asphalt mixture it is important to achieve a gradation that has some air voids. This is on account of the fact that a typical dense graded asphalt mixture also includes typically 4–7% air voids by design. Although intuitively one may argue that the addition of air voids may weaken the mix, the air voids are necessary to provide room within the mixture for the asphalt binder to expand at high temperatures. Reducing or eliminating the presence of these air voids can often result in a distress referred to as flushing or bleeding. In this type of distress, the asphalt binder expands and in combination with densification due to the wheel loads, the binder oozes out of the asphalt mixture onto the top of the pavement surface. This creates a potentially dangerous and slick driving surface. The next question would then be that if one were to avoid exactly following the maximum density line, then how much deviation from this line would still result in an acceptable dense graded mix? This brings us to the second aspect that has to do with the controls exercised during the selection of an aggregate gradation.

In order to help select an aggregate gradation by blending aggregates of different sizes, most agencies provide specification limits or control points for dense graded aggregates. Figure 3.4 shows a typical example of the control points from Superpave specification along with the maximum density line and one possible aggregate gradation. Note that the control points carefully force the aggregate gradation to stay above the maximum density line for the larger aggregate size (19 mm in this example) and then again force the gradation to stay below the maximum density line for the aggregate fines (#200 in this example). This ensures that the aggregate gradation does not closely follow the maximum density line and results in the creation of air voids in the asphalt mixture. Another thing to note from Fig. 3.4 is the region marked between the number 50 and number 8 sieve sizes, referred to as the restricted zone. The restricted zone was introduced as a part of the Superpave mix design based on research conducted during the SHRP. The consensus during the SHRP research was that the aggregate gradation should be required to remain within the specification limits but not pass through the restricted zone. The rationale for this was that packed aggregate fines following this maximum density line might create dense particle-to-particle contacts with a lubricating effect and result in a weak or unstable mix. However, certain state highway agencies reported having historical experience with designing successful mixes that went right through this restricted zone. More recent research has also raised the question on the validity of the restricted zone. As of this writing, there is no consensus on the validity or significance of the restricted zone.

It is evident from the above explanation and Fig. 3.4 that even with the maximum density line and control points, it is possible to develop an infinite number of different aggregate structures that can all be claimed to be dense graded as per our previous definition. One of the methods to help identify a dense graded aggregate structure using a more systematic approach is referred to as the “Bailey method”

named after an engineer from the Illinois Department of Transportation (Vavrik et al. 2001). This method also has the advantages associated with achieving a more optimal level of voids in mineral aggregates or VMA, which will be discussed in a later chapter on asphalt mixture design. The selection of an aggregate gradation based on this method can be summarily described as follows.

The coarse and fine aggregate fractions are not defined based on an arbitrary cutoff size. Rather coarse aggregates are defined as aggregates that create voids when packed together and fine aggregates are aggregates that fill these voids. Therefore, the size that differentiates coarse from fine aggregates is largely dependent on the maximum aggregate size of the mix. Also, the sieve size that differentiates between coarse and fine aggregates is referred to as the "primary control sieve" or PCS for short. The PCS is defined as 0.22 times the nominal maximum aggregate size. Note that the choice of this factor 0.22 is not arbitrary, rather it is related to the diameter of the circle that can be drawn in a void created by three closely packed circles.

The most important feature of the Bailey method is that once the PCS is determined, the volume of coarse aggregates is not determined by arbitrarily selecting a point near the maximum density line. Rather, a sample of dry coarse aggregates is loosely packed in a container to determine the loose unit weight (LUW). The procedure outlined in AASHTO T-19 (AASHTO 1997) can be used to do this more systematically. Once the loose unit weight is determined, the percentage of coarse aggregates in the gradation is then set as approximately 95–105% of the volume occupied by the aggregates from this exercise. Once the volume or percentage passing for the coarse aggregate fraction is determined, the process is repeated for the fine aggregate fraction. In fact, the fine aggregate fraction is broken down into two ranges using a secondary control sieve (SCS) defined as before as 0.22 times the cutoff used for the primary control sieve. The volume at the secondary control sieve is proportionally determined using the same process as before. Readers are strongly encouraged to review some of the reports with working examples as cited in the additional reading material to familiarize themselves with this method. Several practitioners and industry experts have agreed that this approach is a more systematic way to arrive at an aggregate gradation as opposed to selecting any arbitrary line close the maximum density line within the control points (Vavrik et al. 2001; Underwood 2011) (Fig. 3.5).

3.3.1.2 Gap Graded or Stone Matrix Asphalt

Gap graded mixes, also known as stone matrix asphalt (SMA), are also commonly used in the design of asphalt mixtures for the main structural layer of a flexible pavement. Unlike the dense graded aggregate structure where the primary load transfer mechanism is attributed to particle-to-particle contact between both coarse and fine aggregates, the load transfer in a SMA relies primarily on particle-to-particle contact between coarse aggregate particles. Based on this primary difference, a few considerations are important when selecting and using a SMA mix.

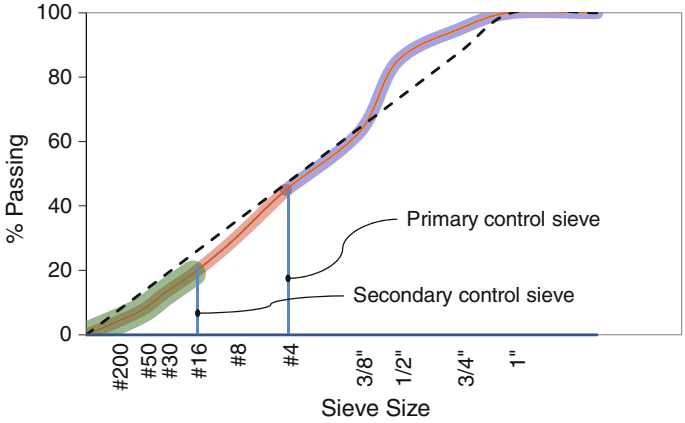


Fig. 3.5 A typical dense graded curve showing the primary and secondary control sieves defined based on the Bailey method and selection of gradation in different regions to develop an overall gradation

First, the aggregate structure in the SMA must maximize the number of contact points between coarse aggregate particles. This is primarily achieved by developing an aggregate gradation with a small percentage of the smaller sized aggregates (Fig. 3.4). As before, the Bailey method can be used as a more systematic method to select an aggregate gradation. However, since the objective is to maximize the contacts between coarse aggregate particles, the selection of the volume of the coarse aggregates is done based on a laboratory measurement of the compacted unit weight (CUW) of the coarse aggregates. This is in contrast to the loose unit weight (LUW) used in the design of the graded aggregate structure.

The second consideration while designing the SMA is that the coarse aggregate structure creates a larger volume of voids that must be filled by a matrix of fine aggregates (that do not participate in the majority of the load transfer unlike the dense mixes) and asphalt binder. The large volume of the binder and available void space in the mix creates a scenario in which the binder and the matrix render the mix susceptible to segregation at elevated mixing and compaction temperatures. In simplistic terms, the binder tends to flow and separate from the mix resulting in a condition referred to as drain down. To prevent this from happening, fibers that tend to prevent the free flow of the asphalt binder are added to the mix. Typically cellulose, mineral, or polyester fibers are used in concentrations that vary from 0.2 to 0.3% by weight of the full asphalt mixture. Long recycled tire rubber fibers have also been used to some extent to achieve this benefit. The efficacy of using such fibers is determined by a simple test referred to as the drain down test. As the name suggests, in this test a sample of the loose asphalt mixture is held in a wire basket at elevated temperatures, and the mass of binder that flows out of the basket over time is measured (Fig. 3.6).

The third and perhaps the most important consideration in the design of the SMA is the toughness of the aggregate particles. In this context, toughness of aggregate

Fig. 3.6 Typical fibers used in the production of SMA mixtures to prevent drain down of asphalt binder (cellulose is shown to the *left* and long tire rubber fiber on the *right*) (Chowdhury et al. 2006).

refers to its ability to withstand high amounts of stress without breaking. Since the SMA relies primarily on coarse aggregate particle-to-particle contact for load transfer, it is expected that the contact points between the coarse aggregate particles will experience very high levels of stress. Consequently, relatively soft aggregates or aggregates with low toughness can break under such high stresses thus reducing the capacity of the mixture to carry loads. Typically, aggregates from igneous sources are considered to be tough, whereas aggregates such as sandstone or limestone may not be suitable for the design of SMA. Engineers should avoid designing a SMA-type aggregate structure if the only available aggregate source does not demonstrate adequate toughness. In fact, it must be noted that most of the aggregate breakage and damage can occur during compaction of the loose mix under a steel wheel roller and long before the roadway is opened to traffic. Figure 3.7 illustrates an X-ray computed tomography image of coarse aggregate

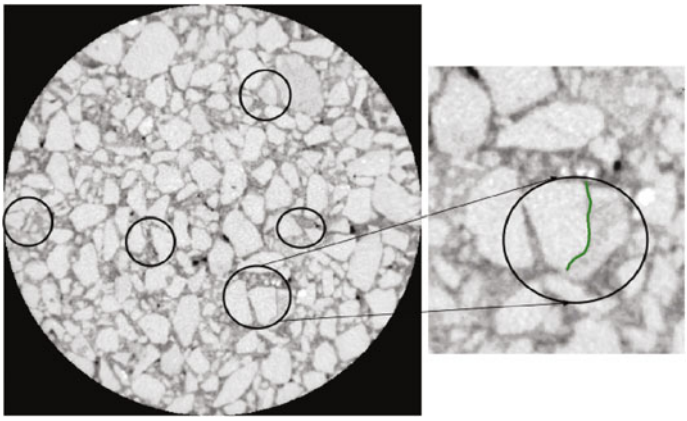


Fig. 3.7 A X-ray computed tomography image that shows low toughness coarse aggregate particles that have broken inside the mix (Gatchalian et al. 2006)

particles that have broken down under the action of the external load. The aggregate source in this particular image was a soft limestone and was not appropriate for a SMA-type mix.

3.3.1.3 Open Graded or Permeable Friction Course

Open graded mixes are also referred to as open graded friction course (OGFC) or permeable friction course (PFC) mixes. There are some minor differences between OGFC and PFC but we are using them interchangeably here in this discussion. The PFC aggregate structure is typically achieved by using only a very small percentage of mid-size aggregates, essentially creating a porous and highly permeable mixture (Fig. 3.4). By design, PFC mixes have much higher air void content (typically over 15%) as compared to dense graded mixes that have air voids in the range of 4–7%. Unlike dense graded or SMA mixes, high porosity renders the PFC mixes to be unsuitable for use as the main structural layer to transmit loads through the pavement structure. Instead, PFC mixes are primarily used on the surface of an asphalt pavement to preserve the existing pavement structure with the following three main benefits: (1) improved drainage from the surface of the pavement, (2) reduced noise, and (3) improved frictional characteristics or skid resistance. For these reasons, PFC mixes are only laid as a thin layer with a thickness that varies from 20 to 30 mm.

PFC mixes provide potential safety benefits over other surfacing materials under wet conditions in terms of drainage to prevent hydroplaning and to improve the skid resistance of vehicles. They reduce splash and spray—the most often cited reasons for their popularity by road users. On wet pavements, PFC reduces headlight glare from oncoming vehicles and improves nighttime wet weather visibility. Several agencies have reported reduction in wet weather accidents with the use of PFCs. Interestingly, there has also been some indication that the benefits associated with the use of PFCs may be negated due to increased speeds as a result of better visibility, the increased confidence exhibited by drivers on porous asphalt, and the perceived safety of PFCs in wet conditions. PFCs also attenuate noise at the pavement–tire interface and have also been shown to provide additional environmental benefits by reducing the charge in pollutants in storm water runoff.

All of the benefits mentioned above are contingent on the porosity of the layer and would be short-lived if clogging occurs in the porous structure of the mix. The design of a well-performing and long-lasting PFC mix is not trivial. For example, clogging can be prevented by designing an aggregate structure that will result in increased porosity. This would allow smaller debris particles to be essentially washed off instead of clogging the pores. However, a larger volume of air voids and a larger pore structure would essentially weaken the mix and reduce its durability. Therefore, the design of a PFC mix requires careful attention in optimizing the pore structure and permeability versus durability. Also, owing to the high void and binder content, the production and placement of PFC mixes faces challenges such

as segregation and drain down similar to the SMA mixes. These problems are addressed using similar remedial strategies such as incorporating a small percentage of fibers in the mix. Readers interested in a more detailed mix design approach are encouraged to refer to the additional reading material provided at the end of this chapter.

3.3.1.4 Uniform or Single Size Aggregates

A predominantly single-sized aggregate, also referred to as a uniform aggregate gradation, cannot be used to provide a dense or other form of load bearing aggregate structure. Single-sized aggregates are typically used only in surface treatment applications such as chip seals that are designed primarily to preserve the pavement structure and restore the frictional and skid resistance characteristics of the pavement. A typical application procedure involves spraying the pavement surface with hot liquid asphalt binder or an asphalt emulsion followed by spraying a blanket of aggregates. The layer of binder or emulsion serves to seal the surface of the pavement preventing or slowing down the ingress of oxygen and water into the pavement structure through microcracks or other surface defects that might have formed during its service life. The blanket of aggregate particles primarily serves to provide texture and frictional characteristics of the pavement surface and enhance its skid resistance. Therefore, a single-sized aggregate is considered to be most appropriate to create such texture. Figure 3.6 illustrates a typical chip seal being placed on an existing surface. Single-sized aggregates are sprayed on the surface to provide skid resistance (Fig. 3.8).

Fig. 3.8 Single-sized aggregates being sprayed on a pavement surface coated with asphalt emulsion (Photograph by A Bhasin)

3.3.1.5 Summarizing Aggregate Gradation

In summary, it must be recognized that in most cases due to economic considerations engineers are constrained to use locally available aggregates. However, engineers do have the flexibility to design an aggregate structure by choosing an appropriate gradation. The choice of gradation has a direct influence on the ultimate performance of the asphalt mixture.

The dense graded mix is one of the most commonly used aggregate structures for the design of asphalt mixtures that primarily serve to transfer traffic loads to the foundation of the pavement. This aggregate structure closely follows the maximum density line but at the same time ensures that the resulting asphalt mixture has room for air voids within the mix. *The primary load transfer mechanism within such a mix is through particle-to-particle contact between both fine coarse and fine aggregate particles.*

Similar to the dense graded mix, the aggregate structure for the stone matrix asphalt or SMA mixes can be used to efficiently transfer loads through the asphalt layer in a pavement structure. However, unlike the dense graded mix, *the primary load transfer mechanism within a SMA mix is* via *stone-to-stone contact between coarse aggregate particles*. This internal load transfer mechanism makes such mixes very efficient in withstanding deformation under repeated and heavy traffic loads. On the other hand, owing to the stone-to-stone contact mechanism, such mixes must only be considered when engineers have access to high quality and tough aggregates. The relatively higher binder content and use of fibers to prevent drain down increase the overall initial cost of the mix compared to a typical dense graded mix. However, the life cycle cost of SMA mixes can be lower particularly in regions with heavy traffic and higher risk of mixture deformation.

Open graded friction course (OGFC) or permeable friction course (PFC) mixes are used primarily in the form of a thin layer on top of an existing pavement structure. The primary use of such mixes is to improve drainage on the surface of the pavement and consequently improve safety and reduce pavement–tire noise. PFC mixes need to be designed carefully so as to ensure that the mixes have large and connected voids that prevent clogging, reduce maintenance, and improve permeability and yet at the same time are not large enough to adversely affect the durability of the mix. Such mixes are only suitable for warm and wet weather climatic conditions and are not suitable for use on unsound pavement structures prone to rutting or cracking, roads subjected to icy conditions where studded or chain tires are used, or roads where dirt, sand, or debris can be tracked on the surface (e.g., farm and beach roads).

Finally, uniformly graded or single-sized aggregates are not used in the production of asphalt mixtures as such but are commonly used for surface treatment applications such as chip seals.

Also note that in most cases, the selection of an aggregate gradation is guided by control points along the maximum density line specific to the particular mixture type (e.g., dense graded, SMA, or PFC). However, there are an infinite number of gradations that can be defined to fall within these control points. In order to

approach aggregate gradation selection in a more systematic way, the Bailey method is commonly used. In this method, the packing of dry aggregates is measured in the laboratory and used as a basis to define percentage passing of different fractions of coarse and fine aggregates at specific points referred to as the primary and secondary control sieves. The Bailey method has been used to systematically design dense graded mixes, SMAs, and fine mixes.

3.3.2 Cleanliness

It is important to have clean aggregates used in the production of asphalt mixtures. In this context, "unclean" aggregates can be considered as stockpiles of aggregates that are contaminated with organic debris or inorganic materials such as excessive dust or clay in the form of particles or lumps. The presence of unwanted organic materials and debris perhaps has the most deleterious impact on the overall quality of the mixture produced. However, in the case of aggregates produced from opencast mines, this type of debris is easily avoided by removing the top surface of the mining area. Organic debris may be an issue in case of aggregate sources such as loose glacial or river gravels that are collected in the form of boulders or loose particles. Organic debris can easily be avoided by taking care during the aggregate mining operations.

Excess dust (other than clay) that coats larger aggregate particles in an aggregate stockpile can sometimes result from crushing operations with relatively soft aggregates such as certain types of limestones. This kind of excess dust can be a bit problematic in one or two different ways but can easily be rectified. First, if the amount of fines or dust coating coarse aggregate particles is excessive, the fines concentration in the asphalt mixture can be more than what the mix was designed for. In a typical drum-type hot mix plant, these fines will be coated with the liquid asphalt binder at high temperatures and incorporated into the mastic portion of the mixture. These fines would be in addition to what the mixture was designed for (more details on fines content and mixture design will be covered in Chap. 7). As a result, the asphalt mixture may tend to be slightly stiffer compared to the design and as such more susceptible to distresses such as low-temperature cracking. However, the magnitude of this impact can easily be estimated by conducting a wet sieve analysis and can also be easily corrected by adjusting the amount of fines added to the asphalt mixture.

Second, dust on aggregate particles can have an adverse impact on applications other than hot mix asphalt, such as chip seals. For example, in a typical chip seal application used for pavement preservation or restoration of skid resistance (see Fig. 3.6), binder in the form of hot liquid or emulsion is sprayed on the pavement surface followed by a blanket of uniform-sized aggregates (the application rate of both of these materials and design of other forms of chip seals such as cape seals is beyond the scope of this book). Unlike a hot mix plant, in this case, the aggregate

particles are not intimately mixed with the binder but are expected to bond with the binder under the application of pressure from pneumatic tire compactors followed by traffic loads. Dust on aggregate particles can prevent such bonding, especially during the early age of the treatment, and result in aggregate loss. Contractors have reported washing aggregates with water as an effective remedy for such situations when an asphalt emulsion is used with dusty aggregates (Fig. 3.9).

Inorganic fines such as clay particles or even clay lumps in the aggregate source can have a more deleterious impact and are a relatively more difficult to deal with. Individual clay particles typically have a plate-like structure with a very high surface area. Such individual particles rarely exist in nature, instead dry clay typically exists in the form of stacks of these plates. The stacks are formed because of the strong intermolecular force of attraction between the surfaces of these individual plates. However, in the presence of water, which is a highly polar liquid, these secondary bonds break and allow water to permeate between the plate structures resulting in the expansion of the stacked plates. The details of the different types of clays, atomic structures in the mineral, and intermolecular forces between these plates in the presence and absence of water are a subject of a more detailed study, and readers are encouraged to refer to additional reading material for more information. In the present context, suffice it to say that small particles of clay, which in turn comprise stacked plate-like structures, expand significantly in volume in the presence of water and shrink upon dehydration. This mechanism is not only just deleterious for asphalt mixtures containing aggregate particles coated with clay or

Fig. 3.9 A photograph showing aggregate particles placed on an emulsion surface; the dust particles can be seen interfering with bonding with the emulsion (Photograph by A Bhasin)

plastic fines but also broadly explains the vast majority of structural problems associated with foundations on expansive soils.

Two questions arise in the context of aggregates contaminated with clay fines. First, what methods can be used to detect whether such fines are in sufficient proportions so as to result in a deleterious impact on the asphalt mix? Second, what remedial actions can be taken when it is not possible for economical or other reasons to use a different source of aggregate?

To address the first question, there are two tests that are typically used to examine the quality of the aggregates in terms of potential contamination with plastic fines or clay: the sand equivalency test and the methylene blue test. The sand equivalency test is simple and intuitive to understand; a more detailed description of the test can be found in ASTM standard D2419. In summary, this test entails taking a sample of the fine aggregate fraction of the aggregate and thoroughly mixing it with a calcium chloride solution in a graduated cylinder. The sample is allowed to rest for approximately 20 min, and the height of the flocculated clay is measured along with the height of the sand or fine aggregate sediment. The sand equivalent is measured as the percentage of the total height of the flocculated clay and sand to the height of the sedimented sand. The higher the value, the greater the content of plastic fines or clay. Typically, a maximum percentage of sand equivalency can be specified to ensure that the aggregates are not contaminated.

The second method to measure clay contamination is the methylene blue test. The test is described in more detail in ASTM C837. In summary, this is a titration-type test in which a sample of the fine aggregate fraction of the aggregates is mixed with water. A few drops of the methylene blue from an aqueous solution are added to this mix. A small drop of the mix is then dispensed onto a filter paper and examined for the presence of a blue halo. The process is repeated until a blue halo can be observed on the filter paper. The rationale for this procedure is that the methylene blue preferentially adsorbs onto the surface of the clay particles. As more volume of the methylene blue is added (titrant), the absorption continues until all the clay particle surfaces are saturated. At this point, excess methylene blue is available in the solution that shows up in the form of a halo when a drop of the mix is dispensed on a filter paper. Based on this brief explanation, it is clear that the higher the concentration of plastic fines or clay, the higher the volume of methylene blue absorption before a halo can be observed in a test sample.

Finally, it is important to briefly discuss remedial methods that can be used when there is an undesirable concentration of plastic fines in the aggregate. One strategy to avoid the deleterious impact of clay or plastic fines is to chemically separate and/or stabilize the agglomerated plate structures of the clay. One of the most common additives to achieve this is hydrated lime. Calcium ions from hydrated lime displace monovalent cations that render the clay particles susceptible to expansion in the presence of water. This cation exchange also causes flocculation of the plate-like particles. The mechanisms and benefits of adding hydrated lime to remedy the presence of clay in aggregates for asphalt mixtures have been well investigated by several researchers over the past few decades (Bell 1996; Eades and

Grim 1960; Little 1995; Kim et al. 2004). In addition to helping with the clay fines in aggregates, there is also extensive evidence in the literature that demonstrates the multifarious benefits of adding hydrated lime to bituminous materials (Lesueur and Little 1999; Little and Petersen 2005; Plancher et al. 1976; Sung Do et al. 2008). These benefits will be discussed in more detail in a subsequent chapter on asphalt additives (Fig. 3.10).

There are five different methods of adding hydrated lime to bituminous materials (Association 2003). Different states have found different methods of addition as most effective and accordingly have adopted one or more of these methods as a part of their specifications (Little and Epps 2001). In this context, the most important consideration is the method of addition of hydrated lime that mitigates the deleterious action of clay contaminated aggregates. For example, addition of hydrated lime to the asphalt binder in the drum mix will have minimal or no mitigating effect on the presence of clays. However, addition of lime slurry to aggregate or marination of lime-treated aggregates may be more effective to treat clays present in the aggregate. In addition to hydrated lime, other chemical additives have also been used with some success to treat clay contaminated aggregates. For example, wood ash and methylene blue have been used as treatment alternatives, but these materials have shown more success in soil treatment or cement concrete applications. Chapter 5 of this book discusses this issue in more detail.

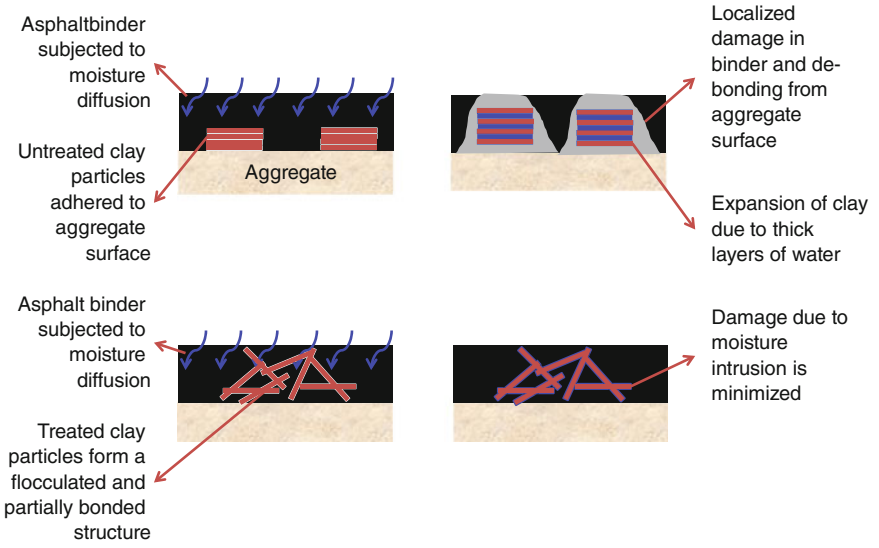


Fig. 3.10 A schematic showing the deleterious impact of clay contaminated aggregate particles as well as the use of chemical additives such as hydrated lime in mitigating this effect

3.3.3 Toughness and Hardness

The terms toughness and hardness of aggregates sound very similar. However, in the context of aggregate properties, toughness and hardness refer to very different properties with different implications in terms of the overall performance of the mix. Specifically, toughness refers to the ability of the aggregate particles to resist fracture under the action of loads, whereas hardness refers to the ability of the aggregate to retain its surface texture over time when subjected to wear. This subsection will discuss the methods that can be used to measure these two characteristics of aggregate particles and implications in terms of asphalt pavement performance.

As mentioned before, toughness refers to the ability of the aggregate to withstand fracture. Aggregate toughness is important in two different scenarios. First, a minimum level of aggregate toughness is required during the placement and compaction of loose asphalt mixtures. Loose hot mix asphalt is transported to the construction site, spread, and compacted to a target density (typically 94% of the theoretical maximum density) using a steel wheel roller. During this process, the binder is in a fluid state and unable to resist any deformation. At the same time, aggregate particles shift and pack themselves by interlocking with each other under the action of the roller. In most cases, the stresses experienced by the aggregate particles are the most intense compared to any point during their subsequent service life. Soft aggregates that do not possess adequate toughness will fracture during this process. This in turn would alter the overall aggregate gradation and negatively affect the ability of the mix to carry loads during its service life.

Second, aggregate toughness is important for SMA (stone matrix asphalt) mixes. As discussed previously, SMA gradation is designed to promote direct load transfer between coarse aggregate particles via contact points. This is also referred to as load transfer from stone-to-stone contact. Soft aggregates with inadequate toughness will tend to fracture under such high contact stresses during the service life of the pavement. Figure 3.7 illustrates an X-ray computed tomography image of an aggregate mixture with fractured aggregate particles. In fact, SMA mixes should be avoided in cases when tough aggregates are not locally or economically available for use.

There are a number of different empirical tests by which the toughness of aggregate particles can be measured. Some of the more common tests are the Los Angeles or LA abrasion test (detailed under AASHTO T96) and the aggregate impact and crushing value tests (detailed under British Standards 812-110 and 112). The LA abrasion test, as it is commonly referred to, is conducted on a sample of coarse aggregate particles (typically retained on No. 12 sieve). The aggregate sample is placed in a cylindrical drum (Fig. 3.11) along with a number of steel spheres (approximately 6–12 spheres each 2 in. in diameter depending on the specification); the steel spheres are referred to as the charge. The drum is then rotated for about 500 revolutions at approximately 30 rpm. This results in the steel spheres and the aggregate particles to impact and colloid with each other.

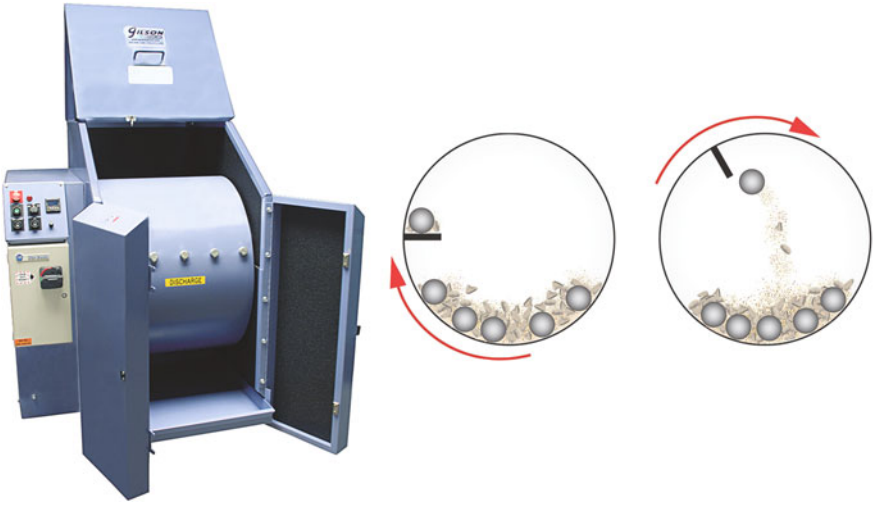


Fig. 3.11 The motorized drum used for LA abrasion or more appropriately impact test (*left*) and a schematic showing degradation of aggregates (*right*) (Image courtesy of Gilson Company Inc., USA)

Aggregates that are not tough will tend to fracture more during this process. After the test is completed, the aggregates are sieved again to determine the mass percentage of the aggregates that were lost. A higher mass loss indicates lower aggregate toughness. Consequently, a maximum mass loss criterion can be specified to ensure that aggregates meet a standard toughness requirement. It is important to discuss one particular aspect about the LA abrasion test. The name "abrasion" is somewhat misleading because the very high mass of the charge (steel spheres) and impact action of the charge with the aggregate particles reflect the ability of the aggregate to resist breakage more than its ability to resist abrasion. Perhaps a better name for this test would be the LA impact test rather than the LA abrasion test.

The aggregate impact and crushing value tests are two other empirical tests that are used to evaluate the toughness of the aggregates. Coarse aggregate samples are used as before and compacted in a metal mold. For the aggregate impact test, a weighted hammer is dropped from a fixed height for a fixed number of times. The aggregate sample is then sieved to determine the mass percentage of aggregates lost due to crushing and used as an indicator of aggregate toughness. The aggregate crushing value is measured in a similar way, except that the aggregates are subjected to a static load instead of an impact load from a fixed height.

Hardness of aggregates refers to the ability of the aggregates to resist abrasion and retain their surface texture. Note that the term hardness used in this context is somewhat different from the standard definition of hardness used in geology and measured on a Mohs scale. In the context of mineral aggregates, this property is particularly important for aggregates that are used in surface mixes and come into contact with the vehicle tires. A more detailed discussion on aggregate texture is

presented in the following subsections. To summarize, aggregate texture plays a very important role in providing skid resistance between the pavement surface and the vehicle tires during braking.

Aggregates used in surface mixes are coated with asphalt binder immediately after construction. However, after opening to traffic, the action of the vehicle tires quickly wears away the coating of the asphalt binder from the aggregate surface. Figure 3.12 shows a photograph of a pavement surface after being opened to traffic for a few weeks. The aggregate surface (light colored limestone) can be clearly seen exposed on the surface. Also notice that in this case, the mix gradation used for the surface layer was a permeable friction course or PFC as discussed earlier. The large pore structure that renders the mix permeable is visible on the surface as well. From this point on the surface of the aggregate plays an important role in providing frictional resistance between the tire and the pavement surface. However, over time the surface of the aggregate may be susceptible to abrasion causing it to loose its surface texture. Aggregates with higher hardness are more resistant to this abrasive action and tend to retain their surface texture longer and consequently provide a safe and skid-resistant driving surface for a longer period of time.

Several different empirical tests are available to measure aggregate hardness. Two more common tests are the British pendulum test and the micro-deval test. In the British pendulum test, the test specimen is prepared by gluing coarse aggregate particles onto a test coupon. The test coupon is then attached to the end of a pendulum that is allowed to swing from a fixed height rubbing over a rubber pad at the bottom point to simulate tire-aggregate wear. Figure 3.13 illustrates a typical

Fig. 3.12 A PFC mix on the pavement surface with exposed aggregate surface a few weeks after it was opened to traffic (Photograph by A Bhasin)

Fig. 3.13 A photograph showing the British pendulum (*left*) and a test coupon before (*top right*) and after (*bottom right*) being swung over a rubber test pad for a specified number of times (British pendulum image courtesy of Gilson Company Inc., USA; Polished coupon: Gatchalian et al. 2006)

British pendulum device along with a test coupon before and after polishing it with several swings of the pendulum. The aggregate wear is clear.

The micro-deval test is also an empirical test that is used to measure aggregate hardness. In many ways, the micro-deval test is very similar to the LA impact test with two exceptions. First, the size of the drum used is much smaller. Second, the steel spheres used as charge are also lighter in weight and smaller. This test is typically conducted in the presence of water. Due to the relatively smaller size of the charge (steel spheres) and lighter weight, the aggregate particles are subjected to more surface abrasion as opposed to an impact breakage compared to the LA impact test. Figure 3.14 shows a typical micro-deval device with a sample of aggregates before and after being subjected to the simulated abrasion in the device (Mahmoud 2007). Figure 3.15 illustrates the loss in texture for select samples of different mineral aggregates. The texture is quantified in terms of a texture index that will be discussed in more detail in the following section. The graph clearly shows that different aggregates not only have different textures to begin with but also have different hardness, i.e., resistance to abrasion. For example, the crushed gravel shows the highest level of hardness with hardly any loss in texture, albeit the texture of this aggregate is very low to begin with. Note that crushed limestones A and B have the same lithology, i.e., limestone, but are from two different sources and have very different texture and hardness characteristics.

A comparison of crushed limestone B and crushed granite also reveals something interesting. Crushed limestone B and crushed granite have very similar and high texture index to begin with and either appear to be suitable for use as an aggregate in surface mixes. On the other hand, crushed granite has higher hardness and thus does not get easily polished over time, while crushed limestone B has lower hardness and tends to get significantly polished over time. As a side note, some mineral aggregates, particularly igneous rocks, retain their surface texture

Fig. 3.14 A photograph showing the micro-deval apparatus (*left*) and a schematic of the abrasion process (*right*) (Image courtesy of Gilson Company Inc., USA)

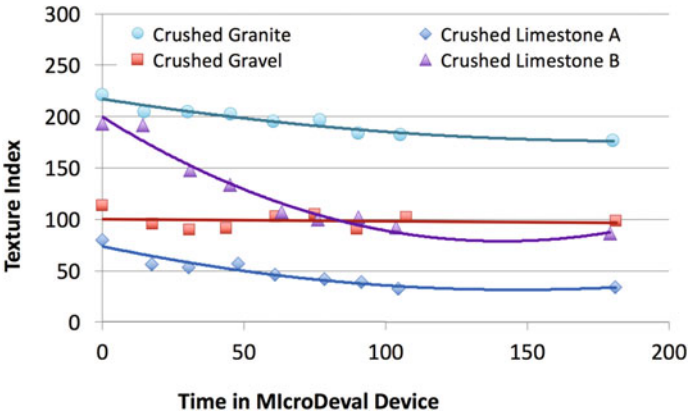


Fig. 3.15 Measured loss in texture for typical mineral aggregates as a function of time used in the micro-deval device (Created using data from Masad et al. 2009)

even after being abraded on the surface. In a very broad sense, this has to do with the manner in which individual microscopic crystals or grains wear off from the surface. The size of these grains, and indirectly the texture, is dictated by the rate at which the rock solidified from its liquid form. It is interesting to note here that a geologic process that occurred millions of years ago has an influence on how soon your vehicle will stop on the road when you hit the brakes.

In summary, toughness of aggregates refers to the ability of the aggregate to resist breakage. A minimum level of toughness is required to ensure that the aggregates do not break under the action of rollers during compaction. Tough aggregates are also critical when aggregate structures such as SMAs are used for mixture design. Hardness refers to the ability of the aggregates to resist abrasion and retain their surface texture. Hardness is particularly important for aggregates that are used in the surface mixes of asphalt pavements to ensure optimal skid resistance.

3.3.4 Durability or Soundness

Durability or soundness of aggregates refers to the ability of the aggregate to resist decomposition due to weathering. More specifically, in the context of asphalt mixtures, durability refers to the ability of the aggregate to resist breakage due to freeze–thaw action. Asphalt binder has historically been used as a water-proofing agent (e.g., as a coating on boat hulls thousands of years ago). Therefore, one school of thought is that aggregate durability, as defined above, is not important for asphalt mixtures since aggregate particles are typically coated with a film of the asphalt binder. While this is true to some extent, aggregate durability does matter for at least two reasons.

First, although asphalt binder may be generally regarded as a waterproofing material, it only implies that it has an extremely low permeability. In other words, it is only a matter of time before water can make its way into the binder and reach the aggregate surface. Since roadways are built to last for several decades, this phenomenon is entirely possible after being exposed to wet weather conditions over a period of several years. In fact, recent studies have shown that the rate with which water diffuses through a film of binder increases as the binder is exposed to cyclic wet and dry events.

Second, consider aggregates that are used in surface mixes. At the time of placement, these aggregates are coated with a film of the asphalt binder. However, in a very short duration of time, the binder coating on the surface wears off leaving behind an exposed aggregate surface. Readers may want to re-examine Fig. 3.12 more closely. The pavement surface in this case comprises a mixture of a dark and a light colored aggregate. The dark colored aggregate particles are igneous in origin, whereas the light colored aggregate particles are a soft porous limestone with low durability. Due to the light color, the exposed limestone particles are clearly visible on the surface. Also, in this particular case, the pavement location was prone to freeze–thaw action during winters. The use of a porous aggregate with low durability and the freeze–thaw action may result in fracture or disintegration of the aggregate particles into several smaller pieces on the surface of the pavement. This may further lead to aggregate loss that could result in windshield damage and also gradually progress into further deterioration and degradation of the pavement structure.

In short, considering the above two factors, durability of aggregates is an important attribute that must be considered during the material selection and mixture design process, particularly for surface mixes. Durability of aggregates can be measured using a few different test methods such as the sodium sulfate or magnesium sulfate soundness test or a simple freeze–thaw test. These tests are very briefly discussed here, and readers can refer to appropriate ASTM or ASSHTO procedures for the specific details on these tests (ASTM C88).

The sodium or magnesium sulfate soundness tests simulate the freeze–thaw process that the aggregate may experience in an accelerated manner. Typically, for these tests a sample of coarse aggregate particles is immersed in a solution of sodium or magnesium sulfate. These salts reduce the surface tension of water and accelerate the absorption of the solution into the aggregate pores. After a certain period of time, the aggregate particles are removed from the solution and allowed to dry. As the water escapes the aggregate particles, the salt within the pores crystalizes and expands in volume. This is similar to expansion of water in the pores as it freezes and turns into ice. The internal stresses can cause the aggregate particles to fracture and disintegrate. The extent of this disintegration is measured by conducting a sieve analysis before and after this process and determining the percentage of coarse aggregate particles that have disintegrated compared to their original size fraction. This percentage then serves as an indicator of the durability of the aggregate (higher percentage loss would imply lower durability).

An alternative to the sodium or magnesium sulfate durability test is a direct freeze–thaw test. In this case, as before, a sample of coarse aggregate particles is used for the test. However, the aggregate sample is immersed in distilled water for a duration of time and then subjected to cyclic freeze–thaw for a specific number of cycles. Aggregate loss is measured as a percentage of the size fraction that disintegrated during this process and was not retained on the sieve size used for the sample. For the example that was cited earlier (Fig. 3.12), a sample of the limestone aggregates was subjected to this type of cyclic freeze–thaw action. Figure 3.16 shows a photograph of the aggregate particles before and after being subjected to the freeze–thaw action. Even without conducting a quantitative analysis on the percentage mass loss of aggregate particles, the picture clearly shows the lack of durability for this particular aggregate.

Some highway agencies generally regard high calcium carbonate content (predominant in most limestones) as an indicator of low durability as well as low hardness. Based on this rationale, a surrogate acid test is also used sometimes to determine the calcium carbonate content in an aggregate. The test involves immersing a sample of the aggregate in hydrochloric acid at a certain concentration and determining the mass of aggregate that is lost due to the reaction. Recall from basic chemistry that calcium carbonate reacts with hydrochloric acid to form calcium chloride salt and carbon dioxide gas. Hard and durable aggregates, such as siliceous aggregates, are typically inert to this test. Figure 3.17 illustrates the carbon dioxide from the reaction between the aggregate particles and the hydrochloric acid.

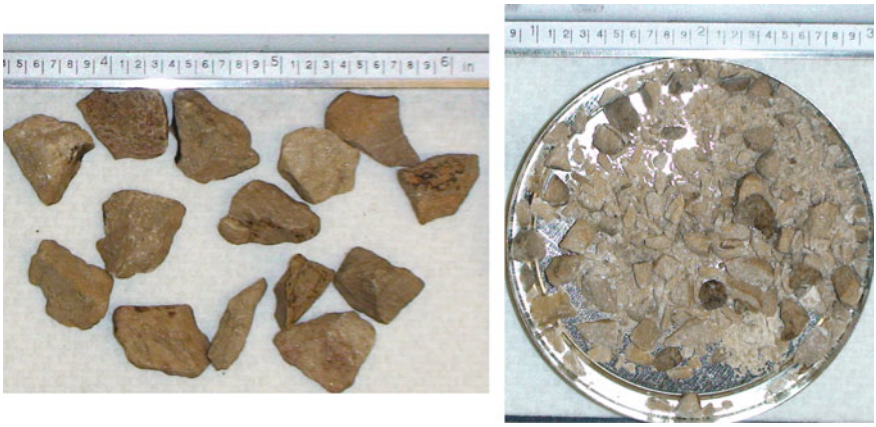


Fig. 3.16 Image showing a sample of low durability coarse aggregate particles before cyclic freeze-thaw (*left*) and the same sample disintegrated after cyclic freeze-thaw (*right*) (Photograph by A Bhasin)

It is important to recall and bear in mind that aggregates typically make up more than 90% by weight of the mix. Therefore, in most cases economies of operations would dictate that aggregate sources closest to the source of the construction site be used. In some cases, such as for surface mixes, it may be worth the cost of procuring and using higher quality aggregates with high durability and hardness when locally available aggregates do not meet set standards.

Fig. 3.17 Image showing a sample of the aggregate in hydrochloric acid; the carbon dioxide bubbles formed due to the reaction are clearly visible (Photograph by A Bhasin)

3.3.5 *Shape, Angularity, and Texture*

Shape, angularity, and texture of aggregates refer to three geometrical characteristics of the aggregate particles at three different length scales. These aggregate characteristics dictate different aspects of the overall performance of asphalt mixtures. In this section, we will first consider the implications of shape, angularity, and texture of aggregates on different aspects of mixture performance. We will then discuss the different methods and metrics that are used to quantify these characteristics.

Let us first consider the shape or form of aggregate particles, which is also a characteristic at the largest length scale. Shape or form of aggregate particles refers to the overall geometric form of the aggregate particles. For example, aggregate particles can be cubical, disk shaped, rod shaped, spherical, or oblong. Typically, a spherical or cubical form is desired compared to a flat or elongated form. Aggregates with flat or elongated form have two direct implications. First, aggregate particles that have a flat and elongated form (as opposed to spherical or cubical) tend to orient themselves so that the smaller dimension is along the direction of compaction. As an extreme example imagine compacting a handful of pennies. Figure 3.18 illustrates a schematic to demonstrate this preferred orientation of aggregates with a flat or elongated form. As a result of this preferred orientation, flat and elongated particles are prone to fracture during compaction that induces very high stresses in the aggregate particles. This in turn reduces the integrity and mechanical stability of the mix.

The second implication due to the preferred orientation of flat and elongated aggregate particles has to do with anisotropic properties of the asphalt mixture. Recall that isotropy refers to the directionally independent nature of a particular property (e.g., stiffness or tensile strength), whereas anisotropy refers to the directionally dependent nature of the property. A simple example is a plank of wood that is anisotropic in terms of its stiffness and tensile strength, i.e., the stiffness or modulus of wood measured by applying a load along the grain of the

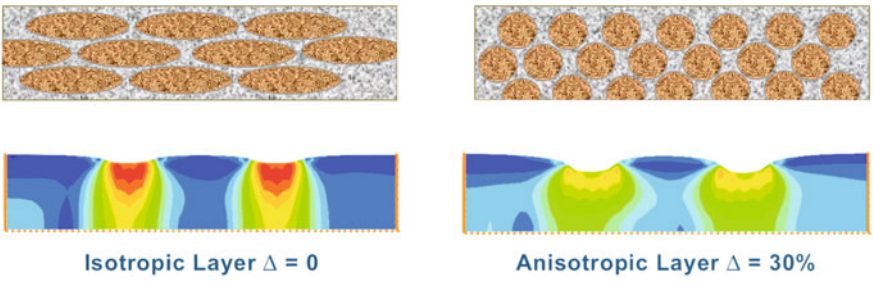


Fig. 3.18 Schematic showing inherent anisotropy due to shape (*top-left*) and isotropy induced due to compaction (*top-right*)

wood is different compared to the modulus measured by applying a load perpendicular to the grain of the wood. The same can be said for tensile strength of wood.

Specifically, in this case the asphalt mixture would have different properties in the vertical direction (direction in which compaction loads are applied) compared to properties in the horizontal direction (plane perpendicular to the direction in which compaction loads are applied). This form of anisotropy has a special name and is referred to as *cross-anisotropy (or transversely isotropic)*. Typically, the mix would have higher stiffness in the vertical direction and relatively lower stiffness in the horizontal direction. Note that the anisotropy can be due to the shape of the aggregate particles (Fig. 3.18 *top-left*) or compaction induced because of the vertical loads from a compactor (Fig. 3.18 *top-right*). The implication of this anisotropic behavior is that the mix would have relatively lower resistance to permanent deformation.

International Center for Aggregate Research (ICAR) evaluated 246 data sets from the AASHTO Road Test, which revealed that surface deflections calculated using isotropic material properties of the AASHTO Road Test aggregate base materials were smaller than the measured deflections under the Benkelman Beam. However, the accuracy of the calculated defections improved when the aggregate base layer was considered to be anisotropic (Masad 2004). Typically, the level of anisotropy is characterized by the ratio of horizontal to vertical modulus (E_x/E_y). Analysis of the errors in the ICAR study showed that the error between measured and calculated responses were minimized when the stiffness in the horizontal direction assumed to be 30% of the stiffness in the vertical direction ($E_x/E_y = 0.3$). Values of anisotropy (E_x/E_y) of 0.4, 0.5, and 1 yielded progressively less accurate predictions.

Figures 3.19 and 3.20 verify the importance of aggregates angularity, shape, and textural properties on the level of cross-anisotropy. Ashtiani (2010) showed clearly that increased angularity, a decrease in the quantity and degree of flat and elongated aggregates particles, and an increase in surface texture all lead to lower anisotropy. Lower levels of anisotropy (or when there is a smaller difference between vertical and horizontal moduli) result in a more efficient distribution of stresses induced by traffic loading. Therefore, one would normally hope to produce aggregates with properties that minimize the degree of anisotropy.

The results presented in Figs. 3.19 and 3.20 clearly demonstrate the influence of aggregate shape features on the degree of anisotropy in unbound aggregates systems. Figure 3.19 shows the effect of aggregate texture (texture is discussed in more detail later) and form on the level of anisotropy as characterized by the shear modular ratio (G_{xx}/G_{xy}). As evidenced in this plot, unbound aggregate systems with more cubical particles and rougher texture had higher shear modular ratios (G_{xx}/G_{xy}) and therefore had lower levels of anisotropy. In general, more isotropic materials produce a more favorable orthogonal load distribution when subjected to traffic loads.

Figure 3.20 demonstrates the impact of particle texture and aggregate angularity on the level of anisotropy characterized by modular ratios (i.e., E_x/E_y). ICAR research showed that aggregate systems containing particles with rougher texture

and more crushed surfaces (more angular) result in less anisotropic systems. Particle surface texture and angularity impact interparticle frictional forces and therefore orthogonal load distribution capacity of the aggregate layer. Collectively, aggregate systems consisting of roughly textured and more angular particles result in systems that more efficiently distribute traffic loads and are less prone to rutting during service.

Although the effects described in the preceding paragraphs are more pronounced in unbound materials, aggregate shape also influences the behavior of asphalt mixtures in a flexible pavement. For example, Masad (2004) performed analyses of fatigue and rutting in asphalt pavements of six different asphalt pavement cross sections with the asphalt thickness ranging from 2 to 6 in. (50–150 mm) and with varying aggregate base thicknesses. The hot mix asphalt (HMA) layer, the aggregate base, and the subgrade in each section were modeled as linear elastic and isotropic (LIS), linear elastic and anisotropic (LAN), nonlinear elastic and isotropic (NIS), and nonlinear elastic and anisotropic (NAN). The results with respect to the asphalt layer are summarized as follows:

a. For the asphalt layer, permanent deformation was always greater for the anisotropic model compared to its isotropic counterpart, Fig. 3.21.
b. Fatigue life predicted using the nonlinear, anisotropic model for the asphalt layer was considerably higher than that predicted for the nonlinear, isotropic material, Fig. 3.22.

One length scale smaller than the aggregate form is the aggregate angularity. Aggregate angularity reflects the number of fractured faces and/or sharp edges of

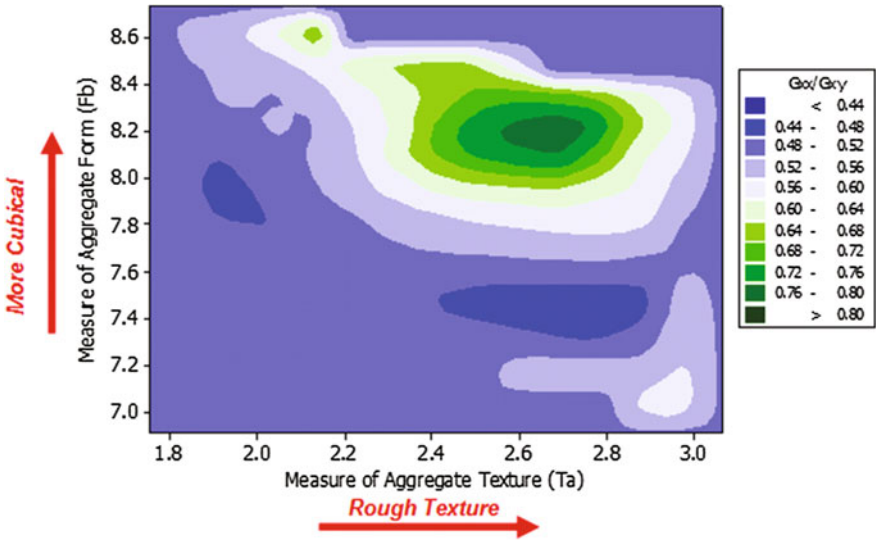


Fig. 3.19 Impact of form and texture on shear modulus ratio (G_{xx}/G_{xy}) (Ashtiani and Little 2010)

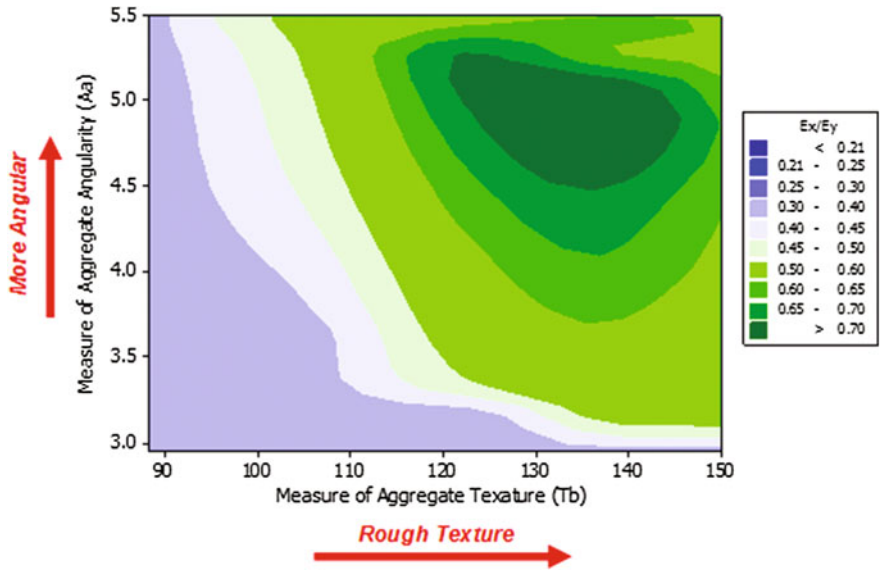


Fig. 3.20 Impact of particle angularity and texture parameters on modular ratio (E_x/E_y) (Ashtiani and Little 2010)

individual aggregate particles. Typically, angular aggregates with more fractured faces are desired as opposed to smooth and rounded particles. The reason for this is that angular aggregate particles tend to interlock better with each other, thus resulting in much higher mechanical friction and resistance to deformation. Aggregate angularity is typically of concern with aggregates from sources such as river or glacial gravel. Unlike other aggregates that are typically obtained from blasting out large pieces of rock in an opencast mine followed by crushing to a desired size, gravel is obtained by simply collecting naturally occurring boulders and stones of different sizes (sometimes as small as a fraction of an in.). The naturally occurring boulders and stones are typically smooth and rounded on the surface due to geologic erosion. Once crushed and fractionated into different sizes, several aggregate particles retain the natural smooth surface either completely or partially. The more the overall percentage of smooth surface on the aggregate particles, the less the mechanical interlock, and consequently there is less resistance to shear deformation. The importance of aggregate angularity can perhaps be best exemplified based on the following case study.

The locally available aggregate in southern parts of the state of Texas, particularly along its coast, is a siliceous river gravel. For this region, this was also the cheapest type of aggregate that could be used for pavement construction. The only problem with the siliceous river gravel was that most of it occurred naturally in smaller sizes (approximately 1–2 in.). As a result, even after some amount of crushing, a number of the aggregate particles would retain a significant portion of

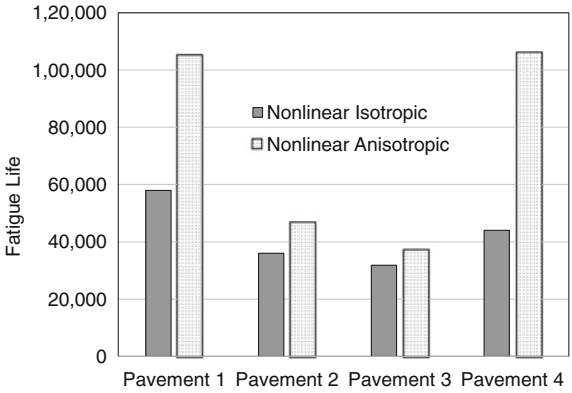


Fig. 3.21 Fatigue life comparison for pavements with asphalt layer considered as either non-linear isotropic or non-linear anisotropic (Based on data from Masad 2004)

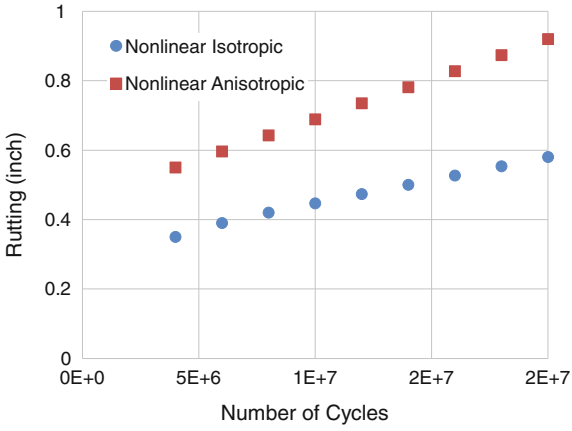


Fig. 3.22 Permanent deformation for a typical pavement section considering the asphalt surface layer non-linear-isotropic (NIS) or non-linear-anisotropic (NAN) (Based on data from Masad 2004)

their original surface that was very smooth. Since the smaller aggregate size fractions were produced by crushing the naturally occurring larger aggregate size, the smaller sized aggregates would have more crushed faces while the larger sized aggregates would have more smooth (natural) and uncrushed faces. The lack of aggregate angularity would result in poor mechanical interlock between the aggregate particles and reduced resistance to shear deformation and increased propensity to rutting.

One solution to the problem was to design an asphalt mixture with a relatively smaller maximum nominal aggregate size. This would ensure that most of the aggregate particles that went in the mix had more crushed and angular surfaces. Figure 3.23 clearly shows this effect, wherein the percentage of angular particles is large for smaller size fractions compared to larger size fractions and vice versa for the percentage of rounded particles. The quantitative methods used to arrive at these

percentages will be discussed later in this section. Figure 3.24 illustrates the rutting resistance in terms of a metric referred to as flow time for three different mixtures produced using these aggregates (more on the flow time test in Chap. 8). These mixtures had different maximum nominal aggregate size. One would expect that as the maximum nominal size of the aggregates decreased, the percentage of aggregates with angular surface (from all size fractions) would increase and result in better particle interlock and increased shear resistance. This phenomenon is very clear when the results from the 19 mm mix are compared to the results from the 12.5 mm mix. A further decrease in the maximum nominal aggregate size to 9.5 mm, however, does not produce this advantage. This is because at this point the overall particle-to-particle contacts have increased resulting in a more indirect pathway via the binder or mastic matrix for load transfer through the mix. This lack of efficiency in internal load transfer is very similar to what was discussed earlier in explaining the benefits of using a stone matrix asphalt. In summary, this case study demonstrates the importance of using angular aggregates to ensure adequate mixture resistance to shear deformation.

The next smaller length scale after angularity is the surface texture of the aggregate particles. In a very basic sense, surface texture of the aggregates refers to its roughness at a micrometer length scale. Roughness is a very length scale-dependent term and is always tied to the sensitivity of the instrument or means that are used to measure it. For example, consider a surface that may seem very smooth to human eye or touch (e.g., by rubbing the surface with a finger). This same surface could turn out to be very rough when an optical microscope with a micrometer resolution is used to examine its surface. Similarly, extremely smooth surfaces at a micrometer length scale would appear to be rough on an atomistic length scale. In the context of aggregate surface texture, the roughness or texture being referred to is in the range of hundreds of micrometer to a millimeter length scale. This length scale is important for two different reasons. For aggregate particles used on the surface mixes, aggregate texture at this length scale is important in dictating the frictional characteristics of the pavement surface and hence its skid resistance. In fact, studies have shown empirical evidence that correlates surface texture of aggregates to pavement skid resistance (Masad et al. 2008). For the bulk of the mix, aggregate texture at this length scale is also important in providing a mechanical bond or adhesion between the binder and the aggregate surface. This is important from an overall durability point of view and also in terms of the resistance of the mixture to moisture-induced damage or stripping.

Having discussed the general definitions and implications of aggregate shape, angularity, and texture, the following paragraphs will briefly discuss the different methods that are used to measure these properties. Consistent with other sections in this chapter, only broad descriptions and approaches to these test methods are presented here. The methods used to measure shape, angularity, and texture can be broadly classified within two groups: manual and automated.

Manual methods to measure shape and angularity are focused on the measurement of aspect ratio and crushed faces on individual aggregate particles, respectively. A manual method to measure the aspect ratio involves the use of a ratio

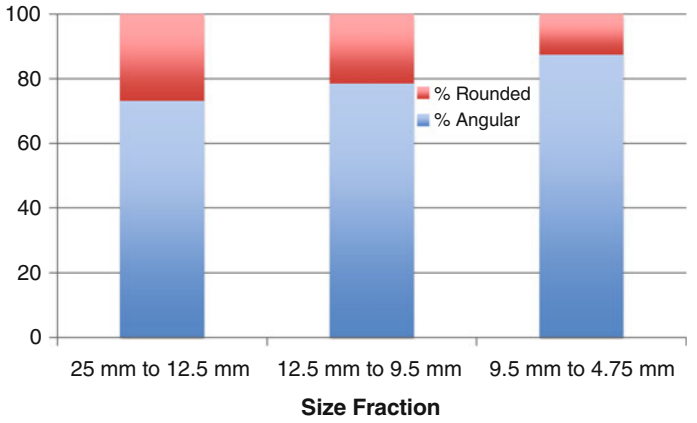


Fig. 3.23 The percentage decrease in rounded aggregates and increase in angular aggregates for different size fractions; note that small size fractions have fewer rounded aggregates and more angular aggregates

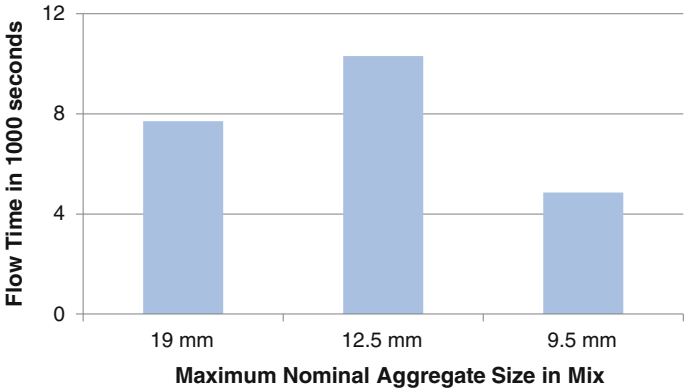


Fig. 3.24 Results from a flow time test that is used to measure rutting resistance of asphalt mixtures with higher flow time values indicating better rutting resistance

caliper with a hinge (Fig. 3.25). The hinge is set at a desired ratio (e.g., 1:3 or 1:5), the smallest dimension of the aggregate is used to open the caliper from the shorter end, which proportionally also opens the caliper at the larger end. The gap in the caliper is compared to the largest dimension of the aggregate particle to determine whether the particle has an aspect ratio above or below this set limit. The process is repeated for every single coarse aggregate particle in a representative sample from a stockpile. Typically, specification limits require that the percentage of flat and elongated particles (i.e., particles that have an aspect ratio above a specified proportion such as 1:5) be below a certain limit (e.g., below 10%). Although this procedure requires hardly any investment in terms of capital equipment, it is

associated with three major drawbacks: The method is extremely time consuming and labor intensive, it only provides a discrete breakup of aggregate angularity (e.g., a certain percentage of aggregate particles are above or below the limit), and the method is prone to human errors. These shortcomings will be addressed later with the automated methods.

A different manual method is used to measure aggregate angularity for coarse and fine aggregate particles. For coarse aggregate particles, the method involves counting the crushed faces on the aggregate particle. Typically, a sample of the coarse aggregate particles is taken from a stockpile. Each particle from the sample is examined for the number of crushed faces. The particles are then separated into different groups with no crushed faces, one crushed face, and two or more crushed faces. The percentage of aggregates in each stockpile is used as a metric for aggregate angularity. As in the case with aspect ratio, this method is prone to human errors, time consuming, and discrete in terms of the metrics it provides for aggregate angularity. Also, although aggregate angularity is important for fine aggregates, it is not feasible to use this method with fine aggregate particles.

For fine aggregates, an empirical test is used to assess the fine aggregate angularity. Figure 3.26 illustrates a photograph of the device used to make such measurements. Briefly, a sample of the fine aggregate particles is placed in a container with a funnel-shaped opening at a fixed height. The aggregates are allowed to flow through the funnel into a cylindrical container below it and self-compact. The mass of the fine aggregate in the cylindrical container is then

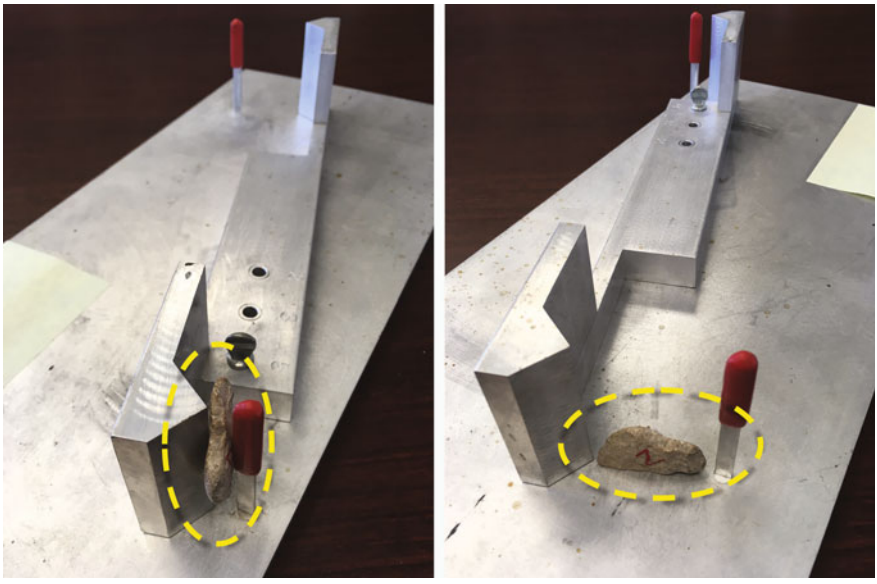


Fig. 3.25 Photograph of the caliper used to measure flat and elongated particles (Photograph by A Bhasin)

measured to determine the density of the self-compacted fine aggregate particles. Fine aggregate particles with angular surfaces tend to create more voids and result in a lower density compared to rounded fine aggregate particles. Therefore, the density of the fine aggregate sample obtained in this standardized method is used as a surrogate indicator of its angularity. Readers are referred to the additional reading material for a more detailed description of these test methods (Rogers et al. 1991; Lee et al. 1999; Masad et al. 2005). Note that there is no standard method to measure texture of aggregate particles manually. The following paragraphs discuss the use of automated imaging methods to measure the shape, angularity, and texture of fine aggregate particles.

Most automated methods to measure aggregate shape, angularity, and texture involve some form of image acquisition and analysis. Image acquisition involves the use of a digital camera and some other mechanical means to capture the images of individual aggregate particles (or small samples of fine aggregate particles). This is what makes up the hardware of such automated methods. Image analysis involves analyzing the acquired images using different mathematical tools to arrive at metrics to quantify the shape, angularity, and texture of aggregates.

Figure 3.27 shows three different devices that are used to capture the images of aggregate particles. The two devices on the left have some form of a conveyor belt system that carries aggregate particles in front of a fixed camera that captures the images of these particles. The third device has aggregate particles on a tray that can either be bottom lit or top lit and a camera that captures images of individual particles. The bottom light creates a dark image of the aggregate particle highlighting its form and angularity (Fig. 3.28) for further analysis. The top light is used

Fig. 3.26 Photograph of the empirical method used to measure angularity of fine aggregate particles (Image courtesy of Gilson Company Inc., USA)

to capture details on the surface of the aggregate particle for texture analysis (Fig. 3.29). This particular device, referred to as the Aggregate Imaging System (AIMS), has been further developed and implemented for routine use. As a side note, one may question that these images are two dimensional in nature and therefore may not accurately characterize the three-dimensional shape characteristics of the aggregate particles. To this end, it must be noted that devices such as AIMS also estimate and record the third dimension of the aggregate particle. This is typically achieved by comparing the focal distance between the camera and surface of the aggregate particle to the focal distance between the camera and the surface on which the aggregate particles rest. Other devices have implemented the use of dual cameras to capture the three-dimensional shape characteristics of the aggregate.

In terms of analysis, several different algorithms have been developed and can be used to characterize the three attributes of interest: shape, texture, and angularity. The general philosophy of obtaining these metrics is described here. For example, one metric to quantify the overall shape or form of the aggregate particle would be to compare the deviation of the aggregate particle from a perfect sphere. Figure 3.28 illustrates this idea in two-dimensions for two particles with very distinct forms or shape. The departure from the shape can be quantified in terms of the departure from the circular shape or other metrics such as the ratio of the shortest to the longest dimension. The angularity of the aggregate particles can also be measured using any number of different mathematical algorithms. For example, the rate at which the angles of the tangents from different points on the surface of the aggregate change can serve as an indicator of aggregate angularity. Similarly, the texture of the aggregate surface can be assessed based on the variations in the gray scale intensity on the digital image. A rough texture would result in several peaks and shadows or noise on the digital image of the surface. This pattern can then be analyzed to arrive at an index that is a quantitative reflection of the aggregate texture.

The AIMS method uses wavelet transformation of the two-dimensional gray scale images to arrive at a texture index. Very broadly speaking this transformation progressively blurs the image and quantifies the loss of information when compared to the previous image sequence. A greater loss in information over several blurring

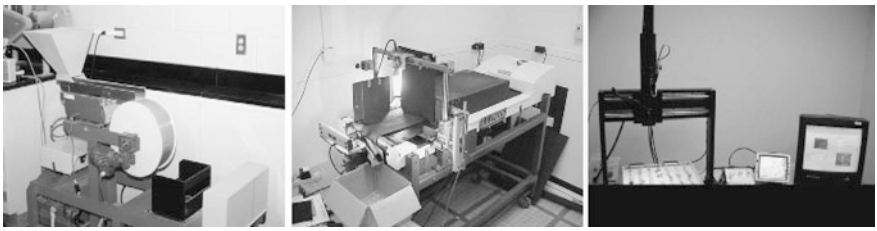


Fig. 3.27 Three different devices used to capture aggregate images to determine shape, angularity and/or texture (With permission from TRB through the National Academy of Sciences, Prowell, B., and Brown, R., Aggregate Properties and the Performance of Superpave-Designed Hot-Mix Asphalt, NCHRP Report 539)

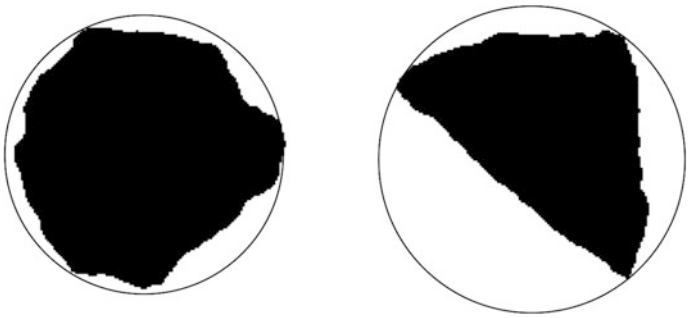


Fig. 3.28 Outline image of two different aggregate particles showing deviation from a circular form; also notice that the aggregate particle on the left is circular in form yet angular

sequences would indicate a rougher texture. A smooth surface would have hardly any information loss before and after blurring, while a rough textures surface would have significant loss of information before and after blurring. Figure 3.29 illustrates the digital images from two different aggregates; the figure also includes the distribution of the gray scale intensity on the two images. Note the difference in distribution of the gray scale; the location of these distributions is dependent on the external lighting conditions but the width of these distributions is related to the texture. Figure 3.30 illustrates the AIMS device along with the texture index distribution for typical aggregate types. Similar distributions for shape and angularity index can also be obtained. Note that unlike the manual methods, a distribution for the shape, angularity, and texture is obtained using this method.

Readers are referred to additional reading material for a more thorough and detailed description of the mathematical models and tools used to quantify shape, angularity, and texture (Masad et al. 2005). Particularly, the NCHRP report 555 (Masad et al. 2005) documents in detail the different mathematical algorithms that were investigated and ultimately adopted for use with the AIMS.

3.3.6 Impact of Aggregate Characteristics on Engineering Properties

We have talked about size and gradation early in this chapter. Qualitatively, we have shown that size and gradation influence the overall behavior of the asphalt mixture. In this section, we briefly discuss the quantitative relationship between the aggregate size and gradation and one of the most basic engineering properties of the asphalt mixture, i.e., its dynamic modulus.

The dynamic modulus of the asphalt layers is the single most important engineering property that is necessary to design a flexible pavement, particularly using

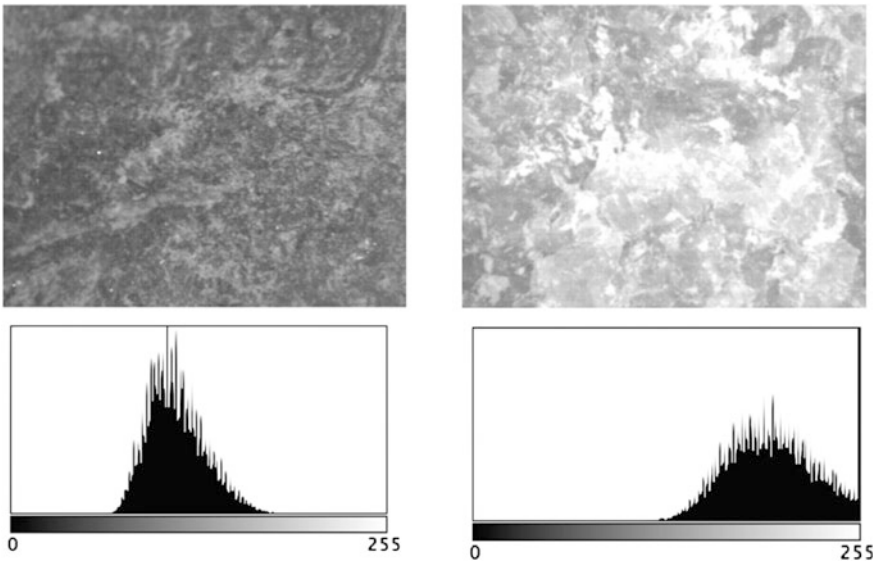


Fig. 3.29 Digital image of two different aggregate surfaces along with a distribution of the gray scale intensity in the image; note that the aggregate on the right is smoother and has a much broader distribution in gray scale as opposed to sharper distribution for the rough aggregate on the left (absolute location of the distribution is not as critical as distribution)

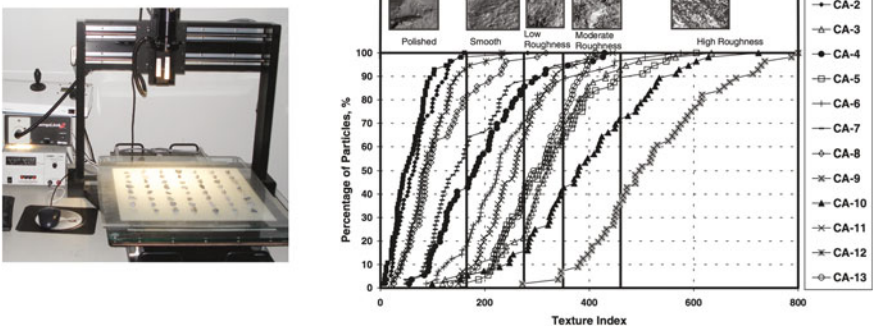


Fig. 3.30 Typical texture index obtained using the aggregate imaging system for different aggregate types (With permission from TRB through the National Academy of Sciences, Al-Rousan, T., Masad, E., Little, D.N., Button, J.W., and Tutumluer, E., Test Methods for Characterizing Aggregate Shape, Texture, and Angularity, NCHRP Report 555)

mechanistic or semimechanistic methods. This is because in order to compute pavement responses to load, which are used in computing fatigue and rutting damage within all pavement layers, reliable layer moduli are required. Furthermore, in the constitutive relationships required for mechanistic analysis (e.g., layered elastic analysis or viscoelastic analysis), layer moduli along with layer Poisson's ratios are required.

Since the modulus of the asphalt layer and/or sublayers is time and temperature dependent, the dynamic modulus is used to represent and model asphalt layers. The dynamic modulus is measured by imparting a sinusoidal stress pulse to a cylindrical asphalt sample. The test is conducted over a range of combinations of frequency of the sinusoidal pulse and temperature in order to cover the range of frequencies and temperatures representative of the pavement environment. This is very similar to what was described for asphalt binders in Chap. 2 on "Time Dependency," except that in this case the time-dependent properties are measured using an asphalt mixture instead of an asphalt binder. Similar to asphalt binder, data obtained from the dynamic moduli tests over a range of frequencies and temperatures are used to construct a master curve for the asphalt mixture. Typically, a sigmoidal function can be used to describe the master curve for the complex modulus $|E^*|$ (recall that this is very similar to Eq. 2.3 used to develop the master curve for the binder).

$$log|E^*| = \delta + \frac{\alpha}{1 + e^{\beta + \log t_r}} \tag{3.2}$$

Similar to Eq. 2.1, in Eq. 3.2, t_r is reduced time (where time is the reciprocal of the test frequency), $\delta + \alpha$ is the maximum value of $|E^*|$, and β and γ are parameters which describe the shape of the sigmoidal function. In the context of the role of aggregate properties on the dynamic modulus of asphalt mixtures, these parameters can be a function of mixture volumetrics and gradation. For example, values of δ and α in Eq. 3.2 can be calculated based on certain properties of the aggregate and mixture as follows:

$$\begin{aligned}\delta = {} & -1.249937 + 0.02932\rho_{200} - 0.001767(\rho_{200})^2 \\ & - 0.00284\rho_4 - 0.05809V_a - 0.802208\left[\frac{V_{b_{eff}}}{V_{b_{eff}} + V_a}\right]\end{aligned} \tag{3.3}$$

$$\begin{aligned}\alpha = {} & 3.871977 - 0.0021\rho_4 + 0.003958\rho_{3/8} - 0.0000\rho_{3/8} \\ & + 0.005470\rho_{3/4}\end{aligned} \tag{3.4}$$

where

Va	Air void content, %
V_{beff}	Effect bitumen content, % by volume
$\rho_{3/4}$	Cumulative % retained on the 3/4 in. sieve
$\rho_{3/8}$	Cumulative % retained on the 3/8 in. sieve

ρ_4 Cumulative % retained on number 4 sieve
ρ_{200} Cumulative % passing the number 200 sieve

Also the dynamic modulus can be predicted over a range of temperatures, rates of loading, and aging conditions from information that is readily available from material specifications of volumetric design of the mixtures:

$$\begin{aligned}\log|E^*| = {} & -1.249937 + 0.02932\rho_{200} - 0.001767(\rho_{200})^2 \\ & - 0.002841\rho_4 - 0.05809V_a - 802208\left[\frac{V_{b_{eff}}}{V_{b_{eff}}}\right] \\ & + \frac{3.871977 - 0.0021\rho_4 + 0.003958\rho_{3/8} + 0.005470\rho_{3/4}}{1 + e^{(-0603313 - 0.3115\log(f) - 0.393532\log(n))}}\end{aligned} \tag{3.5}$$

where

n Viscosity, $\times 106$ Poises
f Loading frequency, Hz

In summary, Eqs. 3.2–3.5 demonstrate the direct relationship between aggregate gradation and the dynamic modulus of the asphalt mixture (for a given binder), which is one of the most important properties required to carry out the thickness design of different layers in a flexible pavement.

3.3.7 Absorption

The last aggregate property that is briefly discussed here is absorption. Certain kinds of aggregates, such as soft limestones, are somewhat porous in nature and tend to absorb water or moisture when left in stockpiles and also tend to absorb binder during mixture production. The absorption of water (in a stockpile) or binder (during production) has serious implications on the cost and durability of the asphalt mixture. For example, absorptive aggregates that are left in a stockpile can absorb significant amounts of water from rain. In this case, it would take much longer to completely dry these aggregates prior to mixing them with the liquid asphalt binder. This increases fuel consumption and adds to the production cost of the mixture.

The second mechanism by which absorptive aggregates add to the cost of mixture production is due to loss of the binder that is absorbed by the aggregates. In other words, some of the asphalt binder will be absorbed by the aggregates and will not contribute to the effective asphalt binder that is used to bind the aggregate particles together in the mix. As a result more binder needs to be added to the mix to compensate for the loss in absorbed binder and achieve the optimum or target effective binder to aggregate ratio. The concept of effective binder content and absorbed binder will be discussed in Chap. 12.

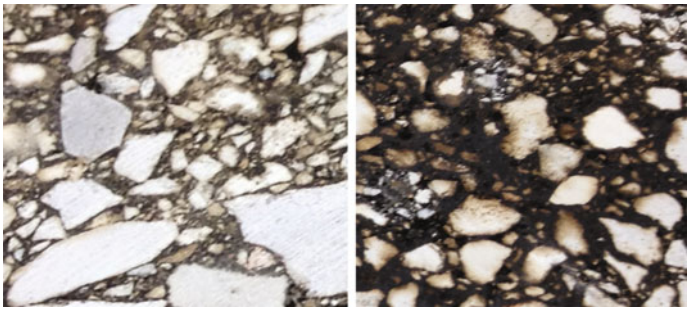


Fig. 3.31 Images of the sections of two very similar asphalt mixtures with absorptive aggregates; the image on the *left* uses a very stiff binder that was not readily absorbed by the aggregates during mixing and the image on the *left* uses a very soft binder that was readily absorbed during mixing (Photograph by A Bhasin)

Figure 3.31 shows the cross section of two different asphalt mixtures with a porous aggregate. The image on the left uses a very stiff and high viscosity binder that does not get easily absorbed by the porous aggregate even at the high mixing temperatures. However, the binder on the right uses a soft and low viscosity asphalt binder that was easily absorbed by the porous aggregates. The dark discoloration of the aggregate edges clearly shows this absorption. Another consideration with the use of absorptive aggregates that has been a subject of some research is that the binder is not homogeneously absorbed by the aggregate. Recall from the previous chapter that an asphalt binder is an ensemble of organic molecules with different molecular weights and polarities. As a result, some of the molecules in the binder are more mobile than others. When the liquid binder at high temperature is in contact with an absorptive aggregate surface, depending on the charge (on the asphalt molecules and the aggregate surface) and mobility some of the binder molecules may preferentially be absorbed in the aggregate pores. In fact, this phenomenon is likely to occur even on the surface of nonporous aggregates. The only difference being that in the case of nonporous aggregates, this process would be referred to as preferential adsorption and for porous aggregates this preferential surface adsorption would continue on as preferential absorption into the pores of the aggregate particles. The net result of this preferential absorption is that the binder in the immediate proximity of the aggregate surface would be deprived of lighter fractions that are likely to be absorbed by the aggregate and would likely have different stiffness and relaxation ability from the bulk binder and result in the formation of damage susceptible zones. While the exact mechanism of such near-surface failure is a subject of ongoing research, it is clear that this phenomenon will be dictated by the properties of both the binder and the aggregate.

In summary, absorptive aggregates absorb a small fraction of asphalt binder. As a result, typically higher binder contents are required to achieve the optimum binder-aggregate ratio, which in turn increases the cost of the mixture. Also,

Table 3.1 Control points based on Superpave specification

Sieve number	Sieve size in mm	Superpave control points for 19 mm nominal max. aggregate size		Restricted Zone	
		Lower	Upper	Lower	Upper
	25		100		
	19	90	100		
	12.5				
	9.5				
4	4.75				
8	2.36	23	49	34.6	34.6
30	0.6			16.7	20.7
50	0.3			13.7	13.7
200	0.075	2	8		
pan	0				

absorptive aggregates may preferentially absorb some of the lighter fractions from the binder resulting in near-surface zones that behave differently from the bulk and may nucleate some of the distresses.

3.4 Exercises

3.1 The Strategic Highway Research Program (SHRP) during early 1990s developed a database of typical aggregates that were used in the USA. The database includes typical aggregate types referred to by their generic names used commonly by engineers as well the specific mineral composition of each aggregate type. The database is documented in a report (Number SHRP-A UIR-91-509) that can be downloaded from www.trb.org. Use this report to identify two pairs of aggregates that have a similar generic mineral classification but different mineralogy. Briefly explain why you think generic mineral classification can be misleading.

3.2 Use the following sieve sizes to plot the 0.45 maximum density line: 19, 12.5, 9.5, 4.75, 2.36, 1.18, 0.6, 0.3, 0.15, 0.075 mm.

3.3 The table below provides the control points based on the Superpave specification for a dense graded mix with 25 mm maximum aggregate size (maximum density line will have 100% passing at 25 mm to the power 0.45) along with the restricted zone. Superimpose the specification limits (including the restricted zone) on the 0.45 maximum density line from 3.2 (Table 3.1).

3.4 Based on your reading from the chapter, briefly discuss about the restricted zone.

Table 3.2 Available aggregate gradation from different stockpiles

Sieve Size	Stock pile #1	Stock pile #2	Stock pile #3	Stock pile #4	Stock pile #5	Stock pile #6
	% Passing	% Passing	% Passing	% Passing	% Passing	% Passing
25.000	100.0	100.0	100.0	100.0	100.0	100.0
19.000	100.0	100.0	100.0	100.0	100.0	100.0
9.500	12.2	57.7	94.7	99.8	99.8	100.0
4.750	0.6	0.9	7.0	67.0	98.0	99.9
2.360	0.4	0.3	0.2	10.6	85.2	94.0
0.600	0.4	0.2	0.2	1.8	40.1	39.4
0.300	0.4	0.2	0.2	1.3	28.9	19.9
0.075	0.3	0.1	0.1	0.7	15.7	3.8

3.5 You have been supplied with aggregates from different stockpiles. The table below presents the gradation of aggregate from these stockpiles. Propose a gradation for a dense graded mix that satisfies the limits in Table 3.2 by blending different stockpiles (i.e., using different weight percentages of each stockpile such that the final gradation meets the specification). Tabulate your results and also plot the final gradation along with the specification limits and maximum density line from 3.2.

References

AASHTO. (1997). *T 19-93-Standard test method for unit weight and voids in aggregate*.

Ashtiani, R.S. (2010). *Anisotropic characterization and performance prediction of chemically and hydraulically bounded pavement foundations*.

Ashtiani, R. & Little, D.N. (2010). *Methodology for designing aggregate mixtures for base courses*, Final Report ICAR/508, International Center for Aggregate Research.

Association, N.L. (2003). *How to add hydrated lime to asphalt: An overview of current methods*.

Bell, F. G. (1996). Lime stabilization of clay minerals and soils. *Engineering Geology, 42,* 223–237.

Chowdhury, A., Button, J.W., & Bhasin, A. (2006). *Fibers from recycled tire as reinforcement in hot mix asphalt*, Report SWUTC/06/167453-1 for US Department of Transportation.

Eades, J.L., & Grim, R.E. (1960). Reactions of hydrated lime with pure clay minerals. *Highway Research Record, 262*.

Fuller, W. B., & Thompson, S. E. (1907). The laws of proportioning concrete. *Transactions of the American Society of Civil Engineers, 59,* 67–172.

Gatchalian, D., Masad, E., Chowdhury, A. & Little, D. (2006). *Characterization of aggregate resistance to degradation in stone matrix asphalt mixtures*, Final Report ICAR/204-1F, International Center for Aggregate Research.

Kim, Y. R., Little, D. N., & Lytton, R. L. (2004). Effect of moisture damage on material properties and fatigue resistance of asphalt mixtures. *Transportation Research Record: Journal of the Transportation Research Board, 1891,* 48–54.

Kohata, Y., et al. (1997). Modelling the non-linear deformation properties of stiff geomaterials. *Géotechnique, 47*(3), 563–580.

Lee, C.J., Pan, C., & White, T.D. (1999). Review of fine aggregate angularity requirements in superpave. *Journal of the Association of Asphalt Paving Technologists,* 68.

Lesueur, D., & Little, D. N. (1999). Effect of hydrated lime on rheology, fracture, and aging of bitumen. *Transportation Research Record: Journal of the Transportation Research Board, 1661,* 93–105.

Little, D.N. (1995). *Handbook for stabilization of pavement subgrades and base courses with lime*. Lime Association of Texas.

Little, D.N., & Epps, J.A. (2001). *The benefits of hydrated lime in hot mix asphalt*. National Lime Association.

Little, D. N., & Petersen, J. C. (2005). Unique effects of hydrated lime on the performance related properties of asphalt cements: Physical and chemical interactions revisited. *Journal of Materials in Civil Engineering (ASCE), 17*(2), 207–218.

Little, D.N., & Bhasin, A. (2006). *Using surface energy measurements to select materials for asphalt pavements*, NCHRP Web Only Document 104.

Mahmoud, E. M. (2007). Development of experimental methods for the evaluation of aggregate resistance to polishing, abrasion, and breakage (Doctoral dissertation, Texas A&M University).

Masad, E. et al. (2005). *Test methods for characterizing aggregate shape, texture, and angularity*.

Masad, E. A., et al. (2008). *Predicting asphalt mixture skid resistance based on aggregate characteristics*. TX: College Station.

Masad, E., Rezaei, A., Chowdhury, A. & Harris, P. (2009). Predicting asphalt mixture skid resistance based on aggregate characteristics, Final Report Number FHWA/TX-09/0-5627-1, Texas Transportation Institute.

Masad, S.A. (2004). *Sensitivity analysis of flexible pavement response and AASHTO 2002 design guide for properties of unbound layers*. College Station, TX: Texas A&M University. Retrieved June 20, 2015, from https://repository.tamu.edu/handle/1969.1/528

Plancher, H., Green, E.L., & Petersen, J.C. (1976). Reduction of oxidative hardening of asphalt by treatment with hydrated lime—A mechanistic study. In *Proceedings of the Association of Asphalt Paving Technologists* (vol. 45, pp. 1–24).

Rogers, C.A., Bailey, M.L., & Price, B. (1991). *Micro-deval test for evaluating the quality of fine aggregate for concrete and asphalt*.

Sung Do, H., Hee Mun, P., & Suk keun, R. (2008). A study on engineering characteristics of asphalt concrete using filler with recycled waste lime. *Waste Management,* 28(1), 191–199.

Tatsuoka, F., et al. (1999). Non-linear resilient behaviour of unbound granular materials predicted by the cross-anisotropic hypo-quasi-elasticity model. In G. Correia (Ed.), *Unbound granular materials—Laboratory testing, in-situ testing and modelling* (pp. 197–206). Rotterdam: Balkema.

Underwood, B. S. (2011). *Multiscale constitutive modeling of asphalt concrete*. Raleigh, North Carolina: North Carolina State University.

Vavrik, W. R., et al. (2001). The Bailey method of gradation evaluation: The influence of aggregate gradation and packing characteristics on voids in the mineral aggregate. *Journal of the Association of Asphalt Paving Technologists, 70,* 132–175.

Chapter 4
Chemical and Mechanical Processes Influencing Adhesion and Moisture Damage in Hot Mix Asphalt Pavements

4.1 Background

Moisture damage can be defined as the loss of strength and durability in asphalt mixtures due to the effects of moisture. Moisture damage can occur because of a loss of bond between the asphalt cement or the mastic (asphalt cement plus the mineral filler—74 μm and smaller aggregate) and the fine and coarse aggregate. Moisture damage also occurs because moisture permeates and weakens the mastic, making it more susceptible to damage during cyclic loading. The literature (Taylor and Khosla 1983; Kiggundu and Roberts 1988; Terrel and AlSwailmi 1994) refers to at least five different mechanisms of stripping: detachment, displacement, spontaneous emulsification, pore pressure, and hydraulic scour. Kiggundu and Roberts (1988) suggest additional mechanisms that may well play a part in moisture damage. These include pH instability and the effects of the environment or climate on asphalt–aggregate material systems.

4.1.1 Detachment

Detachment is the separation of an asphalt film from an aggregate surface by a thin film of water without an obvious break in the film (Majidzadeh and Brovold 1968). Theories that explain adhesive bond energy provide the rationale to understand detachment. Several factors are involved in detachment. First of all, it is necessary to develop a good bond between the asphalt and the aggregate. Such a bond is initially dependent on the ability of the asphalt to wet the aggregate. Wettability of aggregate increases as surface tension or free surface energy of adhesion decreases (Majidzadeh and Brovold 1968). According to Majidzadeh and Brovold (1968), if a three-phase interface consisting of aggregate, asphalt, and water exists, water reduces the free energy of the system more than asphalt to form a thermodynamically

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_4

stable condition of minimum surface energy. Surface energy measurements have established that when the free energy at the asphalt–aggregate interface is calculated in the presence of water, energy is generally released, meaning that the aggregate surface has a strong preference for water over asphalt. The negative values of free energy in Column 5 of Table 4.1 establish that this is true for each of four asphalt–aggregate combinations listed in Table 4.1. The more negative the value, the stronger the preference for detachment of asphalt from aggregate is in the presence of water. The thermodynamic basis for these calculations is presented by Cheng et al. (2002). Work at the Road Research Laboratory in 1962 suggests that most asphalts have relatively low polar activity and that the bond that develops between the aggregate and asphalt is chiefly due to relatively weak dispersion forces. Water molecules are, on the other hand, highly polar and can replace the asphalt at the asphalt–aggregate interface. Cheng et al. (2002) has established this to be the case and will be discussed later.

4.1.2 Displacement

Displacement differs from detachment because it involves displacement of asphalt at the aggregate surface through a break in the asphalt film (Tarrer and Wagh 1991; Fromm 1974a, b). The source of the break or disruption may be incomplete coating of the aggregate surface, film rupture at sharp aggregate corners or edges, pinholes originating in the asphalt film because of aggregate coatings, and so forth. Scott (1978) states that chemical reaction theory can be used to explain stripping as a displacement mechanism. Some authors describe film rupture as a separate mechanism of moisture damage, but it can be incorporated as part of the displacement mechanism. The process of displacement can proceed through changes in the pH of the water at the aggregate surface that enters through the point of disruption. These changes alter the type of polar groups adsorbed, leading to the buildup of opposing, negatively charged, electrical double layers on the aggregate and asphalt surfaces. The drive to reach equilibrium attracts more water and leads to physical separation of the asphalt from the aggregate (Scott 1978; Tarrer and Wagh 1991).

Table 4.1 Comparison of free energy of adhesion (ergs/gm) and rate of damage under repeated load triaxial testing. After Cheng et al. (2001)

Mix	Cycles to accelerated damage	Loss of contact area (debonding) (%)	Free energy of adhesion (dry)	Free energy of adhesion (wet)
AAD + Texas limestone	275	34	141	−67
AAM + Texas limestone	550	27	205	−31
AAD + Georgia granite	250	35	150	−48
AAM + Georgia granite	455	26	199	−30

4.1.3 Spontaneous Emulsification

Spontaneous emulsification is an inverted emulsion of water droplets in asphalt cement. Fromm (1974a, b) demonstrated how an emulsion forms and that once the emulsion formation penetrates the substrata, the adhesive bond is broken. Some research indicates that the formation of such emulsions is further aggravated by the presence of emulsifiers such as clays and asphalt additives (Asphalt Institute 1981; Fromm 1974a, b; Scott 1978). Fromm (1974a, b) observed that spontaneous emulsification occurs when asphalt films are immersed in water and that the rate of emulsification depends on the nature of the asphalt and the presence of additives. The cationic emulsifiers that produce the traditional oil-in-water emulsions are obtained by reacting fatty amines with dilute hydrochloric or acetic acid to produce an amine salt (Morgan and Mulder 1995). However, commercial amine-based asphalt additives, which are organic amine compounds, are chemically different from cationic asphalt emulsifiers, and they cannot function as an emulsifier in their amine form to make normal oil in water–asphalt emulsions.

Organic amines, which are basic nitrogen compounds, bond strongly to aggregates in the presence of water (Robertson 2000). Kiggundu (1986) demonstrated how the rate of emulsification is dependent on the nature and viscosity of asphalt, with an AC-5 (lower viscosity) emulsifying in distilled water much faster than an AC-10 (higher viscosity). He also demonstrated that the process is reversible upon drying.

4.1.4 Pore Pressure

Pore pressure development in water that is entrapped can lead to distress. Stresses imparted to the entrapped water from repeated traffic load applications will worsen the damage as the continued buildup in pore pressure disrupts the asphalt film from the aggregate surface or can cause the growth of microcracks in the asphalt mastic. Bhairampally et al. (2000) used a tertiary damage model developed by Tseng and Lytton (1987) to demonstrate that well-designed asphalt mixtures tend to "strain harden" on repeated loading. This "strain hardening" is of course not classical strain hardening that occurs when metals are cold-worked to develop interactive dislocations to prevent slip but is the "locking" of the aggregate matrix caused by densification during repeated loading. On the other hand, some mixtures exhibit microcracking in the mastic under heavy repeated loading. This results in progressive cohesive or adhesive failure, or both, and is evident in a plot of accumulated permanent strain versus number of load cycles as the rate of damage dramatically increases as microcracking progresses. The rate of this accelerated or tertiary damage is exacerbated in the presence of water as the pore pressure developed in the microcrack voids increases the rate of crack growth and damage through the development of higher pressures at the crack tip and through a weakening of the mastic and of the adhesive bond between the mastic and the aggregate.

Terrel and Al-Swailmi (1994) described the concept of pessimum air voids, which is the range of air void contents within which most asphalt mixtures are typically compacted (between about 8 and 10%). Above this level air voids become interconnected and moisture can flow out under a stress gradient developed by traffic loading. Below this value the air voids are disconnected and are relatively impermeable and thus do not become saturated with water. In the pessimum range, water can enter the voids but cannot escape freely and is thus subjected to pore pressure buildup upon repeated loading.

Figure 4.1 shows the distribution of deviator stress and concomitant damage around a void that is either unfilled or filled with water under the same applied load in a micromechanical analysis.

4.1.5 *Hydraulic Scour*

Hydraulic scour occurs at the pavement surface. Here stripping results from the action of tires on a saturated surface. Water is sucked under the tire into the pavement by the tire action. Osmosis and pullback have been suggested as possible mechanisms of scour (Fromm 1974a, b). Osmosis occurs in the presence of salts or salt solutions in aggregate pores and creates an osmotic pressure gradient that actually sucks water through the asphalt film. Researchers are mixed on support of

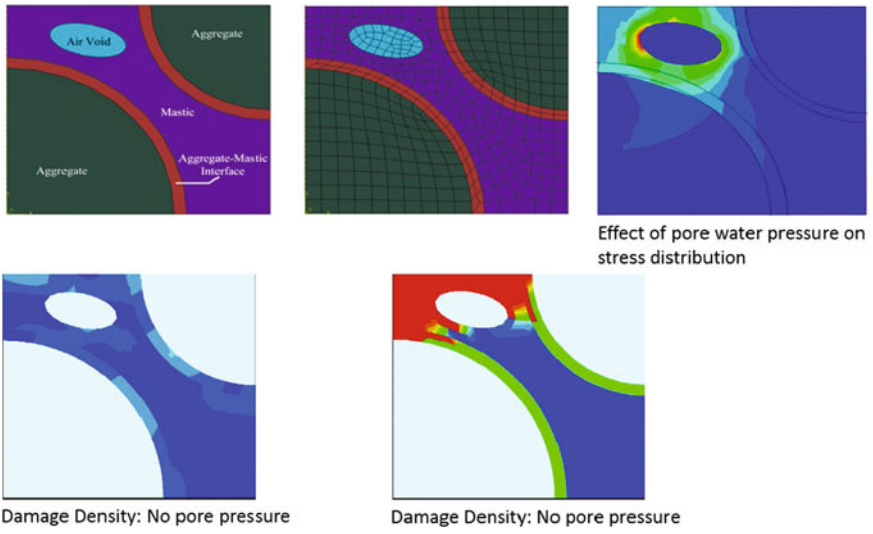


Fig. 4.1 Micromechanical model of pore pressure in an air void and concomitant damage due to mechanical loading. Damage increases as colors move from cool (*blue*) to hot (*red*)

this process. Mack (1964) supports it, while Thelen (1958) feels it is too slow to be valid. However, several factors support the potential occurrence of this mechanism, including the fact that some asphalts are treated with caustics during manufacture, some aggregates possess salts (compositionally), and asphalt films are permeable. In fact, Cheng et al. (2002) demonstrated that the diffusion of water vapor through asphalt cement itself is considerable and that asphalt mastics can hold a rather surprisingly large amount of water. Table 4.2 shows the comparison of the water vapor diffusion rates and the amount of water that can be accommodated by two compositionally very different asphalts (AAD-1 and AAM-1). Cheng et al. (2002) also showed that the amount of water held by these asphalts is related to the level of moisture damage that occurs in the mixtures using these asphalts.

4.1.6 pH Instability

Scott (1978) and Yoon (1987) demonstrated that asphalt–aggregate adhesion is strongly influenced by the pH of the contact water. Kennedy et al. (1984) investigated the effect of various sources of water on the level of damage that occurred in a boiling test. Fehsendfeld and Kriech (undated) observed that the pH of contact water affects the value of the contact angle and the wetting characteristics at the aggregate–asphalt interface region. Scott (1978) observed that the value of interfacial tension between asphalt and glass peaked at intermediate pH values, up to about 9, and then dropped as pH increased. Yoon (1987) found that the pH of contact water increased with duration of contact and was aggregate-specific and that the values stabilized after about 5–10 min of boiling. Yoon determined that the coating retention in boiling tests decreased as pH increased. Kiggundu and Roberts (1988) point out that these results indicate that stabilization of the pH sensitivity at the asphalt–aggregate interface can minimize the potential for bond breakage; provide strong, durable bonds; and reduce stripping.

Tarrer (1996) concluded that (a) the bond between asphalt and aggregate depends on surface chemical activity; (b) water at the aggregate surface (in the field) is at a high pH; (c) some liquids used as antistrips require a long curing period

Table 4.2 Effect of moisture-holding potential of asphalt on moisture damage in triaxial testing. After Cheng et al. (2002)

Parameter	Binder		Ratio, AAD-1/AAM-1
	AAD-1	AAM-1	
Diffusivity (m^2/s)	0.0008	0.0029	3.62
Water-holding potential, W_{100}, parts per 100,000	153	114	1.34
Percent debonding of binder from aggregate	23 (AAD/limestone) 35 (AAD/granite)	27 (AAM/limestone) 26 (AAM/granite)	1.26 1.35

(in excess of about 3 h) to achieve resistance to loss of bond at higher pH levels; and (d) it is possible to achieve a strong chemical bond between aggregate and asphalt cement that is resistant to pH shifts and a high pH environment. This strong chemical bond can be achieved by the formation of insoluble organic salts (such as calcium-based salts), which form rapidly and are not affected by high pH levels or pH shifts.

Although pH shifts affect chemical bonds, it is important to keep the magnitude of the pH shifts in proper perspective. Normally pHs as high as 9 or 10 will not dislodge amines from the acidic surfaces of aggregates, nor will they affect hydrated lime. Values of pH greater than 10 are not normally developed in asphalt mixtures unless a caustic such as lime is added. However, pHs below about 4 can dislodge amines from an aggregate surface and can dissolve lime depending on the type of acid used; these low pHs are not found in hot mix asphalt.

4.1.7 Environmental Effects on the Aggregate–Asphalt System

Terrel and Shute (1989) report that factors such as temperature, air, and water have a profound effect on the durability of asphalt concrete mixtures. In mild climates where good-quality aggregates and good-quality asphalt cements are available, the major contribution to deterioration is traffic loading and the resulting distress manifestations. Premature failure may result when poor materials and traffic are coupled with severe weather. Terrel and Al-Swailmi (1994) identify a number of environmental factors of concern: water from precipitation of groundwater sources, temperature fluctuations (including freeze–thaw conditions), and aging of the asphalt. They identify traffic and construction techniques, which are external to the environment, as important factors. Factors considered by Terrel and Shute to influence response of mixtures to water sensitivity are given in Table 4.3.

4.2 Adhesion Theories

Terrel and Shute (1989) describe four theories that are often used to explain the adhesion between asphalt and aggregate: (a) chemical reaction, (b) surface energy, (c) molecular orientation, and (d) mechanical adhesion. Most likely a combination of mechanisms occurs synergistically to produce adhesion, and no one theory describes adhesion. Terrel and Shute explain that the four theories are affected by the following factors: surface tension of the asphalt cement and aggregate, chemical composition of the asphalt and aggregate, asphalt viscosity, surface texture of the

Table 4.3 Factors influencing response of mixtures to water sensitivity. After Terrel and Shute (1989)

Variable	Factor
Existing condition	Compaction method
	Voids
	Permeability
	Environment
	Time
	Water content
Materials	Asphalt
	Aggregate
	Additives/modifiers
Conditioning	Curing
	Dry versus wet
	Soaking
	Vacuum saturation
	Freeze–thaw
	Repeated loading
	Drying
Other	Traffic
	Environment
	Age

aggregate, aggregate porosity, aggregate cleanliness, and aggregate moisture content and temperature at the time of mixing with asphalt cement.

4.2.1 Chemical Reaction

Chemical reaction is based on the premise that acidic and basic components of both asphalt and aggregate react to form water-insoluble compounds that resist stripping. Rice (1958) suggests the possibility of selective chemical reactions between aggregate and asphalt species. Jeon et al. (1988) described chemisorption of asphalt functionalities on aggregate surfaces and quantified the amount of coverage using a Langmuir model. Thelen (1958) had previously proposed that a bond formed by chemical sorption might be necessary in order to minimize stripping potential in asphalt–aggregate mixtures.

Robertson (2000) stated that overall polarity or separation of charge within the organic molecules promotes attraction of polar asphalt components to the also polar aggregates. He explains that while neither asphalt nor aggregate has a net charge, components of both form nonuniform charge distributions and behave as if they have charges that attract the opposite charge of the other material. As established by Curtis et al. (1992), this is confounded by the fact that aggregates vary substantially

in charge distribution and this charge distribution is affected by the environment. Robertson (2000) goes on to explain the types of reactions that might occur between the polar aggregate surface and asphalt cement. He states that, at a molecular level, basic nitrogen compounds (pyridines) adhere tenaciously to aggregate surfaces. He also describes carboxylic acids in asphalt cement. While they are quite polar and adhere strongly to dry aggregate, they tend to be removed easily from aggregate in the presence of water; but this varies with the type of acid. Plancher et al. (1977) explain that monovalent cation salts, such as sodium and potassium salts of carboxylic acids in asphalt, can be easily removed from the aggregate surface because they are essentially surfactants or soaps, which debond under the "scrubbing" action of traffic in the presence of water. On the other hand, Robertson (2000) indicates that divalent or doubly charged salts of acids (such as calcium salts from hydrated lime) are much more resistant to the action of water. This is also discussed by Scott (1978), Plancher et al. (1977), and Petersen et al. (1987). Robertson (2000) explains that aged asphalts may be much more prone to moisture damage than unaged asphalts. In some but not all asphalts, a very strongly acidic material appears with oxidation. Robertson (2000) explains that if asphalt acids are converted to sodium salts (as can happen with some aggregates), a detergent will form. However, calcium salts of detergents are far less moisture sensitive or are deactivated with lime.

4.2.1.1 Chemical Bonding Theory

Preceeding discussions on fundamental forces of adhesion introduced the concepts of physical and chemical bonding. Although the processes of chemisorption and physisorption are often characterized based on several features, they are not as distinguishable as it may seem (Butt et al. 2003; Kolasinski 2002). Broader terms such as primary and secondary interactions are sometimes used to refer to these stronger and weaker interactions, respectively. In this section, chemical bonding theory will include typical interactions between bitumen and aggregate that are specific in nature, as opposed to nonspecific interactions. The traditional view of chemical bonding in bitumen–aggregate systems is that these bonds produce new compounds upon formation.

Molecular species in bitumen primarily consist of long carbon chains and rings saturated with hydrogen, which are essentially nonpolar in character (Little and Jones 2003). These are the lightweight, oily or waxy fraction in bitumen (Robertson, 2000) called alkanes, paraffins, or aliphatic compounds. The inert character of these molecules stems from the fact that they are saturated (with hydrogen), and made up exclusively from single C–H and C–C bonds, with relatively balanced electron distributions and therefore little tendency to move around. These nonpolar molecules interact mainly through van der Waals forces (McMurry 2000). Because van der Waals forces are additive, their contribution in these large molecules is significant. While active chemical sites on aggregate surfaces promote bonding of particularly polar species in bitumen, covering of these sites by nonpolar

hydrocarbons can completely mask their activity (Curtis 1992). During mixing at high temperatures, however, more polar species should equilibrate with the surface and displace the more weakly adsorbed and nonpolar components on the aggregate surface (Petersen et al. 1982).

Bitumen–aggregate interaction chemistry is highly complex and variable among different systems primarily due to the complex and variable composition of the materials involved (Petersen et al. 1982). The polar molecules in bitumen exhibit specific points (sites), which interact with specific sites within the bitumen (Robertson 2000) and on aggregate surfaces. "Active sites" is a term used to describe reactivity of macromolecules and also the surfaces of minerals and implies a process wherein a surface chemical reaction of interest is promoted by a molecular-scale feature on the organic or mineral surface, i.e., surface functional groups (Johnston 1996).

4.2.1.2 Bitumen Functional Groups

Although bitumen is composed of nonpolar hydrocarbons, heteroatoms such as nitrogen (N), sulfur (S), and oxygen (O) may also exist as part of these molecules. Trace amounts of metals are also present and are considered "fingerprints' of the crude source (Robertson 2000). These atoms introduce polarity into bitumen molecules and, although only present in small amounts, have a controlling effect on the properties of the bitumen and its interaction with aggregate surfaces (Petersen et al. 1982). Petersen (1986) has identified polar, strongly associated functional groups in bitumen. Figure 4.2 shows the chemical structures of important functional groups in natural bitumen including those formed during oxidation. Benzene-like, unsaturated ring structures are common hydrocarbon compounds in crude oil and, along with alkanes, are typical constituents in the molecular make-up of bitumen (McMurry 2000). According to the historical micellar model, resins and ultimately asphaltenes represent the more polar fractions in bitumen. These fractions are also the higher molecular weight, or larger molecular size fractions. An asphalt (bitumen) model based on nonpolar and polar species, derived from molecular size distribution, has largely replaced the historic micellar model during Strategic Highway Research Program Research (SHRP) (Robertson 2000).

Many of the studies on bitumen–aggregate interaction focused on the affinity of different bitumen functional groups for aggregate surfaces. Adsorption of bitumen model compounds, representing specific functional groups in bitumen, has been investigated by researchers including Plancher et al. (1977), Petersen et al. (1982), Curtis et al. (1989), Arderbrant and Pugh (1991), Brannan et al. (1991), and Park et al. (2000). Relative affinity of functional groups for aggregate surfaces and their relative displacement by water are generally presented in these studies. Although rankings differ depending on the type of aggregate, or sometimes model-aggregate, similar general trends have been obtained. Table 4.4 summarizes selected findings for the average affinity of bitumen functional groups for aggregates, as well as their susceptibility for displacement by water. The importance of Table 4.4 is not the

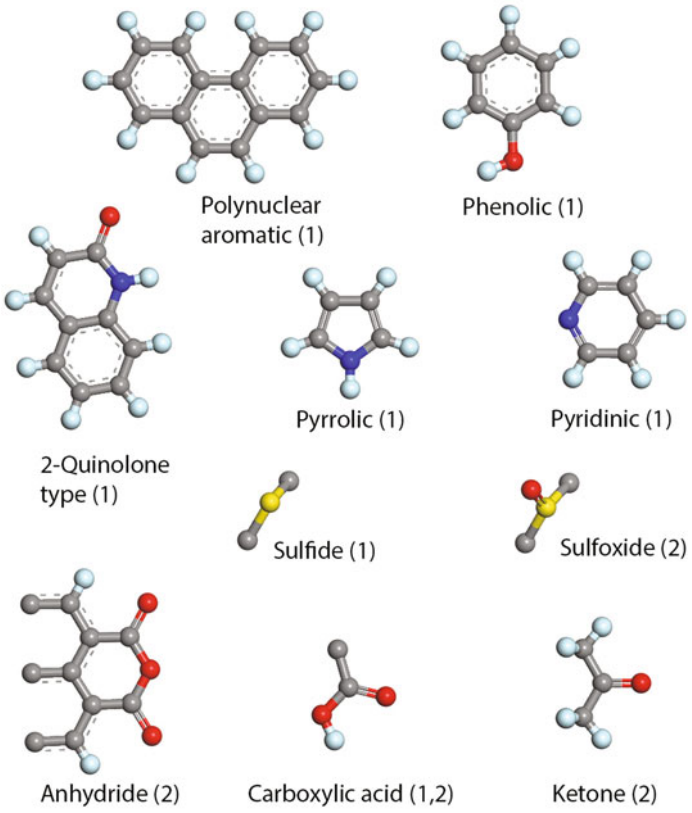


Fig. 4.2 Examples of important chemical functional groups (*1*) naturally occurring and (*2*) formed on oxidative aging. After Petersen (1986)

exact order, but rather the observation that these groups, which are most strongly adsorbed on aggregate surfaces, are also those displaced most easily by water. This observation emphasizes the fact that the influence of moisture on the durability of bitumen–aggregate bonds should be of primary concern from the starting point. Although minor components in bitumen, as a whole, it is essentially the acidic components such as carboxylic acids, anhydrides, and 2-quinolone types that are the most highly concentrated in the adsorbed fraction. This is in line with the fact that bitumen generally exhibits an excess amount of acidic compounds compared to the amount of basic organic compounds (Morgan and Mulder 1995).

Petersen and Plancher (1988) state that the two chemical functionalities, carboxylic acids and sulfoxides, account for almost half of the total chemical functionality in the strongly adsorbed fractions. These compounds are hydrophilic (water loving) with aliphatic structures (zigzag chains as opposed to aromatic ring structures) with no other polar functional groups on the same molecule (mono-functional as opposed to poly-functional). This may contribute to their ease

Table 4.4 General affinity for bitumen functional groups for aggregate surfaces

Plancher et al. (1977)[a]	Petersen et al. (1982)[a]	SHRP (Curtis and co-workers)[b]
Most strongly adsorbed functional groups (decreasing order)		
Carboxylic acids	Carboxylic acids	Carboxylic acids
Anhydrides	Anhydrides	Sulfoxides
2-Quinolones	Phenols	Pyrridine types
Sulfoxides	2-Quinolones	Phenolic
Pyrridine types	Sulfoxides	Pyrrolic
Ketones	Ketones	Ketones
	Pyrridine types	
	Pyrrolic	
Susceptibility of adsorbed functional groups for water displacement (decreasing order)		
Carboxylic acids	Anhydrides	Sulfoxides
Anhydrides	2-Quinolone types	Carboxylic acids
Sulfoxides	Carboxylic acids	Pyrrolic
Pyrridine types	Pyrridine types	Ketones
2-Quinolones	Sulfoxides	Pyrridine types
Ketones	Ketones	Phenolic
	Phenolic	
	Pyrrolic	

[a]Review by Petersen and Plancher (1998)
[b]Reported by Jamieson et al. (1995)

of displacement by water. Poly-functional bitumen molecules are more strongly adsorbed. Typical poly-functional molecules contain ketones, anhydrides, and nitrogen. Certain nitrogen compounds such as pyridine and pyrrolic types as well as phenolics are strongly adsorbed as suggested by water displacement studies. Desorption by solvents such as pyridine may not always produce reliable extraction of some of these strongly adsorbed chemical functionalities. The significance of adsorbed pyridine on aggregate surfaces was demonstrated by Petersen and Plancher (1988) by thermal desorption experiments. Pyridine could neither be desorbed at 240 °C, nor be displaced by water. In these studies, pyridine mimicked the basic nitrogen compounds in bitumen as well as the actions of common amine anti-stripping agents. Results suggest that it does matter which functional group bonds with the aggregate surface first. When the carboxylic acid model compound was coated onto the aggregate first, pyridine was readily displaced by water. Jamieson et al. (1995) report that competitive adsorption studies by quadric-component solutions of model compounds did not have a significant effect on the order of rank presented in Table 4.4.

4.2.1.3 Aggregate Functional Groups

Aggregates are composed of an assemblage of one or more minerals, which have a definite chemical composition and an ordered atomic arrangement. In these arrangements, or atomic lattices, each atom is bound to neighboring atoms through electrostatic coordination bonds. When aggregates are crushed or cleaved, the new surface atoms are deprived of some of their neighboring atoms and some of the coordination bonds are broken. These atoms seek to form new coordination bonds to replace the old broken ones (Thelen 1958). This general molecular description serves as the basis for the development of surface charge, previously described, as well as the concept of surface free energy, discussed subsequently as part of thermodynamic theory. New coordination bonds can be formed by directing some of the forces inward with consequent orientation adjustment of the crystal lattice as the atoms are pulled closer together. Another way to satisfy the broken bonds is to attract contaminants such as water and organics to the surface also serving as bonding sites for functional groups in bitumen.

Functional groups, bonding sites, or active sites on aggregate surfaces have not received as much attention as bitumen by researchers in this field. Silica is the most widely used surface in these studies due to its natural abundance and its association with moisture damage (Curtis et al. 1989; Ardebrant and Pugh 1991; Ernstsson and Larsson 1999; and Park et al. 2000). Petersen and Plancher (1998), in their review of the literature, provided insightful reference in this regard. Active sites on these surfaces range from surface hydroxyl groups of varying acidities to hydrogen bonding sites of high acidity. Strong electropositive Lewis acid bonding sites are typically formed on minerals where metals, such as magnesium, iron and calcium, are present. Oxygen in silica and other minerals can act as Lewis base sites. In addition to providing adsorption sites for many bitumen functionalities, oxide and hydroxyl surface functional groups act as strong adsorption sites for water molecules. In turn, water serves as bonding sites for organic molecules as mentioned previously. The latter can be bitumen or organic contaminants that may replace or cover adsorbed water molecules when aggregates are exposed to the environment over time. It has been reported that aggregate stored for long periods were found to resist stripping better than freshly cleaved rock (Endersky et al. 1947 and Thelen 1958).

Jamieson et al. (1995) report that in general, chemical sites on aggregate surfaces associated with high affinity for bitumen include elements such as aluminum, iron, magnesium, and calcium. Elements associated with low bonding affinity include sodium and potassium. The role of aggregate physical and chemical properties on net adsorption test results, including eleven aggregates and three bitumen types from the SHRP Materials Reference Library (MRL), was evaluated by stepwise regression analysis. Five aggregate variables were investigated as summarized in Table 4.5. The correlation coefficient of 0.9 indicates that the net adsorption is primarily a function of aggregate properties. The increased order of impact is therefore, potassium oxide, surface area, calcium oxide, zeta potential, and sodium oxide. While these results corroborate the impact of alkali metals (K and Na) and

Table 4.5 Physical and chemical properties of aggregates on net adsorption. After Jamieson et al. (1995)

Variable influencing net adsorption	Correlation coefficient
Potassium oxide	0.48
Surface area	0.71
Calcium oxide	0.75
Zeta potential	0.87
Sodium oxide	0.90

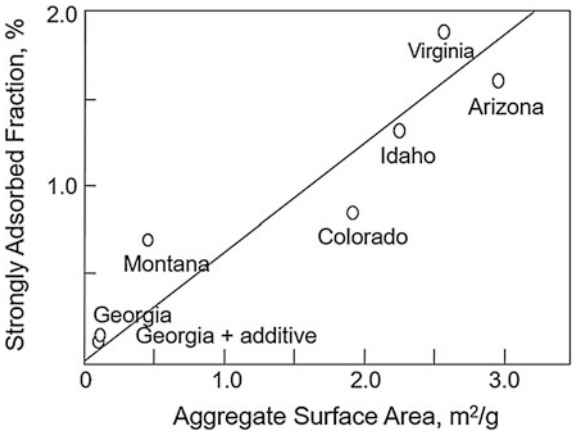


Fig. 4.3 Importance of active site density, in addition to surface area. After Petersen et al. (1982)

calcium, it also shows that part of the adsorption–desorption action can be explained by zeta potential, which will be discussed subsequently.

The contribution of surface area suggests the importance of availability of more active sites per unit mass of aggregate for interaction. In this regard, Petersen et al. (1982) illustrated that surface site density on aggregates is also important. The relationship between aggregate surface area and percentage strongly adsorbed fraction is furnished in Fig. 4.3. Although the Montana and the Colorado aggregates adsorbed almost the same amount of strongly adsorbed functionalities, the Colorado aggregate has a surface area four times that of the Montana aggregate. The researchers concluded that the Montana aggregate has a higher density of adsorption sites. This may also be observed when adsorption sites are not accessible due to steric hindrance of the bitumen functional groups.

4.2.2 Surface Energy and Molecular Orientation

From a simplistic viewpoint, surface energy may be described in terms of the relative wettability of aggregate surfaces by water or asphalt. Water is a better wetting agent than asphalt due to lower viscosity and a lower surface tension. However, the concept of using surface energy to calculate the cohesive strength of

the asphalt mastic and the adhesive bond energy between aggregate and the asphalt cement or between aggregate and the mastic is a much more complex subject that is worthy of a more detailed discussion. This is presented in sect. 4.6 entitled "*Nature of Asphalt–Aggregate Interaction*." Molecular orientation is coupled with surface energy because both are a part of a theory that considers structuring of asphalt molecules at the asphalt–aggregate interface and assumes that adhesion between asphalt and aggregate is facilitated by a surface energy reduction at the aggregate surface as asphalt is adsorbed to the surface (Hubbard 1938; Rice 1958; Sanderson 1952). Kiggundu and Roberts (1988) describe molecular orientation and surface energy as synergistic processes. They also mention the observations of researchers regarding surface phenomena. For example, Yoon (1987) and Tarrer (1996) described how aggregates that impart a relatively high pH to contact water or that have a relatively high zeta potential have a high stripping or debonding potential. Scott (1978) stated, "If water penetrates the asphalt film to the mineral aggregate surface under conditions where micro droplets form, the pH reached may be sufficient to ionize and dissociate adsorbed water molecules."

4.2.3 Mechanical Adhesion

This physical form of adhesion relies on physical aggregate properties, including surface texture, porosity or absorption, surface coatings, surface area, and particle size (Terrel and Al-Swailmi 1994). The philosophy is rather simple—to produce an aggregate surface capable of maximizing the surface area and texture to facilitate a strong physical bond that can synergistically improve the nature of the chemical bond between the asphalt and aggregate even in the presence of water.

Particle form, angularity, and surface texture of aggregates were analyzed in recent years using imaging techniques. An aggregate imaging system (AIMS) has been used at Texas A&M Transportation Institute for the characterization of these properties irrespective of size. A unique feature of this system is the ability to capture texture through wavelet analysis of black and white images. In wavelet analysis, signals or images are decomposed at different resolutions to obtain different scales of texture. Surface texture essentially captures the surface irregularity at a scale that is too small to affect the overall shape (Chadan et al. 2004). Masad et al. (2004) obtained the texture of limestone, gravel, and granite utilizing this imaging technique. The texture for fine fractions and coarse fractions was found to be different with all aggregates. The researchers presented an anisotropic viscoplastic continuum mechanics model, which was used to fit experimental data from compressive triaxial tests conducted at different confining pressures. One of the model parameters, a work hardening parameter, $\overline{\kappa}$, represents the effect of adhesion between the binder and aggregate or cohesion within the binder. In Figs. 4.4 and 4.5, the relationship between this parameter and aggregate texture

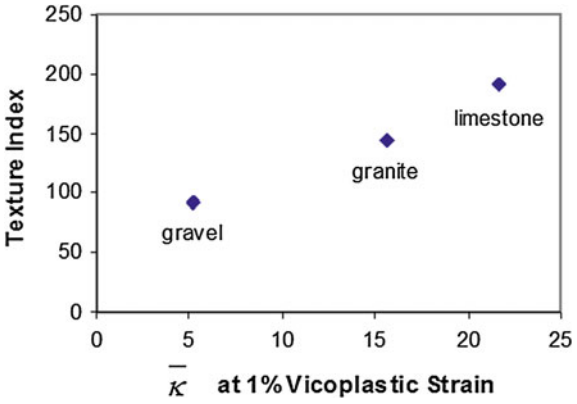


Fig. 4.4 Relationship between the hardening model parameter and fine aggregate surface texture. Data from Masad et al. (2004)

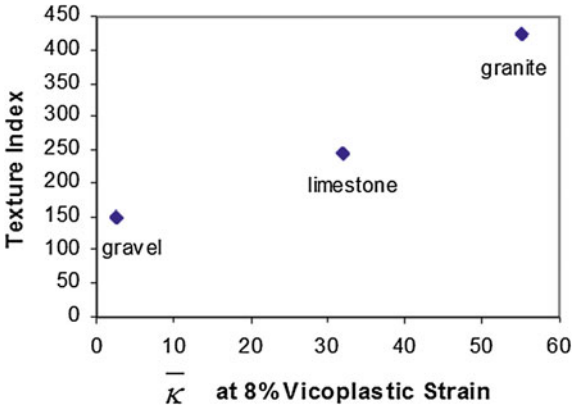


Fig. 4.5 Relationship between the hardening model parameter and coarse aggregate surface texture. Data from Masad et al. (2004)

suggests that it primarily represents adhesion in the mixture. In addition, it is shown that $\overline{\kappa}$ increases with an increase in applied pressure; the texture index of the fine fraction correlates with the initial $\overline{\kappa}$ at 1% viscoplastic strain, while the texture for the coarse aggregate correlates with the final $\overline{\kappa}$ at 8% viscoplastic strain. The coarse fraction of granite exhibited the highest texture index and, accordingly, the highest final $\overline{\kappa}$ value.

Although the effect of surface chemistry of the materials cannot be excluded, these correlations suggest that the fine fraction, with higher surface area, dominates initial resistance to lower induced shear strains, while coarse aggregate supposedly forms a more stable phase at higher induced shear strains and controls adhesion under these conditions.

4.3 Cohesion Theories

Cohesion develops in the mastic and is influenced by the rheology of the filled binder. Kim et al. (2002) describe how the resistance of a mastic to microcrack development is strongly influenced by the dispersion of mineral filler. Thus, the cohesive strength of the mastic is controlled not by the asphalt cement alone, but by the combination and interaction of the asphalt cement and the mineral filler. Terrel and Al-Swailmi indicate that water can affect cohesion in several ways, including weakening of the mastic due to moisture saturation and void swelling or expansion. Cohesion properties would logically influence the properties in the mastic beyond the region where interfacial properties dominate. A classic reminder of this is the work of Schmidt and Graf (1972) who showed that an asphalt mixture will lose about 50% of its modulus upon saturation. The loss may continue with time, but upon drying, the modulus can be completely recovered. This is shown graphically in Fig. 4.6

Cheng et al. (2002) describe the severe weakening of asphalt mixtures when they are subjected to moisture conditioning. Figure 4.6 illustrates the strength loss in a repeated load triaxial test when subjected to 85% saturation. Cheng et al. (2002) indicates that this strength loss is predictable when one compares the wet adhesive bond strength between the asphalt and the aggregate with the much higher dry adhesive bond strength. But Cheng et al. (2002) go on to demonstrate that the rate of damage in various mixtures is also related to the diffusion of water into the asphalt mastic, and that the asphalts that hold the greatest amount of water accumulate damage at a faster rate (Fig. 4.7).

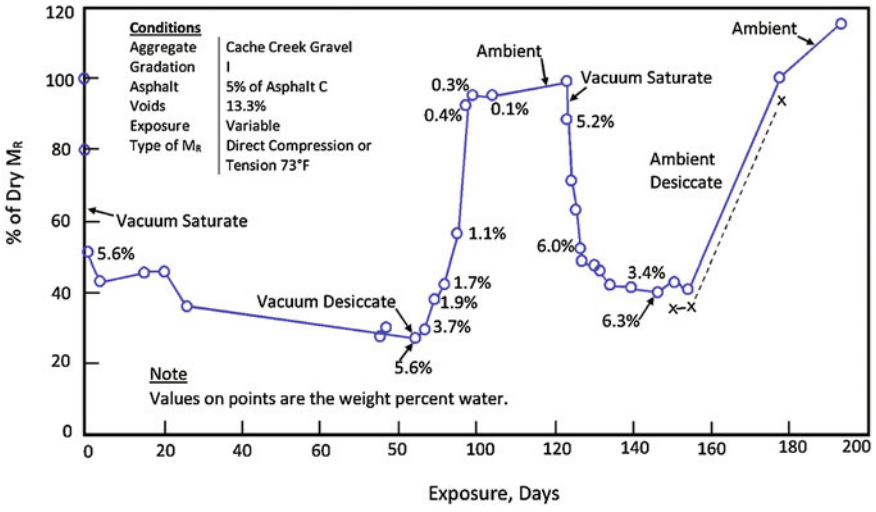


Fig. 4.6 Effect of moisture on resilient modulus is reversible. After Schmidt and Graf (1972)

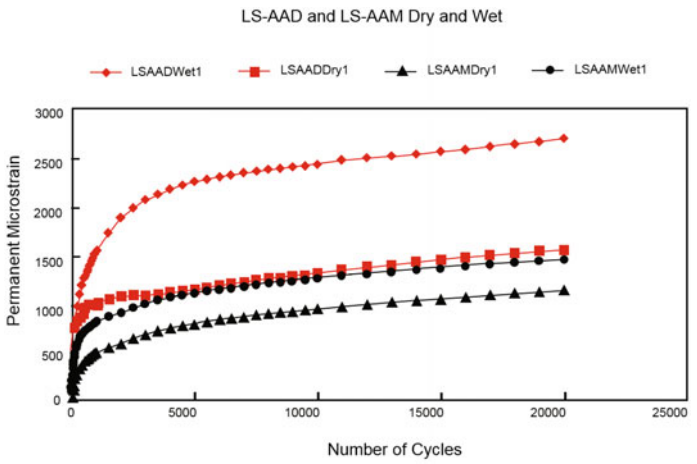


Fig. 4.7 Repeated load permanent deformation experiments for AAD-limestone (LSAAD) and AAM-limestone (LSAAM) in dry and wet conditions. After Cheng et al. (2002)

4.4 Combining Theories

Kiggundu and Roberts (1988) attempted to combine some of the theories discussed above. They realized that no single theory properly explains moisture damage. They provided the primary and secondary contribution relationships in Table 4.6. This table attempts to relate theories that explain loss of adhesion to stripping mechanisms. For example, the mechanism of pH instability is, according to Kiggundu and Roberts, explained by both chemical reaction theory and physical and chemical components of interfacial energy theory. Detachment, as a second example, is

Table 4.6 Proposed relationship between theories of adhesion bond loss and stripping mechanisms. After Kiggundu and Roberts (1988)

Proposed operating mode (mechanism)	Mechanical interlock			Chemical reaction			Interfacial energy		
Detachment	P	C	P–C	P	C	P–C	P	C	P–C
Displacement	S						S	W	
Spontaneous emulsification					S		S		
Film rupture	S			S	W				
Pore pressure	S								
Hydraulic scouring	S								
pH instability					S				S

P Physical
C Chemical
P–C Physical and chemical
S Primary contributor
W Secondary contributor

believed by Kiggundu and Roberts to be explained by physical and chemical aspects of interfacial energy theory as well as physical aspects of mechanical interlock theory. The physical aspects are manifested, according to Kiggundu and Roberts, by surface energy, while the chemical aspects are contributed by the effect of polarity of the molecules present at the common boundary. Even with this attempt to simplify the interaction of different theories and mechanisms, the interactive complexity of the processes becomes clearly evident. For example, surface bond is not solely a physical process because surface bond is dictated by the chemical nature of bonding at the asphalt and aggregate surface as well as by the presence of broken bonds or incomplete coordination of atoms due to broken bonds resulting in an increase in free energy.

4.5 Nature of Asphalt–Aggregate Interaction

4.5.1 Adhesive Failure Versus Cohesive Failure

Damage in asphalt mixtures can occur within the mastic (cohesive fracture) or at the aggregate–mastic interface (adhesive fracture or failure). Whether or not a cohesive or adhesive failure occurs depends on the nature of the mastic and the relative thickness of the mastic around the coarse and fine aggregate. Lytton (personal communication, 2002) used micromechanics to assess the "thickness" of the asphalt film at which adhesive failure gives way to cohesive failure. Figure 4.8 is a plot of the cohesive and adhesive bond strength determined from cohesive and adhesive surface energies versus thickness of the asphalt binder or mastic. The theory essentially states that asphalt mixtures with thin asphalt films fail in tension by adhesive bond rupture, while those with thicker asphalt films (or mastic films) fail because of damage within the mastic (cohesive failure) as opposed to interfacial

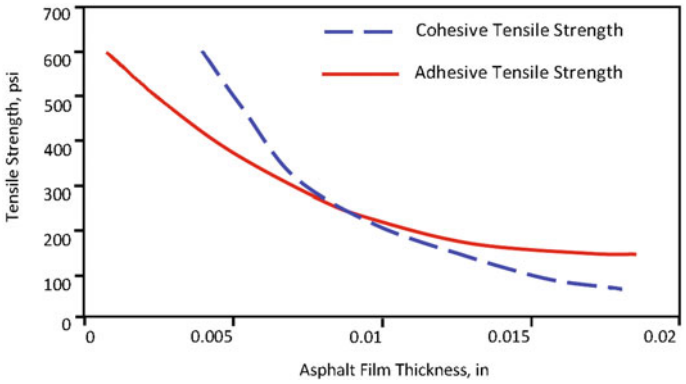


Fig. 4.8 Adhesive versus cohesive bond failure based on asphalt film thickness

debonding. The thickness that differentiates these two types of failure is dependent on the rheology of the asphalt (or mastic), the amount of damage the asphalt or mastic can withstand prior to failure, the rate of loading, and the temperature at the time of testing.

Consider an example. According to Fig. 4.8, when asphalt or mastic coatings are thin, adhesive strength controls performance. In this stage, the adhesive bond strength in the presence of water determines mixture strength and is the critical condition. On the other hand, when asphalt or mastic coatings are relatively thick, thicker than the transition point of Fig. 4.8, cohesive properties limit or control damage resistance. Therefore, in this situation, the impact of moisture intrusion into the mixture may be the key to assessing moisture damage of the mixture. In this case it may be more important to consider the impact of how much moisture the mastic holds and the impact on rheology of this infused water (Table 4.2) than to consider adhesive bond strength in the presence of water.

Realistically, it is necessary to consider both adhesive bond and cohesive strength properties of the mixtures as mastic coatings range from relatively thin to relatively thick throughout the mixture. Mixture strength then becomes a question of the statistical distribution of conditions where adhesion or cohesion strengths control. Thus, the effects of moisture on adhesive bond strength and cohesive mastic strength are perhaps equally important. Kim et al. (2002) used dynamic mechanical analysis (DMA) testing to evaluate the rate of damage in asphalt binders and mastics. DMA subjects a cylindrical asphalt mastic to cyclic, torsional loading until failure occurs. Failure is due to the development and propagation of microcracks, which begin at the outer circumference of the cylindrical sample. Kim et al. (2002) demonstrated that the rate of damage and the amount of damage various mastics can accumulate before failure depend on the nature of the mastic. This is critically important because it essentially states that a mastic that is well designed will tolerate more damage prior to failure than one that is not. Kim et al. showed that mastics with the proper amount of and type of filler can accommodate more damage prior to failure than unfilled systems and that polymer-modified systems can accommodate more damage prior to failure than unmodified systems. This indicates that the nature of the mastic (and the impact of the filler or modifier) strongly affects moisture damage because it helps control whether a cohesive or an adhesive bond failure occurs. Figure 4.9 is a plot of accumulated dissipated pseudostrain energy (DPSE) versus number of torsional DMA load cycles to failure of representative mastics. The filled asphalts or mastics (with hydrated lime (HL) or limestone (LS)) allow a higher accumulation of DPSE prior to failure than neat asphalt (AAD or AAM). The type of filler and its physicochemical interaction with the asphalt have a strong impact (Lesueur and Little 1999). The proposed mechanism of fatigue life extension is that properly designed mastics or modified asphalts affect the rate of microdamage accumulation and resist rapid, catastrophic failure via microcrack coalescence. The process may be a redistribution, redirection, or "pinning" of crack energy.

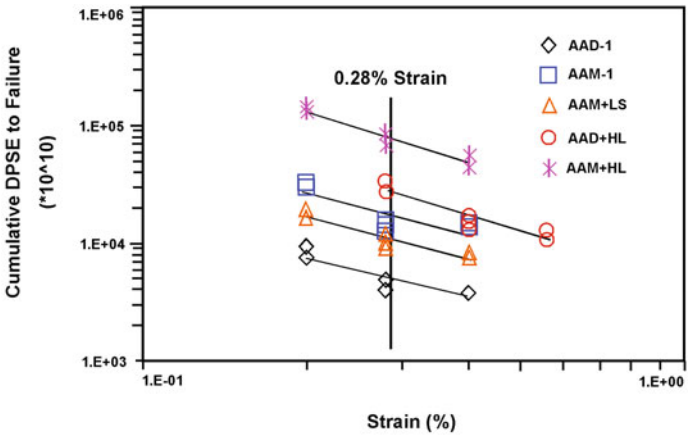


Fig. 4.9 Relationship between cycles to failure (DMA) and DPSE for various mastics: asphalts AAD-1 and AAM-1 and asphalts AAD with hydrated lime (AAD-HL) and AAM with hydrated lime (AAM-HL) and with limestone (AAM-LS). After Kim et al. (2002) units for dissipate pseudo-strain energy (DPSE) are Pacals

Little et al. (1999) have shown that microcrack damage rates are related to cohesive mastic surface energies based on Schapery's viscoelastic fracture theory. Cohesive bond strength can be calculated on the basis of cohesive surface energy measurements of the dry or water-saturated mastic (Cheng et al. 2002). Moisture intrusion weakens the cohesive bond and makes the resulting mixture more susceptible to damage (Table 4.2).

4.5.2 Effect of Aggregate Characteristics

A general hypothesis has been that acidic aggregates are hydrophobic while basic aggregates are hydrophilic. However, there are notable exceptions (e.g., Majidzadeh and Brovold 1968; Maupin 1982). The general conclusion is that few if any aggregates can completely resist the stripping action of water (Tarrer and Wagh 1991).

Tarrer and Wagh (1991) list a number of factors that influence the asphalt–aggregate bond: surface texture, penetration of pores and cracks with asphalt, aggregate angularity, aging of the aggregate surface through environmental effects, adsorbed coatings on the surface of the aggregate, and the nature of dry aggregates versus wet aggregates.

Surface texture of the aggregate affects its ability to be properly coated, and a good initial coating is necessary to prevent stripping (Maupin 1982). Cheng et al. (2002) demonstrated that the adhesive bond, calculated from basic surface energy measurements of the asphalt and aggregate, between certain granites and asphalt was higher than between limestone aggregate and asphalt when the bond was quantified as energy per unit of surface area. However, when the bond was

quantified as energy per unit of aggregate mass, the bond energy was far greater for the calcareous aggregates than for the siliceous. These results agreed well with mechanical mixture testing. They point out clearly the importance of the interaction of the physical and the chemical bond. Besides the importance of a good mechanical bond promoted by an amenable surface texture, stripping has been determined to be more severe in more angular aggregates (Gzemski et al. 1968) because the angularity may promote bond rupture of the binder or mastic, leaving a point of intrusion for the water. Cheng et al. (2002) have shown that, regardless of the strength of the bond between the asphalt and aggregate, the bond between water and aggregate is considerably stronger. Table 4.7 shows adhesive bond strengths calculated in ergs/cm^2 for five different liquids or semisolids (four binders and water) and three different aggregates. Note that the bond between water and either of the aggregates is at least 30% stronger than for any of the asphalts.

The effects of crushing of the aggregate are very interesting. One might expect that a freshly crushed aggregate surface would have a greater free energy than an uncrushed aggregate surface. This is because broken bonds due to fracture should substantially increase the internal energy even though having somewhat of a counter effect on Gibb's free energy due to an entropy increase. However, there is another side to consider. Tarrer and Wagh (1991) point out that sometimes newly crushed faces tend to strip faster than stockpiled aggregates. They state that it is characteristic of many aggregates that one or more layers of water molecules become strongly adsorbed on the aggregate surface as a result of electrochemical attractions. Thelen (1958) states that upon aging, the outermost adsorbed water molecules may become partially replaced or covered by organic contaminants present in air (e.g., fatty acids and oils) that reduce stripping potential. However, this seems unlikely because these fatty acids are relatively heavy and are not likely to volatilize. A general oxidation process reduces free radicals at the oxidation sites and may make weathered aggregates more resistant to stripping than freshly crushed aggregates. On the other hand, if the freshly crushed aggregate can be effectively coated with asphalt and the adsorption of the water layer can be prevented from the outset, the asphalt–aggregate bond developed may be the most effective. Certainly there is much room for advancement in the state of knowledge here.

Tarrer and Wagh (1991) and Hindermann (1968) discuss the effect of crushed aggregate surface on bond strength in light of the ways the aggregate surface may react to broken bonds created by crushing or cleavage. They discuss two potential

Table 4.7 Adhesive bond energy per unit area of sample (ergs/cm^2). After Cheng et al. (2002)

Binder	Aggregate		
	Georgia granite	Texas limestone	Colorado limestone
AAD-1	153	141	124
AAM-1	198	205	179
Rubber asphalt	212	189	166
Aged rubber asphalt	171	164	145
Water	256	264	231

reactions. In one, new coordination bonds may be formed by redirection inward to the atomic lattice. If this were the case, the aggregate would have no affinity for asphalt or water. This is a very unlikely process. In the second and more likely process, water, oil, or other contaminants in the air are attracted to the fresh surface to satisfy broken bonds. Since water is normally available, the driving force for the adsorption of water on the freshly crushed aggregate faces is that it reduces the free energy of the system. Although asphalts and other organics may also spread over the crushed faces of the aggregate, the rate at which they spread depends largely on their viscosity. Water is more prevalent and spreads much more quickly (Tarrer and Wagh 1991). Apparently, asphalt and organic materials spread over water films on an aggregate surface and tend to be stripped from water films by water (Tarrer and Wagh 1991), further complicating the process.

Clearly, Tarrer and Wagh 1991 make the case that heating aggregates that contain free water and adsorbed water films may remove free water and the outermost adsorbed water molecules, causing the interfacial tension between the aggregate and the asphalt to decrease (Thelen 1958; Majidzadeh and Brovold 1968), resulting in a reduction in stripping potential. The heating effect probably also reduces asphalt viscosity and allows better penetration into the aggregate surface, promoting a more effective physical bond.

According to Tarrer and Wagh 1991, the asphalt–aggregate bond is enhanced by three processes: (a) preheating the aggregate, (b) weathering the aggregate, and (c) removing aggregate coatings. When the aggregate surface is heated, the outermost adsorbed water layer is released, improving the state of interfacial tension between the asphalt and aggregate and, in turn, improving the bond between asphalt and aggregate. The weathering process results in a replacement of the adsorbed water layer with organic fatty acids from the air. This results in an improved asphalt–aggregate bond (Fromm 1974a, b). A dust coating on the aggregate surface promotes stripping by preventing intimate contact between the asphalt and aggregate and by creating channels through which water can penetrate (Castan 1968).

Cucalon (2016) used a microcalorimeter to measure the enthalpy of immersion, ΔH, when asphalt binder was introduced to a uniform gradation of fine aggregate. The enthalpy of immersion can be considered the bond energy between the binder and aggregate and is calculated by integrating to find the area under a heat flow versus time plot from the microcalorimeter. Figure 4.10 presents heat of reaction versus time of reaction for two different mineral aggregates (gabbro and limestone) with the same asphalt binder at two different relative humidities (RH): dry with a RH of approximately zero and RH = 33%. Based on the measured specific surface area of each aggregate, the RH of 33% was found to provide sufficient moisture to allow the development of a monolayer of water over the aggregate particles.

It is clear from Fig. 4.10 that a strong exotherm develops when the aggregate is dry representing a good and strong bond between the binder and the aggregate surface. However, at RH = 33%, the exotherm is virtually nonexistent presumably due to the monolayer of water. This is verification of the impact of hydropilicity of the aggregate surface and the effect of the presence of even a minute amount of water on binder-aggregate bond.

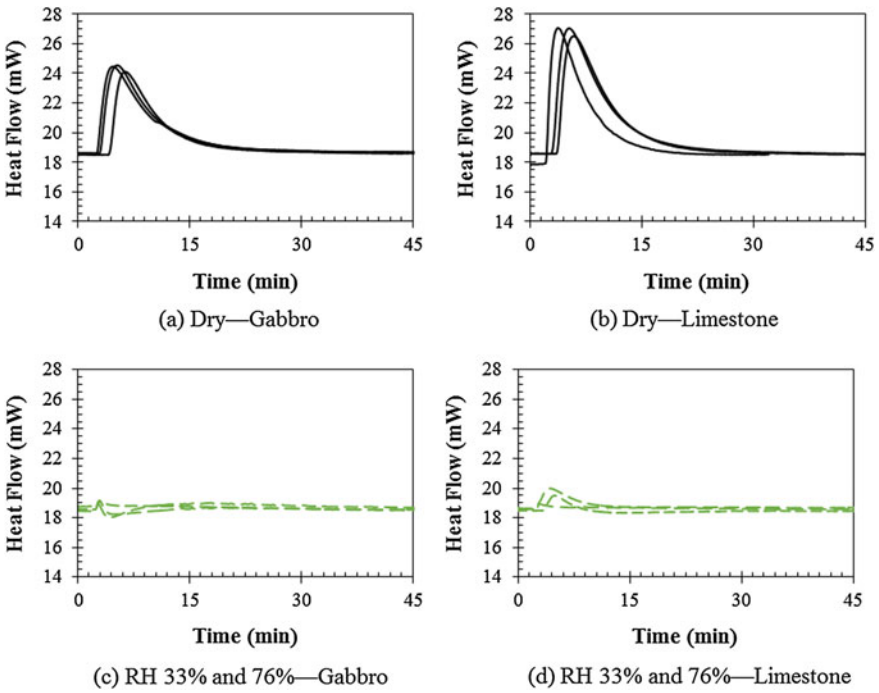


Fig. 4.10 Enthalpy of immersion of binder and aggregate with aggregate subjected to a specific thermal regime. After Cucalon (2016)

This is further verified by Cucalon 2016 when she compared water affinity for two aggregates (Gabbro and Limestone) with four different additives: PG 64-22 control grade, warm mix asphalt containing a wax additive, WMA1; warm mix asphalt containing an amine additive, WMA2; and water. Water is by far the strongest work of adhesion, Fig. 4.11.

4.5.3 Calculation of Asphalt–Aggregate Bond Strength

4.5.3.1 Fundamental Mechanisms

In 1984 Schapery proposed a basic viscoelastic fracture theory, which was derived from first principles of materials science and based on an energy balance. This theory states that the load-induced energy that causes fracture damage is balanced by the energy stored on newly formed crack faces. The energy imparted to the system can be quantified as the product of two properties of the materials in question: tensile creep compliance over the time of loading and the strain energy per unit of crack area produced from one tensile load to the next. The energy stored on fracture faces can be quantified by surface energy measurements of the material.

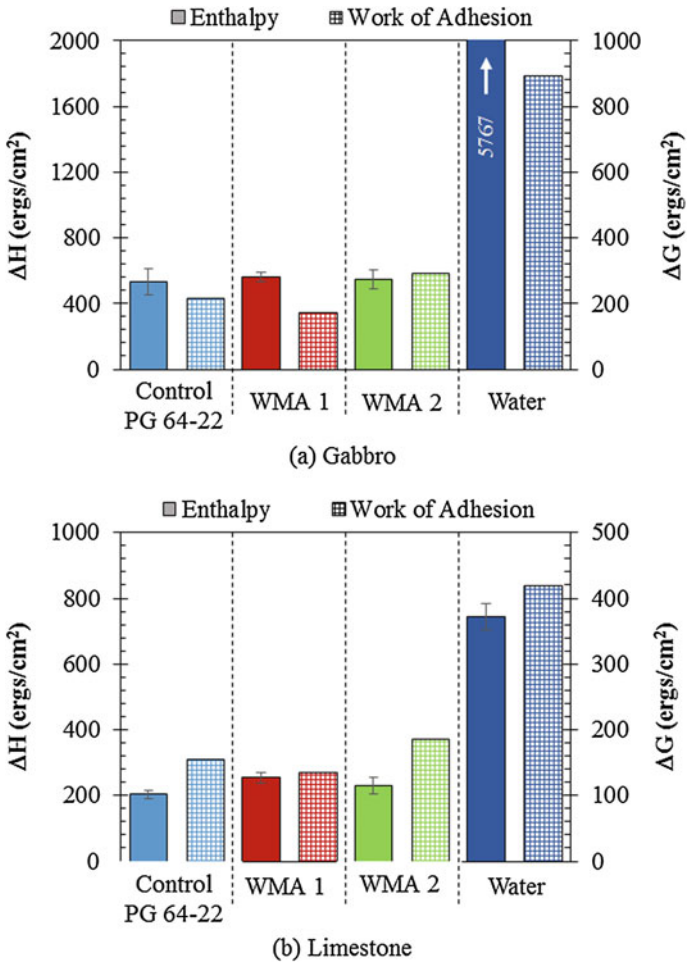


Fig. 4.11 Enthalpy of immersion (ΔH) and work of adhesion (ΔG) for gabbro and limestone with four binders. After Cucalon (2016)

Fortunately, the material properties required to assess this energy balance can be effectively measured. Si et al. (2002) and Kim et al. (2002) demonstrate how to measure tensile creep compliance and the strain energy that causes damage (pseudostrain energy) during cyclic fatigue testing of asphalt mixtures. This concept of pseudostrain energy is not mysterious; it is merely a mathematical calculation that allows one to separate the dissipated energy that actually causes damage from the energy that is recovered over time and does not cause damage. Surface energies can also be measured. Cheng et al. (2001, 2002) have demonstrated how surface energy measurements on the aggregate and asphalt cement can be used to calculate surface energies of cohesion (related to fracture within the mastic—asphalt and filler) and adhesion (related to fracture at the asphalt–aggregate interface).

Using this fundamental look at fracture damage, it is easy to relate surface energy to pavement distress and to understand the wide-ranging importance of surface energy as an indicator of distress in asphalt pavements. Obviously, surface energy can be used to directly assess fracture potential: both cohesive and adhesive. But surface energy is also related to permanent deformation distress, the fatigue failure process, and cohesive strength reduction and adhesive failure (stripping) in the presence of moisture. It is important to briefly develop this connection. As previously described, Bhairampally et al. (2000) used a tertiary damage model developed by Tseng and Lytton (1987) to demonstrate that well-designed asphalt mixtures tend to "strain harden" upon repeated loading.

As previously discussed, microcracking or tertiary damage leads to a departure from the typical "strain hardening" stress–strain curve representing an accelerated rate of damage due to the development of microcracking and the ultimate acceleration of microcrack growth. Cheng et al. (2002) showed that the acceleration in damage, or tertiary damage, is related to cohesive and adhesive bond strengths of the mastic and asphalt mixtures in question. Table 4.1 presents the strong relationship between the number of cycles to failure in repeated load permanent deformation testing and cohesive and adhesive bond energies (which were calculated from bitumen and aggregate surface energy measurements). In Table 4.8 the free energy of adhesion in the presence of water is calculated. The negative sign indicates a preference of the aggregate for water over asphalt, and a less negative value represents a lower driving force to replace the asphalt in question with water. Thus, it is consistent that asphalt AAM bonds more strongly with either the limestone or granite aggregate than asphalt AAD and that it is less likely to strip.

Two back-to-back studies for the Federal Highway Administration performed at Texas A&M University's Texas A&M Transportation Institute have established the importance of the healing phenomenon in the fatigue damage process. Field validation of healing that occurs during rest periods was presented by Williams et al. (1998). Here the researchers measured a substantial recovery in modulus via surface wave measurement following rest periods. Little et al. (1999), Kim et al. (1997), (2002), and Si et al. (2002) measured the healing effect during repeated load tensile and torsional shear fatigue testing. They quantified the effect of healing in terms of recovery of dissipated energy during the rest period and in terms of extended fatigue life due to the cumulative effect of a number of rest periods. Little et al. (1999) further established that the healing process.

Table 4.8 Gibbs free energy per unit mass (ergs/gm $\times$ 10^3). After Cheng et al. (2002)

Binder	Aggregate		
	Georgia granite	Texas limestone	Colorado limestone
AAD-1	158	614	375
AAM-1	206	889	536
Rubber asphalt	219	819	497
Aged rubber asphalt	178	714	435

The fact that surface energy of dewetting is fundamentally related to fracture and that surface energy of wetting is fundamentally related to healing is discussed by Little et al. (1998) and Little et al. (1999). In fact, Schapery presented a corollary to his viscoelastic fracture theory for healing in which he related healing to surface energy and found that an increase in surface energy resulted in better or faster crack closure. After studying the results of a large experimental matrix comparing surface energy with healing rate plots, Lytton proposed a short and long term component of healing, where Liftshitz-van der Waals controlled short term healing and acid base component controlled long term healing. When this is coupled with Schapery's theory of viscoelastic fracture, a much more complete understanding of the entire fracture fatigue process is achieved, because the fatigue process consists of fracture during loading and healing during rest periods between load applications. Lytton et al. (1993) showed that the healing process is responsible for the major component of the laboratory-to-field fatigue shift factor. Since this shift factor historically ranges between about 3 and more than 100, healing is indeed a significant part of the fatigue damage process.

A logical extension can be made from understanding adhesive fracture based on surface energy to understanding the debonding process between bitumen and aggregate in the presence of moisture (stripping). Cheng et al. (2002) presented a detailed methodology by which to measure the surface energies (all components) of asphalt using the Wilhelmy plate method and the surface energies of aggregates (all components) using the universal sorption device (USD). They then showed how to compute the adhesive bonding energy between the bitumen and the aggregate both in a dry state and in a wet state (in the presence of a third medium—water). Table 4.7 (Cheng et al. 2002) demonstrates that the adhesive bond calculated per unit area of aggregate is highly dependent on the aggregate and asphalt surface energies and that the values of the adhesive bond vary over a significant range. They further point out that the affinity of the aggregate for water is far greater than it is for asphalt, so that if water has access to the aggregate interface it is likely to replace the asphalt (strip), and the rate of replacement is a function of the aggregate–asphalt bond strength. In Table 4.8 the same results are presented in terms of energy per unit mass instead of energy per unit area. Energy per unit mass takes the surface area into account. This is shown to be very important as the rank order of adhesive bond energy changes when this conversion is made. The far greater surface area of the limestone ranks it ahead of the granite in terms of bonding energy per unit mass even though this particular granite actually has a higher bonding energy per unit area.

Cheng et al. (2002) also measured the diffusivity and moisture-holding potential of various bitumens using the USD. Lytton developed a method by which to solve Fick's second law to differentiate adsorption from absorption in the sorption process so that diffusivity and moisture-holding potential could be determined. Using this approach, Cheng et al. (2002) found that diffusivity and water-holding potential vary significantly among bitumens and that the ultimate water-holding potential is related closely to damage. For example, asphalt AAD-1 has a lower diffusivity than

asphalt AAM-1, but it has much greater water-holding potential (Table 4.2). This leads to a much higher level of damage in mixtures with AAD-1 than in mixtures with AAM-1. This damage may be due to two factors: the weakening of the mastic due to diffusion of water into the bitumen and the migration of water through the mastic to the mastic–aggregate interface causing stripping.

4.5.3.2 Fundamental Principles Shared by Material Processes

The preceding section points out that surface energy can be fundamentally related to material processes and failure mechanisms. From this background a set of principles can be established that can be used to measure material properties required to assess the basic distress processes. These principles are as follows:

1. All materials have surface energies, both asphalts and aggregates.
2. All surface energies have three components, all of which can be measured.
3. The theory of adhesive and cohesive bonding has been developed in industrial surface chemistry and chemical engineering and is used reliably and with confidence.
4. Fracture and healing involve two types of material properties: chemical and physical. Neither fracture nor healing can be properly described without the use of both properties: chemical—surface energies; physical—modulus and tensile strength and the way they change with age.
5. The presence of moisture at the asphalt–aggregate interface interrupts the bond and accelerates the rate of fracture damage. The presence of moisture in the mastic reduces cohesive strength and fracture resistance and, therefore, reduces the potential for microcracks in the mastic to heal.
6. On the basis of their surface energy characteristics, some combinations of aggregates and asphalts can be determined by calculations to bond well and heal well, whereas other combinations will not. In selecting materials for an asphalt pavement mixture from among several available alternatives, it is possible to select the best combination of all of the available aggregates and asphalts to resist fracture, heal effectively, bond well, and resist moisture damage. Predicting their performance will require the measurement of physical properties as well.

4.6 Thermodynamic Approach

An extended discussion of the thermodynamic approach is presented by Hefer et al. (2007). This section summarizes that explanation. This theory, also called adsorption theory, is the most widely used approach in adhesion science as indicated by most comprehensive references on this subject, such as by Schultz and

Nadrin (1994). Thermodynamic theory is based on the concept that an adhesive will adhere to a substrate due to established intermolecular forces at the interface provided that intimate contact is achieved. The magnitude of these fundamental forces can generally be related to thermodynamic quantities, such as surface free energies of the materials involved in the adhesive bond. The orientation of polar molecules in bitumen as part of the process to minimize the free energy at the interface has been recognized and discussed in previous reviews on stripping in bitumen–aggregate systems (Hicks 1991; Kiggundu and Roberts 1988; Little and Jones, 2003; Rice 1958).

Thermodynamics is the study of energy changes. A spontaneous process is one that occurs on its own, without external assistance. Such a process occurs due to an imbalance between two natural tendencies. The first tendency is the spontaneous conversion of potential energy into work and heat, known as enthalpy. The second tendency is the spontaneous increase in randomness of the system, known as entropy (Kittrick 1977). In order to relate enthalpy and entropy, the Gibbs free energy (G) is defined so that, at constant temperature (Kelvin) and pressure (atm),

$$\Delta G = \Delta H - T\Delta S \tag{4.1}$$

Gibbs free energy therefore represents the difference between the initial and final energy state of a system and predicts whether a process, carried out at the defined constant temperature and pressure conditions, can occur or cannot occur.

Enthalpy (ΔH) is the total heat given off or absorbed during this process. Consider the interaction between two chemical entities, such as two ions. The potential energy of attraction between two isolated ions with opposite charge will decrease as they approach each other at a temperature of absolute zero. This change in energy as a function of separation distance is illustrated in Fig. 4.12. A negative sign indicates that energy is given off in the process. When the two entities are sufficiently close, their electron clouds overlap and cause potential energy of repulsion to increase rapidly. The resultant of the attractive and repulsion curves indicates the existence of a state of minimum energy, where the entities will reside in a more stable, equilibrium state. The equilibrium distance (r_o) is therefore indicated by the position of the "potential energy well," while the energy given off is given by the depth (d) of the well.

For $\Delta G < 0$, the energy will be released to perform work and the process will occur spontaneously. If, on the other hand $\Delta G > 0$, then energy will have to be absorbed from the environment and cannot occur on its own. For $\Delta G = 0$, the process is in equilibrium. Equation (4.1) indicates that the absolute temperature, T, must be specified due to temperature dependency of entropy. Fortunately, the magnitudes of H and S do not vary much over a limited temperature range. The free energy, however, may vary considerably with temperature unless S is very small. At relatively low temperatures, such as room temperature, ΔH usually dominates while ΔS becomes more important at high temperatures (Kittrick 1977).

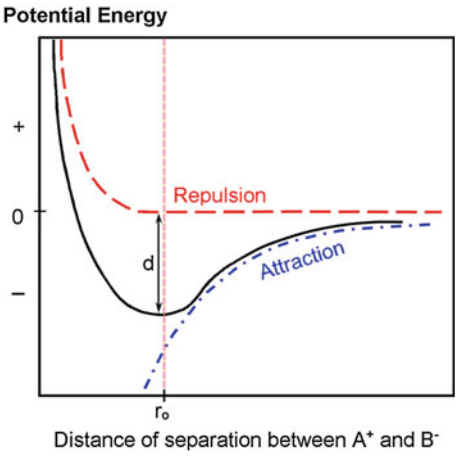


Fig. 4.12 Potential energy as a function of separation distance at a temperature of absolute zero

Gibbs free energy is an important thermodynamic parameter in quantifying adhesive bonding. It should be pointed out that the Gibbs free energy in this context is the excess free energy of the system associated with the surface or interface. A detailed account of the derivation of thermodynamic parameters related to surfaces is presented by Adamson and Gast (1997). It is common to refer to the Gibbs free energy as the free energy of adhesion or, when applicable to the binding between similar phases, the free energy of cohesion.

In a thermodynamic sense, surface energy, γ, is the reversible work required to create a unit area of new surface. While researchers such as Good and van Oss (1991) base their theory on the definition of the Gibbs free energy, the work of adhesion (W_a) is more commonly used in the literature pertaining to the thermodynamic theory of adhesion. Although equal in magnitude, the work of adhesion and the Gibbs free energy of adhesion should be interpreted as follows:

$$W^a = -\Delta G^a \tag{4.2}$$

Consider a brittle material of unit cross-sectional area subjected to a tensile force. Then, if the material is completely brittle, the work done on the sample is dissipated only through propagation of a crack, thereby creating two new surfaces. The total work expended per unit of surface area in forming the two surfaces (W_a) is then equal to twice the surface energy per unit of surface area of the material under consideration. Under these conditions,

$$W^c = 2\gamma \tag{4.3}$$

or,

$$\Delta G^c = -2\gamma \tag{4.4}$$

When two dissimilar materials form an interface by being in intimate contact, a tensile force can be applied to split the materials into dissimilar parts. For a completely brittle interface of unit cross-sectional area, the energy expended should be the sum of the individual surface energies for the two materials involved. However, because the dissimilar materials are separated, some of the intermolecular forces present during intimate contact are now missing. That is, an interfacial energy may have existed before separation, which should be accounted for by subtracting it from the energy required to create the two new surfaces (Pocius 1997). Dupré, in 1867, postulated the following relationship, which plays a central role in the study of adhesion.

$$W^a = \gamma_1 + \gamma_2 - \gamma_{12} \tag{4.5}$$

or,

$$\Delta G^a = \gamma_{12} - \gamma_1 - \gamma_2 \tag{4.6}$$

where γ_i is the surface energy of the ith material and γ_{12} is the interfacial energy between the two materials in contact.

The terms wetting, spreading, and contact angle have become synonymous with adhesion. The shape of a liquid drop on a surface, usually described through the contact angle between them, provides information on the intermolecular forces of the individual phases involved as well as the interfacial forces between them. Although the great utility of contact angle measurements stems from their interpretations based on thermodynamic considerations, they can also be used to provide a measure of wetting and spreading on a macroscopic scale. Wetting of a surface (or spreading of the liquid over the surface) is the process where the adhesive comes into intimate molecular contact with the surface and establishes fundamental forces of adhesion. Wetting is a prerequisite for good adhesion. For complete spreading, the contact angle equals zero and the adhesive spreads spontaneously over the surface (Myers 1991). Naturally, wetting and spreading depend also on the viscosity of the liquid and roughness and heterogeneity of the solid surface involved. From a thermodynamic point of view, wetting and spreading depend on the competition between adhesive forces and cohesive forces, which can be used to define a spreading coefficient, S_p.

$$S_p = W^a - W^c \tag{4.7}$$

Thus, the higher and more positive the value of S_p, the greater the work of adhesion is compared to the cohesive energy. A negative value of S_p represents a finite contact angle and zero corresponds to final equilibrium (Shanahan 1992).

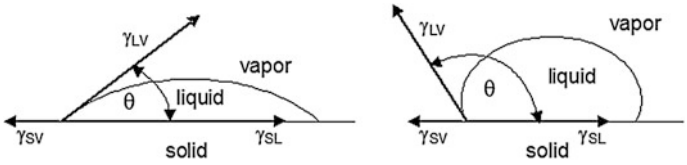


Fig. 4.13 Three-phase boundary of a liquid drop on a *solid surface* on vapor

The properties of interfaces can normally be described as triple junctions, or a three-phase boundary. Young, in 1805, proposed an equation to obtain surface tension from the contact angle formed when a drop of liquid is placed on a perfectly smooth, rigid solid. A schematic of the contact angle experiment is presented in Fig. 4.13, where θ is the contact angle between the solid–liquid (SL) interface and the tangent of the liquid–vapor (LV) interface. If water is the liquid under consideration, the case on the left side would be an example of a hydrophilic (water-loving) surface, evident from the drop that tends to spread over the solid due to favorable interaction between interfacial forces. The contact angle on the right side, however, is greater than 90º, indicating that water does not spread readily over this surface, and is thus hydrophobic (water-hating).

To each interface, forming the triple line, a surface or interfacial free energy or tension can be attributed. Thus, γ_{SV} is the surface free energy of the solid in equilibrium with the saturated vapor of the liquid, γ_{LV} is the surface tension of the liquid in equilibrium with the solid, and γ_{SL} is the solid–liquid interfacial free energy. Young's equation is the result of the triple line in equilibrium (tension in force per unit length).

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV} \cos\theta \tag{4.8}$$

Although this equation was originally based on a mechanical (tension) definition, it has since been verified by rigorous thermodynamic derivations based on minimizing the overall free energy (Pocius 1997). The solid–vapor interfacial energy in this equation, however, is not the true surface free energy of the solid. When vapor is adsorbed onto the surface of the solid, its free energy is reduced so that,

$$\gamma_{SV} = \gamma_S - \pi_e \tag{4.9}$$

where γ_S is the surface free energy of the solid only in contact with its own vapor, and π_e the equilibrium spreading pressure, which is a measure of energy released during vapor adsorption. Although equilibrium spreading for the liquid phase should also be possible, this value has historically been assumed equal to zero. It has been shown that π_e is negligible for high-energy liquids on low-energy solids, which is usually the case for contact angle measurements on polymeric surfaces (Good and van Oss 1991). Equations (4.5) and (4.8) can therefore be combined in the following way to obtain the Young–Dupré equation.

$$W_a^{LS} = \gamma_{LV}(1 + \cos\theta) \tag{4.10}$$

The Young–Dupré equation is the starting point for any method that utilizes contact angles to obtain surface free energies by relating the contact angle to the work of adhesion. This is also the way to tie contact angles back to adhesion.

Fowkes (1964) suggested that surface energy is composed of a polar and non-polar component. Since that time, several models have been proposed to calculate the work of adhesion based on surface energy components. Several researchers and practitioners from different sectors of the adhesion science community report that the theory developed by van Oss et al. (1988) was the best available at the time (Adão et al. 1998 and Woodward 2000). Although the theory also received much critique during the past decade (Kwok 1999), its application has not always been well understood (Della Volpe and Siboni 2000).

Following the form suggested by Fowkes (1964), the surface energy of a single phase is given by,

$$\gamma_i = \gamma_i^{LW} + \gamma_i^{AB} \tag{4.11}$$

where LW denotes Lifshitz-van der Waals and AB denotes acid–base. The Lifshitz theory was introduced conceptually in the section addressing fundamental forces. The London, Keesom, and Debye components together form the apolar (LW) component of attraction. Based on the Lifshitz theory, Chaudhury and Good (1991) showed that the contribution of the Keesom and Debye van der Waals forces are small relative to the London forces. The acid–base component represents polar, or specific, interactions primarily due to hydrogen bonding (Good and Van Oss 1991).

It follows that the free energies of cohesion and adhesion likewise have two components. Therefore, the Gibbs free energy of cohesion is,

$$\Delta G_i^c = -2\gamma_i = \Delta G_i^{cLW} + \Delta G_i^{cAB} \tag{4.12}$$

and the Gibbs free energy of adhesion is,

$$\Delta G_{ij}^a = \gamma_{ij} - \gamma_i - \gamma_j = \Delta G_{ij}^{aLW} + \Delta G_{ij}^{aAB} \tag{4.13}$$

The van der Waals forces represent the interaction between two non-polar molecules and its treatment can be traced back to first principles. For the LW component, the Berthelot geometric mean rule therefore holds,

$$\Delta G_{ij}^{aLW} = -2\sqrt{\gamma_i^{LW}\gamma_j^{LW}} \tag{4.14}$$

The acid–base component, however, cannot be treated in the same fashion, although some theories assume that the geometric mean rule holds for this component as well. These interactions are specific and are only possible between

interaction partners with complementary acid–base properties. The relationship describing the AB component of free energy was derived empirically by van Oss et al. (1988),

$$\Delta G_{ij}^{aAB} = -2\left(\sqrt{\gamma_i^+ \gamma_j^-} + \sqrt{\gamma_i^- \gamma_j^+}\right) \tag{4.15}$$

Van Oss and his co-workers presented the full version of the Young–Dupré equation by inserting ΔG^{LW} and ΔG^{AB},

$$\begin{aligned} -\Delta G^a = W^a &= \gamma_L^{Tot}(1 + \cos\theta) \\ &= 2\left(\sqrt{\gamma_L^{LW}\gamma_S^{LW}} + \sqrt{\gamma_L^+ \gamma_S^-} + \sqrt{\gamma_L^- \gamma_S^+}\right) \end{aligned} \tag{4.16}$$

where the subscript L represents a liquid and the subscript S the solid under consideration. Equation (4.16) implies that if the contact angles of three liquids with different and known polarities are measured on an unknown surface, then the three unknown surface energy components can be solved. These polarities have been defined as monopolar basic, monopolar acidic, bipolar (basic and acidic), or apolar (only van der Waals forces). In addition, the use of this equation is not restricted to contact angle measurements, but can be applied with any technique able to quantify the work of adhesion between the unknown surface and probe substances with known surface energy components. An important practical application of this theory is that it can be used to predict the work of adhesion between two materials if their surface energy components are known. Similarly, the work of cohesion can be predicted within the bulk phase of a material. By adapting the Dupré equation for the interaction between two condensed phases, 1 and 2, within a liquid, 3,

$$\Delta G_{132}^a = \gamma_{12} - \gamma_{13} - \gamma_{23} \tag{4.17}$$

and with the components of the free energy of interfacial interaction additive,

$$\Delta G_{132}^a = \Delta G_{132}^{aLW} + \Delta G_{132}^{aAB} \tag{4.18}$$

van Oss et al. (1988) and Good and van Oss (1991) proposed the following form for the interaction between two materials, 1 and 2, submersed in a polar liquid, 3.

$$\Delta G_{132}^a = 2\left[\begin{array}{r} \sqrt{\gamma_1^{LW}\gamma_3^{LW}} + \sqrt{\gamma_2^{LW}\gamma_3^{LW}} - \sqrt{\gamma_1^{LW}\gamma_2^{LW}} - \gamma_3^{LW} \\ + \sqrt{\gamma_3^+}\left(\sqrt{\gamma_1^-} + \sqrt{\gamma_2^-} - \sqrt{\gamma_3^-}\right) \\ + \sqrt{\gamma_3^-}\left(\sqrt{\gamma_1^+} + \sqrt{\gamma_2^+} - \sqrt{\gamma_3^+}\right) \\ -\sqrt{\gamma_1^+\gamma_2^-} - \sqrt{\gamma_1^-\gamma_2^+} \end{array}\right] \tag{4.19}$$

When the liquid is water, this interaction is called the "hydrophobic interaction" where $\Delta G_{132}^{a} < 0$. For $\Delta G_{132}^{a} > 0$, the interaction between 1 and 2 becomes repulsion, which is the driving force for phase separation of adhesives in aqueous media (Van Oss et al. 1988).

Thermodynamic concepts have been used by many researchers to elucidate adhesion between bitumen and aggregate materials. Ensley et al. (1984) determined the heat of immersion using microcalorimetry. Curtis et al. (1992) determined Gibbs free energies from adsorption isotherms. Lytton (2004) utilized surface energies measured on bitumen and aggregates surfaces to calculate free energies of adhesion and cohesion by applying modern surface energy theories as discussed in the preceding paragraphs.

These concepts, especially surface energy, have played an important role in discoveries of the rules governing the microfracture and healing in bitumen–aggregate mixtures (Lytton 2000). Thermodynamic theory lends itself to quantifying the relationship between fundamental adhesion and practical adhesion. Since real materials are not completely brittle, especially polymer-like materials such as bitumen, energy dissipation can result from molecular chain disentanglement and stretching, and conformational changes (rotation about flexible intermolecular bonds) under applied loads, i.e., viscoelastic effects, as well as plastic deformation. Merril and Pocius (1991) report that the apparent or practical work of adhesion is much larger than the fundamental work of adhesion based on adhesive peel test experiments. However, the experiments showed that the practical work of adhesion is a multiplication of some dissipation factor with the fundamental work of adhesion. From fracture mechanics, the well-known Griffiths crack growth criterion suggests that for a completely brittle material, the energy necessary to break a adhesive (or cohesive) bond is equal to the fundamental work of adhesion, W_a. This energy can be considered the minimum amount of practical adhesion that one can expect from a bond. If other modes of dissipating energy exist, then this minimum value increases (Pocius 1997). Schapery (1984) extended this crack growth theory to viscoelastic materials based on the fundamental laws of fracture. Schapery (1989) and Lytton et al. (1998) developed similar relationships for healing. Healing is the opposite of fracture and is pronounced when longer rest periods are introduced between applied loads, manifested in extended fatigue life. Lytton et al. (1998) applied these fundamental relationships of fracture and healing to asphalt mixtures.

The fundamental law of fracture for viscoelastic materials is presented in Eq. (4.20). This expression essentially describes an energy balance between the energy required for fracture and the energy released, expended in crack propagation and extension. An expression similar to this equation exists for healing.

$$W = E_R D(t_\alpha) J_R \tag{4.20}$$

where W is the work of adhesion (W^a) or cohesion (W^c) per unit of each crack surface area created, i.e., the minimum energy required to cause fracture. The product of terms on the right side is termed pseudostrain energy because the energy required in overcoming nonlinear and viscoelastic effects is eliminated. The

reference modulus E_R, is an arbitrary value used during this correction; $D(t_\alpha)$ is the viscoelastic creep compliance over a period t_α, the time required for a crack to move a distance equal to the length (α) of the process zone ahead of the crack tip; and J_R is the pseudo strain energy release rate, i.e. the change of available energy per unit crack area, from one load cycle to the next. For the situation where the pseudostrain energy released is greater than the required minimum energy for bond breakage, crack extension occurs.

From Eqs. (4.2) and (4.12), the work of adhesion (W) for a dry interface is given by,

$$W_{12}^{a} = -\left(\Delta G_{12}^{aLW} + \Delta G_{12}^{aAB}\right) \tag{4.21}$$

where ΔG_{12}^{aLW} and ΔG_{12}^{aAB} the Lifshitz van der Waals and acid–base components of the free energy of interfacial interaction, respectively, between materials one and two.

According to van Oss (1994), the generalized form of Eq. (4.18) includes another term, namely the free energy of electrostatic interaction (ΔG_{132}^{aEL}).

$$W_{132}^{a} = -\left(\Delta G_{132}^{aLW} + \Delta G_{132}^{aAB} + \Delta G_{132}^{aEL}\right) \tag{4.22}$$

The primary parameter required to calculate the free energy of electrostatic interaction is surface charge (or zeta potential) of the two materials under consideration. The concept of zeta potential was introduced previously and will be developed further in Sect. 4.10.3. Hefer et al. (2007) hypothesized that this term, ΔG_{132}^{aEL}, holds the key to incorporate the effect of pH of the interface water on the adhesive bond. Hefer et al. (2007) recently calculated this term from data obtained during the Strategic Highway Research Program (SHRP) research. Surprisingly, these calculations revealed that the contribution of ΔG_{132}^{aEL} to the total work of adhesion is very small.

The fact that the electrostatic term is insignificant based on relative magnitude, however, does not mean that the pH of the contacting water can be neglected. The detrimental effect of a fluctuating pH on the adhesive bond in the presence of water has been known for at least 50 years (Gzemski 1968). The ability of the electrostatic term to distinguish among different mixtures suggests that electrokinetic potential (zeta potential) is important. Hefer et al. (2007) postulated that another, primary, mechanism exists that is related to electrical phenomena at the interface. Labib and Williams (1986) showed that a fundamental relationship exists between the *pH scale* and the *donicity scale*, i.e., zeta potential as a function of pH and electron donor/acceptor properties, respectively. This relationship implies that pH of the interface water will influence the electron donor–acceptor properties of the two surfaces under consideration. Both scales are energy scales and ought to be compatible with the surface energy theory proposed by van Oss et al. (1988).

4.7 Application of Surface Energy to Predict Moisture Damage in Asphalt Mixtures

The most evident form of moisture damage is stripping of asphalt binder from the aggregate surface due to exposure to moisture. The correlation between the surface properties of these materials and their tendency to strip in the presence of water is relatively well established in the literature. The three quantities based on the surface energies of asphalt binders and aggregate that are related to the moisture sensitivity of an asphalt mixture are:

- work of adhesion between the asphalt binder and aggregate (W_{AB}),
- work of debonding or reduction in free energy of the system when water displaces asphalt binder from a binder-aggregate interface (W_{ABW}^{wet}), and
- work of cohesion of the asphalt binder or mastic (W_{BB}).

The above three quantities are computed using the surface free energy components of the individual materials. For an asphalt mixture to be durable and have a relatively low sensitivity to moisture, it is desirable that the work of adhesion, W_{AB}, between the asphalt binder and the aggregate be as high as possible. Furthermore the more negative the magnitude of the absolute value of the work of debonding when water displaces the asphalt binder from the binder-aggregate interface, W_{ABW}^{wet}, the greater the thermodynamic potential that drives moisture damage will be. Therefore, it is normally desirable that this quantity be as small as possible. This is because this term is usually (almost always) a negative quantity, and the more negative it is, the more likely it is for water to replace bitumen at the interface with mineral or aggregate. Little and Bhasin (2006) developed and energy ratio (ER) based on the two energy terms W_{AB} and W_{ABW}^{wet} and which combines these terms into a single, dimensionless parameter. They later adjusted this ratio to also consider the affinity of the bitumen to develop a cohesive interaction (W_{BB}). The resulting energy ratio (ER_2) is expressed by Eq. (4.23):

$$ER_2 = \left| \frac{W_{AB} - W_{BB}}{W_{ABW}^{wet}} \right| * SSA \tag{4.23}$$

Equation (4.23) defines ER_2 as a parameter that can be used to estimate the moisture sensitivity of asphalt mixtures based on the hypothesis that moisture sensitivity is directly proportional to the dry adhesive bond strength, and inversely proportional to the work of debonding or the reduction in free energy during debonding. The latter term is determined as follows:

$$W_{ABW}^{wet} = \gamma_{AW} + \gamma_{BW} - \gamma_{AB} \tag{4.24}$$

where γ_{AW} is the bond strength between the aggregate and water, γ_{BW} is the bond strength between bitumen and water, and γ_{AB} is the interfacial bond energy between the bitumen and the aggregate. In Eq. 4.23 the value ER_2 is the product of the

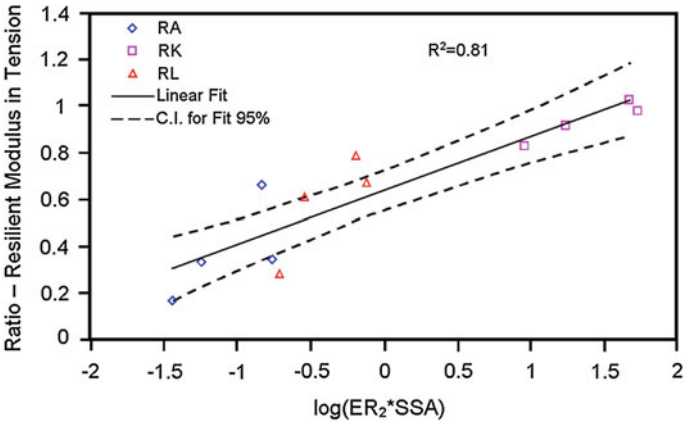


Fig. 4.14 Laboratory performance (tensile modulus) of mixtures compared to ER_2 * SSA parameter. After Little and Bhasin (2006)

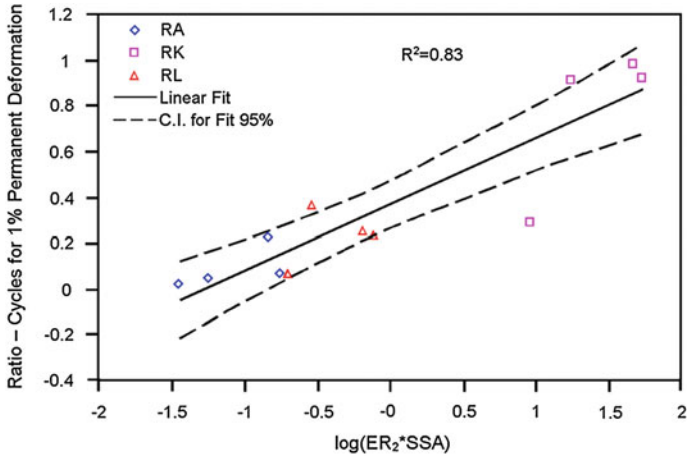


Fig. 4.15 Laboratory performance (plastic deformation) of mixtures compared to ER_2 * SSA parameter. After Little and Bhasin (2006)

absolute value of the ratio of dry work of adhesion to the work of debonding multiplied by the specific surface area (SSA) of the aggregate.

Figures 4.14 and 4.15 show evidence of the utility of Eq. 4.23 in predicting moisture damage based on laboratory experiments (tensile modulus, Fig. 4.14, and plastic deformation, Fig. 4.15) with twelve different mixtures (twelve combinations of three aggregates and four bitumens). In the tensile test (Fig. 4.14), the ratio of tensile resilient modulus in the wet (moisture conditioned) state compared to the tensile resilient modulus in the dry state is plotted on the ordinate v. log/ER2 * SSA on the abscissa. In Fig. 4.15, repeated, compressive loading was continued until the sample accumulated one percent permanent strain. The ratio of the number of

cycles to accumulate one percent strain for each mixture under wet (moisture conditioned) and dry conditions is plotted on the ordinate v. the energy ratio on the abscissa. In both figures, the dashed lines represent the 95% confidence range and demonstrate the strong relationship between moisture damage, as recorded by these ratios, and ER_2..

4.8 Effect of Asphalt Composition on Adhesion

4.8.1 Asphalt Composition

The chemistry of asphalt is complex. This brief overview is certainly a simplification of the complex nature of asphalt and is meant to provide (a) definitions of basic terms and (b) descriptions of basic asphalt components, which are used in discussion throughout this chapter.

4.8.2 Elemental Composition

Asphalt molecules are comprised primarily of carbon and hydrogen (between 90% and 95%) by weight. However, the remaining atoms, called heteroatoms, are very important to the interaction of asphalt molecules and hence the performance of asphalt. They include oxygen, nitrogen, sulfur, nickel, vanadium, and iron.

4.8.3 Molecular Structure

Asphalt atoms are linked together to form molecules. Perhaps the simplest is the aliphatic carbon–carbon chain saturated with hydrogen bonds. Carbon–carbon bonds can also form rings saturated with hydrogen. These saturates are essentially nonpolar and interact primarily through relatively weak van der Waals forces. A second class of asphalt molecules involves aromatics. This molecule has six carbon atoms in the form of a hexagonal ring. This ring possesses a unique bond with alternating single and double bonds between carbon atoms. Figure 4.16 shows representative examples of saturates (aliphatic and cyclic) and aromatic structures.

The unique bonds in the aromatic or benzilic rings facilitate polyaromatic atomic structures allowing extension in two dimensions and stacking in the third dimension.

4.8.4 Bonds Among Asphalt Molecules

Strong covalent bonds link atoms together to form asphalt molecules. These molecules interact with one another through much weaker bonds (Jones 1992): pi–pi bonds, hydrogen or polar bonds, and van der Waals bonds. These are represented

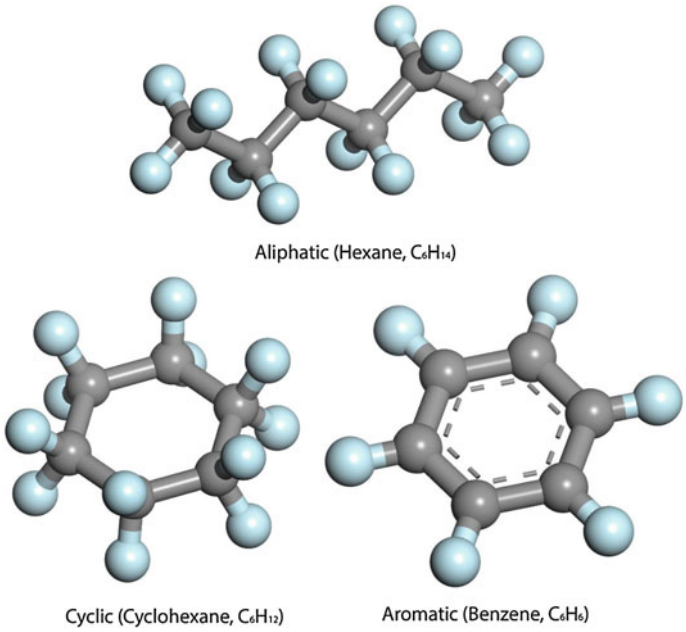


Fig. 4.16 Types of asphalt molecules. After Jones (1992)

in Fig. 4.16. Pi–pi bonds are unique to aromatic molecules. They provide polarity and the ability of aromatic molecules to link together in unique configurations, including a stacked arrangement as shown in Fig. 4.17. Heteroatoms among asphalt molecules develop polarity and link together by forming hydrogen bonds. Figure 4.17 shows a hydrogen bond between two very important asphalt functional groups: a sulfoxide and a carboxylic acid. van der Waals bonding is the weakest of the secondary bonds. They form when molecules equilibrate at lower temperatures or stress is removed. van der Waals bonding is responsible for the free-flowing nature of asphalt at high temperatures versus the semisolid nature at lower temperatures (Jones 1992). As a point of reference, it is important to understand that covalent primary bonds within the molecule are from 10 to 100 times stronger than secondary bonds.

4.8.5 Polar Versus Nonpolar Molecules

Polar molecules form "networks" through hydrogen and pi–pi bonds that give asphalt its elastic properties. Nonpolar materials form the body of the material in which the network is formed and contribute to the viscous properties of asphalt (Jones 1992). Degree of polarity is the most important property of polar molecules,

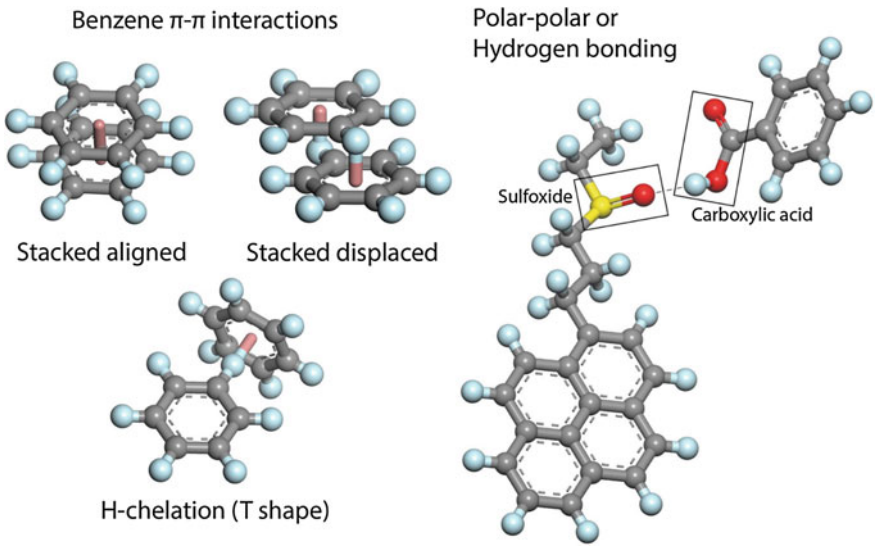


Fig. 4.17 Types of intermolecular asphalt bonds and multifunctional molecules. After Jones (1992)

while degree of aromaticity is the second most important. Highly polar and highly aromatic molecules form the most interactive and strongest molecular networks.

Nonpolar molecules do not interact strongly enough to form networks, but they do substantially influence asphalt performance. The molecular weight of nonpolar molecules is related to low-temperature performance (Jones 1992). A preponderance of high-molecular weight nonpolar molecules will lead to asphalts that stiffen and perform poorly at low service temperatures. If nonpolars are waxy, they will crystallize at low temperatures and become crack susceptible. Nonpolar and polar molecules must interact in an acceptable manner or be "compatible." If polar and nonpolar molecules are relatively similar in chemistry, they will be compatible; however, if they are very different, the polar network will not stay in solution, and phase separation can be a substantial problem.

4.8.6 Asphalt Model

Jones (1992) explains the history of development of an asphalt model. He describes how analytical techniques including size exclusion chromatography and ion exchange chromatography have led to viewing asphalt as a two-phase system. The polar molecules interact with each other through polar–polar or hydrogen bonding. These bonds form associations that create a network within the nonpolar solvent molecules. However, as explained by Jones, both phases make a significant contribution to asphalt performance. Figure 4.18 illustrates the model described by

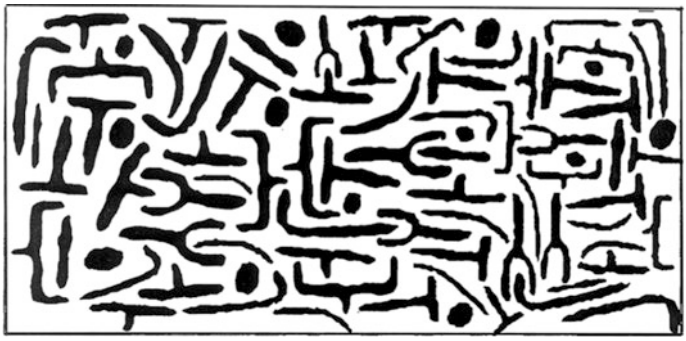


Fig. 4.18 SHRP asphalt model. After Jones (1992)

Jones (1992)—the Strategic Highway Research Program (SHRP) model. The associations of polar molecules are due to polar sites on the asphalt through hydrogen bonding. Other interactions take place through pi–pi bonding and van der Waals bonding. These interactions provide the major contribution to viscoelastic properties of the asphalt. Actually, the term phase is not accurate in the description of polar versus nonpolar components because the mixture is homogeneous and the bonds between the polar molecules are rather weak and form and break constantly.

4.8.7 Multifunctional Organic Molecules

Data show that having two or more functional groups on the same molecule makes it possible to form chains of weak polar–polar interactions. According to Jones (1992), these chains are the foundation of the polar networks. An example of a multifunctional organic molecule is one with both a carboxylic acid (R–COOH) and a sulfoxide (S=O) on the same molecule. Figure 4.19 is an example of networks and chains conformed ofbi- and multi-functional molecular structures. Multifunctional organic moleculeshave a major impact on aging. This is because for polar molecules to generatesignificant physical changes, it is necessary for them to interact in chainlikestructures or form networks. Figure 4.19 illustrates how oxidation of molecules with a single active site results in a “quenching” of the effect of oxidation, while the oxidation of molecules with multiple active sites develops a continuous network.

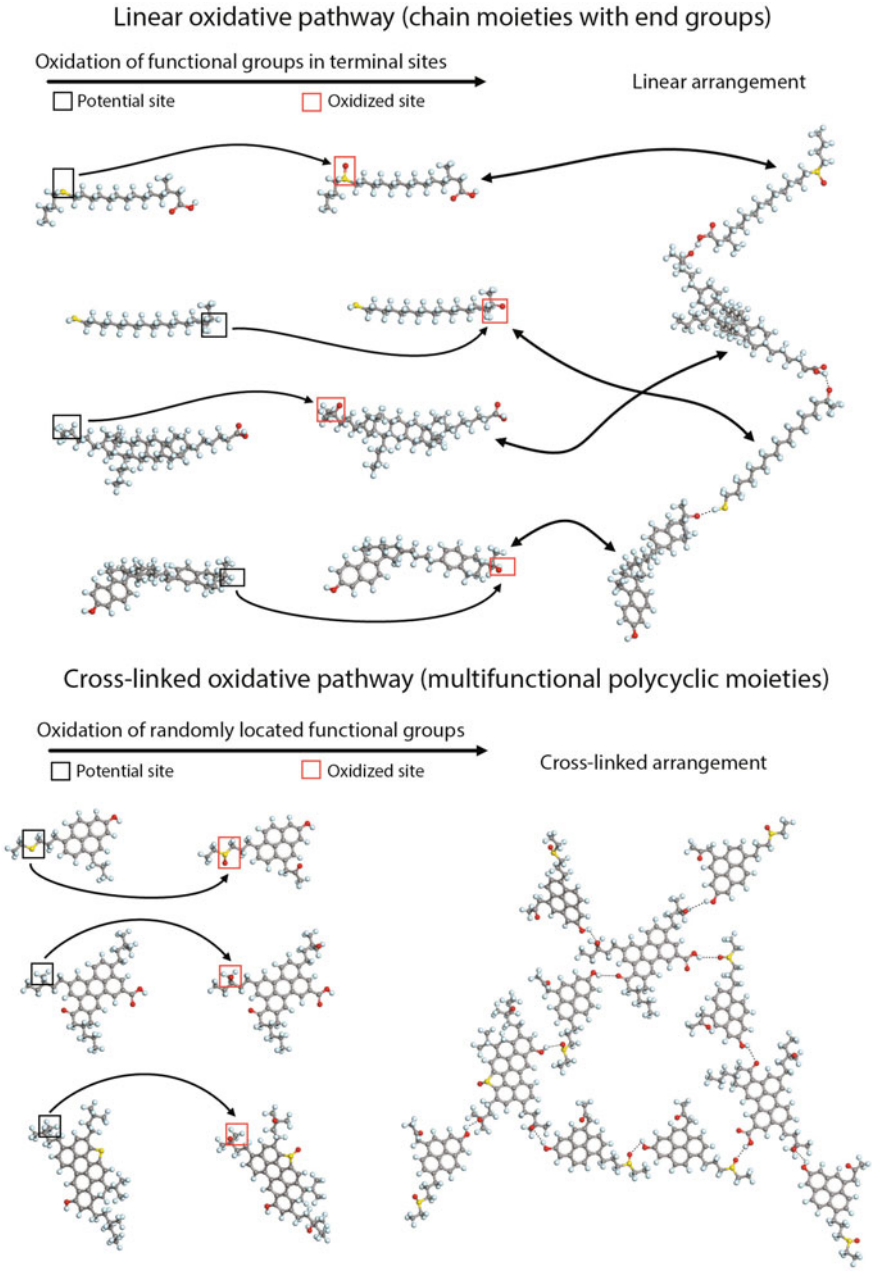


Fig. 4.19 Asphalt with simple active sites. After Jones (1992)

4.9 Asphalt Chemistry and Adhesion

Polarity or separation of charge within the organic molecules promotes attraction of polar asphalt components to the polar surfaces of aggregates. Although neither asphalt nor aggregate has a net charge, components of both have nonuniform charge distributions, and both behave as if they have charges that attract the opposite charge of the other material. Curtis et al. (1992) have shown that aggregates vary widely in terms of surface charge and are influenced by environmental changes. Robertson (2000) points out that adhesion between asphalt and aggregate arises between the polars of the asphalt and the polar surface of the aggregate. He also points out that polarity alone in asphalt is not sufficient to achieve good adhesion in pavements because asphalt is affected by the environment. Robertson (2000) further states that asphalt has the capability of incorporating and transporting water. Absorption of water varies with asphalt composition and changes further as asphalt is oxidized. Cheng et al. (2002), as discussed previously, have shown that a substantial quantity of water can diffuse through and be retained in a film of asphalt cement or an asphalt mastic, substantially changing the rheology of the binder. Robertson (2000) states that at the molecular level in asphalt, basic nitrogen compounds (pyridines) tend to adhere to aggregate surfaces tenaciously. Carboxylic acids are easily removed from aggregate in the presence of water if the acids form a monovalent salt by interaction at the aggregate surface, but divalent (calcium) salts of acids are much more resistant to the action of water.

Curtis (1992) ranked the affinity of various asphalt functional groups to bond to aggregate surfaces by using adsorption isotherms (UV adsorption spectroscopy). In general she found acidic groups, carboxylic acids, and sulfoxides to have the highest adsorptions, while ketone and nonbasic nitrogen groups had the least. However, the sulfoxide and carboxylic acids were more susceptible to desorption in the presence of water. According to Curtis (1992), the general trend of desorption potential of polar groups from aggregate surfaces is sulfoxide > carboxylic acid > nonbasic nitrogen $\geq$ ketone > basic nitrogen > phenol.

4.9.1 Effect of Aggregate Properties on Adhesion

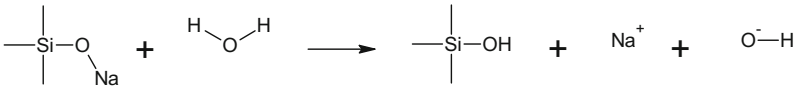
A number of aggregate properties affect the adhesive bond between asphalt and aggregate: size and shape of aggregate, pore volume and size, surface area, chemical constituents at the surface, acidity and alkalinity, adsorption size surface density, and surface charge or polarity.

4.9.2 Pore Volume and Surface Area

Yoon and Tarrer (1988) investigated five aggregates (granite, dolomite, chert gravel, quartz gravel, and limestone). They measured pore volume, surface area, average pore size, and percentage coating after boiling. Their study showed that stripping resistance is defined by the level of physical bond that is achieved between the asphalt and aggregate, and this is, in turn, defined by surface area, pore volume, and pore size. The optimal resistance to stripping was developed in aggregates that provide a large surface area for bonding as well as a favorable pore size for adequate (deep) asphalt penetration. This is probably because when asphalt cement coats a rough surface with fine pores, air is trapped and the asphalt has difficulty penetrating the fine pores (Yoon and Tarrer 1988). However, the penetration of asphalt cement into pores is synergistically dependent not only on the pore structure but also on the viscosity of the asphalt cement at mixing temperatures. Yoon and Tarrer (1988) also determined that aggregates with approximately equal physical properties (e.g., pore volume and structure and surface area) can have very different properties depending on their basic chemistry and mineralogy, which define surface activity. Yoon and Tarrer found substantially higher bonding power for limestone than for quartz gravel even though both had similar physical surface structures. The results of Cheng et al. (2002) were very similar; they found that a certain granite aggregate has a much higher surface energy per unit area than a certain limestone, but when bonding energy was computed in terms of unit mass instead of unit surface area (incorporating effects of surface area), the limestone was predicted to have a much greater potential to resist damage in repeated loading tests of asphalt samples at 85% saturation.

4.9.3 pH of Contacting Water

Hughes et al. (1960) and Scott (1978) reported that adhesion between asphalt cement and aggregate in the presence of water became weakened when the pH of the buffer solution was increased from 7.0 to 9.0 (Scott 1978). Yoon and Tarrer (1988) showed that if different aggregate powders (chert gravel, quartz sand, quartz gravel, granite, limestone, and dolomite) were added to water and allowed to react with water for up to about 30 min, the pH of the blend would increase to some asymptotic value (Fig. 4.22). Even granite, known to be acidic, showed an increase in pH over time to about 8.8. The granite reaction in water, which leads to this gradual pH increase, is, according to Yoon and Tarrer, due to the silicate lattice reaction with the water to impart excess hydroxyl ions as follows:



This is a typical hydrolytic reaction of the salt of a weak acid.

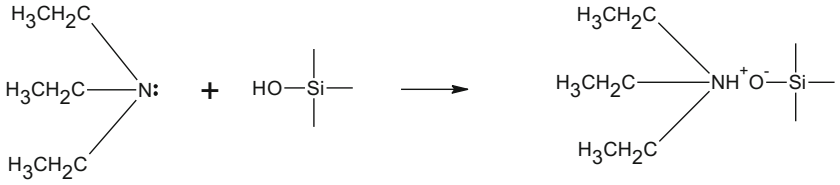
Yoon and Tarrer (1988) assessed the sensitivity of stripping to changes in pH of water in contact with the aggregate surface. They performed boiling stripping tests to verify the sensitivity. The pH of the water was modified by using a solution of HCl or NaOH. The stripping became more severe as the pH increased. Yoon and Tarrer explain that when an aggregate is being coated with asphalt, the aggregate selectively adsorbs some components of the asphalt—the more polar compounds and hydrogen bonds or salt links are formed. Conspicuously, the type and quantities of the adsorbed components affect the degree of adhesion. Yoon and Tarrer state that the presence of ketones and phenolics is thought to improve stripping resistance, whereas carboxylic acids, anhydrides, and 2-quinolenes are thought to increase stripping sensitivity because of the substantial water susceptibility of the associated bonds.

According to Yoon and Tarrer (1988), the water susceptibility of the hydrogen bonds and salt links between the adsorbed asphalt components and the aggregate surface increase as the pH of the water at the aggregate surface increases. Therefore, it seems reasonable to conclude that stripping sensitivity will increase as the pH of the water increases. Experimental results of Yoon and Tarrer (1988) substantiate this hypothesis. However, they warn that other surface aggregate properties also play a role. Different types of metal ions affect stripping potential. For example, alkaline earth metals in limestone associate strongly with the asphalt components in carboxylic acids to form alkaline earth salts, and the bonds formed are not dissociated easily in water even at a high pH. In other words, the adsorption is strong because of the insolubility of the alkaline earth salts formed between the limestone and the bitumen acids.

The addition of hydrated lime offers a mechanism to tie up carboxylic acids and 2-quinolenes so they cannot interact with hydrogen bonding functionalities on the aggregate surface to produce moisture-sensitive bonds. Thomas (2002) points out that the interaction of lime with components in the asphalt not only prevents the formation of moisture-sensitive bonds but also subsequently allows more resistant bonds (e.g., with nitrogen compounds from the asphalt) to proliferate. He points out that an additional benefit of the use of lime is to react with or adsorb compounds that can be further oxidized and enhance the increase in viscosity as a result of oxidation. In fact, experiments at Western Research Institute (2002) show a substantial improvement in moisture resistance after severe freeze–thaw experiments when lime is added directly to the bitumen and before the bitumen is coated on the aggregate.

In a manner similar to the reaction between acidic compounds such as carboxylic acids in asphalt and alkaline aggregate or with lime, an amine compound either if present in asphalt or added in the form of an anti-stripping additive will react with

acidic surfaces as in the case of siliceous aggregates to form a surface compound. Evidence of the formation of such a surface compound between siliceous surfaces and amine compounds was demonstrated by Titova et al. (1987). The following is silica, triethylame:



Different from pure Coulombic interaction, is the electrostatic interaction between surfaces in the presence of an aqueous medium that contains dissolved ions, i.e., an electrolyte. Most surfaces are charged in the presence of water. This is due to the high dielectric constant of water, which makes it a good solvent for ions. Ions of opposite charge, or counter-ions, bind directly to the surface to neutralize the surface charges, known as the Stern layer. Thermal motion prevents ions from accumulating on the surface so that a diffuse layer of counter-ions and co-ions exists. The "fixed" Stern layer and mobile diffuse layer together form an electric double layer. This concept is illustrated in Fig. 4.20. The electric potential at the surface (ψo) decreases with distance into the bulk water.

The electric potential at the shear plane between the fixed and mobile layers is measurable and called the zeta potential or electrokinetic potential. Zeta potential is commonly used as a representation of surface charge. The study of charge particles in motion in an electric field is called electrophoresis.

Electrophoresis have been applied to bitumen–aggregate systems in the past to investigate moisture damage, or stripping, in these systems (Labib 1992; Yoon and Tarrar 1988; and Scott 1978). However, the existence of bulk water, hydrogen bonded to the aggregate surface in the "dry" state, should not be overlooked. Thelen (1958) conducted experiments on quartz to demonstrate how fresh aggregate adsorb a water layer several molecules thick. This illustrates the importance of aggregate preconditioning as part of the hot mix asphalt manufacturing process. Nevertheless, at conventional plant mix temperatures at least a molecular layer of water remain, Fig. 4.21.

Whenever water is present, pH can be expected to play a role by influencing aggregate surface charge. Two aspects should be considered with regard to bitumen–aggregate systems. The first is diffusion of external water to the bitumen–aggregate interface (Scott 1958; Nguyen et al. 1992; Cheng 2002). The pH of this water will differ depending on the environment. Secondly, several studies indicate that the water is influenced by the aggregate surface as illustrated in Fig. 4.22 (Labib 1992; Yoon and Tarrar 1988; Scott 1978; and Huang et al. 2000).

If two surfaces approach each other and electric double layers overlap, an electrostatic double-layer force arises. Labib (1992) qualitatively illustrated this

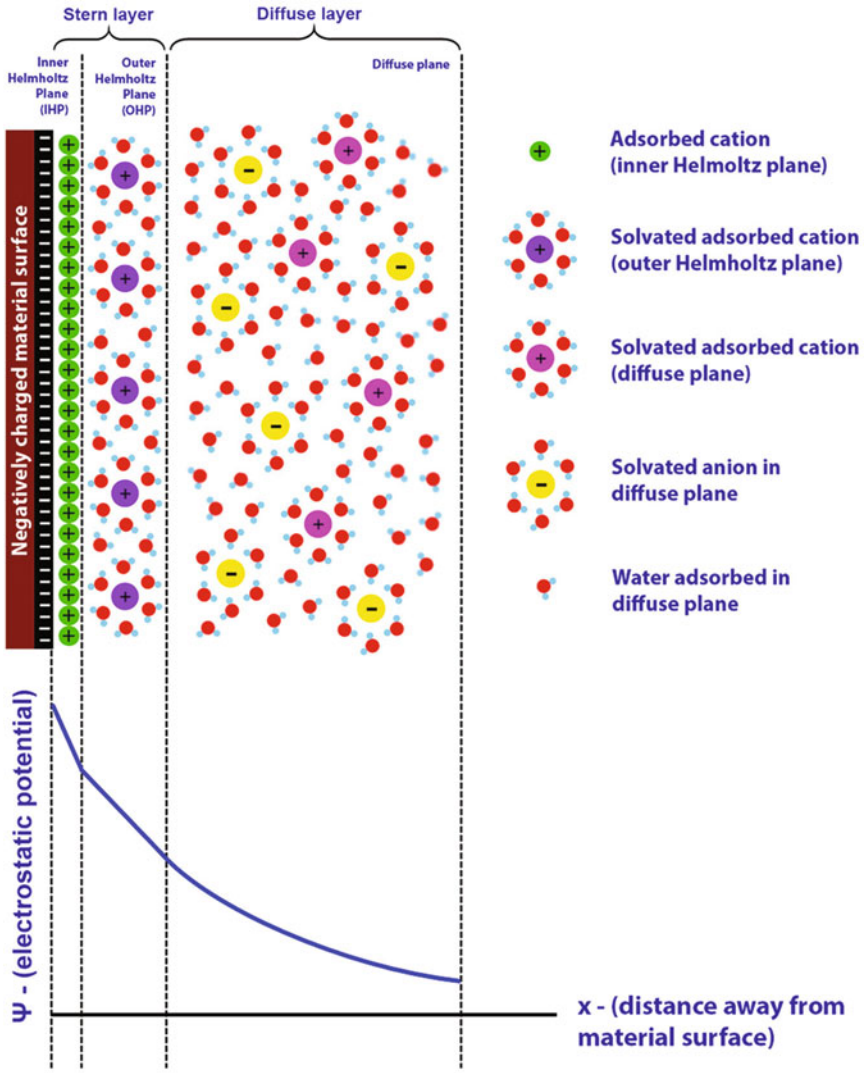


Fig. 4.20 Schematic illustration of the Stern layer with thickness. After Butt et al. (2003)

concept by superposition zeta potential versus pH curves for bitumen and aggregates, respectively. Figure 4.23 is an interaction diagram for a siliceous conglomerate, SHRP Materials Reference Library code RJ, and four types of bitumen. The pH at zero zeta potential is known as the isoelectric point (IEP). The IEP for all the bitumen types as well as the granite is around three; i.e., both bitumen and aggregate are negatively charged at pH > 3. More importantly, this situation will

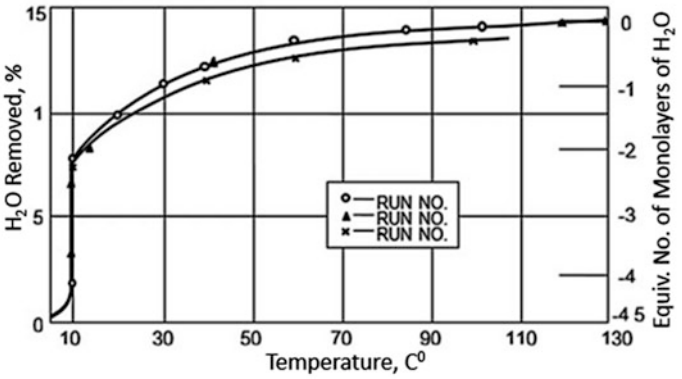


Fig. 4.21 Removal of adsorbed water molecules by heat from quartz. After Thelen (1958)

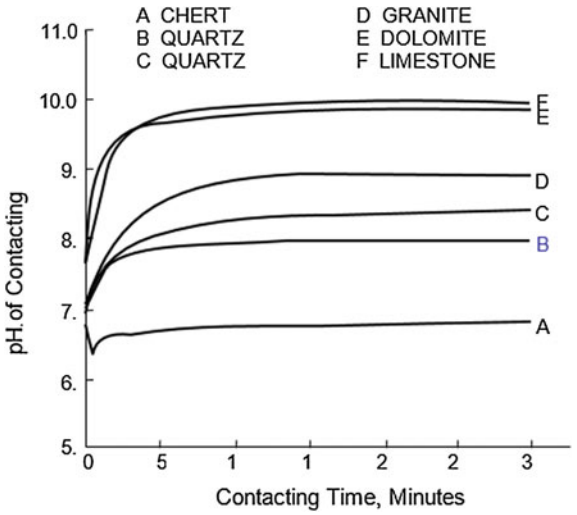


Fig. 4.22 Changes in pH of water in which aggregates were immersed. After Tarrar and Wagh (1992)

result in repulsion and therefore stripping. Figure 4.24 is an interaction diagram for calcite-based RD limestone and four bitumen types. In this case adhesion is favored due to oppositely charged surfaces over a wide range of pH conditions.

Hefer et al. (2007) considered this approach to be of fundamental importance toward the quantification of adhesion in the presence of water, with recognition of the effect of pH. Further discussions on this subject are presented under *Thermodynamic Theory*.

For a detailed description of mechanisms of charge development on aggregate surfaces, the reader is referred to Mertens and Wright (1959). A general, and traditional, classification of aggregates based on surface charge is presented in Fig. 4.25. This classification corresponds to zeta potential data for the two major aggregate types presented above.

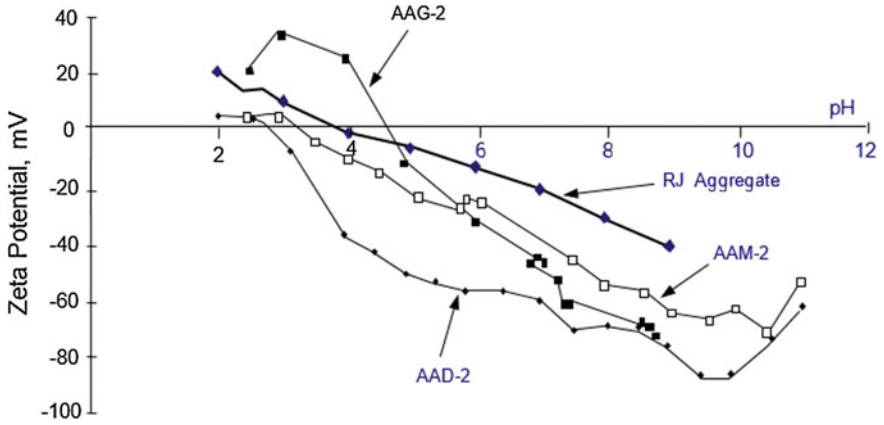


Fig. 4.23 Interaction diagrams for granite aggregate (RJ) with bitumen. After Labib (1992)

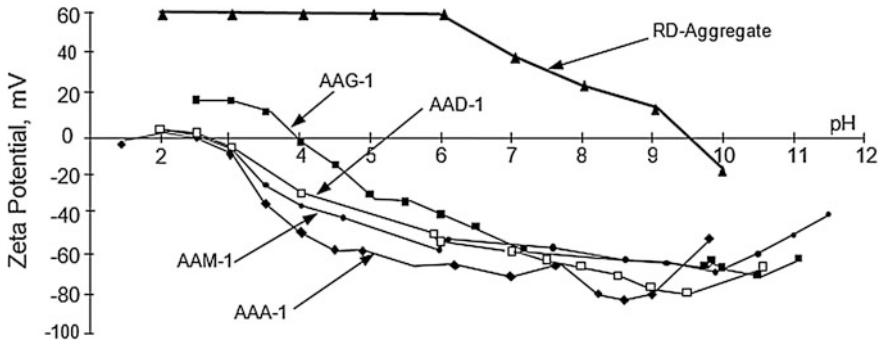


Fig. 4.24 Interaction diagram for limestone aggregate (RD) with bitumen. After Labib (1992)

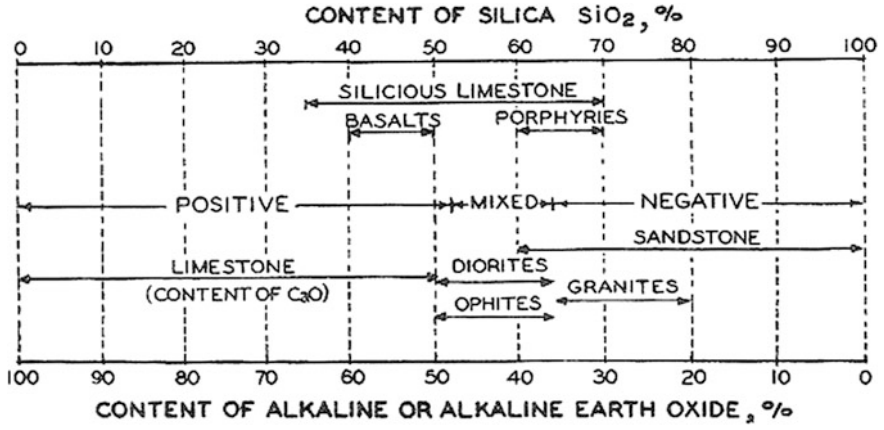
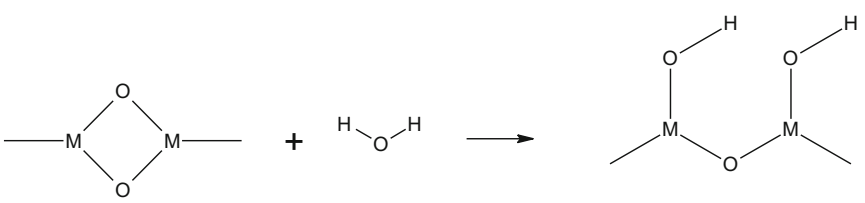


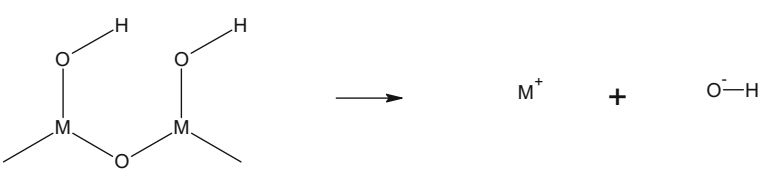
Fig. 4.25 Classification of aggregates. After Mertens and Wright (1959)

4.10 Surface Potential

Interfacial activity between asphalt cement and the aggregate surface is fundamentally important in assessing stripping potential. Yoon and Tarrer state that functional groups of asphalt that are adsorbed on the aggregate surface come mainly from the acid fraction of the asphalt. Yoon and Tarrer offer the example of carboxylic acid (R–COOH), which in the presence of water separates into the carboxylate anion (R–COO–) and the proton (H+). This causes the asphalt surface to have a negative polarity at the interface. Aggregates with water present are negatively charged, and as a result, a repulsive force develops between the negatively charged aggregate surface and the negatively charged asphalt surface at the interface. Payatakis (1975) states that solid surfaces in contact with water usually acquire changes through chemical reactions at the solid surface and adsorption of complex ions from the solution. For example, metal oxide surfaces in water hydrolyze to form hydroxyl groups:



After subsequent dissociation



A high pH value of the water in contact with the mineral surface will cause the surface to be more negatively charged. Yoon and Tarrer (1988) report that the intensity of the repulsion developed between the asphalt and aggregate depends on the surface charge of both the asphalt and aggregate. They used zeta potential as a method to measure aggregate surface charge and found a general trend that aggregates that have a relatively high surface potential in water are more susceptible to stripping (see Fig. 4.26).

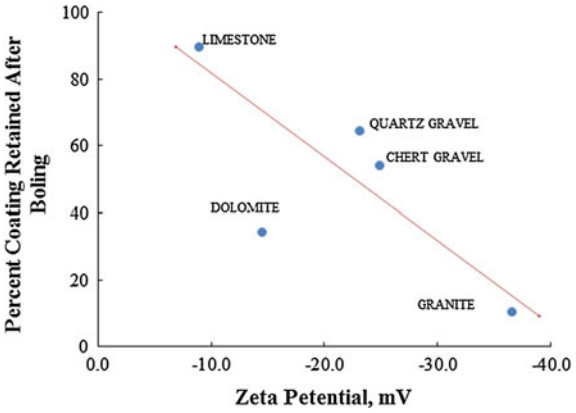


Fig. 4.26 Comparison of aggregate surface potential and stripping propensity as determined by boiling water tests. After Yoon and Tarrar (1988)

4.11 SHRP Research on Aggregate Surface Chemistry

Labib (1991) confirmed the existence of a range of acid–base types among various SHRP aggregates using zeta potential measurements and electrophoretic mobility. He reported that it is significant that the initial pH of aggregates was greater than 9.0, irrespective of aggregate type. This would neutralize the bitumen carboxylic acids at the interface and cause hydrolysis of bitumen–aggregate bonds. The high pH was attributed to basic soluble salts even in acidic aggregates. Labib (1991) documented the sensitivity of the bitumen–aggregate bond to pH. He identified three pH regions. At pHs above about 8.5 (Region 3), dissolution of the surface silica occurred in quartz or silica aggregates. In carbonate-based aggregates at pHs between about 1 and 6 (Region 1), calcium ion dissolution occurred, and the presence of carboxylic acids enhanced stripping in this region through cohesive failure in the aggregate. Podoll et al. (1991) used surface analysis by laser ionization to confirm that bitumen–aggregate bond disruption occurs within the aggregate and not at the interface. They found notably less sodium, potassium, and calcium in the top monolayer of aggregate in stripped areas than in unstripped areas. This indicates that dissolution of the cations was greater where bitumen had been stripped away. Scott's (1978) work on bitumen-coated glass slides supports Podoll et al. He found that debonding occurred in the more water-soluble glasses and not in the more stable opal glasses. Jamieson et al. (1995) conclude that net adsorption of bitumen on aggregate is a function of five aggregate variables: potassium oxide, surface area, calcium oxide, zeta potential, and sodium oxide. Alkali earth metals (sodium and potassium) are detrimental to adhesion. Higher surface area provides more active sites per unit mass for interaction. Calcium forms water-resistant bonds, and aggregates with a more negative surface charge may provide more potential for adsorption.

4.12 SHRP Adhesion Model

The SHRP adhesion model concludes that aggregate properties have a greater impact on adhesion than do various binder properties. Adhesion is achieved mainly by polar constituents in the bitumen bonding with active aggregate sites through hydrogen bonding, van der Waals interaction, or electrostatic forces. The general trend is that sulfoxides and carboxylic acids have the greatest affinity for aggregates. However, in the presence of water, sulfoxides and carboxylic acid groups are more susceptible to debonding, whereas phenolic groups and nitrogen bases are more effective in providing a durable bond (Jamieson et al. 1995). It is also apparent that aromatic hydrocarbons have much less affinity for aggregate surfaces than the polar groups.

4.13 SHRP Stripping Model

The SHRP view is that stripping is controlled by cohesive failure within the aggregate rather than at the bitumen–aggregate interface (Jamieson et al. 1995). Surfaces rich in alkali metals are more susceptible to debonding than surfaces rich in alkaline earth metals because the latter form water-insoluble salts with acid and other groups with the bitumen. Podoll et al. (1991) state that stripping of siliceous aggregate may be associated with the presence of water-soluble cations and aluminosilicates. The mechanism is probably dissolution of salts, dissociation of silica due to the high pH environment generated by solubilization of alkaline earth cations, electrostatic repulsion between negatively charged aggregates and ionic components of the bitumen at the surface, and dissolution of soaps formed between acid anions on the bitumen surface and alkali metal cations on the aggregate surface. The superior stripping resistance of some limestones is due to the formation of water-insoluble (covalent) bonds between calcium sites on the aggregate and bitumen constituents, but stripping of calcareous aggregate can occur where their water solubility is high.

4.14 Ways to Improve Adhesion

4.14.1 Interaction of Acidic Aggregates and Asphalt with Alkaline Amine Compounds

Amines have a long hydrocarbon chain. The chain is compatible in asphalt cement, and, in the presence of water, the amine is ionized to form an amine ion, $R–NH_3$, which has a positive charge (cationic). The physical properties of fatty amines can be altered by changing the nature of the hydrocarbon chain while the chemical

nature can be altered by changing the number of amine groups and their positions in the molecule (Porubszky et al. 1969). Taken together, the chain length and number of amine groups greatly influence the adhesion of the asphalt. Optimum performance is typically achieved with 14–18 carbon chain amines with one or two amine groups (Porubszky et al. 1969; Tarrer and Wagh 1991). Fatty amines enable asphalt to wet aggregate surfaces. The amine group reacts with the aggregate surface while the hydrophobic hydrocarbon chain of the fatty amine is anchored in the asphalt. The net effect (Tarrer and Wagh) is that the long hydrocarbon chain acts as a bridge between the hydrophilic aggregate and the hydrophobic bitumen surface, encouraging a strong bond.

4.14.2 Effect of Hydrated Lime on Adhesive Bond

Plancher et al. (1977) hypothesized that hydrated lime improved binder–aggregate adhesion by interacting with carboxylic acids in the asphalt and forming insoluble salts that are readily adsorbed at the aggregate surface. This is an important reaction because hydroxyl (OH) groups are found on the surfaces of siliceous aggregates. These SiOH groups form hydrogen bonds with carboxylic acid groups from asphalt and strongly affect the adhesion between the asphalt and aggregate (Hicks 1991). But this hydrogen bond is quickly broken in the presence of water, and the two groups dissociate and reassociate with water molecules through hydrogen bonding. This means that the hydrogen bonding between the water molecules and the SiOH group and between the water molecule and the COOH group is preferred over the bond between SiOH and COOH. When lime is added, some dissociation of the $Ca(OH)_2$ molecule occurs, resulting in calcium ions (Ca^{++}). These ions interact with the carboxylic acids (COOH) and 2-quinolenes (Petersen et al. 1987) to form rather insoluble calcium organic salts. This leaves the SiOH molecule free to bond with nitrogen groups in the asphalt (Petersen et al. 1987). These bonds are strong and contribute to adhesion. Figure 4.2 illustrates some of the important functional groups in asphalt.

Schmidt and Graf (1972) state that the effectiveness of hydrated lime as an anti-stripping agent cannot be completely explained by the reaction between calcium from lime and the acids in the asphalt. They state that lime provides calcium ions that migrate to the aggregate surface and replace hydrogen, sodium, potassium, or other cations. In 1997 the Western Research Institute provided an excellent explanation of the hypothesized aggregate–asphalt interaction. Susceptibility to water stripping depends, at least in part, on the water solubility of organic salts formed from the reaction of carboxylic acids in the bitumen with carbonates in the aggregates. High-molecular weight magnesium and calcium salts are relatively hydrophobic and not very soluble in water. Sodium salts, being more soluble, lead to stripping. Further, it was found in SHRP research that carboxylic acids in bitumen hydrogen bond very strongly with hydroxyl groups on siliceous aggregates, these being highly concentrated on the aggregate surface. However, this

hydrogen bond is highly sensitive to disruption by water, thus accounting, at least in part, for the high moisture sensitivity to moisture damage of pavement mixtures containing siliceous aggregates. Conversion of carboxylic acids to insoluble salts (e.g., calcium salts) prior to use in pavement mixtures could prevent adsorption of water-sensitive free acids on the aggregate in the first place. When pavement containing surface active materials is wet and is subjected to mechanical action of traffic, it is predictable that the surface activity of the sodium carboxylates (soaps) in the bitumen will help scrub the oil (bitumen) away from the rock. The practical, perhaps conservative, solution to the historical problem of stripping is to convert all acidic materials in asphalt to water-insensitive (nonsurface active) calcium salts at the time of production. This would require lime treatment at the refinery. Some refineries do this today (SHRP bitumen AAG). The recommendation here is that conversion of acids to calcium salts be made a universal requirement. The process recommended here reduces moisture susceptibility of the whole asphalt rather than just at the interface. Lime treatment of the aggregate is desirable. Yoon and Tarrer (1988) discuss the effect of water pH on stripping potential in asphalt mixtures with respect to anti-stripping additives. Their analysis showed that as the pH of the water increases, the adsorptive bonds between amine-type additives and aggregate surfaces are weakened. As a result, water can more easily displace asphalt from the aggregate surface. They point out that this is not the case with hydrated lime, where the resistance to stripping is independent of the pH of the contacting water. However, other research has shown that normally pHs as high as 10 will not dislodge amines from the aggregate, and pHs greater than 10 are very unusual. The effectiveness of the polyamine additives increased with curing time in studies by Yoon and Tarrer (1988). They found that by storing asphalt–aggregate mixtures for a few hours at 300 °F, the effectiveness of some additives improved considerably even at a high pH value of contacting water. Yoon and Tarrer (1988) hypothesize that the reason for the improved performance with curing might be the development of a film of polymerized asphalt.

4.14.3 Other Chemical Treatments

Jamieson et al. (1995) describe three possible treatments to improve adhesion: addition of cations to the aggregate surface, addition of anti-stripping agents to the bitumen, and aggregate pretreatment with organosilanes. Jamieson et al. (1995) point to research that shows that enhanced bonding is associated with relatively large concentrations of iron, calcium, magnesium, and aluminum at the aggregate surface. Jamieson et al. describe that the principal role of anti-stripping agents is to trigger the dissociation of aggregations of bitumen components, thereby increasing the availability of bitumen functional groups for active sites on the aggregate surface. Bonding energy measurements indicate that the effectiveness of aggregate pretreatment with modifiers is dependent on aggregate type, probably because antistrip agents are usually amines with relatively similar properties, whereas

aggregates vary widely (Jamieson et al. 1995). Organosilane pretreatment of aggregate increases the number of polar adsorption sites on the aggregate surface (Divito and Morris 1982; Graf 1986). Research during SHRP ranked the overall performance of organosilane treatments as a function of hydrophobic bonding enhancements and determined the order of ranking to be amino > hydrocarbon > thiol.

4.15 Dusty and Dirty Aggregates

4.15.1 General Mechanisms of Bond Disruption with Dirty or Dusty Aggregates

Dusty aggregates may generally be referred to as aggregates coated with materials smaller than 75 μm. This may cause a problem in developing an acceptable bond between fine and coarse aggregate because the asphalt binder tends to coat the dust and not the aggregate, leading to a greater probability for bond interruption and hence displacement. Dirty aggregates normally refer to aggregates coated with clay mineral fines. While clay-sized materials are soil particles smaller than 2 μm, true clays are not only very small particles but also have a unique mineralogy and morphology. Clay minerals are made up of alternating layers of silca and alumina, which comprise particles that have a great affinity to adsorb water. This is why clay fines are plastic in nature and have a large plasticity index [range of moisture content between the plastic limit (where the soil acts as a plastic semisolid) and the liquid limit (where the soil acts as a liquid)]. The presence of clay particles on the aggregate surface is similar to that of dust. The asphalt bond with the fine and coarse aggregate is disrupted by the presence of the dust or clay. In fact, the situation is worse with clay fines because these particles have a tendency to swell when they take on water, and this swelling mechanism can break or disrupt an existing bond with asphalt. Furthermore, clay is more active than other soil particles. This can lead to other complex reactions between asphalt, water, and the clays, including emulsification. Clay particles adsorb cations because of their strong negative surface charge and their enormous specific surface area. The amount and nature of the cations adsorbed can affect bond interactions and emulsification potential. In summary, aggregates coated with dust or clay disrupt the asphalt–aggregate bond and can also lead to more complex reactions among water, asphalt, and aggregate, such as emulsification. Kandhal et al. (1998) evaluated aggregate tests to assess the potential for aggregate fines to cause stripping in asphalt mixtures. They considered the sand equivalent test, the plasticity index test, and the methylene blue test. They evaluated a set of 10 asphalt mixtures using a common coarse limestone aggregate but with different fine aggregates. They used two validation tests to assess moisture damage: American Association of State Highway and Transportation T283 and the Hamburg wheel-tracking test. After a careful

statistical analysis of results, they found that the methylene blue test did the best job of identifying moisture sensitivity of the mixtures.

4.15.2 Modification of Dusty and Dirty Aggregates to Improve Asphalt–Aggregate Interaction

Hydrated lime has been used to treat dusty and dirty aggregates. The mechanism is partially because hydrated lime reacts with clay to change its properties. Two basic mechanisms are involved: cation exchange or molecular crowding of calcium hydroxide molecules at the surface of the clay and pozzolanic reaction. Cation exchange or calcium hydroxide crowding provides an abundance of divalent calcium ions, which, because of their high concentration and divalent nature, replace the normally available cations in the clay environment. This leads to a substantial reduction in clay plasticity (Little 1995) and causes clay colloids to flocculate into larger aggregates (Little 1995). However, the most important reaction is the pozzolanic reaction, where caustic calcium hydroxide raises the pH of the lime-water-clay system to more than 12. At this high pH, clay minerals are altered and the silica and alumina become soluble. Soluble silica and alumina then combine with free calcium cations to form calcium silicate and calcium aluminate hydrates, further reducing plasticity, stabilizing the clay, and forming more well-cemented agglomerates (Little 1995). One might expect that the lime-modified clay coatings will "peal" from the aggregates and no longer remain as coatings but as "cemented" small aggregates of flocculated clay that can be separately coated with asphalt.

4.16 Exercises

4.16.1 Describe how hydrated lime impacts the moisture resistance between asphalt mixtures. In your answer address the impact of calcium cations at the surface of siliceous mineral aggregates and the impact of hydrated lime particles dispersed within the asphalt binder.

4.16.2 Describe the mechanism by which hydrated lime used as filler in asphalt binder reacts with asphalt binder to reduce the impact of oxidative aging. How would this impact asphalt performance over a 10-year surface life of the pavement with respect to fatigue cracking, low-temperature cracking and rutting?

4.16.3 Describe the mechanism by which SBS polymer added at the rate of 4–6% by weight of asphalt binder may form a continuous matrix that protects the asphalt from excessive plastic deformation. How does this matric impact fatigue damage and fracture toughness.

4.16.4 Which of the following functional groups are most susceptible to be de-bonded in the presence of water: anhydrides, 2-quinolone types, carboxylic acids, pyridine types, sulfoxides, ketones, phenolics, and pyrrolics? How can you lessen the potential for such debonding?

4.16.5 What do Figs. 4.23 and 4.24 tell us about the security of the asphalt-aggregate bond over the life of an asphalt mixture? Which aggregate would use select for your asphalt pavement based on these two figures and why?

4.16.6 List and describe the seven moisture damage mechanisms.

4.16.7 Describe the dispersed polar fluids model.

4.16.8 Using Fig. 4.1 as a guide, describe how pore pressure in a saturated asphalt mixture will lead to rapid degradation of the mixture under traffic loading.

4.16.9 Describe the steps you would take as a pavement engineer to guard against moisture damage through the following phases: aggregate-binder selection, mineral filler selection, modifier (polymer or other), pavement structural design, consideration of moisture infiltration or capillary rise.

4.17 Summary and Conclusions

Although several separate mechanisms have been identified to explain the process of moisture damage in asphalt pavements, it is more likely that most asphalt pavements suffer moisture damage as a result of a synergy of several processes. From a chemical standpoint, the literature is clear that neither asphalt nor aggregate has a net charge, but components of both have nonuniform charge distributions, and both behave as if they have charges that attract the opposite charge of the other material. Researchers point out that certain polar asphalt compounds develop more tenacious and moisture-resistant bonds with the aggregate surface than others and that the development of the more tenacious and long-lasting bonds can be promoted by treatment of the asphalt mixtures with additives. The most durable bonds appear to be formed by interaction of phenolic groups and nitrogen bases from the bitumen. These form insoluble salts. While sulfoxides and carboxylic acids have a greater affinity for the aggregate surfaces, they are most susceptible to dissolution on water. The asphalt–aggregate bond is affected by aggregate mineralogy, adsorbed cations on the aggregate surface, and the surface texture and porosity. Favorable chemical bonding between asphalt and aggregate alone will not optimize the adhesive bond and minimize moisture damage. The bond is part physical, and, therefore, the asphalt must be able to wet and permeate the aggregate surface. This process is dependent on asphalt rheology at mixing temperatures and the nature of

the aggregate surface, pore size, pore shape, and aggregate mineralogy. To complicate matters somewhat, the ability to bond asphalt to aggregate is dynamic and changes with time. This is largely affected by the shift in pH at the aggregate–water interface, which can be triggered by dissociation of aggregate minerals near the surface or by the nature of the pore water (cation type and concentration). Moisture damage is certainly not limited to adhesive failure, but weakening of the cohesive strength of the mastic due to moisture infiltration is equally important. Recent research has shown that water can diffuse into asphalt of mastics and that each can hold an appreciable amount of water. Research over many years has clearly shown that this water can weaken the asphalt mixture, making it more susceptible to damage. Thus, the logical view is that the deleterious effects of moisture on the adhesive and cohesive properties, both of which influence asphalt mixture performance, must be considered.

References

Adamson, A. W., & Gast, A. P. (1997). *Physical Chemistry of Surfaces* (6th ed.). New York: Wiley.

Adão, M. H., Saramago, B., & Fernandes, A. C. (1998). Estimation of the surface tension components of thiodiglycol. *Langmuir, 14,* 4198–4203.

Asphalt Institute. (1981). Cause and prevention of stripping in asphalt pavements. Educational Series No. 10, College Park, Md.

Arderbrant, H., & Pugh, R. J. (1991). Surface acidity/basicity of road stone aggregates by adsorption from non-aqueous solutions. *Colloids and Surfaces, 53,* 101–116.

Brannan, C. J., Jeon, Y. W., Perry, L. M., & Curtis, C. W. (1991). Adsorption behavior of asphalt models and asphalts on siliceous and calcareous aggregates. *Transportation Research Record, 1323,* 10–21.

Bhairampally, R. K., Lytton, R. L., & Little, D. N. (2000). Numerical and graphical method to assess permanent deformation potential for repeated compressive loading of asphalt mixtures. *Transportation Research Record: Journal of the Transportation Research Board*, *1723*, 150–158.

Bhasin, A., Little, D. N., Vasconcelos, K. L., & Masad, E. (2007). Surface free energy to identify moisture sensitivity of materials for asphalt mixes. *Transportation Research Record, 2001,* 37–45.

Bhasin, A., & Little, D. N. (2007). Characterization of aggregate surface energy using the universal sorption device. *Journal of Materials in Civil Engineering, 19,* 634–641.

Bikerman, J. J. (1978). Surface energy of solids. *Topics in Current Chemistry, 77,* 1–66.

Boyd, G. E., & Livingston, H. K. (1942). Adsorption and the energy changes at crystalline solid surfaces. *Journal of the American Chemical Society, 64,* 2383–2388.

Butt, H.-J., Graf, K., & Kappl, M. (2003). *Physics and chemistry of interfaces*. Weinheim: WILEY-VCH Verlag GmbH & Co. KgaA.

Castan, M. (1968). Rising of binder to the surface of an open-graded bituminous mix. *Bulletin de liaison des laboratoires routiers*, *33*, 77–84.

Carpick, R. W., Agrait, N., Ogletree, D. F., & Salmeron, M. (1996). Measurement of interfacial shear (friction) with an ultrahigh vacuum atomic force microscope. *Journal of Vacuum Science and Technology, 14,* 1289–1295.

Chadan, C., Sivakumar, K., Masad, E., & Fletcher, T. (2004). Application of imaging techniques to geometry analysis of aggregate particles. *Journal of Computing in Civil Engineering* (*ASCE*), 75–82.

Chaudhury, M. K., & Good, R. J. (1991). Chapter 3. In L. H. Lee (Ed.), *Fundamentals of adhesion*. New York: Plenum Press.

Cheng, D. Z., Little, D. N., Lytton, R. L., & Holste, J. C. (2001). Surface free energy measurement of aggregates and its application on adhesion and moisture damage of asphalt–aggregate system. In *Proceeding of the 9th International Center for Aggregate Research Symposium*, Austin, Texas.

Cheng, D., D. N. Little, R. L. Lytton, and J. C. Holste. (2002). Surface energy measurement of asphalt and its application to predicting fatigue and healing in asphalt mixtures. *Transportation Research Record: Journal of the Transportation Research Board*, *1810*, 44–53.

Chibowski, E., & Waksmundzki, A. (1978). A relationship between the zeta potential and surface free energy changes of the sulfur/n-Heptane-Water system. *Journal of Colloid and Interface Science, 66,* 213–219.

Clint, J. H. (2001). Adhesion and components of solid surface energies. *Current Opinion in Colloid and Interface Science, 6,* 28–33.

Cucalon, M. G. (2016). Physico-chemical interactions at binder-aggregate interface. Ph.D. dissertation, Texas A&M Univesity, College Station, Texas.

Curtis, C. W. (1992). Fundamental properties on asphalt aggregate interactions adhesion and adsorption. Final report on contract A-003B. Strategic highway research program, National Research Council, Washington, D.C.

Curtis, C. W., Lytton, R. L., Brannan, C. J. (1992). Influence of aggregate chemistry on the adsorption and desorption of asphalt. In *Transportation Research Record 1362, TRB* (pp. 1–9). National Research Council, Washington, D.C.

Curtis, C. W., Clapp, D. J., Jeon, Y. W., & Kiggundu, B. M. (1989). Adsorption of model asphalt functionalities, AC-20, and oxidized asphalts on aggregate surfaces. *Transportation Research Record, 1228,* 112–127.

Della Volpe, C., & Siboni, S. (2000). Acid-base surface free energies of solids and the definition of scales in the Good-van Oss-Chaudhury theory. *Journal of Adhesion Science and Technology, 14*(2), 235–272.

Divito, J. A., & Morris, G. R. (1982). Silane pretreatment of mineral aggregate to prevent stripping in flexible pavements. *Transportation Research Record*, *843*, 104–111.

Dixon, J. B. & Schulze, D. G. (2002). Soil mineralogy with environmental applications. In Y. Deng & J. B. Dixon (Eds.), *Soil organic matter and organic minerals interactions*. Madison, Wisconsin: Soil Science Society of America, Inc.

Endersby, V. A., Griffin, R. L., & Sommer, H. J. (1947). Adhesion between asphalts and aggregates in the presence of water. *Proceedings of the Association of Asphalt Paving Technologists, 16,* 411–451.

Ensley, J. C. Petersen, & Robertson, R. E. (1984). Asphalt-aggregate bonding energy measurements by microcalorimetric methods. *Thermochimica Acta, 77,* 95–107.

Ernstsson, M. M., & Larsson, A. (1999). A multianalytical approach to characterize acidic adsorption sites on a quartz powder. *Colloids and Surfaces, 168,* 215–230.

Fehsendfeld, F. M., & Kriech, A. J. (Undated). *The effect of plant design changes on hot mix asphalt*. Heritage Research Group.

Fowkes, F. M. (1964). Attractive forces at interfaces. *Industrial and Engineering Chemistry, 56,* 40–52.

Fromm, H. J. (1974a). The mechanisms of asphalt stripping from aggregate surfaces. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 43, pp. 191–223). 68 Moisture Sensitivity of Asphalt Pavements: A National Seminar.

Fromm, H. J. (1974b). The mechanisms of asphalt stripping from aggregate surfaces. *Proceedings of the Association of Asphalt Paving Technologists, 43,* 191–223.

Good, R. J., & van Oss, C. J. (1991). The modern theory of contact angles and the hydrogen bond component of surface energies. In M. E. Schrader & G. Loeb (Eds.), *Modern approach to wettability: theory and applications* (pp. 1–27). New York: Plenum Press.

Graf, P. E. (1986). Factors affecting moisture susceptibility of asphalt concrete mixes. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 55, p. 175).

Gregg, S. J. & Sing, K. S. W. (1967). *Adsorption, surface area and porosity*. London: Academic Press,

Gzemski, F. C., McGlashan, D. W., & Dolch, W. L. (1968). Highway research circular 78: Thermodynamic aspects of the stripping problem. HRB, National Research Council, Washington, D.C.

Harkins, W. D., & Boyd, G. E. (1942). The binding energy between a crystalline solid and a liquid: The energy of adhesion and emersion. Energy of emersion of crystalline powders. II. *Journal of the American Chemical Society, 64,* 1195–1204.

Hefer, A. W., Little, D. N., & Herbert, B. E. (2007). Bitumen surface energy characterization by inverse gas chromatography. *Journal of Testing and Evaluation, 35,* 233.

Hicks, G. R. (1992). *Moisture damage in asphalt concrete, national cooperative highway research program synthesis of highway practice 175*. Washington, D.C.: Transportation Research Board.

Hicks, R. G. (1991). NCHRP synthesis of highway practice 175: Moisture damage in asphalt concrete. TRB, National Research Council, Washington, D.C.

Hindermann, W. L. (1968). The swing to full-depth: The case for laying asphalt on the raw subgrade. Information Series No. 146, Asphalt Institute, College Park, Md.

Hubbard, P. (1938). Adhesion of asphalt to aggregates in presence of water. *Highway Research Board Proceedings*, *18*(Part 1), 238–249.

Huang, S-C., Branthaver, J. F., & Robertson. (2000). International Centre for Aggregate Research, Denver, Colorado.

Hubbard, P. (1938). Adhesion of asphalt to aggregate in the presence of water. *Highway Research Board*, *8*(Part 1).

Hughes, R. I., et al. (1960). Adhesion in Bitumen Macadam. *Journal of Applied Chemistry*, *10*.

Jamieson, I. L., Moulthrop, J. S., & Jones, D. R. (1995). *SHRP results on binder—Aggregate adhesion and resistance to stripping. Asphalt yearbook 1995*. United Kingdom: Institute of Asphalt Technology.

Jeon, W. Y., Curtis, C. W., & Kiggundu, B. M. (1988). Adsorption behavior of asphalt functionalities on dry and moist silica. Submitted to TRB.

Johnston, C. T. (1996). Sorption of organic compounds on clay minerals: A surface functional group approach. In B. Sawhney (Ed.), *Organic pollutants in the environment* (Chapter 1). Boulder, CO: Clay Minerals Society.

Jones, D. R. (1992). An asphalt primer: Understand how the origin and composition of paving grade asphalt cements affect their performance. SHRP Technical Memorandum No. 4. Strategic Highway Research Program, National Research Council, Washington, D.C.

Jura, G., & Harkins, W. D. (1944). Surfaces of solids. XI. Determination of the decrease (p) of free surface energy of a solid by an adsorbed Film. *Journal of the American Chemical Society, 66,* 1356–1362.

Kandhal, P. S., Lynn, C. Y., & Parker, F. (1998). Test for plastic fines in aggregates related to stripping in asphalt paving mixtures. Report 98-3. National Center for Asphalt Technology.

Kennedy, T. W., Roberts, F. L., & Lee, K. W. (1984). Evaluating moisture susceptibility of asphalt mixtures using the texas boiling test. *Transportation Research Record*, *968*, 45–54.

Kiggundu, B. M. (1986). Effects of submergence in distilled water on the surface coloration of Asphalt. Unpublished data, NMERI.

Kiggundu, B. M., & Roberts, F. L. (1988). The success/failure of methods used to predict the stripping potential in the performance of bituminous pavement mixtures. Submitted to TRB.

Kim, Y. R., Lee, H. J., & Little, D. N. (1997). Fatigue characterization of asphalt concrete using viscoelasticity and continuum damage mechanics. *Journal of Association of Asphalt Paving Technologists, 66,* 520–569.

Kim, Y.-R., Little, D. N., & Lytton, R. L. (2002). Fatigue and healing characterization of asphalt mixtures. *Journal of Materials in Civil Engineering, American Society of Civil Engineers*.

Kittrick, J. A. (1977). Mineral Equilibria and the Soil System. In J. B. Dixon & S. B. Weed (Eds.), *Minerals in soils environments* (pp. 1–25). Wisconsin: Soil Science Society of America.

Kolasinski, K. W. (2002). *Surface science: Foundations of catalysis and nanoscience chichester*. Wiley.

Kwok, D. Y. (1999). The usefulness of the lifshitz van der Waals/acid-base approach to surface tension components and interfacial tension. *Colloids and Surfaces A: Physiochemical and Engineering Aspects, 156,* 191–200.

Labib, M. (1991). End of Phase II report: Evaluation of donor-acceptor properties of asphalt and aggregate materials and relationship to asphalt composite performance. Strategic Highway Research Program, National Research Council, Washington, D.C.

Labib, M. E. (1992). Asphalt-aggregate interactions and mechanisms for water stripping. *American Chemical Society, Fuel, 37,* 1472–1481.

Lesueur, D., & Little, D. N. (1999). Effect of hydrated lime on rheology, fracture, and aging of bitumen. *Transportation Research Record: Journal of the Transportation Research Board*, *1661*, 93–105.

Little, D. N. (1995). *Handbook for stabilization of bases little, subbases with lime*. New York: Kendall-Hunt Publishing Co.

Little, D. N., Lytton, R. L., & Williams, D. (1998). Propagation and healing of microcracks in asphalt concrete and their contributions to fatigue. In A. Usmani (Ed.), *Asphalt science and technology* (pp. 149–195). New York: Marcel Dekker Inc.

Little, D. N., Lytton, R. L., Williams, D., & Kim, R. Y. (1999). Analysis of the mechanism of microdamage healing based on the application of micromechanics first principles of fracture and healing. *Journal of Association of Asphalt Paving Technologists, 68,* 501–542.

Little, D. N. & Bhasin, A. (2006). *Using surface energy measurements to select materials for asphalt pavement*. National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C.

Little, D. N., & Jones, J. R. (2003). *Chemical and mechanical mechanisms of moisture damage in hot mix asphalt pavements*. San Diego, CA: National Seminar in Moisture Sensitivity.

Lytton, R. L., Uzan, J., Fernando, E. G., Roque, R., Hiltmen, D., & Stoffels, S. (1993). Development and validation of performance prediction models and specifications for asphalt binders and paving mixes. SHRP Report A-357. Strategic Highway Research Program, National Research Council, Washington, D.C.

Lytton, R. L., Chen, C. W., & Little, D. N. (1998). Microdamage healing in asphalt and asphalt concrete. Volume 3: A micromechanics fracture and healing model for asphalt concrete. Report FHWA-RD-98-143. College Station, Texas: Texas Transportation Institute.

Lytton, R. L. (2000). Characterizing asphalt pavements for performance. *Transportation Research Record*, *1723*, 5–16.

Lytton, R. L. (2004). *Adhesive fracture in asphalt concrete mixtures*. Chapter in J. Youtcheff (Ed.).

Mack, C. (1964). *Bituminous materials* (A. Holberg, ed.) (Vol. 1). New York: Interscience Publishers.

Majidzadeh, K., & Brovold. F. N. (1968). Special Report 98: State of the Art: Effect of Water on Bitumen–Aggregate Mixtures. HRB, National Research Council, Washington, D.C.

Masad, E., Tashman, L., Little, D. N., & Zbib, H. (2004) Viscoplastic modeling of asphalt mixes with the effects of anisotropy, damage and aggregate characteristics. Submitted for Publication in the *Journal of Mechanics of Materials*.

Maupin, G. W. (1982). The use of antistripping additives in Virginia. Submitted at the 51st Association of Asphalt Paving Technologists, Kansas City, MO. Morgan, P., and P. Mulder. 1995. *The Shell Bitumen Industrial Handbook* (p. 120).

McMurry, J. (2000). *Organic chemistry* (5th ed.). USA: Brooks/Cole.

Merril, W. W., & Pocius, A. V. (1991). Direct measurement of molecular level adhesion forces between biaxially oriented solid polymer films. *Langmuir, 7,* 1975–1980.

Mertens, E. W., & Wright, J. R. (1959). Cationic emulsions: How they differ from conventional emulsions in theory and practice. *Highway Research Board Proceedings, 38,* 386–397.

Morgan, P., & Mulder, A. (1995). *The shell bitumen industrial handbook*. London: Thomas Telford Publishing.

Myers, D., (1991). *Surfaces, interfaces, and colloids: Principles and applications*. New York: VCH Publishers, Inc.

Neu, T. R. (1996). Significance of bacterial surface-active compounds in interaction of bacteria with interfaces. *Microbiology and Molecular Biology Reviews, 60,* 151–166.

Nguyen, T., Byrd, E., & Bentz, D. (1992). Insitu measurement of water at the asphalt/siliceous aggregate interface. *Americal Chemical Society, Fuel, 37,* 1466–1471.

Park, S., Jo, M. C., & Park, J. B. (2000). Adsorption and thermal desorption behaviour of asphalt-like functionalaities on silica. *Adsorption Science & Technology, 18,* 675–684.

Payatakis, A. C. (1975). Surface chemistry applied to solid–liquid separations. In *Theory and practice of solid–liquid separation*.

Petersen, J. C. (1988). Lime-treated pavements offer increased durability. *Roads & Bridges Magazine*.

Petersen, C. J., & Plancher, H. (1998). Model studies and interpretive review and the competitive adsorption and water displacement of petroleum asphalt chemical functionalities on mineral aggregate surfaces. *Petroleum Science and Technology, 16,* 89–131.

Petersen, J. C., Plancher, H., & Harnsberger, P. M. (1987). Lime Treatment of Asphalt—Final Report. National Lime Association.

Petersen, J. C., Plancher, H., Ensley, E. K., Miyake, G., & Venable, R. L. (1982). Chemistry of asphalt-aggregate interaction: Relationship with moisture damage prediction test. *Transportation Research Record, 843,* 95.

Petersen, C. J. (1986). Quantitative functional group analysis of asphalts using differential infrared spectrometry and selective chemical reactions—Theory and application. *Transportation Research Record, 1096,* 1.

Plancher, H., Dorrence, S., & Petersen, J. C. (1977). Identification of chemical types in asphalts strongly absorbed at the asphalt-aggregate interface and their relative displacement by water. *Proc Association of Asphalt Paving Technologists, 46,* 151–175.

Pocius, A. V. (1997). *Adhesion and adhesives technology*. Ohio: Hanser/Gardner Publications, Inc.

Podoll, R. T., Becker, C. H., & Irwin, K. C. (1991). Phase II progress report: Surface analysis by laser ionization of the asphalt–aggregate bond. Strategic Highway Research Program, National Research Council, Washington, D.C.

Porubszky, I., Csizmadia, M., Szebenyi, E., Dobozy, O. K., & Simon, M. (1969). Bitumen Adhesion to Stones. Chimie, physique et applications pratiques des agents de surface: compte-rendus du 5ème Congrès International de la Détergence, Barcelona, Spain, 9–13 Sept (Vol. 2, Part 2, pp. 713–725).

Rice, J. M. (1958). Relationship of aggregate characteristics to the effect of water on bituminous paving mixtures. In *Symposium on Effect of Water on Bituminous Paving Mixtures*, ASTM STP 240 (pp. 17–34).

Robertson, R. E. (2000). Transportation research circular 499: Chemical properties of asphalts and their effects on pavement performance. TRB, National Research Council, Washington, D.C.

Sanderson, F. C. (1952). Methylchlorosilanes as antistripping agents. *Highway Research Board Proceedings, 31,* 288–300.

Schmidt, R. J., & Graf, P. E. (1972). The effect of water on the resilient modulus of asphalt treated mixes. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 41, pp. 118–162). 70 Moisture Sensitivity of Asphalt Pavements: A National Seminar.

Scott, J. A. N. (1978). Adhesion and disbonding mechanisms of asphalt used in highway construction and maintenance. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 47, pp. 19–48).

Schapery, R. A. (1984). Correspondence principles and a generalized J-integral for large deformation and fracture analysis of viscoelastic media. *International Journal of Fracture, 25,* 194–223.

Schapery, R.A. (1989) On the mechanics of crack closing and bonding in linear viscoelastic media. *International Journal of Fracture*, *39*, 163–189.

Schultz., & Nardin, M. (1994). Theories and mechanisms of adhesion. In A. Pizzi & K.L. Mittal (Eds.), *Handbook of adhesive technology* (pp. 19–33). New York: Marcell Dekker, Inc.

Shanahan, M. E. R. (1992). Wetting and spreading. In D. E. Packham (Ed.), *Handbook of adhesion* (pp. 506–509). Essex, England: Longman Group UK Ltd.

Shuttleworth, R. (1950). The surface tension of solids. *Proceedings of the Physical Society Section A, 63,* 444–457.

Si, Z., Little, D. N., & Lytton. R. L. (2002). Evaluation of fatigue healing effect of asphalt concrete by pseudostiffness. *Transportation Research Record: Journal of the Transportation Research Board*, *1789*, 73–79.

Tarrer, A. R. (1996). Use of hydrated lime to reduce hardening and stripping in asphalt mixtures. *Presented at the 4th Annual International Center for Aggregate Research Symposium*, Atlanta, GA.

Tarrer, A. R., & Wagh, V. (1991). The effect of the physical and chemical characteristics of the aggregate on bonding. Strategic Highway Research Program, National Research Council, Washington, D.C.

Taylor, M. A., & Khosla, N. P. (1983). Stripping of asphalt pavements: State of the art. *Transportation Research Record*, *911*, 150–158.

Terrel, R. L., & S. Al-Swailmi. 1994. Water sensitivity of asphalt–aggregate mixes: Test selection. SHRP Report A-403. Strategic Highway Research Program, National Research Council, Washington, D.C.

Terrel, R. L., & J. W. Shute. (1989). Summary Report on Water Sensitivity. SHRP-A/IR-89-003. Strategic Highway Research Program, National Research Council, Washington, D.C.

Thelen, E. (1958). Surface energy and adhesion properties in asphalt–aggregate systems. Bulletin 192, HRB (pp. 63–74). National Research Council, Washington, D.C.

Thomas, K. (2002). Quarterly Technical Report (pp. 25–62). Western Research Institute. FHWA Contract No. DTFH61-99C-00022. Laramie, Wyo.

Titova, T. I., Kosheleva, L. S., & Zhdanov, S. P. (1987). IR study of hydroxylated silica. *Langmuir, 3*(6), 960–967.

Tseng, K. H., & Lytton, R. L. (1987). Prediction of permanent deformation in flexible pavement materials. In H. G. Schreuders & C. R. Marek (Eds.), *Implication of aggregates in design, construction, and performance of flexible pavements*, ASTM STP 1016, American Society for Testing and Materials, Philadelphia, PA. Western Research Institute. 2002. Fundamental Properties of Asphalts and Modified Asphalts.

Task 11-5, FHWA Quarterly Technical Report, Contract No. DTFH61-99C-00022, Nov.

Van Oss, C. J., Good, R. J., & Chaudhury, M. K. (1988). Additive and nonadditive surface tension components and the interpretation of contact angles. *Langmuir, 4,* 884–891.

Van Oss, C. J. (1994). *Interfacial forces in aqueous media*. New York: Marcel Dekker Inc.

Van Oss, C. J., Wu, W., Docoslis, A., & Giese, R. F. (2001). The interfacial tensions with water and the Lewis acid-base surface tension parameters of polar organic liquids derived from their aqueous solubilities. *Colloid Surf. B-Biointerfaces, 20,* 87–91.

Van Oss, C. J. (2006). *Interfacial forces in aqueous media*. Taylor and Francis Group: CRC Press.

Van Oss, C. J., & Giese, R. F. (1995). The hydrophilicity and hydrophobicity of clay minerals. *Clays and Clay Minerals, 43,* 474–477.

Williams, D., Little, D. N., Lytton, R. L., Kim, Y. R., & Kim, Y. (1998). Fundamental properties of asphalts and modified asphalts. Task K, FHWA Final Report (Vol. 2), Contract No. DTFH61-92C-00170.

Woodward, R. P. (2000). Prediction of adhesion and wetting from Lewis acid base measurements. As Presented at TPOs in Automotive, http://www.firsttenangstorms.com.

Yoon, H. J. (1987). *Interface phenomenon and surfactants in asphalt paving materials*. Dissertation, Auburn University.

Yoon, H. H., & Tarrer, A. R. (1988). Effect of aggregate properties on stripping. In *Transportation Research Record 1171*, TRB, National Research Council, Washington, D.C., (Vol. 71, pp. 37–43).

Chapter 5
Modifiers and Fillers

5.1 Introduction

Additives to asphalt mixtures have been categorized in many ways. Roberts et al. (1996) broke them down as shown in Table 5.1. A detailed treatise on each type or category of additive deserves an individual chapter to adequately cover the subject. In this chapter, we focus on extenders, antistripping agents, fillers, and rubbers as they are the most commonly used to alter the properties and improve performance of asphalt mixtures.

Before going into depth regarding mechanisms of stabilization or modification of asphalt cement, it is appropriate to review some of the highlights of the asphalt structure presented in Chap. 2, Asphalt Binder. Through various analytical techniques including small angle X-ray scattering, SAXS; ion-exchange chromatography, IEC; atomic force microscopy, AFM; and confocal scanning electron microscopy, CLSM, a colloidal picture of asphalt microstructure is becoming more widely accepted. Lesueur (2009) states that asphaltene micelles could be viewed as a more open molecular assembly of low fractal dimension similar to polymeric coils, as shown in Fig. 2.8. Not being too far afield from the historical spherical description, this crude approximation has continued to be used by Lesueur and others as a pedagogical aid. A more precise view of the aggregation of asphaltenes can be gained from Fig. 5.1. Generally, it has been found that asphaltenes precipitate from the oily or saturate fraction without the resins component (Koots and Speight 1975).

Although the role of resins with regard to stabilization of asphaltenes remains a topic of some dispute, some level of molecular interaction permits the resin to enter the asphaltene micelles. This helps shield the differences in polarity between asphaltenes and maltenes, and therefore creates a surfactant-like behavior, Lesueur (2009). Atomic force microscopy (AFM) studies (Lober et al. 1996; Allen et al. 2012; and Jahanagir 2015) have identified "bee structures" with average heights of between 22 and 85 nm and typical distances between ridges of the bee-backs on the

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_5

Table 5.1 Asphalt cement and hot mix asphalt (HMA) binder additives, from Roberts et al. (1996)

Category	Type	General purpose or use	Generic examples
1	Extender	Substituted for a portion of asphalt cement (typically between 20 and 35% by weight of total asphalt binder) to decrease the amount of asphalt cement required	Sulfur Lignin
2	Filler	Fill voids and therefore reduce optimum asphalt content Meet aggregate gradation specifications Increase stability Improve the asphalt cement–aggregate bond	Mineral filler Crusher fines Lime Portland cement Fly ash Carbon black
3	Antistripping Agents	Minimize stripping of asphalt cement from aggregates	Amines Lime
4	Rubber	Increase HMA stiffness at high service temperatures Increase HMA elasticity at medium service temperatures to resist fatigue cracking Decrease HMA stiffness at low temperatures to resist thermal cracking	Natural latex Synthetic latex (e.g., polychloroprene latex) Block copolymer (e.g., styrene-butadiene-styrene (SBS)) Reclaimed rubber (e.g., crumb rubber from old tires)
	Plastic		Polyethylene/polypropylene Ethylene acrylate copolymer Ethylene vinyl acetate (EVA) Polyvinyl chloride (PVC) Ethylene propylene or EPDM Polyolefins
	Rubber–plastic combinations		Blends of rubber and plastic
5	Fiber	Improving tensile strength of HMA mixtures Improving cohesion of HMA mixtures Permit higher asphalt content without significant increase in draindown	*Natural*: Asbestos Rock wool *Manufactured*: Polypropylene Polyester Fiberglass Mineral Cellulose

(continued)

Table 5.1 (continued)

Category	Type	General purpose or use	Generic examples
	Oxidant	Increase HMA stiffness after the HMA is placed	Manganese salts
	Antioxidant	Increase the durability of HMA mixtures by retarding their oxidation	Lead compounds Carbon Calcium salts
	Hydrocarbon	Restore aged asphalt cements to current specifications Increase HMA stiffness in general	Recycling and rejuvenating oils Hard and natural asphalts
	Waste Materials	Replace aggregate or asphalt volume with a cheaper waste product	Roofing shingles Recycled tires Glass

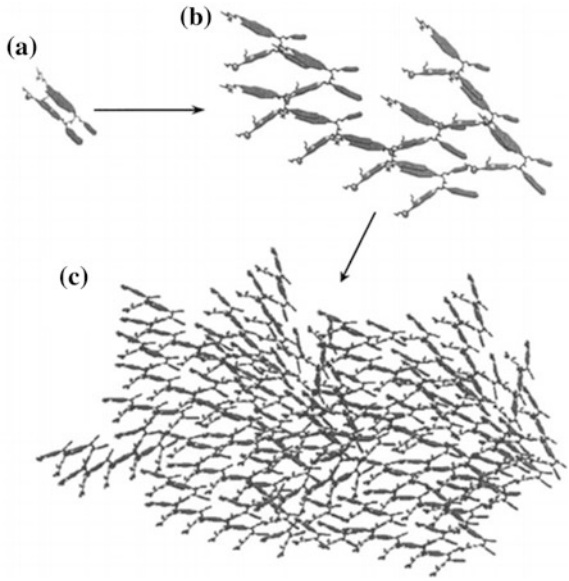


Fig. 5.1 Model of asphaltenes aggregation: **a** asphaltene molecule, **b** micelle with size ca. 3–4 nm, and **c** asphaltene aggregate with size ca. 12–15 nm. After Gawrys and Kilpatrick (2005)

order of 150 nm. A link between the "bee" phase and asphaltenes has been proposed in several studies (Mason et al. 2007; Allen et al. 2012; and Jahanagir et al. 2015), while other studies have identified "bee structures" as crystalline wax structures (Claudy et al. 1992 and Pauli et al. 2015).

Resins are known to stabilize asphaltenes by acting as a surfactant, creating a solvation layer to help keep asphaltenes in suspension. Storm et al. (1993, 1995) identified a solvation parameter, K, which quantifies the increase in volume fraction of the solid phase due to adsorbed resins. In this relationship, Eq. 5.1, x_{asp} is the mass fraction of asphaltenes; Kx_{asp} represents the effective volume fraction of solid phase, ϕ_{eff}. This is illustrated in Fig. 5.2.

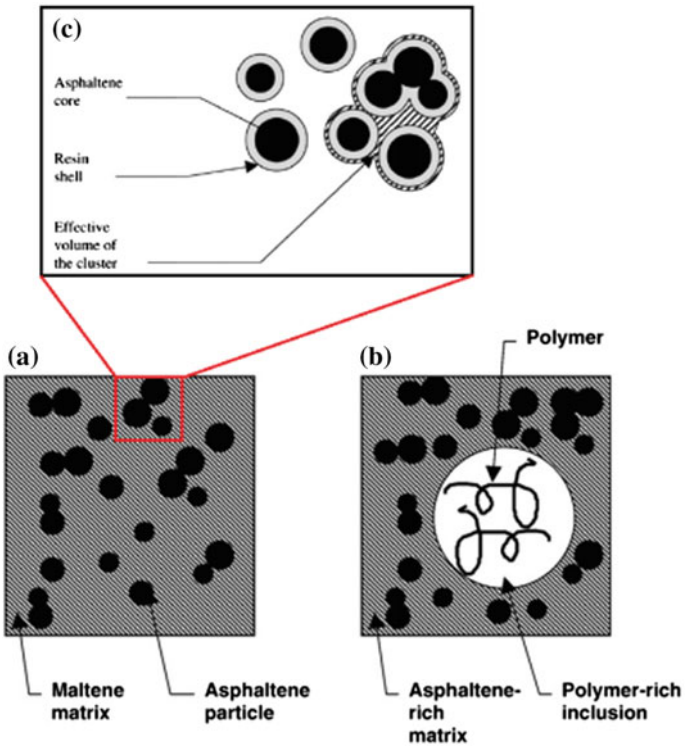


Fig. 5.2 A simplified view of colloidal structure of bitumen illustrating the effect of the solvation layer. After Lesueur (2009)

$$\phi_{eff} = Kx_{asp} \tag{5.1}$$

Lesueur (2013) explains that a solvation constant, $\frac{K}{\phi_m}$, where ϕ_m is the effective volume packing fraction, accounts for the solvent entrapped inside the asphaltene micelles. This proportionality constant ranges between 1.7 and 12.8 at 60 °C. Table 5.2 from Lesueur (2009) illustrates typical values of solvation constants for various bitumen sources. Higher values of this solvation constant represent a greater capacity to adsorb solvent among asphaltene molecules.

Linear alkanes present in asphalt can crystallize forming molecules typically between about 24 and 40 carbon atoms long (Gawel and Baginska 2004; Claudy et al. 1992). Waxes typically crystallize in bitumen at temperatures below about 90 °C. High-melting-point waxes are among the candidate additives with the capability of decreasing viscosity during mixing for warm mix asphalts. They are thought to act as a "plasticizer" at temperatures above the melting or dissolution temperature while acting as a "filler" at temperatures below dissolution temperature.

Table 5.2 Typical values of solvation constant, $\frac{K}{\phi_m}$, for various bitumen sources. After Lesueur (2009)

Bitumen Source	Temperature for $\frac{K}{\phi_m}$, °C	$\frac{K}{\phi_m}$	E_a^a, kj/mol	Temp. range, °C	Reference(s)
Ratawi	65	3.1	7.5	25–400	Storm et al. (1995)
	150	1.7		35–400	Storm et al. (1995)
Venezuela 1	60	5.1	4.0	35–400	Lesueur et al. (1997)
Saudi Arabia	60	5.6	2.7	65–135	Lesueur et al. (1997)
	60	5.6	4.0	65–135	Lesueur (1996)
France	60	6.5	3.0	65–135	Lesueur (1996)
Venezuela 2	60	7.8	3.0	65–135	Lesueur (1996)
Mexico	60	5.2	2.0	65–135	Lesueur (1996)
Kuwait	60	8.0	6.0	65–135	Lesueur (1996)
Black Sea	60	12.8	4.1	65–135	Lesueur (1996)

[a]E_a = activation energy

Aging has a notable impact on asphalt chemistry and microstructure. Aging leads to a decrease in aromatic content and increase in resin content with an associated increase in asphaltene content, while saturates remain essentially unchanged. Aging results in the formation of sulfoxides and ketones first that finally yield anhydrides and carboxylic acids (Petersen et al. 1993).

5.2 Principles of Modification

5.2.1 Acid Modification

Acid modification of asphalts is not a new process and according to Lesueur (2009) extends back for over 100 years where it was observed that bitumen reacted with sulfuric acid, fatty acids, nitric acid and acid sludges. The reaction result in rheological changes similar to the process of air blowing (Siegmann 1950a, b). Recent acid modification has focused on use of polyphosphoric acid (PPA) as it is economical and mimics the effects of air blowing in an easily controlled manner. Generally, adding as little as 1% PPA can increase the hardness of a bitumen by one paving grade classification (Orange et al. 2004). However, the reactivity of PPA is certainly bitumen-dependent (crude source-dependent) (Orange et al. 2004; Baumgardner et al. 2005). Orange et al. (2004) propose that PPA modification takes place via neutralization of polar interactions between stacked asphaltene molecules by either protonation of basic sites or esterification. The net effect, according to

Lesueur (2009), is an increase in the solvation of asphaltenes, which increases this fraction together with viscosity of the bitumen–PPA blend. Baumgarderner et al. (2005) suggest that other mechanisms occur as well, such as co-polymerization of saturates, alkyl aromatization of saturates, formation of ionic clusters, and cyclization of alkyl aromatics. Regardless, the end result is an increase in the effective volume of asphaltenes through a change in solvation content.

5.2.2 *Palierne Model*

Lesueur (2009) explains that the impact of the addition of polymers and fillers to bitumen rests with understanding the multi-phase structure these additives generate, which can best be understood by borrowing from the science of emulsion rheology. Lesueur further explains that suspensions can be viewed as special cases of emulsions and that in a general sense emulsions are two-phase systems where one phase is dispersed within the matrix of the other. If the matrix is liquid and the dispersed phase consists of nondeformable hard spheres, the system is referred to as a suspension. On the other hand, if the matrix is a liquid and the dispersed phase is another liquid, the system is referred to as emulsion in the classic sense. Lesueur (2009) states that the viscoelastic properties of emulsion can be modeled by the Palierne model in the general case of viscoelastic spherical inclusions in a viscoelastic matrix (Palierne 1991). The Palierne model derives from the Einstein model in which the inherent viscosity of a suspension is greater than that of the solvent, η_0, due to stress concentration effects in the vicinity of the spherical, nondeformable particles. As Lesueur explains, the total stress in suspension is the sum of the stress in the absence of the particle plus the excess stress associated with the presence of the particles. The equation that describes this relationship is:

$$\eta = \eta_0(1 + 2.5\phi) \tag{5.2}$$

where ϕ is the volume fraction of the dispersed phase and the stress concentrations are accounted for through the term 2.5 $\eta_0\phi$. The value 2.5 is referred to as the intrinsic viscosity of hard spheres and the relationship is only valid for a few percentages of the disperse phase, which is generally in the range of the amount of mineral filler or polymer added to bitumen to form a modified material or mastic fraction. Landau et al. (1958) described the impact of aspect ratio on the intrinsic viscosity, [η], of elliptical particles. They showed that elongated particles have a higher intrinsic viscosity and a more pronounced thickening effect than spheres. This will be discussed later in this chapter for what are referred to as "active" fillers.

Lesueur (2009) points out that in the case of emulsion of two liquids, dispersed particles still deform but in a viscous manner in which Taylor (1932) showed the induced flow to be a function of the viscosity ratio between the two liquids. When

the dispersed liquid viscosity is much higher than the matrix the Einstein relation used, [η] = 2.5. On the other hand, if the matrix is much more viscous, the intrinsic viscosity falls down to one. Palierne (1991) also proposed a more general model that accounts for interfacial deformation of droplets or a liquid suspension and delayed elastic deformation of the droplets.

5.2.3 Suspension Limit

As previously defined mastics comprise bitumen and suspended filler, normally mineral filler, smaller than 0.075 mm in the USA, and smaller than 0.064 mm in Europe. The typical range of mineral filler is between about 0.010 and 0.050 mm and the typical content is between about 2 and 12% by weight of bitumen. Therefore, mastics are suspensions as defined by Mitchell and Lee (1939) and later Rigden (1947). Heukelom and Wijga (1971) derived the relationship between viscosity and the filler volume fraction ϕ:

$$\eta = \eta_0 \left(1 - \frac{\phi}{\phi_m}\right)^{-2} \tag{5.3}$$

where η is viscosity of the mastic and η_0 is the viscosity of the neat bitumen and ϕ_m is the maximum packing factor. Lesueur (2009) points out that mastics are suspensions because the asphaltenes do not compete with mineral fillers as they occur at the nanometer scale while mineral fillers occur at the millimeter scale. Furthermore, because mineral fillers are introduced at low enough volumes to be considered suspensions, the intrinsic viscosity [η] can be predicted by the Einstein relationship:

$$[\eta] = \frac{2}{\phi_m} \tag{5.4}$$

Table 5.3 presents values of intrinsic viscosity for various fillers. This table is based on empirical data and is reproduced from Lesueur (2009). Of particular interest here is that most mineral fillers have values of intrinsic viscosity in the range of 2.4–4.9 regardless of the base bitumen, but certain fillers, e.g., lime and fly ash, provide a substantially broader range of intrinsic viscosities and considerably higher median viscosities. Also certain fillers, such as kaolin, have higher intrinsic viscosity due to its high aspect ratio and based on its unique, colloidal morphology.

Table 5.3 Values for the intrinsic viscosity [η] of different fillers in bitumen (after Lesueur 2009)

Filler type	[η][a]	Reference(s)
Limestone	3.8	Heukelom and Wijga (1971)
Limestone	2.6–3.9 (25 °C)	Shashidhar et al. (1999)
	3.0–3.7 (70 °C)	Lesueur et al. (1995)
	2.5 (65 °C)	
	2.4 (135 °C)	
Dolomitic limestone	4.9 (25 °C)	Shashidhar et al. (1999)
	4.4 (70 °C)	
Hydrated Lime	3.2–10	Lesueur and Little (1999)
Lime	7	Tunnicliff (1967)
Sandstone	2.8 (25 °C)	Shashidhar et al. (1999)
	4.0 (70 °C)	Lesueur et al. (1995)
Siliceous filler	2.4 (65 °C)	
	2.4 (135 °C)	
Granite	2.7–4.2 (25°)	Shashidhar et al. (1999)
	3.5–4.1 (70 °C)	
Fly ash	10.2 (25 °C)	Shashidhar et al. (1999)
	14.1 (70 °C)	
Slate dust	4.2	Heukelom and Wijga (1971)
Ball clay	3.2	Lesueur et al. (1995)
Kaolin	3.7	Heukelom and Wijga (1971)
Carbon black	2.6 (65 °C)	Lesueur et al. (1995)
	3.9 (135 °C)	
Asbestos	16.5	Tunnicliff (1971)
Polyester fibers	26–34	Chin et al. (2005)
Mineral fibers	26	

[a]Note that according to Lesueur (2005) some authors did not explicitly give the value of [η] and it was then calculated from the published data. The range found for different filler/bitumen combinations is given when appropriate, corresponding to different binder or filler origin, or measuring temperature. Since [η] defines the effect of the solute in the suspension, it is a dimensionless parameter

5.3 Application of Modification to Bitumen

5.3.1 Compatibility

The addition rate for polymers in bitumen is typically between about 2 and 12%. However, the polymer-rich phase of the modified bitumen occupies a considerably larger volume (between about 4 and 10 times the volume of the added polymer) due to a process referred to as swelling in which light aromatic fractions of the bitumen occupy space within the polymer matrix. Swelling is greatest in styrene-butadiene (SB) polymers and ethylene vinyl acetate (EVA) polymers and considerably less in

polyolefins. Even for compatible systems, according the Lesueur (2005), the equilibrium situation is a macroscopic phase separation of two phases. Lesueur (2005) refers to one phase as the asphaltene-rich phase (ARP) and the other as the polymer-rich phase (PRP). The ARP is the denser phase and the PRP occurs at a rate controlled by the Stokes sedimentation rate (Tadros 1983). The larger the density difference and the larger the particle size, the faster the solvation rate will be (Tadros 1983). Phase separation of the two phases can be prevented or controlled within reasonable limits by "lightly" cross-linking the PRP to prevent coalescence of this phase.

In summary, bitumens with a high asphaltene content generally have lower compatibility and for acceptable compatibility the aromaticity of the maltenes should fall between certain limits (Lesueur 2005; Laval and Quivoron 1973; and Laval and Brule 1974).

5.3.2 *Structure of Polymer-Modified Bitumen*

The phase separation that occurs because polymers and asphaltenes do not mix partitions part of the polymer swollen by the aromatic compounds and leaves the other part comprising asphaltenes remaining in the rest of the maltene fraction. According to Lesueur (2005), the low polymer contents found in paving asphalts create an emulsion with a PRP dispersed within the ARP matrix. For additive contents greater than about 6% by weight (depending on the polymer and bitumen characteristics), phase inversion occurs and the inverse morphology is seen with asphaltene-rich domains dispersed within a polymer-rich matrix (Lesueur 2005; Brion and Brule 1986; and Brule 1996). As a consequence, large quantities of aromatics are needed to swell the polymer and the matrix can become depleted in maltenes and rich in asphaltenes. This is referred to as physical distillation (Lesueur et al. 1998).

5.3.3 *Practical Consequences*

Lesueur (2005) summarizes that the Palierne model applies to polymer-modified bitumens and has two major implications. First, the volume fraction of the polymer-rich inclusions is a fundamental parameter and is governed by physical distillation, which is defined by properties of the polymer-rich phase and the ability to enhance the contribution of the polymer-rich particles. Second, the exact morphology of the polymer-modified bitumen is highly temperature dependent and will not affect the modulus of the system in the temperature range to which paving mixtures are subjected.

5.4 Extenders

5.4.1 Sulfur

Between 1975 and 1984 the Federal Highway Administration, FHWA, invested in demonstration projects in 18 states (about 26 projects) to evaluate the utility of sulfur in asphalt as both an extender and as an additive. The driving force during this period was the fear of disruption in crude oil supply and as a result a shortage of asphalt. Asphalt mixtures containing sulfur are referred to as sulfur extended asphalt (SEA). In this process elemental sulfur is blended with bitumen in weight ratios ranging from 20/80 (sulfur/bitumen) to 40/60 (sulfur/bitumen). When added at a 20/80 ratio, the sulfur reacts and is virtually totally dissolved within the bitumen while that in excess of 20% ultimately solidifies and acts as a filler and/or stabilizer in the mixture. Figure 5.3 illustrates the sulfur addition process starting with the elemental sulfur added as a solid in which S_8 rings, which are bonded with one another covalently, are added to the bitumen. When sulfur is heated above about 113 °C, it melts into a low viscosity liquid and is easily mixed with bitumen. At the melting point, S_8 rings separate from the solid mass and react at least to some extent with components in the bitumen to form polysulfides (Fig. 5.4). Simultaneously the melting sulfur that does not react with bitumen solidifies into a considerably more viscous liquid. Ultimately, the quantity of sulfur not soluble in bitumen takes again the crystalline form and acts as a filler within the voids of the aggregate matrix. This filler effect stiffens and can substantially lower the plastic deformation potential of the binder (bitumen plus sulfur). Consequently, SEA is popular in high-temperature environments where rutting and shoving are critical distress mechanisms.

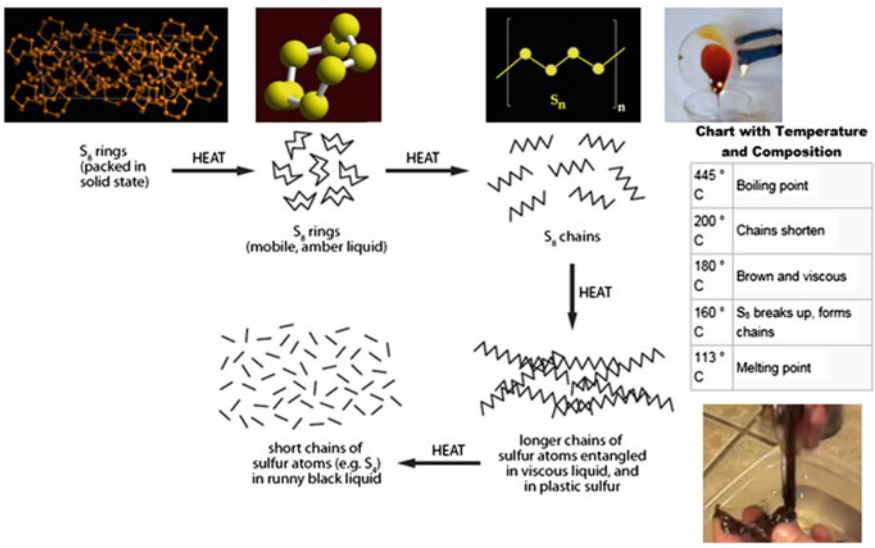


Fig. 5.3 Changes in sulfur morphology and rheology as it is incorporated into bitumen

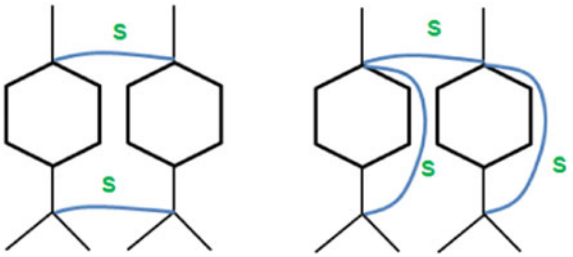


Fig. 5.4 Polysulfides of sulfur and bitumen. Sulfur chain linkages reacting with cyclic moieties of the asphalt molecular makeup

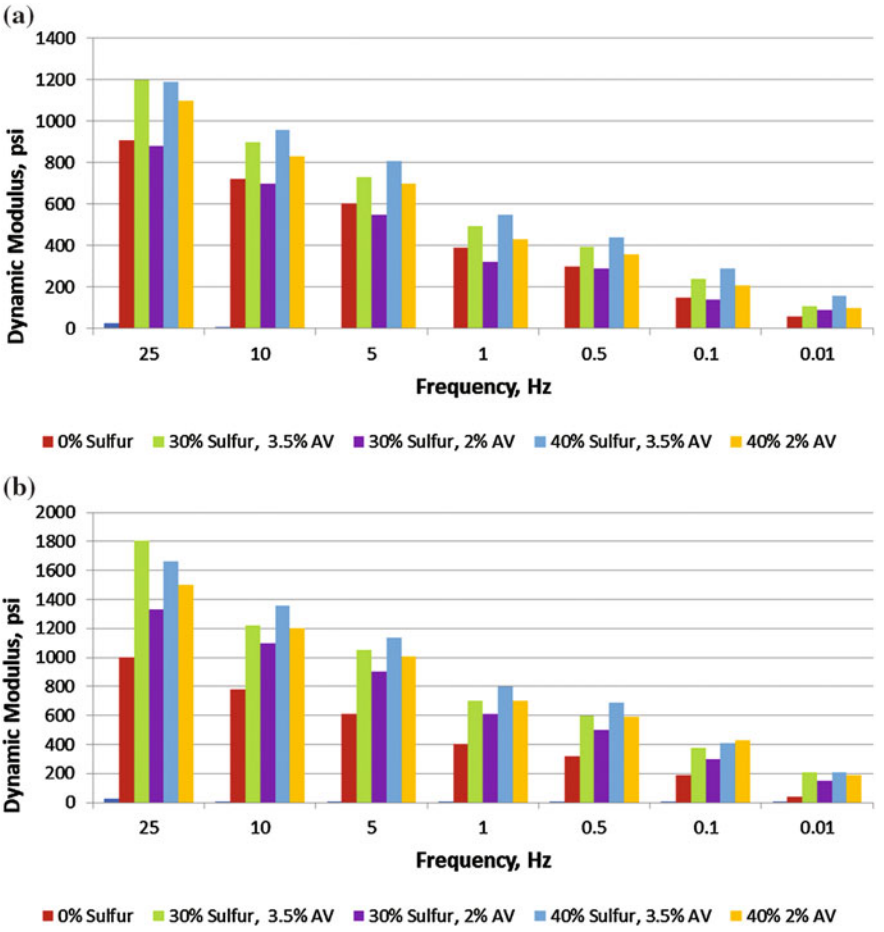


Fig. 5.5 Dynamic moduli at 21 °C aged for one day and 14 days and compared to a control asphalt mixture bound with PG 67-22 binder with Thiopave sulfur extended asphalt (SEA) mixtures of the same asphalt with 30 and 40% of the binder replaced with elemental sulfur following the Thiopave process, after Timm et al. (2012)

The mechanical impact of the sulfur addition is demonstrated in Figs. 5.5 through 5.7. In Fig. 5.5, the impact of sulfur addition at the rates of 30 and 40% by weight of binder and at two air void contents (2.0 and 3.5% for the SEA compared

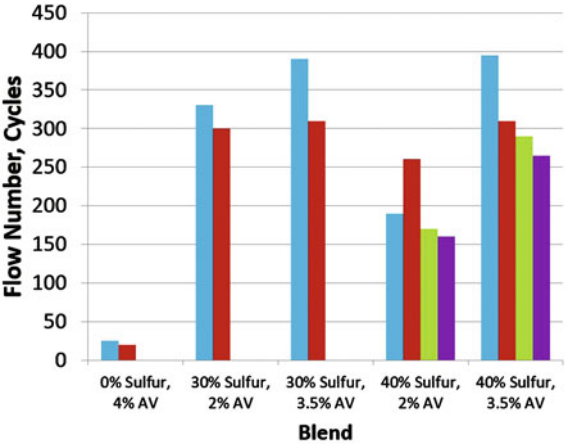


Fig. 5.6 Flow number results comparing a control asphalt mixture bound with PG 67-22 binder with Thiopave sulfur extended asphalt (SEA) mixtures of the same asphalt with 30 and 40% of the binder replaced with elemental sulfur following the Thiopave process, after Timm et al. (2012)

to 4.0% for the unmodified bitumen) is demonstrated. In this figure it is clear that sulfur addition increases stiffness and that the level of change increases with time during which sulfur crystallization takes place.

In Fig. 5.6 the impact of sulfur content on the flow number is seen. Here flow is defined as the number of cycles at which the rate of plastic deformation markedly increases. One would expect this type of effect as the rheology of the base bitumen is altered by the formation of polysulfides and by the filler effect of the nonreactive sulfur, above about 20% by weight of the bitumen. Finally, Fig. 5.7 demonstrates the fatigue response of a control asphalt mixture compacted to 4.0% air voids and bound with a typical bitumen, performance grade (PG) 67-22, compared to Thiopave sulfur asphalt mixtures at air void contents of 3.5 and 2.0% and at sulfur contents of 30 and 40% by weight of bitumen.

The National Center for Asphalt Technology (NCAT) study form which Figs. 5.5 through 5.7 were constructed, used the Thiopave sulfur extended asphalt protocol for adding sulfur to modify bitumen and demonstrated the effect of sulfur

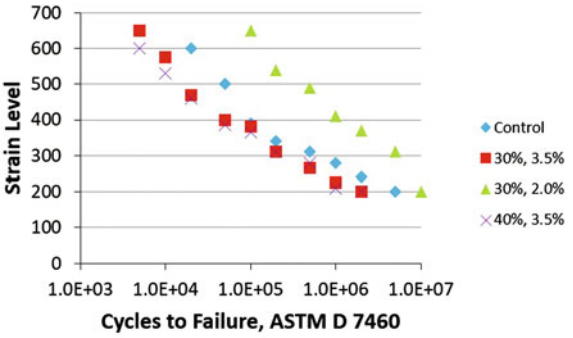


Fig. 5.7 Flexural fatigue results comparing a control asphalt mixture bound with PG 67-22 binder with Thiopave sulfur extended asphalt (SEA) mixtures of the same asphalt with 30 and 40% of the binder replaced with elemental sulfur following the Thiopave process, after Timm et al. (2012)

which is to alter the properties of the bitumen via modification of the properties of the bitumen as well as through the filler effect of the solidified sulfur that does not combine or dissolve within the bitumen. The net result is SEA (with 30% sulfur by weight or less) with relatively similar fatigue properties to bitumen without sulfur but with greater resistance to plastic deformation. The NCAT study concluded that:

1. Dynamic moduli (E*) increased for all combinations of frequency and temperature due to sulfur addition.
2. SEA mixtures (Thiopave) showed superior rutting resistance.
3. The control mixture demonstrated superior fatigue resistance at relatively high strains (i.e., greater than about 400 με) but nearly the same fatigue life at strains of about 200 με or less, the traditional level for design for pavement expected to carry traffic at approximately 10 million equivalent single axle loads.

Environmental concerns due to H_2S and/or SO_2 emissions were not considered by Timm et al. (2012) to be a significant factor as long as plant production and construction temperatures are maintained at below about 149 °C, which is within the practical working limits for warm mix asphalt production and construction.

5.5 Additives that Promote Improved Bond Between Aggregate and Binder

As discussed in Chap. 4, carboxylic acids readily bond with siliceous aggregate surfaces, but they are also easily displaced by water. This has been established experimentally by a number of researchers, i.e., Simmons and Beard (1987) and Ardebrant and Pugh (1991a, b) who show how carboxylic acids and a tertiary amine interact with metal hydroxyl groups. These interactions between the mineral functional group and the organic functional group can range from completely covalent to completely ionic. Bond energy increases as the equilibrium shifts to the right and the bond becomes more ionic. Maximum strength occurs when the surface hydroxyl group acts as a strong base and the organic functionality acts as a strong acid (Hefer et al. 2005). The interaction between a silanol group and a carboxylic acid is depicted in the reaction shown below (Hefer et al. 2005).

$$-SiOH + RC(=O)-OH \rightleftharpoons -SiO\cdots HO-C(=O)R \rightleftharpoons -SiOH_2^{+}\,{}^{-}O-C(=O)R$$

where "•••" denotes hydrogen bonding. Carboxylic acids are weak acids as they are not easily dissociated to form ionic species, and the right side of the reaction is less likely to occur. Hydrogen bonds, however, are susceptible to displacement or interruption by water according to the following reaction (Hefer et al. 2005):

$$-SiO\cdots HO-C(=O)R + 2H_2O \rightleftharpoons -SiO\cdots HOH + H_2O + RC(=O)-OH$$

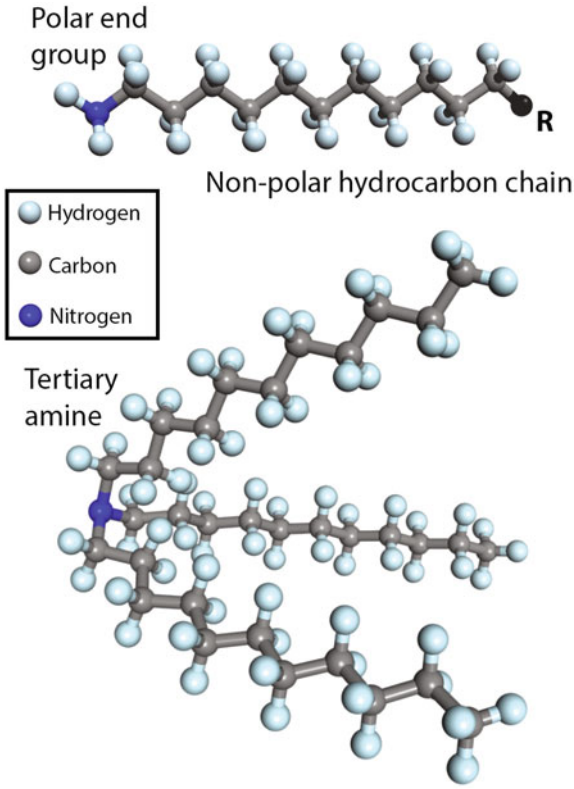


Fig. 5.8 Schematic illustration of typical amine groups. After Hefer et al. (2005)

In this hydrolysis reaction, water, acting as either a base or an acid, forms a stronger hydrogen bond with the silanol group than the hydrogen bond between the functional group and the silanol group.

Basic, amine anti-strip agents are generally used to improve adhesion of bitumen with siliceous aggregates. Logaraj (2002) describes the two main characteristics of anti-strip additives. They have a polar amine end group which will chemically bond with the siliceous aggregate surface, and they have a hydrocarbon chain with similar properties to that of the bitumen so that they interact and become part of the bitumen. Hefer et al. (2005) illustrated a primary amine (RNH_2) and tertiary amine (NR_3) as schematically illustrated in Fig. 5.8. The number of amine groups and length of the chains can greatly influence the adhesion of the bitumen, while optimum performance is typically achieved with 14–18 carbon chain amines with one or two amine groups (Tarrar and Wagh 1992).

Ardebrant and Pugh (1991a, b) presented the interaction between a tertiary amine and a metal hydroxyl group. The following interaction shows the interaction of these basic molecules with surface silanol groups,

$$-SiOH + R_3N \rightleftharpoons -SiOH \cdots NR_3 \rightleftharpoons -SiO^{-}{}^{+}HNR_3$$

In this interaction, the lone pair electrons on the nitrogen of the amine possess a partial negative dipole and exert an attraction to the partial positive hydrogen dipole on the silanol group (Hefer et al. 2005). The strength of the basic amine is determined by the type and length of the R groups. For strong bases, protonation of the amine occurs leaving the ionic species as shown on the right-hand side of the reaction. As described by Hefer et al. (2005), Titova et al. (1987) showed that a tertiary amine with three ethyl (CH_2CH_3) groups, triethylamine, interacts with a silanol group to form an insoluble silica triethylamine compound that is stable at temperatures greater than 484 °F, making the hydrogen bonding less likely to occur with this reaction.

Hefer et al. (2005) explain that species of lower basicity in bitumen might form hydrogen bonds, which would then continue to be susceptible to water disruption or displacement. In general, the hydrolysis reaction for organic bases that form hydrogen bonds with silanol groups is shown in the following equation, where X denotes some O or N-containing basic functional groups. Therefore, if water acts as a base or an acid and forms a stronger hydrogen bond with the surface hydroxyl or with the basic functional group than the existing hydrogen bond, displacement of this functionality occurs.

$$-SiOH \cdots XR + 2H_2O \rightleftharpoons -SiO \cdots OH_2 + HOH \cdots XR$$

The pH of water at the interface also affects the hydrogen bonds formed between metal hydroxyl groups on aggregate surfaces and basic or acidic functionalities in the bitumen. Two general cases are discussed by Hefer et al. (2005). The pH range over which these bonds would be expected to retain adhesion is a function of the acidity of the surface hydroxyl groups and the functional groups in the adhesive bitumen (Simmons and Beard 1987; Hefer et al. 2005).

Hefer et al. state that the generation of covalent bonds at interfaces has become an industry of its own. Coupling agents are chemicals used to "couple" an organic and inorganic phase through covalent bonding, producing durable adhesive bonds in adverse environmental conditions. Silane coupling agents, or organo-functional silanes, form the largest and most successful group of these materials (Saville and Axon 1937; Walker 1994). The silane molecule is generally composed of a silicon (Si) atom at the center, with an organic chemical functional group (R) compatible with the organic matrix, and (R') which is a hydrolysable group. The coupling mechanism consists of three steps: hydrolysis, coupling, and condensation. Hydrolysis takes place when the organo-functional silane is brought into contact with moisture to generate active silanol groups (SiOH), and usually an alcohol,

$$RSi(OR')_3 + H_2O \rightarrow RSi(OH)_3 + 3R'OH$$
Silane Trisilanol Alcohol

Silanol molecules then form a hydrogen bond with the hydroxylated metal oxide surface (MOH). In addition, similar reactions occur between silanol groups themselves (not shown), to form hydrogen bonds.

$$-MOH + RSi(OH)_3 \rightarrow R-Si(OH)_2-O-H\cdots O-M-$$

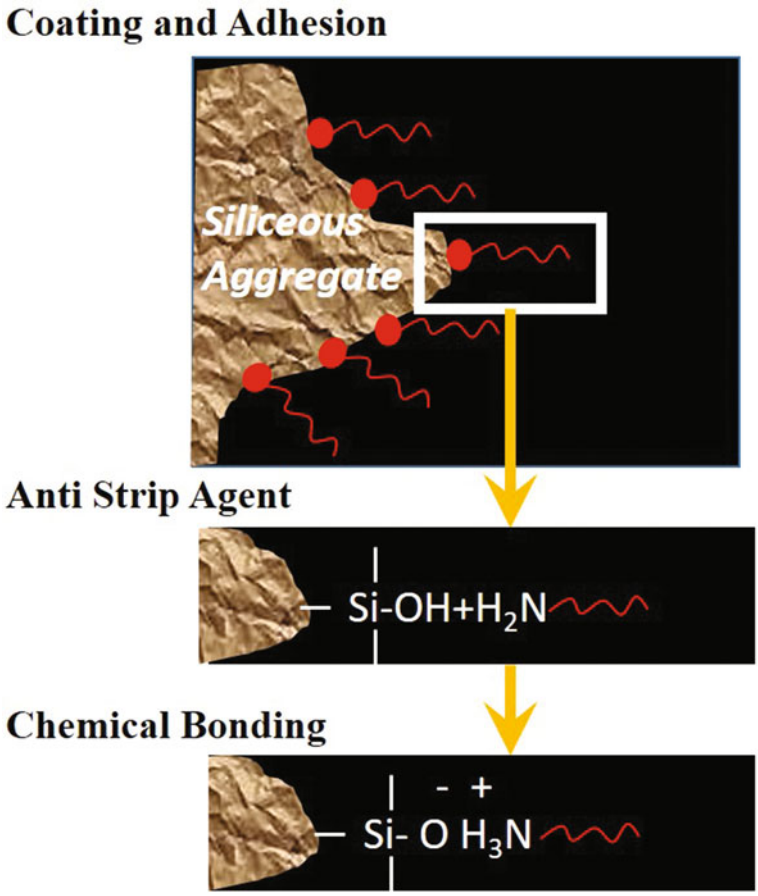
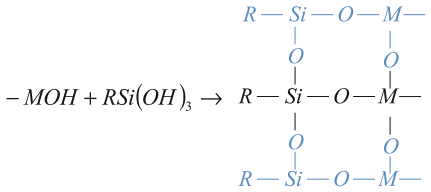


Fig. 5.9 Water-resistant chemical bond of amine to siliceous aggregate, after Logaraj (2002)

Upon heating or drying, these hydrogen bonds are condensed (or collapse) to produce H_2O and a covalently bonded metallosiloxane (MOSi) and a cross-linked siloxane (SiOSi) film structure over the surface as illustrated below (Fig. 5.9).



5.6 Fillers

5.6.1 *Active Filler: Hydrated Lime*

Hydrated lime (HL) is produced by hydrating calcium oxide (CaO) through a carefully controlled hydration process. The quicklime is produced by calcination of high-purity, high-calcium limestone ($CaCO_3$) or dolomitic limestone ($CaCO_3 \cdot MgCO_3$). When hydrated the dolomitic limestone forms $Ca(OH)_2 \cdot Mg(OH)_2$. Care must be taken when hydrating dolomitic lime as the hydration kinetics of dolomitic lime is slower than that of high-calcium limestone. Unhydrated MgO in asphalt can slowly hydrate in the pavement layer and disrupt the pavement surface as it increases in volume during hydration.

Hydrated lime may be introduced into the asphalt mixture in a variety of ways. Probably the most common method is to blend hydrated lime into a pug mill with the aggregate to be treated prior to introducing the aggregate into the drum dryer of either a drum or batch plant. Sufficient moisture is added to allow interactions between the aggregate and hydrated lime to take place (Little and Epps 2001; Sebaaly et al. 2003). Lime can also be added just after the hot asphalt is added downstream of aggregate drying in the drum of the drum plant or it can be applied in a slurry form to stockpiled aggregate and allowed to marinate over time to help insure a more complete reaction between the lime slurry and the aggregate surface.

Over about the past 40 years (Lesueur et al. 2013; Hicks 1991) hydrated lime has been used primarily as an additive to reduce sensitivity to moisture damage and has been used in approximately 10% of asphalt mixtures produced in the USA (Hicks and Scholtz 2003). Lime has been claimed to provide multi-functional effects and is seen as an "active" as opposed to an "inert" filler (Lesueur 2010; Little and Epps 2001; Little and Petersen 2005; Sebaaly et al. 2003). Field experience for the use of HL in North America among state transportation agencies (Lesueur et al. 2012) estimates a typical additive rate for HL at 1–1.5% based on the dry weight of the aggregate. Hicks and Scholz (2003) estimated that use of hydrated lime increases the pavement life by 20–50%. Although the use of HL in asphalt mixtures

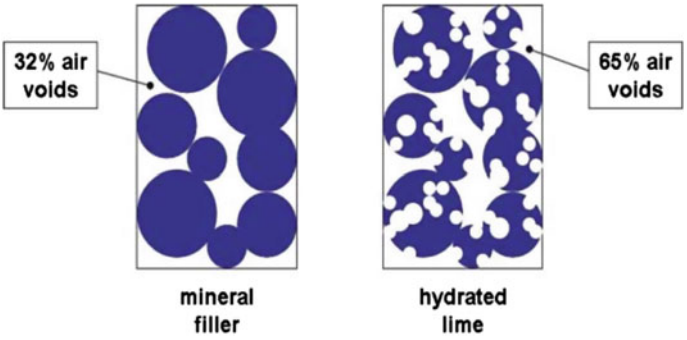


Fig. 5.10 Difference in the dry porosity of HL (*right*) and that of a mineral filler (*left*). The dry porosity of HL is higher than that of the mineral filler because the porosity inside the particles, which is negligible in the mineral filler, sums up to the porosity between the particles. After Lesueur et al. (2009)

is not as well developed in Europe, its beneficial effects have been reported (Lesueur 2010) and life extension estimates between 20 and 25% have been reported (Raynaud 2009; CROW 2007; Voskuilen and Verhoef 2003).

The impact of HL on asphalt mixtures may be addressed under several categories: (a) as a filler, (b) as a modifier of the aggregate surface, (c) as a physical filler, and (d) as a filler with chemical interactions with the asphalt binder. Lesueur (2013) states that HL has more voids in a dry, compacted state (Rigden air voids) than other mineral fillers with 65% being typical for lime as compared to 30–34% for other mineral fillers. This difference is partially due to the voids among the particles. The other porosity comes from porosity within the particles leading to an even higher value of overall porosity (Fig. 5.10).

Lesueur (2013) explains that it is well known in suspension rheology that the increase in viscosity of a liquid upon the addition of solid particles is mainly governed by the volume fraction of the solid particles (Coussot 2005; Lesueur 2009). The higher the volume fraction of solid particles the higher the viscosity of the mastic becomes. When particle movement is blocked because of contact among particles, the maximum packing fraction of the particles is reached which, at least partially, explains why dry porosity of a filler relates to the stiffening impact of that filler in the mastic.

A wealth of data supports the curious impact of HL as a filler over the range of temperatures normally encountered in the pavement. The stiffening effect of HL is prominent at high temperature due in part to the physical filler effect but diminishes below about 25 °C (Wortelboer et al. 1996; Hopman et al. 1999; Khattak and Kyatham 2008; Lesueur and Little 1999; Pilat et al. 2000; Seebaly et al. 2005; Johansson and Isacsson 1998; and Lackner et al. 2005). See Fig. 5.11. Although no definitive, mechanistic explanation has been articulated at this time, Lesueur (2013)

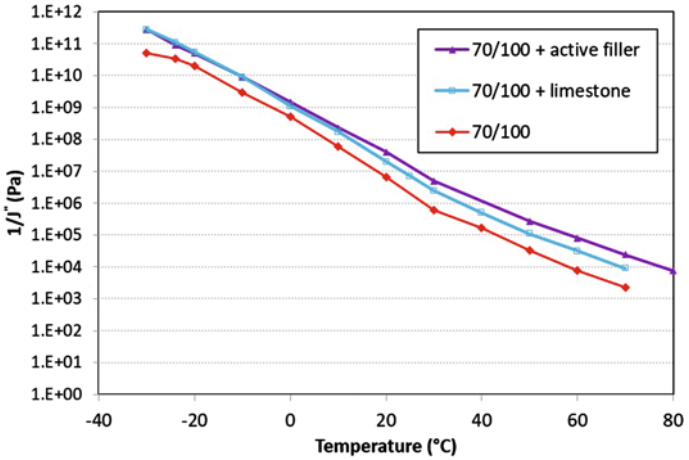


Fig. 5.11 Comparison of HL and limestone filler based on imaginary compliance at 10 radians per second versus temperature. Base asphalt is 70/100 penetration; filler content was 50% by weight and active filler contained 40% replacement of limestone with HL. After Wortelboer et al. (1996)

proposes a possible explanation. He suggests that at high temperatures the internal voids of the HL particles are filled with bitumen, increasing the apparent volume fraction of the solids, or as Lesueur defines them, bitumen-filled hydrated lime particles (BFHLs). However, below about 25 °C these BFHLs are deformable, and the mechanical contrast diminishes between the matrix and the inclusion.

5.6.2 Hydrated Lime: Aggregate Surface Modifier

For reasons previously discussed, siliceous aggregates have been challenging as it is difficult to achieve an effective, durable bond between the aggregate and bitumen (Curtis et al. 1993). Curtis et al. (1993) explain that the reasons for this are that both anionic and cationic surfactants typically present in the bitumen strongly bond with calcium cations, but only cationic surfactants bond strongly with silica. This means that anionic surfactants are easily displaced by water.

Lesueur (2013) explains that like all compounds, HL has a solubility level primarily dependent on its composition but also on certain external variables. Hydrated lime is sufficiently soluble to allow precipitation of calcium ions on the aggregate surface after which the accumulation of cations on the aggregate surface can, combined with acids from the bitumen, form a water-insoluble salt. This results in favorable moisture damage resistance in lime-treated siliceous sands and gravels. This is one reason that HL has been effective as a surface treatment of aggregate either by margination and/or pug mill mixing with damp aggregate (Blazek et al. 2000). In addition, Ramond and Lesueur (2004) report that surface treatment of

aggregate with HL can result in precipitation of calcium carbonate as HL interacts with carbon dioxide from the atmosphere. The result is a high roughness, and therefore higher surface area, surface that favors a more secure bond with the bitumen based on the greater bonding area. Schiffner (2003) reported difficulty in extracting bitumen for mixtures due to the tenacity of the bitumen–aggregate bond. He reported only being able to recover about 90% of the bitumen from a basalt aggregate mixture.

Lesueur (2013) points out that one would not necessarily expect HL to modify limestone aggregate surfaces in the same way it does for siliceous aggregate surfaces, but HL has been shown to be quite effective in improving bond strength between bitumen and the aggregate surface (Huang et al. 2005; Mohammad et al. 2008).

In California, Colorado, Nevada, and other western U.S. states HL has been used to pretreat or treat the aggregate surface when the aggregate is plagued with clayey coatings. The ability of HL to stabilize and change the properties of clay is well established (Little 1995). HL reacts with clay by saturating the clay lattice with calcium ions and thereby reducing the ability of the clay to sorb or hold water. In addition, the high pH produced by the presence of HL produces a chemical reaction between the clay and HL to stabilize the clay into a material with surface properties that are more amenable to a durable bond and with volumetric stability and essentially preventing the clay from developing an sorbed barrier of water and hydrated cations (Eulitz et al. 1998; Schellenberg and Eulitz 1999).

5.6.3 *Rheology of Filler Stiffening Effect*

Lesueur and Little (1999) evaluated the stiffening effect of fillers using the Marion-Pierce approach as quantified by Heukelom and Wija (1971):

$$\eta_0 = \eta_{0,m}\left(1 - \frac{\phi}{\phi_m}\right)^{-2} \tag{5.5}$$

where $\eta_{0,m}$ = unmodified asphalt viscosity; η_0 = viscosity of the filled asphalt; ϕ_m = maximum packing factor; and ϕ = volume fraction of filler added to the bitumen.

They used the maximum packing fraction, ϕ_m, which is related to the Rigden air voids measured on the dry filler, to quantify the filler effect. We emphasize here that ϕ_m is related to Rigden air voids and not equal to it. The ϕ_m value is not determined on a dry material, but instead is determined for a filler that is dispersed in and affected by the medium within which it is dispersed, and is potentially affected by intervening adsorbed layers. According to Lesueur and Little (1999), most mineral fillers have a ϕ_m value of about 63%. However, this value can be complicated by the presence of an adsorbed layer around the filler particles. If a layer–building interaction does exist between the particles and the bitumen, then the value of ϕ_m

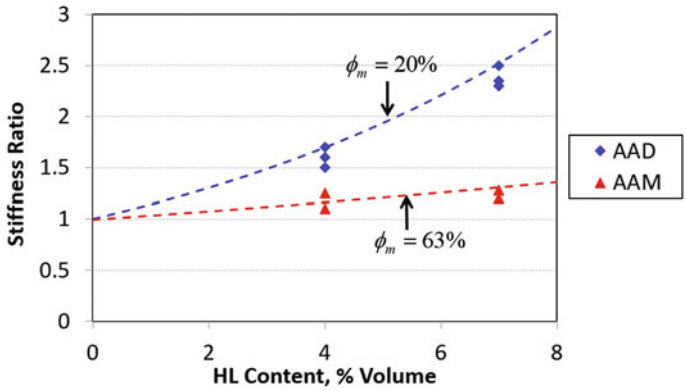


Fig. 5.12 Stiffening ratio: defined as zero shear viscosity (ZSV) of blend divided by base bitumen as a function of volume fraction. After Little and Petersen (2005)

will be lower than the typical value of 63% due to the interactive effects. Lesueur and Little (1999) measured the effect of adding 4 and 7.5% by volume of limestone (LS) and HL fillers to Strategic Highway Research Program (SHRP) bitumens AAD-1 and AAM-1 at high temperatures (between 50 and 100 °C) and in the low-temperature range (between 16 and −10 °C). Asphalts AAD-1 and AAM-1 are substantially different in chemical composition. Asphalt AAD-1 is a bitumen with high asphaltene and polar chemical functional group content; AAM-1 has very low asphaltene content and is less polar. The fillers substantially stiffened the bitumens at temperatures between 50 and 100 °C; however, the ϕ_m for HL used as a filler in bitumen AAD-1 was only 20% whereas the ϕ_m value for limestone (LS) used as a filler in AAD-1 was approximately 63%. In the case of bitumen AAM-1, both the HL and LS fillers responded with a ϕ_m value of approximately 63%, which is typical for traditional fillers. Thus, only AAD-1 with HL filler showed the effect of adsorbed layer buildup. This effect is shown graphically in Fig. 5.12 for the HL in bitumens AAD-1 and AAM-1. The basis for the stiffness comparisons in Fig. 5.12 is Newtonian viscosity, which was derived for each material from master curves developed by applying the time–temperature superposition principle using a set frequency range and temperatures ranging from 50 to 100 °C.

Rationale for the differences in the physicochemical response of bitumens AAD-1 and AAM-1 with HL can be found in their chemical compositional differences. Petersen et al. (1987a, b, c, d) showed that bitumen components containing strongly interacting chemical functional groups are irreversibly adsorbed on the surface of HL. AAD-1 is rich in asphaltenes and polar components that adsorb on the surface of the HL particle, thus increasing its "effective" size. On the other hand, AAM-1 is relatively deficient in both asphaltenes and polar chemical functionality. Limestone does not have the high concentration density of reactive bonding sites for asphalt components on its surface that HL does, and thus, it does not show the effects characteristic of HL.

Better estimates of viscosity versus volume fraction can be found using the Marion-Pierce equation rather than ϕ_m. Lesueur and Little (1999) used it merely to compare effects of fillers and not as a definitive model. In an attempt to evaluate the effects of filler on mastic properties of bitumens AAD-1 and AAM-1, Kim et al. (2003) employed a rheology-based composite model. The model is a modified version of the Kerner equation by Lewis and Nielsen (1968) and Nielsen (1970). Shashidhar et al. (1999) studied various mastic systems by introducing Nielsen's model and found that the role of fillers can be described based on their physical properties. The Nielsen model is expressed as:

$$\frac{G_c}{G_m} = \frac{1 + ABV_p}{1 - B\psi V_p} \tag{5.6}$$

where A = K_E – 1; and

$$B = \frac{G_p/G_m - 1}{G_p/G_m + A} \tag{5.7}$$

$$\psi = 1 + \frac{1 - \phi}{\phi^2} V_p \tag{5.8}$$

G_c = effective shear modulus of composite such as mastic; G_m = shear modulus of matrix such as asphalt binder; G_p = shear modulus of particle such as filler; V_p = volume fraction of filler; K_E = generalized Einstein coefficient; and ϕ = maximum volumetric packing fraction.

The constant B becomes 1.0 for a very large G_p/G_m ratio, i.e., when filler particles are much stiffer than the matrix, such as in asphalt mastics. The parameters KE and ϕ are determined by nonlinear regression fitting techniques from Eq. 5.6. In this model, the physical meaning of the maximum volumetric packing fraction is the maximum amount of filler that can be added to the matrix without the emergence of air voids. It typically denotes the mechanical volume-filling contribution to mastic stiffening. The generalized Einstein coefficient is the stiffening rate of the mastic as a function of filler addition. It is an indicator of the physicochemical contribution to stiffening. A more detailed discussion can be found in Shashidhar et al. (1999) and Shashidhar and Romero (1998).

Kim et al. (2003) used the dynamic shear rheometer (DSR) to compare the stiffening effect of LS and HL filler when added to bitumens AAD-1 and AAM-1. They used these data to calculate model parameters KE and ϕ. Figure 5.13 shows the comparison of the relative stiffening effect of LS and HL on bitumens AAD-1 and AAM-1. It is obvious that HL has a higher stiffening effect in both bitumens than LS, but that the effect of HL is much more pronounced in AAD-1 than in AAM-1, supporting the previous work of Lesueur and Little (1999). In fact, the three curves in Fig. 5.14 are Nielsen model fits to the data for AAD-1 with LS, AAD-1 with HL, and AAM-1 with LS. The model parameters KE and ϕ determined from this regression analysis reflect a considerably higher KE for AAD-1

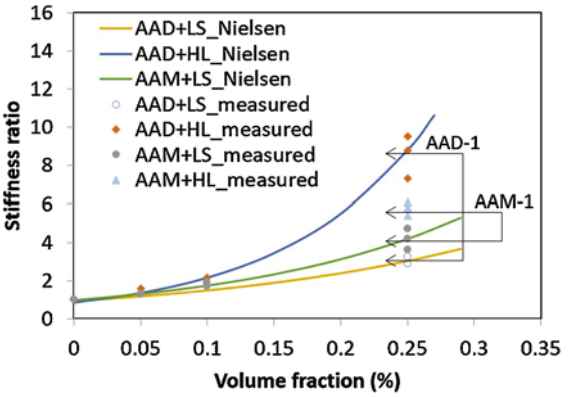


Fig. 5.13 Stiffening ratio versus filler volume fraction. After Little and Petersen (2005)

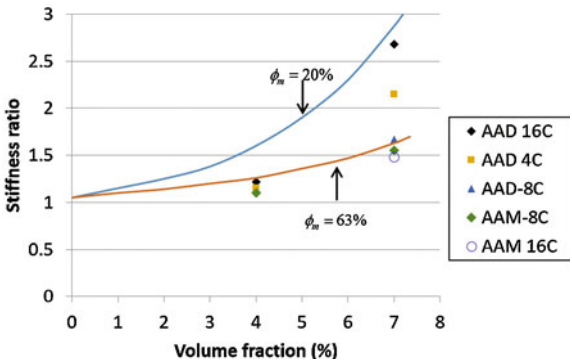


Fig. 5.14 Stiffening ratio: defined as G* of blend divided by G* of base bitumen as a function of volume fraction

with HL than for AAD-1 with LS, which infers that HL has a greater physico-chemical interaction with the binder AAD-1 than does LS. On the other hand, the *KE* values between AAM-1 with LS and AAM-1 with HL were similar.

Of relevance to the present work are rheological data taken from an earlier aging study by Petersen et al. (1987a, b, c, d) and shown in Table 5.4. In this study, West Texas-Maya blend and a Boscan bitumen were treated with 20% HL and LS of similar particle size distribution. For the unaged bitumens, consistent with the present stiffness data, HL produced a greater increase in the 60 °C dynamic shear modulus (G^*) of the Boscan bitumen than did LS. The stiffening effect of the two fillers was similar for the West Texas-Maya blend; however, HL produced a significant increase in the tan δ with this bitumen, while the LS produced a decrease. The data on the bitumens after the thin film accelerated aging test (TFAAT), which ages bitumen in the laboratory to a level typical of pavement aging, are particularly significant (Petersen 1989). In this procedure, 4.0 g of asphalt was aged as a 160-μm-thick film at 113 °C (235.4 °F) for three days in a standard rolling thin film over (RTFO) bottle in a standard RTFO oven. The opening of the RTFO bottle was fitted with a capillary tube to restrict volatile losses to the same approximate level

Table 5.4 Comparison of rheological data[a] at 60 °C for unaged and TFAAT-aged asphalts before and after treatment with limestone power and high-calcium lime [based on data from Petersen et al. (1987a, b, c, d)]

Asphalt	Treatment	G' (dynes/cm^2)	G" (dynes/cm^2)	G* (dynes/cm^2)	Tan δ	Viscosity, poise	Aging Index[b]
Unaged							
Boscan	None	1.266×10^3	1.32×10^4	1.33×10^4	10.5	8.38×10^2	–
	20% Limestone	2.47×10^3	2.09×10^4	2.30×10^4	9.3	1.4×10^3	–
	20% High-calcium lime	3.67×10^3	3.18×10^4	3.20×10^4	8.7	2.02×10^3	–
West Texas-Maya blend	None	1.41×10^3	1.84×10^4	1.85×10^4	13.1	1.17×10^3	–
	20% Limestone	2.08×10^3	2.55×10^4	2.56×10^4	12.2	1.61×10^3	–
	20% High-calcium lime	1.63×10^3	2.59×10^4	2.60×10^4	15.9	1.64×10^3	–
TFAAT-aged							
Boscan	None	7.01×10^3	2.14×10^4	2.25×10^4	3.1	1.79×10^5	2014
	20% Limestone	1.96×10^4	5.09×10^4	5.46×10^4	2.6	4.33×10^5	299
	20% High-calcium lime	9.28×10^2	6.74×10^3	6.80×10^3	7.3	5.40×10^4	27
West Texas-Maya blend	None	1.66×10^4	4.70×10^3	4.99×10^4	2.8	3.96×10^5	338
	20 % Limestone	5.73×10^4	1.35×10^5	1.46×10^5	2.3	1.16×10^6	720
	20% High-calcium lime	1.39×10^3	1.06×10^4	1.07×10^4	7.6	8.49×10^4	52

[a]Rheological data on unaged and aged asphalts obtained at 15.85 rad/s and 0.126 rad/s, respectively
[b]Ratio at viscosity aged/viscosity unaged

experienced in compacted mixtures. Note in Table 5.4 that the bitumens containing LS showed both higher aging indexes and lower tan δ than the untreated bitumens. On the other hand, HL significantly lowered the aging index and increased the tan δ relative to both the untreated and LS-treated bitumens. Thus, it is evident that a chemical and/or a physicochemical interaction occurred between HL and the bitumen that did not occur with LS.

Hydrated lime works to reduce the rate of oxidative age hardening by irreversibly adsorbing reactive components from the asphalt onto its surface. These components are primarily reactive pre-asphaltenes and asphaltenes, as will be shown later. Removal of these components from the asphalt matrix improves asphalt component compatibility, thus reducing viscosity sensitivity to the oxidation products produced by aging. The components removed by lime are also reactive to oxidation; therefore, their removal also reduces the rate of formation of oxidation products (Plancher and Peterson 1976; Petersen et al. 1987a, b, c, d). The significantly higher tan δ, a relative measure of the ability of asphalt to dissipate shear stress, of the aged HL-treated asphalts (Table 5.4) strongly indicates that lime–bitumen interaction should be beneficial in reducing cracking in aged pavements.

5.6.4 *Effects of Hydrated Lime on Low-Temperature Flow Properties*

The stiffening effect of fillers was also evaluated at low temperatures using Eq. 5.5. However, instead of viscosity, a complex modulus was used to evaluate the filler effect via Eq. 5.5. The complex modulus accounts for the viscoelastic effects. The effects are presented in Fig. 5.14. The lower the temperature, the more closely HL behaved as a classical filler in bitumen AAD-1, and the effective volume content of the filler at lower temperatures became the true volume fraction.

Lesueur and Little (1999) prepared 12.5 × 25 × 125 mm^3 beams from neat bitumen and bitumen plus filler (mastic). The beams were prepared with a 5-mm notch in the center to meet the requirements of ASTM E 399-90 for fracture toughness testing. Specimens were loaded monotonically until failure at −30°C with a cross-head speed of 1.2 mm/min to ensure brittle failure. Fracture toughness was computed according to ASTM E 399 (Lesueur and Little 1999), and the data are recorded in Table 5.5. The fracture toughness values for unaged samples were in the range of 40 kN $m^{-3/2}$, which is in agreement with values found by Rodriquez et al. (1995) on similar bitumens. Note also in Table 5.5 that the fracture toughness was preserved in the mixtures after TFO-PAV aging. Lesueur and Little (1999) also compared the surface energies of AAD-1 and AAM-1, measured using the Wilhelmy plate test, with the fracture energy and found the fracture energy was about ten times greater than that calculated by Griffith's (1921) theory. This theory states that the energy required for fracture is that required to form two fracture

Table 5.5 Fracture toughness of HL-modified materials before and after TFO-PAV aging testing at (−30 °C). After Little and Petersen (2005)

Bitumen	Unaged		Aged	
	0% HL	20% HL	0% HL	20% HL
AAD	38 ± 12	71 ± 23	40 ± 10	71 ± 6
AAM	34 ± 5	67 ± 4	37 ± 8	69 ± 19

Note Toughness measured kN-m$^{-0.5}$ in three-point bending at −30 °C and 1.2 mm/mm. Uncertainty corresponds to 95% confidence interval (±2 sigma, where sigma = standard deviation)

surfaces. This leads to the conclusion that the true fracture energy is the sum of the surface energy and the energy needed to plastically deform the material. From Table 5.5 it is clear that the addition of 20% HL by weight of binder just about doubles fracture toughness and is similar to the fracture toughness increases measured by Rodriquez et al. (1995) on similar volume fractions of glass spheres or baghouse fines in bitumen. Lesueur and Little (1999) proposed that a complex mechanism including crack pinning (as previously proposed by Rodriquez et al.) was responsible for the substantially higher level of energy needed to cause fracture in filled systems.

Figure 5.15 shows the comparison of the ratio of fracture toughness of filled bitumens (K_{IC}) to neat bitumens (K_{ICO}). It is apparent that the K_{IC}/K_{ICO} ratio is substantially higher for HL-filled bitumens than for bitumens filled with other fillers evaluated, indicating that HL more efficiently dissipates crack tip energy even though the mechanism may be complex and is not precisely known.

The low-temperature flow properties of bitumen modified with HL and LS are particularly intriguing. Petersen et al. (1987a, b, c, d) performed tensile elongation tests on ribbon-shaped specimens (0.18 × 0.16 × 1.3 cm) of oxidatively aged asphalts in which the asphalt was elongated in tension at a very slow elongation rate (2.5×10^{-4} s^{-1}) at temperatures in the brittle-fracture region, where conditions are related to the tensile stresses responsible for low-temperature, thermally induced transverse cracking in pavements. Although the elongation rate chosen is probably

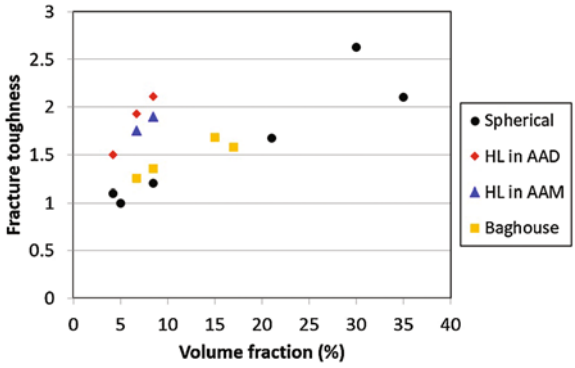


Fig. 5.15 Fracture toughness as a function of filler volume fraction. After Little and Petersen (2005)

Table 5.6 Comparison of tensile elongation data on TFAAT-aged asphalts, both untreated and after treatment with limestone powder and high-calcium lime. After Little and Petersen (2005)

Asphalt	Treatment	Temperature (°C)	Elongation[a] (%)	Tensile Stress[b] (kPa)	Elongation At Maximum Tensile Stress[c]	Stiffness Modulus (kPa)
Boscan	None	−5	10.6	560	10.2/710	2.24×10^4
		−10	4.8	830	4.1/850	3.68×10^4
	20% Limestone	−5	4.0	1,000	3.5/1,100	6.19×10^4
		−10	2.8	1,680	2.8/1,680	1.13×10^5
	20% High-calcium lime	−5	15+	680	15.0/680	2.06×10^4
		−10	11.7	1,170	9.8/1,240	4.37×10^4
West Texas-Maya blend	None	−5	5.5	650	4.5/850	3.24×10^4
		−10	4.4	1,340	4.4/1,340	4.89×10^4
	20% Limestone	−5	4.0	1,580	3.7/1,750	8.74×10^4
		−10	0.75	1,310	0.8/1,310	1.56×10^5
	20% High-calcium Lime	−5	13.0	920	9.8/1,170	4.06×10^5
		−10	8.3	2,170	8.3/2,170	6.12×10^4

[a]Elongation to break or maximum test elongation of 15% at fractional elongation rate of 2.5% $\times 10^{-4}$/s
[b]Tensile stress at break or at maximum elongation of 15%
[c]elongation at break/tensile strength in kPa

higher than the shrinkage rate in pavements during environmental temperature cycling, it was the slowest rate possible within the constraints of the mechanical spectrometer used to make measurements. The elongation data were obtained on the bitumen mastic after aging in the thin film accelerated aging test (TFAAT) (Petersen 1989), which produces a level of oxidation similar to that reached during its service life in pavements where transverse cracking may become a problem. Test results showed that the bitumens containing HL, even though generally stiffer than their untreated counterparts, were elongated in tension to significantly greater amounts before low-temperature break than their untreated counterparts. Thus, the aged, lime-treated bitumens were both tougher (more area under the stress–strain curve) and more ductile in the brittle temperature region. This is illustrated in Table 5.6 (Petersen et al. 1987a, b, c, d). Also of significance is that this benefit was not evident for the aged bitumens containing LS. While the higher tensile strength at break of the HL-treated bitumens suggests stronger pavements, the increased elongation to break is more significant. Pavements that can elongate through ductile flow, of course, have a greater potential to resist fracture and particularly thermally induced low-temperature transverse cracking.

Note in Table 5.6 that, for all asphalts, the elongation to break was about twice as large for the lime-treated asphalts as for their untreated controls. The data for the Boscan asphalt is of particular interest. Even though the lime-treated asphalt showed a higher viscosity and a lower tan δ at the higher 60 °C temperature (Table 5.4), it also showed the greatest benefit from improvement in low-temperature flow properties at −10 °C after aging (Table 5.6). These results suggest that pavements built with this lime-treated asphalt should have relatively greater resistance to rutting at the high temperature and also greater resistance to transverse thermally induced cracking at the low temperature. Based on the data at these two temperature extremes, it is tempting to speculate that, at some intermediate temperature, for example 25 °C, asphalt mixtures containing the lime-treated asphalt should have superior resistance to failure during repeated load cycling during fatigue testing and increased resistance to fatigue cracking in pavements.

5.6.5 *Influence of Filler on Damage in Asphalt Mastic*

Kim et al. (2003) used the dynamic mechanical analysis (DMA) test to evaluate damage in sand asphalt mixtures. The DMA is a controlled-strain, cyclic torsional experiment on a 50-mm-high and 12-mm-diameter specimen of asphalt mastic (about 8% binder) and Ottawa sand compacted to about 6% air voids. The experiment is continued until a transition point (*Nt*) is reached that coincides with a peak in the phase angle. Reese (1997) and Rowe and Bouldin (2000) have identified this as a realistic failure criterion.

Kim et al. (2003) used dynamic mechanical analysis (DMA) to evaluate damage in sand asphalt mixtures. The DMA is a controlled-strain, cyclic torsional experiment on a 50-mm-high and 12-mm-diameter cylindrical sample with about 8%

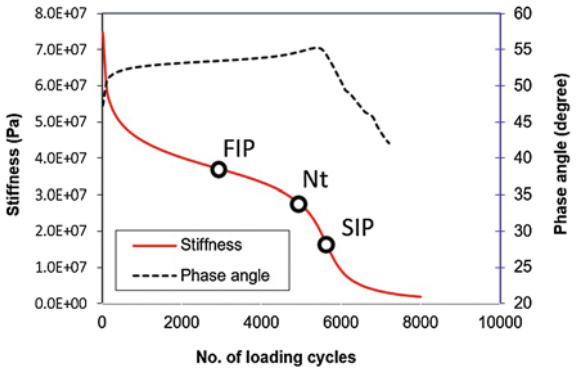


Fig. 5.16 Changes are stiffness and phase angle due to fatigue damage. After Kim et al. (2003)

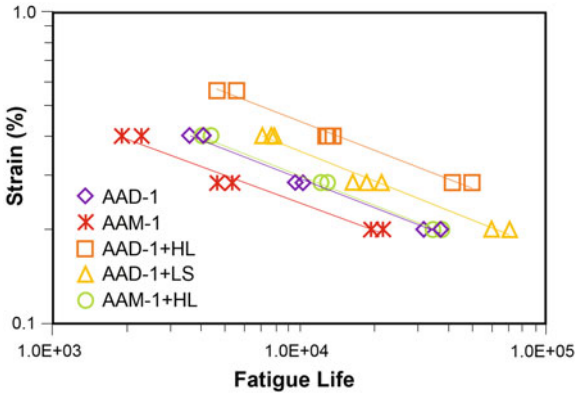


Fig. 5.17 Fatigue life of DMA sand asphalt mixtures with various asphalt binder and filler combinations. After Kim et al. (2003)

binder and Ottawa sand compacted to about 6% air voids. Kim et al. (2003) investigated three failure points described in the literature and shown in Fig. 5.16: the first inflection point (FIP), the second inflection point (SIP), and the transition point (N_t). After careful evaluation using error analysis, Kim et al. selected the N_t parameter as the most reliable indicator of fatigue failure. Rowe and Bouldin (2000) describe N_t as the maximum value of the product of stiffness and number of cycles.

Figure 5.17 presents the number of cycles to failure at each of three strain levels for each sand asphalt mixture for two asphalts, AAM and AAD, and for the same asphalt with LS and HL as fillers.

The difference between filled binders and unfilled binders can also be identified by monitoring dissipated pseudostrain energy (DPSE) in fatigue tests. The DPSE is defined as dissipated energy calculated from measured physical stress versus the pseudostrain domain. By applying the pseudostrain concept, linear viscoelastic time dependency is eliminated, and additional time dependency can be expressed as a quantity due to nonlinearity and/or damage. The theoretical background for the pseudostrain concept can be found in the literature (Kim et al. 2003).

In typical cyclic loading fatigue tests, the area inside of the hysteresis loop of stress versus pseudostrain at each cycle represents the amount of dissipated energy due to damage. The analytical harmonic representation of the pseudostrain at time t under the strain-controlled torsional shear mode can be expressed as:

$$\gamma^R(t) = \frac{1}{G_R}[\gamma_0]\,|G*|\,\sin(2\pi ft + \theta + \varphi) \tag{5.9}$$

where $\gamma^R(t)$ = pseudostrain at time t; G_R = reference shear modulus, typically equal to 1; γ_0 = shear strain amplitude; $|G*|$ = linear viscoelastic dynamic modulus; f = loading frequency in Hz; θ = regression constant in radian; and φ = linear viscoelastic phase angle in radian.

Figure 5.18 shows typical hysteresis stress–pseudostrain behavior resulting from harmonic cyclic loading without and with damage, respectively. As can be seen in Fig. 5.18a the hysteretic behavior due to both loading–unloading and repetitive

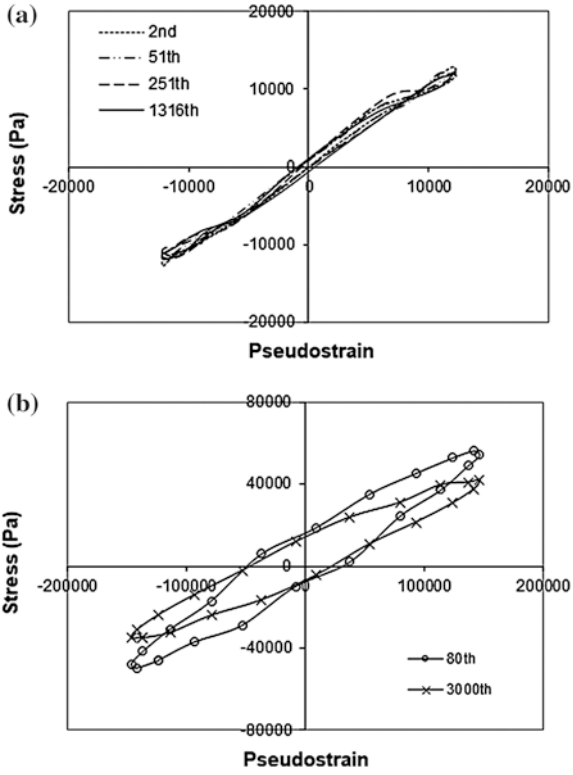


Fig. 5.18 Typical hysteresis stress pseudostrain behavior **a** without damage and **b** with damage. After Kim et al. (2003)

loading disappears in stress versus pseudostrain plots if the sample is tested in the linear viscoelastic region without damage. On the other hand, when a specimen is loaded to a level that causes damage, a hysteresis loop results (Fig. 5.18b). The inside area of the hysteresis loop at each loading cycle was calculated and plotted on a log-linear scale, as shown in Fig. 5.19a, which demonstrates that samples experience damage at an increasing rate in the early stage, but the rate of damage then decreases and eventually decelerates because the sample is close to complete failure. The area below the curve between the DPSE and the number of loading cycles represents cumulative dissipated pseudostrain energy (CDPSE).

Figure 5.19b presents the changes in CDPSE as a function of loading cycles. Correspondingly, CDPSE at fatigue failure can be employed as a quantitative indicator to explain fatigue resistance and/or damage accumulation capability of a material. The CDPSE values corresponding to fatigue failure for each mixture at different strain levels (0.20, 0.28, 0.40, and 0.56%) were determined and are illustrated in Fig. 5.20. Average values of CDPSE for each mixture at a 0.28% strain level were calculated for better clarity. The percent increases in the average CDPSE at a 0.28% strain level for AAD-1 mixed with limestone, AAD-1 mixed with hydrated lime, and AAM-1 mixed with hydrated lime were 133, 588, and 442%, respectively, thus demonstrating the unique effects of HL.

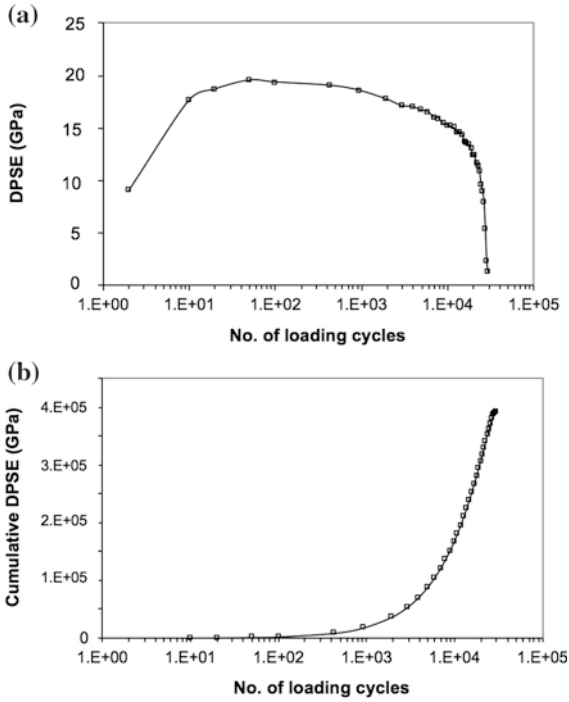


Fig. 5.19 Log-linear plots of: **a** dissipated pseudostrain energy and **b** cumulative dissipated pseudostrain energy versus number of loading cycles. After Kim et al. (2003)

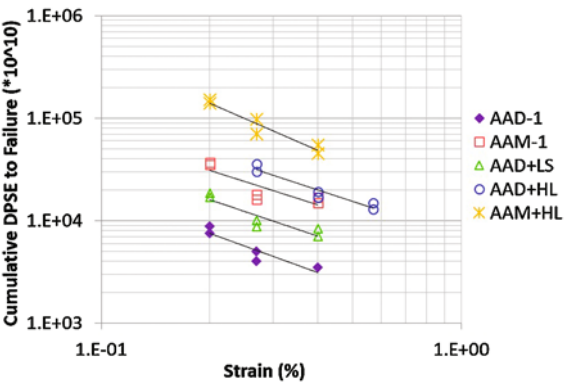


Fig. 5.20 Cumulative DSPE at fatigue failure for each mixture at different strain levels. After Kim et al. (2003)

The effect of filler addition on fatigue life can be easily verified by studying Fig. 5.20. Filled systems show greater damage accumulation capability than do unfilled systems. It is also evident that hydrated lime is a better material than limestone filler in terms of fatigue resistance under the strain-controlled mode of loading. The percent increase in average CDPSE at the 0.28% strain level cited

previously further infers that the degree of interaction and the effectiveness of hydrated lime are dependent on the type of bitumen with which it is used. Hydrated lime is more effective with bitumen AAD-1 than bitumen AAM-1. This is consistent with results from Lesueur and Little (1999).

5.6.6 *Effect of Hydrated Lime on Microstructural Model of Asphalt*

Past work (Plancher and Petersen 1976; Petersen et al. 1987a, b, c, d) clearly demonstrates that the interaction of hydrated lime with asphalt is different from that of typical mineral fillers. Although HL exhibits the typical filler effect, which increases asphalt stiffness, it also chemically reacts with asphalt components to alter both the physical state of the dispersed microstructure (related to the asphaltene component of asphalt) and the component compatibility of the asphalt. As a result, hydrated lime produces changes in the flow properties of asphalt that are different from those produced by typical mineral fillers. Thus, to understand the fundamental mechanisms involved in the asphalt-lime reaction, attention must be focused on the physicochemical composition of the asphalt at the molecular or nanoscale level.

At this point it is instructional to briefly review the microstructural model of asphalt as proposed by Petersen et al. 1987a, b, c, d and by Lesueur et al. 1997. Asphalt can be considered as a dispersion of highly associated molecular agglomerates (loosely referred to as asphaltenes and associated resins) in a solvating phase. The basic building block of this dispersed microstructure is a cluster of very strongly associated, highly aromatic molecules that form the "hard core" nuclei of the associated species. These nuclei are kept dispersed in the solvating phase by adsorption of polar resinous components of the asphalt. When these dispersed molecular agglomerates are in low concentration (asphalts with low asphaltene content) and do not significantly interact with one another, the asphalt tends to exhibit Newtonian flow characteristics. However, when the concentration and/or the size of the molecular agglomerates increases to where they begin to interact with one another, larger agglomerated networks are formed, which cause the asphalt to exhibit non-Newtonian flow characteristics. This change in flow behavior is the result of interactions or weak bonding between the microstructural units within the network structure creating a resistance to flow. These interactions must be disrupted to permit flow within the asphalt matrix. When the interactions between microstructural units become sufficiently strong through an increase in microstructure concentration and/or size, a more stable, three-dimensional structured network is formed. Microfracture can then occur within this network under sufficient shear stress. Lowering the temperature increases the interactions within the microstructure. It is proposed that, if this nanoscale microstructure does not self-heal through interaction with the solvating phase, macrocracks will form and propagate from the interconnecting of microcracks, leading to fracture failure of the

asphalt. This proposed sequence of events might manifest itself in an asphalt pavement mixture as, for example, fatigue cracking.

5.6.7 *Hydrated Lime: Chemical and Physicochemical Interactions*

The discussion that follows focuses on how hydrated lime reacts with asphalt to reduce the interactions between the microstructural units in the asphalt, thus decreasing the deleterious effects of microstructure on the flow properties of the asphalt.

Unlike typical mineral fillers, HL is highly reactive chemically and is of relatively low molecular weight; thus, it has a high relative concentration of reactive chemical functionalities. It is also chemically a relatively strong base that causes acidic materials such as carboxylic acids to form calcium salts that are virtually insoluble. Thus, carboxylic acids and related chemical functionalities such as 2-quinolone types in asphalt react irreversibly with the HL. Because the HL is insoluble in the asphalt, this reaction removes these acidic components from the asphalt phase of the matrix by their adsorption on the surface of the HL particles. Data in Table 5.7 from the study by Petersen et al. (1987a, b, c, d) show the results of the analysis of components that were irreversibly adsorbed from a Boscan asphalt onto the surface of the HL particles. The HL particles containing the adsorbed components were separated from a sample of the lime-treated asphalt by dissolving the asphalt in toluene, followed by filtering and washing the lime particles.

The remaining adsorbed components were separated from the lime by carefully dissolving the lime with dilute hydrochloric acid. As seen in the table, the adsorbed components comprised about 5% of the asphalt. The chemical data in the table show high concentrations of carboxylic acids and 2-quinolone types. The ability of lime to remove these components from the asphalt is evidenced by their very low concentration in the remaining 95% of the asphalt not adsorbed by the lime.

Table 5.7 Analysis of components of Boscan asphalt adsorbed and not adsorbed on hydrated lime. Based on data from Petersen et al. (1987a, b, c, and d)

Asphalt fraction	Percent of total asphalt	Concentration, mol/1,000 g	
		Carboxylic acids	2-Quinolone types
Adsorbed on high-calcium lime	5.6	0.83	0.15
Adsorbed on dolomitic lime	4.7	0.80	0.23
Not adsorbed on high-calcium lime	94.4	<0.005	<0.005
Not adsorbed on dolomitic lime	95.3	<0.005	<0.005

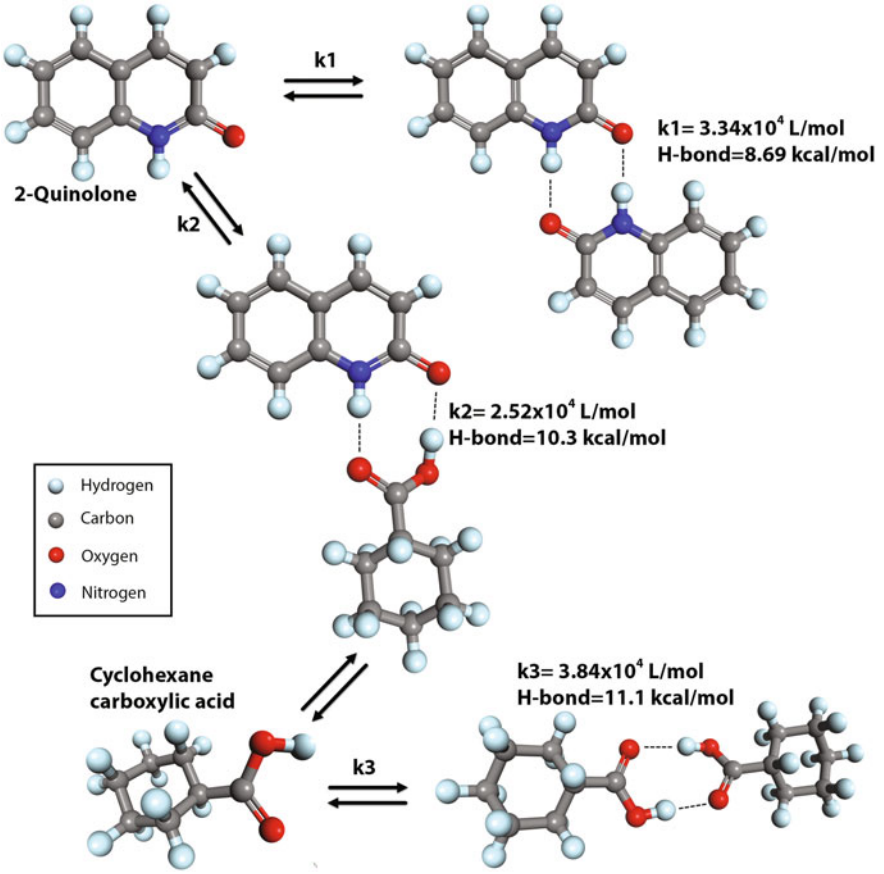


Fig. 5.21 Model study illustrating strong hydrogen bonding interaction of carboxylic acids and 2-quinolone types in asphalts (based on data from Petersen 2011)

Next, it is instructive to consider the question of what effect adsorption of the asphalt components containing carboxylic acid and 2-quinolone types might have on the physical properties of the remaining unadsorbed asphalt. Insight regarding this question is gained by considering the data in Fig. 5.21. The data in this figure are taken from a model thermodynamic study of the hydrogen bonding interactions of cyclohexanecarboxylic acid and 2-quinolone (Petersen 1971). Note in the figure that both the carboxylic acid and 2-quinolone form hydrogen-bonded dimers, and they also interact with each other to form a mixed dimer. Both the dimers and the mixed dimer show strong associations, with hydrogen bond strengths of about 10 kcal/mol and equilibrium constants for dimer formation ranging from 3.84×10^3 to 3.34×10^4. As evidence of the strength of this dimer association in asphalt, Petersen et al. (2011) observed that the acid dimer of naturally occurring carboxylic acids in asphalt is only about 50% dissociated after a 2,000-fold dilution of the asphalt in carbon tetrachloride.

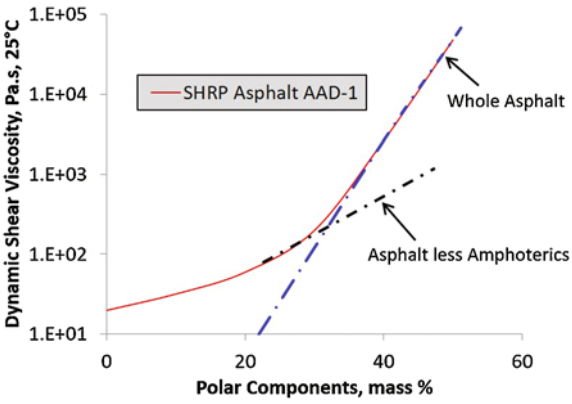


Fig. 5.22 Effect of removal as amphoteric fraction from asphalt AAD-1 on dynamic shear viscosity. Based on data from Branthaver et al. (1993)

It logically follows that the location within the asphalt microstructure of the molecules containing the strongly associating carboxylic acids and 2-quinolone types is of great importance, because microstructure plays a major role in determining the viscoelastic and flow properties of the asphalt. This subject is addressed by the ion-exchange chromatography (IEC) data shown in Fig. 5.22. These data were obtained at Western Research Institute (WRI) by J.F. Branthaver (Branthaver et al. 1993) during the Strategic Highway Research Program (SHRP). The data show that, for SHRP asphalt AAD-1, 60% of the carboxylic acids and more than 90% of the 2-quinolone types were found in the so-called amphoteric IEC fraction. The term amphoteric denotes asphalt components that have both acidic and basic chemical functional groups within the same molecule or indivisible unit. Thus, these polyfunctional molecules can link or bind together associated molecular agglomerates to form larger associations. Because asphalt viscosity is directly related to the size of the agglomerated microstructural units (Petersen et al. 1993), the amphoterics, which contain the 2-quinolone types and the majority of the carboxylic acids, should play a major role in determining the rheological properties of the asphalt. This conclusion is supported by data in Fig. 5.22 for SHRP asphalt AAD-1.

It is evident from Fig. 5.22 that removal of the amphoteric fraction (28% of the whole asphalt) reduced the viscosity of the asphalt by about 1.5 orders of magnitude. It was clearly demonstrated during the SHRP research (Branthaver et al. 1993) that the amphoteric fraction of asphalt contains the highly associated components of the microstructure, is a major contributor to the asphaltene fraction (heptane insoluble fraction), and is primarily responsible for the elastic component of the dynamic shear modulus. Thus, reducing the effects of this fraction by its partial removal from the asphalt by adsorption on the surface of the HL particle should increase the tan δ (ratio of the viscous/elastic moduli), yielding an asphalt phase that has a greater ability to dissipate shear stress through viscous flow. It logically follows that this should also increase the asphalt's ability to heal during repeated load cycling, thus increasing its resistance to fatigue damage.

5.6.8 Other Literature to Support Lime–Bitumen Interaction

Plancher and Peterson (1976) found that 4–6% of the bitumen, consisting almost exclusively of asphalt components containing the carboxylic acid and 2-quinolone type chemical function groups, was irreversibly adsorbed onto the surface of the lime particles. This adsorbed component was also found to be primarily a component of the asphaltene fraction of the bitumen (Petersen et al. 1976, 1987a, b, c, d). Removal of this component from the bitumen matrix by adsorption on the HL particle increased the component compatibility of the bitumen and reduced its sensitivity to viscosity increase upon oxidative aging. In the 1976 study, the HL with its adsorbed components was removed from the bitumen prior to the aging studies to facilitate investigation of the mechanisms of the reaction of HL and its effects on bitumen properties. The 1987 mastic study further verified the results of the previous work. These studies demonstrated that the unique action of HL was highly asphalt source (composition) dependent, with the more polar, high asphaltene-content asphalts generally being most responsive to the action of the HL. Hydrated lime was also shown to have unique effects not exhibited by LS of a similar particle size distribution, the most significant of which were increasing the tan δ and improving low-temperature flow properties and aging characteristics, as previously discussed.

The work by Petersen and Plancher and Petersen is supported by the work of Hoffman et al. (1998), who also concluded that HL interacts with bitumen and that this interaction depends on the chemical composition of the bitumen. Hoffman and coworkers established this interaction using size exclusion chromatography, which indicated that the larger molecular agglomerates in the asphalt interacted with the HL surface. Further documentation of the interaction was established by Hoffman using light absorption methods. Also in agreement with the earlier work of Petersen et al. (1987a, b, c, d), they concluded, based on a study of two bitumens, that HL interacts more intensively with bitumen than LS filler at the same percentage and with a similar overall size gradation. They also stated that the lime–bitumen interaction may result because of an increase in specific surface, the nature of the Ca $(OH)_2$, or a combination of both. Hoffman et al. (1998) even suggested the possibility of larger molecules entering the pores of HL. This effect was established by Johannson (1998). Finally, Hoffman et al. (1998) further confirmed the results of Plancher and Peterson (1976) and Petersen et al. (1987a, b, c, d) that bitumen composition significantly affects the interaction between HL and bitumen. Vanelstraleter and Verhasselt (1998) also evaluated LS and HL fillers of similar gradation. They concluded that the HL was an "active filler" as compared to the LS of comparable size and grading, thus supporting the work of other researchers. The active filler was much more effective in increasing the ring and ball softening point and increasing high-temperature stiffness (based on complex modulus). The HL system was slightly less susceptible to the effects of construction aging than was the system containing LS filler. The authors further stated that low-temperature rheological effects (complex modulus) were much less pronounced because of the active

HL filler when compared with high-temperature effects. These data are in line with those of Johansson (1998) and Little (1996).

Buttlar et al. (1999) used particulate composite micromechanics models to investigate three reinforcement regimes in asphalt mastics: volume filling, physicochemical interactions, and particle interactions. The fundamental nature of their modeling approach allowed the examination of physicochemical stiffening effects through an equivalent rigid layer modeling technique. This technique showed that the development of a rigid adsorbed layer with a thickness of just 2–10 Å can account for differences between the volume-filling reinforcement prediction and the measured values at lower volume concentrations. Buttlar et al. further concluded that physicochemical reinforcement effects are now believed to play a dominant role throughout the range of filler-to-asphalt ratios encountered in practice. Some of the stiffening effects that were originally thought to be a result of particle interaction reinforcement were shown to be largely explained by physicochemical reinforcement. Buttlar et al. (1999) further stated that HL showed a much higher level of physicochemical reinforcement than fillers consisting of baghouse fines, which concurs with previous research. Johansson (1998) have also reported the uniqueness of HL. They concluded that its potential to interact with bitumen is expected on the basis of its rugose surface and very high specific surface area. Johansson and Isacsson (1996) prepared blends of limestone fines and HL in different bitumens. They found that 20% HL and 20% $CaCO_3$ greatly increased initial creep stiffness for each bitumen, but had no apparent effect on the rate or magnitude of physical hardening. They used a Burger model to quantify dissipated energy during bending beam rheometer (BBR) testing. Dissipation of energy at low temperatures through deformation is viewed as a positive effect, because it reduces the potential for fracture.

Little (1996) and Vanelstralete and Verhasslet (1998) reported similar trends in that the filler effect of HL in bitumen at high temperatures is much more pronounced than at low temperatures, and that at low temperatures, the effect of HL does not appear to significantly diminish the ability of the binder to relax. Craus et al. (1978) reported that, in comparative calorimetric tests, the heat released during mixing with bitumen was eight times higher for HL than for any other filler evaluated. Johansson (1998) believes that this is caused by not only release of interfacial heat but also chemical interaction. WRI (1996) documented the formation of organic salts from the reaction of carboxylic acids in the bitumen with calcium in the lime. WRI recommended conversion of carboxylic acids to insoluble salts (e.g., calcium) before use in pavement mixtures to prevent adsorption of water-sensitive free acids on the aggregate. This infers a chemical reaction between a component of the bitumen and the surface of the HL. Several other studies describe the effects of HL as it interacts with bitumen to reduce the effect of age hardening (Petersen et al. 1987a, b, c, d; Elder et al. 1985; Wisneski et al. 1996; Johansson et al. 1995, 1996; Johansson and Isacsson 1996).

5.7 Polymer Modification

During the past 40 or so years polymer modification of asphalt has become popular. Both plastomers including polyethylene (PE), ethylene vinyl acetate (EVA), ethylene butyl acrylate (EBA) and thermoplastic elastomers including styrene-butadiene-styrene (SBS), styrene-isoprene-styrene (SIS), and styrene-ethylene/butylene-styrene (SEBS) have been shown to improve asphalt properties. Perhaps the most important step in determining whether or not a polymer can improve the properties of the binder is to determine whether the polymer is compatible with the asphalt binder. This compatibility depends on the structure of asphalt. Many believe that polymers and asphalt should be considered as multi-phase, viscoelastic emulsions consisting of polymers, asphaltenes and maltenes (Lesueur 2009; Lesueur et al. 1999). Others consider asphalt to be a homogeneous, continuous molecular solution (Redelius 2009).

Zhu et al. (2014) compared advantages and disadvantages of plastomers and thermoplastic elastomers and explained why, based on how the balance favored the use of the elastomer styrene-butadiene-styrene (SBS) over most other polymers. In short it disperses well in asphalt, compared with other polymers, and produces good improvements in rutting resistance without deterioration of fracture properties. However, regardless of the polymer, the controlling factor is normally the base asphalt binder because it dominates in the weight percentage of the blend, making up over 90% of the blend by weight. In general, high asphaltene contents tend to decrease compatibility between asphalt binder and polymer and the maltene fraction should fall within a certain range as well. The SARA (saturates, aromatics, resins, and asphaltene) composition is a good reference for this.

5.7.1 Plastomers

A popular plastomer for asphalt modification has been low-density polyethylene (LDPE). This is primarily due to availability and relatively low cost. When added to asphalt these polyolefins usually swell by light components from the asphalt forming a biphasic structure with a polyolefin-dispersed phase and a continuous, asphalt matrix phase. With an increase in the concentration of the polyolefin, a phase inversion takes place. The interaction of these two phases is critical to the response of the mixture. Two interlocked continuous phases may prove an ideal structure for this polymer-modified system (Zhu et al. 2014). However, the performance of asphalt modified with polyolefins has proven to be somewhat disappointing in improving elasticity and in maintaining adequate storage stability. This is in part due to the susceptibility of regular, long chains to pack closely and crystallize. This leads to a lack of “interaction” with the asphalt and therefore instability in the system. One can also make a case for the crystallized domains of the polyolefins leading toward a more brittle system.

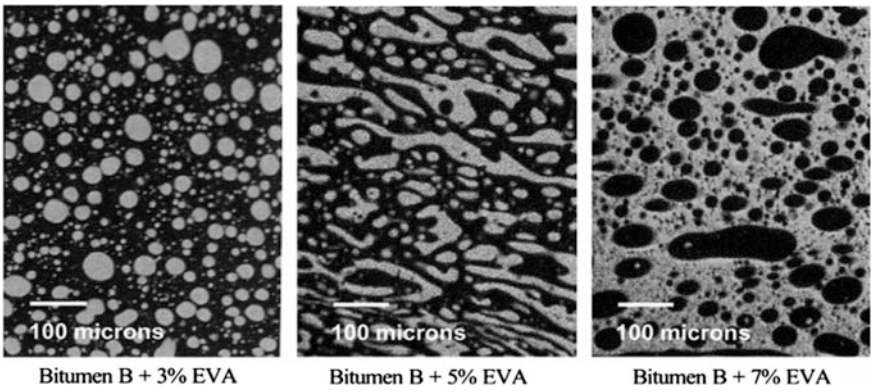


Fig. 5.23 Fluorescent images of EVA modified asphalt with various EVA contents by weight. After Airey (2002)

Ethylene vinyl acetate (EVA) and ethylene butyl acrylate (EBA) are more widely used plastomers compared to polyethylene. The presence of polar acetate in the case of EVA disrupts the ability to closely pack crystalline microstructures and reduces the degree of crystallization. When the vinyl acetate content is too low, the degree of crystallization is high and the properties of the EVA are similar to those of LDPE. Higher vinyl acetate contents lead to a biphase microstructure: crystalline phase and a rubbery, vinyl acetate-rich phase. Once EVA is added to the asphalt light molecular components of the asphalt swell the polymer. When the EVA content is low, a dispersed EVA-rich phase develops within a continuous asphalt matrix. With increase in the EVA content, phase inversion ultimately occurs making the EVA-rich phase the continuous phase (Fig. 5.23).

EVA tends to provide resistance to rutting due to the tough and rigid interlocking network that can develop with the appropriate EVA content. However, EVA is not very effective in improving elastic recovery and low-temperature properties, which are heavily dependent on the EVA network, are not significantly improved (Champion et al. 2000).

5.7.2 Thermoplastic Elastomers

The most popular elastomers are SBS copolymers and SIS copolymers. SBS copolymers comprise SBS triblock chains with a biphase morphology of rigid polystyrene (PS) domains (dispersed phase) in the flexible polybutadiene (PB) matrix (continuous phase) (Fig. 5.24).

At normal pavement temperatures PS blocks provide elasticity due to their rubbery response. The incompatibility of the PS and PB blocks makes it possible to physically crosslink PS blocks as uniformly distributed domains by intermolecular forces at ambient temperatures. This aggregation of PS blocks disappears at high

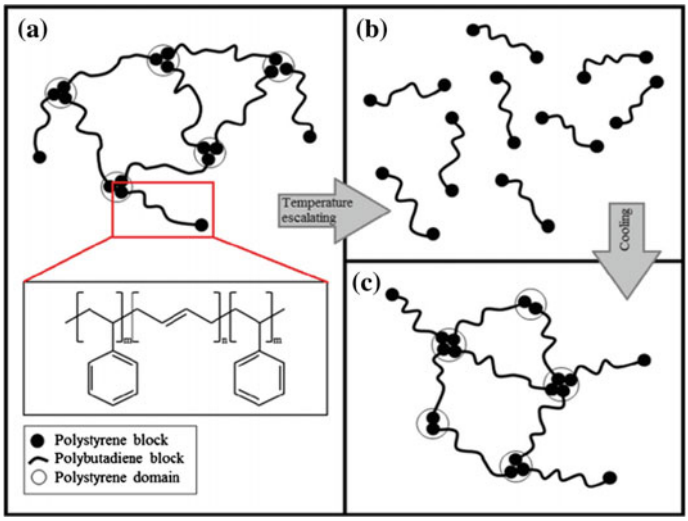


Fig. 5.24 Structure of styrene-butadiene-styrene (SBS) and schematic of reversible crosslink in SBS. After Zhu et al. (2014)

temperatures when kinetic energy of molecular thermodynamic movement is greater than the energy of intermolecular forces. According to Zhu et al. 2014 several interactions are proposed to happen once SBS is introduced into asphalt:

1. PB blocks interact with positively charged groups in asphalt through their π-electrons; PS blocks interact with electron-rich groups in asphalt through aromatic protons.
2. PS blocks in SBS copolymers absorb some saturated branches and a few rings in light components of asphalt, which leads to swelling of PS blocks and the hardening of the asphalt.
3. As SBS concentration increases, phase inversion begins.
4. The ideal microstructure is the formation of two interlocking continuous phases. Within the SBS-rich phase, there are two subphases: a bitumen rich phase and essentially poor PS domains.
5. Once the SBS-rich phase forms, a rubbery supporting network is created in the modified asphalt. This results in increased complex modulus and viscosity, improved elastic response, and enhanced cracking resistance at low temperatures.

Zhu et al. (2014) report that polymer-modified asphalt is susceptible to oxidative aging and poor storage stability. They report several technical developments aimed toward improving aging resistance and storage stability including sulfur vulcanization, use of antioxidants, use of hydrophobic clay minerals, and functionalization. Sulfur vulcanization is a chemical process used in the rubber industry where sulfur works to chemically crosslink polymer molecules to asphalt through sulfide and/or polysulfide bonds (Wen et al. 2002). Antioxidants work through retarding

the impact of oxidation by working as scavengers of free radicals and/or decomposing the hydroperoxides that are generated in the process of oxidation (Ouyang et al. 2006; Li et al. 2010). Hydrophobic clay minerals work to improve aging resistance by forming barrier properties using the disperse clay colloid platelets. Storage capacity is also enhanced by decreasing the density difference between polymer modifiers and the asphalt (Sadeghpour Galooyak et al. 2011). Functionalization of reactive polymers refers to adding specific functional groups to the polymer to obtain specific functions of the polymer-modified asphalt.

Lesueur et al. (1999) explain how polymer modification works from the perspective of the Palierne emulsion model and by knowing the mechanical properties of each phase. This model was developed to describe linear viscoelastic properties of an emulsion made of viscoelastic spherical inclusions in a viscoelastic matrix (Brahimi et al. 1991). Palierne (1991) showed that the complex modulus, G^* of an emulsion can be written as

$$G^* = G_m^* P^*(\lambda^*, Ca^*, \phi) \tag{5.10}$$

where G_m^* is the complex modulus of the matrix and P^* is a complex function of the ratio λ^* of the complex moduli (which is a generalization of the viscosity ratio λ introduced by Taylor), of Ca^*, which is the complex capillary number, and ϕ, which is the volume fraction of the dispersed phase:

$$\lambda^* = \frac{G_i^*}{G_m^*}, Ca^* = G_m^* \frac{a}{\gamma}, \tag{5.11}$$

where γ is the interfacial tension and a is the radius of an inclusion. In the semidilute case (more than 10 volume % of dispersed phase), Palierne showed that P^* is given by

$$P^* = \frac{1 + 1.5\frac{E^*}{D^*}\phi}{1 - \frac{E^*}{D^*}\phi} \tag{5.12}$$

$$E^* = 2(\lambda^* - 1)(19\lambda^* + 16) + 8\frac{5\lambda^* + 2}{Ca^*}, \tag{5.13}$$

$$D^* = (2\lambda^* + 3)(19\lambda^* + 16) + 10\frac{\lambda^* + 1}{Ca^*} \tag{5.14}$$

Lesueur et al. (1999) used the Palierne emulsion model to explain how as a consequence of the colloidal nature of asphalt including a solid phase called asphaltenes and a dispersed phase called the maltenes the mechanical properties of asphalt can be improved through polymer modification. Their explanation is simple and direct and suggests that the improvement occurs for two reasons. First, the polymer swells causing a decrease in the lower molecular weight maltene fraction of the matrix. This leads to an increase in the asphaltene content of the matrix

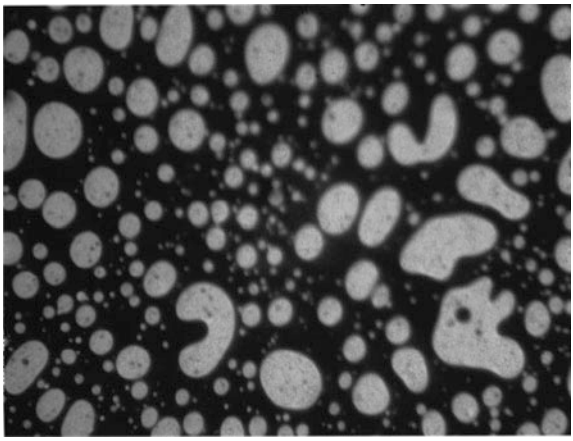


Fig. 5.25 Morphology of ultraviolet fluorescence microscope of a polymer-modified asphalt (x1000). The bright inclusions are the polymer-rich dispersed particles. The matrix, enriched in asphaltenes, remains dark. After Lesueur et al. (1999)

compared to its original condition and a subsequent increase in modulus. Second, the presence of the dispersed phase comprising the swollen polymer modifies the rheology of the material. If the result is a highly swollen polymer, this creates a hard matrix with a high volume fraction of soft inclusions. If the result is a lightly swollen polymer, this generates a blend with a matrix similar to the original asphalt with a low volume fraction of harder polymer particles. This means that the swelling of the polymer controls the ultimate rheological properties of the blend. Figure 5.25 from Lesueur et al. (1999) shows the morphology of a ultraviolet fluorescence microscope of a polymer-modified asphalt.

Figure 5.26 shows the impact of the real and imaginary shear modulus of an asphalt binder with and without modification with styrene-butadiene copolymer

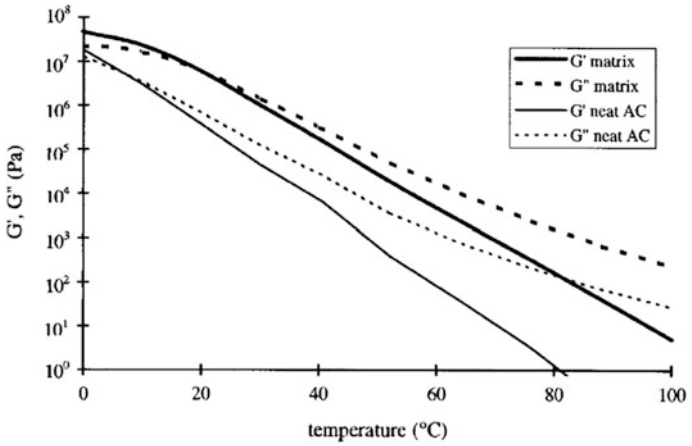


Fig. 5.26 Real (G') and imaginary (G") of complex modulus at 10 rad/s versus temperature for the neat asphalt cement and the matrix of the corresponding asphalt modified with 6% by weight of styrene-butadiene copolymer. After Lesueur et al. (1999)

(6% by weight). The effect is to substantially improve binder properties at the higher temperatures to resist plastic deformation with a lesser effect at the lower temperatures. Nevertheless, some general stiffening across the spectrum of service temperatures in some cases may lead to increased fracture potential, and this must be monitored. However, certain multi-phase effects may also reduce the potential for microcrack development such as crack pinning and improved fracture toughness.

5.8 Summary

A number of additives can be used to enhance the properties of asphalt including: extenders such as sulfur, fillers such as aggregate finer than 75 microns, polymers, fibers, and antistripping agents. This chapter reviews various models used to explain how fillers and polymers react with bitumen and how bitumen chemistry along with a realistic phenomenological models can be used to understand how the additive impacts the rheological properties of the bitumen and how bitumen chemistry impacts this interaction and, in turn, why some asphalts respond favorably with certain additives while others do not.

This chapter focuses on sulfur as primarily an extender but also to some extent as a modifier. The same may be said for hydrated lime, where the fine hydrated lime filler extends and stiffens the bitumen but often with a much greater impact than traditional fillers. Because of the unique properties of hydrated lime and its ability to encapsulate bitumen within it smaller particles and internal voids together with its unique ability to sorb selected polar components of some bitumen, hydrated lime is a "reactive" filler and not an inert filler. Other benefits of hydrated lime are discussed including its ability to reduce the impact of oxidative aging and the ability to form durable, calcium-based organic salt bond between the bitumen and the aggregate surface to reduce moisture sensitivity.

The role of amine additives to reduce moisture sensitivity is explained, and some simplified models are provided to help explain the mechanism of filler bitumen reactions.

This chapter concludes with a brief description of the mechanism of interaction between plastomers and elastomers and bitumen and a general description of why some bitumen is more compatible with a selected polymer than other. The reader should conclude with a general understanding of how a polymer added at a rate of less than about six percent can form a continuous phase with some asphalts.

5.9 Exercises

5.8.1 Use the Palierne model to describe how a suspension in an asphalt matrix can add viscosity to the mixture and how this model may be used to explain how particles with an aspect ratio that defines them as nonspherical may increase the intrinsic viscosity of the suspension.

5.8.2 Describe the mechanisms by which sulfur added to asphalt binder extends the binder and alters the rheology of the asphalt binder during the mixing process and after the asphalt mixture has been compacted and reaches and equilibrium conditions due to a period of curing of about one month.

5.8.3 Describe the interaction between silanol groups and carboxylic acid in the asphalt binder. How can hydrated lime impact this reaction? Using Fig. 5.8, describe how amine groups can alter this reaction to provide a stronger and more durable bond, which is more resistant to moisture damage.

5.8.4 Explain how the addition of hydrated lime has a profound effect on dynamic modulus of asphalt mixtures high test temperatures but a much smaller or even almost indistinguishable effect at temperatures less than about 20 °C.

5.8.5 Figure 5.13 describes the active filler effect of hydrated lime in two very different asphalt binders. What is the compositional difference between asphalt binders AAD and AAM in this figure and as described in the text, and how does this compositional difference impact the relative effect of hydrated lime on these two binders as illustrated in Fig. 5.13.

5.8.6 Using Fig. 5.22 as a basis, describe the impact of the amphoteric fraction of the binder on the dynamic shear viscosity of the asphalt binder over the range of test temperatures.

5.8.7 Describe how Lesueur et al. (1999) used the Palierne model to explain the two reasons why the mechanical behavior of asphalt is improved by polymer modification.

5.8.8 What criteria would you use to select a filler or modifier to achieve a target resistance to (a) fatigue cracking or (b) plastic deformation, rutting?

5.8.9 Using Fig. 5.3 as a basis, describe the rheological changes asphalt binder goes through as temperatures changes during the mixing, compaction and services stages of the asphalt pavement.

5.8.10 Prepare a binder specification of an asphalt binder with a EVA polymer additive.

References

Airey, G. G. (2002). Rheological evaluation of ethylene vinyl acetate polymer modified bitumens. *Construction and Building Materials*, *16*(8).

Allen, R. G., Little, D. N., & Bhasin, A. (2012). Structural characterization of micromechanical properties in asphalt using atomic force microscopy. *Journal of Materials in Civil Engineering, 24,* 1317–1327.

Anderson, D. A. (1996). Influence of fines on performance of asphalt concrete mixtures. In *Proceedings of the 4th Annual Symposium*, International Center for Aggregates Research, Atlanta.

Ardebrant, H., & Pugh, R. J. (1991a). Surface acidity/basicity of road stone aggregates by adsorption from non-aqueous solutions. *Colloids and Surfaces, 53,* 101–116.

Ardebrant, H., & Pugh, R. J. (1991b). Wetting studies on silicate minerals and rocks used in bitumen highways. *Colloids and Surfaces, 58,* 111–130.

Bahia, H. U., Anderson, D. A., & Christensen, D. W. (1992). The bending beam rheometer: A simple device for measuring low- temperature rheology of asphalt binders. *Asphalt Paving Technologists, 61,* 117–148.

Baumgardner, G. L., Masson, J. F., & Hardee, J. R. (2005). Polyphosphoric acid modified asphalt: Proposed mechanisms. *Journal of Association of Asphalt Paving Technologists, 74,* 283–305.

Bhairampally, R. K., Lytton, R. L., & Little, D. N. (2002). A numerical and graphical method to assess permanent deformation potential for repeated compressive loading of asphalt mixtures. Transportation Research Record 1723 (pp. 150–158), Transportation Research Board, Washington, D.C.

Blazek, J., Sebor, G., Maxa, D., Ajib, M., & Paniagua, H. (2000). Effect of hydrated lime addition on properties of asphalt. *Petroleum and Coal, 42*(1), 41–45.

Brahimi, B., Ait-Kadi, A., Ajji, A., Jérôme, R., & Fayt, R. (1991). Rheological properties of copolymer modified polyethylene/polystyrene blends. *Journal of Rheology*.

Branthaver, J. F. J., et al. (1993). Binder characterization and evaluation, Volume 2: Chemistry. Report SHRP-A-368, Strategic Highway Re- search Program, Washington, D.C.

Brion, Y., & Brule, B. (1986). Etudes des melanges bitumes-polymeres—composition—structure —proprieties. Transportation Research Board: Monograph, ISBN: 2-7208-3535-8. [French].

Brule, B. (1996). Polymer-modified asphalt cements used in the road construction industry: Basic principles. Transportation Research Record, 1535 (pp. 48–53), Transportation Research Board, Washington, D.C.

Buttlar, W. G., Bozkurt, D., A-Khateeb, G. G., & Waldorf, A. S. (1999). Understanding asphalt mastic behavior through micromechanics. In *Proceedings of the Annual Meeting*, Transportation Research Board, Washington, D.C.

Champion-Lapalu, L., Planche, J.-P., Maúin, D., Anderson, D. A., & Gerard, J.-F. (2000). 101 low-temperature rheological and fracture properties of polymer-modified bitumens. In *Proceedings, 2nd Eurasphalt and Eurobitume Congress*, Barcelona, Book I.

Chen, J. S., Liao, M. C., & Shiah, M. S. (2002). Asphalt modified by styrene-butadiene-styrene triblock copolymer: Morphology and model. *Journal of Materials in Civil Engineering*, *14*(3).

Claudy, P., Létoffé, J. M., King, G. N., & Planche, J. P. (1992). Characterization of asphalts cements by thermomicroscopy and differential scanning calorimetry: Correlation to classic physical properties. *Fuel Science and Technology International, 10,* 735–765.

Coussot, P. (2005). *Rheometry of pastes, suspensions and granular materials*. New York: Wiley and Sons.

Craus, J., Ishai, I., & Sides, A. (1978). Some physico-chemical aspects of the effect and the role of the filler in bituminous paving mixtures. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 47, pp. 558–588).

CROW. (2007). (Centrum voor Regelgeving en Onderzoek in de Grond-, Water- en Wegenbouw en de Ver- keerstechniek – Dutch technology platform for transport, infrastructure and public space), Deelhoofdstuk 31.2 Asfaltverhardingen, Ede (The Netherlands), CROW (in Dutch).

Curtis, C. W., Ensley, K., & Epps, J. (1993). Fundamental properties of asphalt-aggregate interactions including adhesion and absorption (SHRP Report No. A-341). Washington, DC: National Research Council.

Elder, A. C., Hatting, M. M., Servas, V. P., & Marais, C. P. (1985). Use of aging tests to determine the efficiency of hydrated lime additions to asphalt in retarding its oxidative hardening. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 54, pp. 118–139).

Eulitz, H.-J., Schellenberg, K., Ritter, H.-J., & Schmidt, S.-O. (1998). Verbesserung von Asphalteigen- schaften durch Zugabe von Kalkhydrat (Report No. 2/98/B005). Köln: Forschungsgemeinschaft Kalk und Mörtel e. V (in German).

Gawel, I., & Baginska, K. (2004). Effect of chemical nature on susceptibility of asphalt to aging. *Petroleum Science and Technology, 22,* 1261–1271.

Gawrys, K. L., & Kilpatrick, P. K. (2005). Asphaltenic aggregates are polydisperse oblate cylinders. *Journal of Colloid and Interface Science, 288*(2), 325–334.

Griffith, A. A. (1921). The phenomena of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London, Series A, 221,* 163–198.

Hefer, A., & Little, D. N. (2005). Adhesion in Bitumen-Aggregate Systems and Quantification of the effects of Water on the Adhesive Bond. Research Report ICAR-505-1.

Hefer, A. W., Little, D. N., & Lytton, R. L. (2005). A synthesis of theories and mechanisms of bitumen-aggregate adhesion including recent advances in quantifying the effects of water. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 74).

Heukelom, W., & Wija, P. W. O. (1971). Viscosity of dispersions governed by concentration and rate of shear. *Asphalt Paving Technologists*, *40*.

Hicks, R. G. (1991). *NCHRP synthesis of highway practice 175: Moisture damage in asphalt concrete*. Washington, DC: Transportation Research Board.

Hicks, R. G., & Scholz, T. V. (2003). Life cycle costs for lime in hot mix asphalt (Vol. 3). Arlington, VA: National Lime Association. Retrieved from http://www.lime.org/LCCA/LCCA_Vol_I.pdf, http://www.lime.org/LCCA/LCCA_Vol_II.pdf, http://www.lime.org/LCCA/LCCA_Vol_III.pdf.

Hoffman, P., van Veldhuizen, M., Noor, H., & Smits, R. (1998). Hydroxide in filler, Netherlands Pavement Consultants, Utrecht, The Netherlands.

Hopman, P. C., Vanelstraete, A., Verhasselt, A., & Walter, D. (1999, May). Effects of hydrated lime on the behaviour of mastics and on their construction ageing. In *Proceedings of the 5th International Conference on Durable and Safe Road Pavements* (Vol. 1, pp. 59–68). Warsaw, Poland: Polish Road and Bridge Research Institute.

Huang, S. C., Robertson, R. E., Branthaver, J. F., & Petersen, J. C. (2005). Impact of lime modification of asphalt and freeze-thaw cycling on the asphalt-aggregate interaction and moisture resistance to moisture damage. *Journal of Materials in Civil Engineering, 17*(6), 711–718.

Johansson, L. S. (1998). Bitumen aging and hydrated lime. Report TRITA-IP FR 98-38, Royal Institute of Technology, Stockholm, Sweden.

Johansson, L. S., Branthaver, J., & Robertson, R. (1995). A study of rheological properties of lime treated paving asphalts aged at 60 °C in a pressure aging vessel. *Fuel Science and Technology International, 13*(10), 1317–1343.

Johansson, L. S., Branthaver, J., & Robertson, R. (1996). The influence of metal-containing compounds on enhancement and inhibition of asphalt oxidation. *Fuel Science and Technology International, 14*(8), 1143–1159.

Johansson, L. S., & Isacsson, U. (1996). Influence of testing conditions on the capacity of hydrated lime to inhibit bitumen aging, Division of Highway Engineering, Royal Institute of Technology, Stockholm, Sweden.

Johansson, L. S., & Isacsson, U. (1998). Effect of filler on low temperature physical hardening of bitumen. *Construction Building Materials, 12,* 463–470.

Khattak, M. J., & Kyatham, V. (2008). Viscoelastic behavior of hydrated lime-modified asphalt matrix and hot-mix asphalt under moisture damage conditions. *Transportation Research Record, 2057,* 64–74.

Kim, Y.-R., Little, D. N., & Song, I. (2003). Mechanistic evaluation of mineral fillers on fatigue resistance and fundamental material characteristics. In *Proceedings of the 82nd Transportation Research Board Meeting*, Transportation Research Board, Washington, D.C.

Koots, J. A., & Speight, J. G. (1975). Relation of petroleum resins to asphaltenes. *Fuel, 54,* 179–184.

Lackner, R., Spiegl, M., Blab, R., & Eberhardsteiner, J. (2005). Is low-temperature creep of asphalt mastic independent of filler shape and mineralogy? Arguments from multiscale analysis. *Journal of Materials in Civil Engineering, 17*(5), 485–491.

Landau, L. D., & Lifshitz, E. M. (1958). *Fluid mechanics* (1st English ed.). London: Pergamon Press.

Laval, C., & Quivoron, C. (1973). Mise en évidence d'une corrélation entre le rapport hydrophile/lipophile des résines époxydes et leur compatibilité avec le bitume routier. *Comptes rendus de l'Académie des Sciences IIc, 256,* 743–746.

Laval C., & Brûlé B. (1974). Etude expérimentale de la compatibilité de résines epoxydes avec le bitume. Application à la prévision de systèmes compatibles. Rapport de Recherche LCPC 40, LCPC Eds, Paris (France). [French].

Lesueur, D., Dekker, D. L., & Planche, J. P. (1995). Comparison of carbon black from pyrolized tires to other fillers as asphalt rheology modifiers. Transportation Research Record, 1515 (pp. 47–55), Transportation Research Board, Washington, D.C.

Lesueur, D., Gerard, J. F., Claudy, P., Letoffe, J. M., Planche, J. P., & Martin, D. (1996). A structure-related model to describe asphalt linear viscoelasticity. *Journal of Rheology, 40,* 813–836.

Lesueur, D., Gerard, J. F., Claudy, P., Letoffe, J. M., Planche, J. P., & Martin, D. (1997). Relationships between the structure and the mechanical properties of paving grade asphalt cements. *Proceedings of the Association of Asphalt Paving Technologists, 66,* 486–519.

Lesueur, D., Gerard, J. F., Claudy, P., Letoffe, J. M., Planche, J. P., & Martin, D. (1998). Polymer modified asphalts as viscoelastic emulsions. *Journal of Rheology, 42,* 1059–1074.

Lesueur, D., & Little, D. N. (1999). "Effect of hydrated lime on rheology, fracture, and aging of bitumen. Transportation Research Record, 1661 (pp. 93–105), Transportation Research Board, Washington, D.C.

Lesueur, D. (2009). The colloidal structure of bitumen: Consequences on the rheology and on the mechanisms of bitumen modification. *Advances in Colloid and Interface Science, 145,* 42–82.

Lesueur, D. (2010). Hydrated lime: A proven additive for durable asphalt pavements—Critical literature review. Brussels: European Lime Association (EuLA). Retrieved from http://www.eula.eu.

Lesueur, D., Petit, J., & Ritter, H.-J. (2012). Increasing the durability of asphalt mixtures by hydrated lime addition: What evidence? *European Roads Review, 20,* 48–55.

Lesueur, D., Petit, J., & Ritter, H. J. (2013). The mechanisms of hydrated lime modification of asphalt mixtures: A state-of-the-art review. *Road Materials and Pavement Design, 14*(1), 1–16.

Lewis, T. B., & Nielsen, L. E. (1968). Viscosity of dispersed and aggregated suspensions of spheres. *Transactions. Society of Rheology, 12,* 421–443.

Li, Y., Li, L., Zhang, Y., Zhao, S., Xie, L., & Yao, S. (2010). Improving the aging resistance of styrene-butadiene-styrene tri-block copolymer and application in polymer-modified asphalt. *Journal of Applied Polymer Science, 116,* 754–761. doi:10.1002/app.31458.

Little, D. N. (1995). *Handbook for stabilization of pavement subgrades and base courses with lime*. Dubuque, IA: Kendall/Hunt Publishing Company.

Little, D. N. (1996). Hydrated lime as a multi-functional modifier for asphalt mixtures. In *Proceedings of the European Seminar*, Lhoist Group, Brussels, Belgium.

Little, D. N., & Epps, J. A. (2001). The Benefits of hydrated lime in hot mix asphalt. Arlington, VA: National Lime Association. Retrieved from http://www.lime.org/ABenefit.pdf.

Little, D. N., Lytton, R. L., Williams, D., & Chen, W. (2001). Microdamage healing in asphalt and asphalt concrete. Volume I: Project summary report. FHWA-RD-98-141, Federal Highway Administration, Washington, D.C.

Little, D. N., & Petersen, J. C. (2005). Unique effects of hydrated lime filler on the performance-related properties of asphalt cements: Physical and chemical interactions revisited. *Journal of Materials in Civil Engineering, 17*(2), 207–218.

Logaraj, S. (2002). Chemistry of asphalt-aggregate interaction—Influence of additives. Presented at the Moisture Damage Symposium, Laramie, Wyoming.

Lytton, R. L., Uzan, J., Fernando, E. G., Roque, R., Hiltunen, D., & Stoffels, S. M. (1993). Development and validation of performance prediction model and specifications for asphalt binders and paving mixes. Report SHRP-A-357, Strategic Highway Research Program, Washington, D.C.

Mitchell, J. C., & Lee, A. R. (1939). The evaluation of fillers for tar and other bituminous surfaces. *Journal of the Society of Chemical Industry*, *58*, 299–306.

Mohammad, L., Abadie, C., Gokmen, R., & Puppala, A. (2000). Mechanistic evaluation of hydrated lime in hot-mix asphalt mixtures. Transportation Research Record 1723 (pp. 26–36), Transportation Research Board, Washington, D.C.

Mohammad, L. N., Saadeh, S., Kabir, M., & Othman, A. (2008). Mechanistic properties of hot-mix asphalt mixtures containing hydrated lime. *Transportation Research Record, 2051,* 49–63.

Nielsen, L. E. (1970). Generalized equation for the elastic moduli of composite materials. *Journal of Applied Physics, 41,* 4626–4627.

Orange, G., Dupuis, D., & Martin, J. V. (2004). Chemical modification of bitumen through polyphosphoric acid: Properties–microstructure relationship. In *Proceedings of the 3rd Eurasphalt and Eurobitume Congress*, Vienna, Austria.

Ouyang, C., Wang, S., Zhang, Y., & Zhang, Y. (2006). Improving the aging resistance of styrene-butadiene-styrene tri-block copolymer modified asphalt by addition of antioxidants. *Polymer Degradation and Stability*, *91*(4).

Palierne, J.-F. (1991). Linear rheology of viscoelastic emulsions with interfacial tension. *Rheologica Acta*, *30*(497).

Pauli, T., Grimes, W., Beiswenger, J., & Schmets, A. (2015). Surface structuring of wax in complex media. *Journal of Materials in Civil Engineering*, *27*.

Petersen, J. C. (1971). A thermodynamic study by infrared spectroscopy of the association of 2-quinolone, some carboxylic acids, and the corresponding 2-quinolone-acid dimer. *Journal of Physical Chemistry, 75,* 1129–1135.

Petersen, J. C. (1989). A thin film accelerated aging test for evaluating asphalt oxidative aging. *Asphalt Paving Technologists, 58,* 220–237.

Petersen, J. C., Branthaver, J. F., Robertson, R. E., Harnsberger, P. M., Duvall, J. J., & Ensley, E. K. (1993). Effects of physicochemical factors on asphalt oxidation kinetics. Transportation Research Record 1391 (pp. 1–10), Transportation Research Board, Washington, D.C.

Petersen, J. C., Plancher, H., & Harnsberger, P. M. (1987a). Lime treatment of asphalts to reduce age hardening and improve flow properties. In *Proceedings of the Association of Asphalt Paving Technologists* (Vol. 56, pp. 632–653).

Petersen, J. C., Plancher, H., & Harnsberger, P. M. (1987b). Lime treatment of asphalt. Final Report Prepared for the National Lime Association, Western Research Institute, Laramie, Wyo.

Petersen, J. C., Plancher, H., & Harnsberger, P. M. (1987c). Lime treatment of asphalt to reduce age hardening and improve flow properties. In *Proceedings of the Association of Asphalt Paving Technology* (Vol. 56, pp. 632–653).

Petersen, J. C., Plancher, H., & Harnsberger, P. M. (1987d). Lime treatment of asphalt. Final Report Prepared for The National Lime Association, Western Research Institute, Laramie, Wyo.

Petersen, J. C., & Glaser, H. (2011). Asphalt oxidation mechanisms and the role of oxidation products on age hardening revisited. *Road Materials Pavement Design*, *12*.

Pilat, J., Radziszewski, P., & Kalabiska, M. (2000). The analysis of visco-elastic properties of mineral- asphalt mixes with lime and rubber powder. In *Proceedings of the 2nd Eurasphalt and Eurobitume Congress* (Vol. 1, pp. 648–654), paper 0036. Brussels, Belgium: Eurobitume and European Asphalt Pavement Association.

Plancher, H., Green, E. L., & Petersen, J. C. (1976). Reduction of oxidative hardening of asphalts by treatment with hydrated lime—a mechanistic study. In *Proceedings of the Association Asphalt Paving Technologists* (Vol. 45, pp. 1–24). Lino Lakes, MN: Association Asphalt Paving Technologists (AAPT).

Ramond, G., & Lesueur, D. (2004). Adhesion liant granulat. In H. Di Benedetto & J.-F. Corté (Eds.), *Materiaux routiers bitumineux* (Vol. 1, pp. 177–203). Paris: Lavoisier (in French).

Raynaud, C. (2009). L'ajout de chaux hydratée dans les enrobés bitumineux. *BTP Matériaux*, *22*, 42–43 (in French).

Redelius. (2009). Asphaltenes in bitumen, what they are and what they are not. *Road Materials Pavement Design*, *10*(1).

Reese, R. (1997). Properties of aged asphalt binder related to asphalt concrete fatigue life. *Asphalt Paving Technologists, 66,* 604–632.

Rigden, D. J. (1947). Mineral fillers in bituminous road surfacings: A study of filler/binder systems in relation to filler characteristics. *Journal of the Society of Chemical Industry, 66,* 299–309.

Roberts, F. L., Kandal, P. S., Brown, E. R., Lee, D., & Kennedy, T. W. (1996). *Hot mix asphalt materials, mixture design, and construction* (2nd ed.). Maryland: NAPA Research and Education Foundation.

Rodriquez, M. G., Morrison, G. R., Van Loon, J. R., & Hesp, S. A. M. (1995). Low-temperature failure in particulate-filled asphalt binders and asphalt concrete mixes. *Asphalt Paving Technologists, 69,* 159–187.

Rowe, G. M., & Bouldin, M. G. (2000). Improved techniques to evaluate the fatigue resistance of asphaltic mixes. In *Proceedings of the 2nd Enraphalt and Eurobitume Congress*, Barcelona, Spain.

Sadeghpour Galooyak, S., Dabir, B., Nazarbeygi, A. E., Moeini, A., & Berahman, B. (2011). The effect of nanoclay on rheological properties and storage stability of SBS-modified bitumen. *Petroleum Science and Technology, 29*(8), 850–859. doi:10.1080/10916460903502449.

Saville, V. B., & Axon, E. O. (1937). Adhesion of asphaltic binders to mineral aggregates. *Journal of the Association of Asphalt Paving Technologists, 9,* 86–101.

Sebaaly, P. E., Hitti, E., & Weitzel, D. (2003). Effectiveness of lime in hot-mix asphalt pavements. Transportation Research Record 1832 (pp. 34–41), Transportation Research Board, Washington, D.C.

Shashidhar, N., Needham, S. P., Chollar, B. H., & Romero, P. (1999). Prediction of the performance of mineral fillers in stone matrix asphalt. *Asphalt Paving Technologists, 68,* 222–251.

Shashidhar, N., & Romero, P. (1998). Factors affecting the stiffening potential of mineral fillers. Transportation Research Record 1638 (pp. 94–100), Transportation Research Board, Washington, D.C.

Si, Z. (2001). Characterization of microdamage and healing of asphalt concrete mixtures. PhD thesis, Texas A&M University, College Station, Tex.

Siegmann, M. C. (1950a). Manufacture of asphaltic bitumen. In J. P. Pfeiffer (Ed.), *The properties of asphaltic bitumen* (pp. 121–154). Amsterdam: Elsevier.

Siegmann, M. C. (1950b). Methods of routine investigation. In J. P. Pfeiffer (Ed.), *The properties of asphaltic bitumen* (pp. 155–188). Amsterdam: Elsevier.

Simmons, G. W., & Beard, B. (1987). Characterization of acid-base properties of the hydrated oxides on iron and titanium metal surfaces. *Journal of Physical Chemistry, 91,* 1143.

Storm, D. A., Sheu, E. Y., & De Tar, M. M. (1993). Macrostructure of asphaltenes in vacuum residue by small-angle X-Ray scattering. *Fuel, 72,* 977–981.

Storm, D. A., Barresi, R. J., & Sheu, E. Y. (1995). Rheological study of Ratawi vacuum residue in the 298-673 K temperature range. *Energy and Fuels, 9,* 168–176.

Tadros, T. F., & Vincent, B. (1983). Emulsion stability. In P. Becher (Ed.), *Encyclopedia of emulsion science, vol. 1. Basic Theory* (pp. 129–285). New York: Marcel Dekker.

Tarrar, A. R., & Wagh, V. P. (1992). The effect of the physical and chemical characteristics of the aggregate on bonding. Report SHRP-A/ UIR-91-507, Washington, D.C.: Strategic Highway Research Program, National Research Council.

Taylor, G. I. (1932). The viscosity of a fluid containing small drops of another fluid. *Papers of a Mathematical and Physical Character, 138,* 41–48.

Timm, D. H., Robbins, M. M., Willis, J. R., Tran, N., & Taylor, A. J. (2012). Evaluation of Mixture Performance and Structural Capacity of Pavements Using Shell Thiopave: Phase II—Construction, Laboratory Evaluation, and Full-Scale Testing of Thiopave Test Sections—Final Report. Report 11-03, National Center for Asphalt Technology, Auburn University.

Titova, T. I., Kosheleva, L. S., & Zhdanov, S. P. (1987). IR study of hydroxylated silica. *Langmuir, 3,* 960.

Tunnicliff, D. G. (1967). Binding effects of mineral filler. In *Proceedings of the Association of Asphalt Paving Technologists* (pp. 114–156), Transportation Research Board, Washington, D.C.

Vanelstralete, A., & Verhasselt, A. (1998). Interpretation Report EP 5080, Belgian Road Research Centre, Brussels, Belgium.

Voskuilen, J. L. M., & Verhoef, P. N. W. (2003). Causes of premature ravelling failure in porous asphalt. In *Proceedings of the RILEM symposium on Performance Testing and Evaluation of Bituminous Materials* (pp. 191–197). Bagneux, France: International Union of Laboratories and Experts in Construction Materials, Systems and Structures.

Walker, P. (1994). Silane and other adhesion promoters in adhesive technology. In A. Pizzi & K. L. Mittal (Eds.), *Handbook of adhesive technology* (pp. 50–52). New York: Marcel Dekker Inc.

Wen, G., Zhang, Y., Zhang, Y., Sun, K., & Fan, Y. (2002). Improved properties of SBS-modified asphalt with dynamic vulcanization. *Polymer Engineering and Science, 42,* 1070–1081. doi:10.1002/pen.11013.

Western Research Institute (WRI). (1996). Fundamental properties of asphalts and modified asphalts, annual technical report. Contract DTFH61-92C-00170, U.S. Department of Transportation, Washington, D.C.

Wisneski, M. L., Chaffin, J. M., Davison, R. R., Bullin, J. A., & Glover, C. J. (1996). Use of lime in recycling asphalt. Transportation Research Record 1535 (pp. 117–123), Transportation Research Board, Washington, D.C.

Wortelboer, J. P., Hoppen, H. J., Ramond, G., & Pastor, M. (1996). Rheological properties of bitumen/filler Mixtures. In *Proceedings of the 1st Eurasphalt and Eurobitume Congress*, paper 4.079. Brussels, Belgium: Eurobitume and European Asphalt Pavement Association (EAPA).

Zhu, J., Birgisson, B., & Kringos, N. (2014). Polymer Modification of bitumen: Advances and challenges. *European Polymer Journal*, *54*(1).

Chapter 6
Mastics and Mortars

6.1 Introduction

From a practical standpoint, agencies responsible for pavement construction and maintenance typically control the quality of the asphalt mixture by enforcing tests and specifications for the two main ingredients (asphalt binder and aggregate) as well as the composite (asphalt mixture). Chapters 2 and 3 were focused on the two main ingredients in an asphalt mixture: asphalt binder and aggregates. The subsequent chapter, Chap. 7, will focus on the optimal proportioning of these two ingredients (and air voids) to produce a durable asphalt mixture. The focus of this chapter is to discuss the properties and relevance of the mastic and mortar (also referred to as the fine aggregate matrix or FAM for short).

Mastics and mortars can be considered as materials at the two intermediate length scales between the binder and the asphalt mixture. Recall that in general, a composite has typically two components: matrix and inclusions. The matrix is usually the continuous or interconnected component that binds discrete particles of the inclusion. Based on this generally accepted definition, in the context of an asphalt mixture, asphalt binder is the matrix and aggregate particles are the inclusion. Assuming all particles are fully coated with the asphalt binder (which is generally the case for a good mix), the asphalt binder would form a continuous matrix that binds discrete aggregate particles with each other through a film of the binder in between them.

One approach to fully understand the behavior of asphalt mixtures would be to treat it as a two-component composite comprising binder and aggregate particles. Strictly speaking, air voids must also be regarded as the third component. However, aggregate particles vary in size that span over three orders of magnitude. Therefore, it would be extremely complex and challenging task to understand the behavior of asphalt mixtures as a composite with aggregate particles that vary from micrometers (e.g., fines or filler) to millimeters (e.g., coarse aggregate particles). Such a task would require integrating different dominant particle-to-particle and

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_6

binder-to-particle interaction mechanisms at different length scales (discussed later in more detail). For this reason, in many scenarios it is more practical to evaluate mastics and mortars at two separate intermediate length scales that fall between the asphalt binder and full asphalt mixture.

Mastics are generally defined as a mixture of asphalt binder and aggregate filler or fines, i.e., aggregate particles passing the number 200 sieve and typically finer than 75 μms in size. Mortars, also referred to as fine aggregate matrix or FAM, are generally defined as a mixture of asphalt binder and fine aggregates. In this case, fine aggregates are typically (but not always) defined as aggregates of all sizes that pass the number 16 sieve or are finer than 1.18 mm. In other words, mastic can be considered as a composite that comprises the binder as the matrix and aggregate fines (passing number 200 sieve) as the filler particles. At a slightly higher length scale, one can interpret mortar or FAM as a composite that comprises the mastic as the matrix and fine aggregate particles as the inclusion. Finally, one can treat the asphalt mixture as a composite that comprises the mortar or FAM as the matrix and the coarse aggregate particles as the inclusion. Note that from a practical standpoint, as of this writing, most agencies do not enforce or require specifications that are based on test methods or evaluation of the mastics or mortars. This is simply because contractors producing an asphalt mixture purchase asphalt binder and aggregates to produce a mixture, whereas mastics or mortars are neither procured nor produced in these isolated forms for pavement construction purposes.

Despite the fact that mastics and mortars do not exist in isolation in practice, it is impossible to develop a better understanding of the behavior of asphalt mixtures without fully appreciating the mechanisms of load transfer and distress resistance (or evolution) at these length scales. For example, mastics and mortars can be used to evaluate the performance of new materials, efficiently design modifiers and additives, better understand the mechanisms of failure, and predict the performance of full asphalt mixtures. For all of these reasons, it is important to understand and evaluate the behavior of these two intermediate scaled composites. This chapter will discuss the behavior of mastics and mortars and their relevance to the overall mixture performance.

6.2 Mastics

As mentioned before, mastic refers to asphalt binder combined with aggregate fillers or fines (passing number 200 or 75 μm size sieve). The significance of aggregate fillers to increase the stiffness of the asphalt mixture and influence the overall performance of the mix was recognized as early as the 1940s (Rigden 1947). Mineral fillers interact with the binder in two different ways. First, since filler particles are significantly stiffer than the binder, they tend to increase the mastic stiffness. Second, the specific surface area of mineral fillers (i.e., area per unit mass of the material) is much higher due to the extremely small size of the filler particles. As a result, several physicochemical interactions are possible at the filler–binder

interface that dominate the overall behavior of the asphalt mixture. In fact, certain types of fillers can result in significant changes to the binder chemistry; such fillers are referred to as "active fillers" (e.g., hydrated lime) and were discussed in more detail in Chap. 5 of this book.

6.2.1 Mechanical Role of Filler Particles in Mastic

Intuitively, it is evident that increasing the ratio of filler particles to asphalt binder will result in a significant increase in the stiffness of the composite. However, an optimal level of physical stiffening is desirable to ensure that the resulting mix is both durable and resistant to typical distresses such as rutting and cracking. The concept of physical stiffening of the mastic due to the filler particles is best explained on the basis of Rigden air voids (Rigden 1947). *Rigden air voids* refer to the volume of air voids that remain after compacting the mineral filler. Rigden air voids are determined by compacting a sample of dry aggregate fines to their maximum density.

Rigden (1947) proposed that when asphalt binder is mixed with fillers, a portion of this binder is consumed to fill up these air voids and is referred to as the *fixed binder*. The binder that remains after filling these voids is referred to as the *free binder*. Figure 6.1 illustrates this concept in the form of a phase diagram and also shows a typical device used to measure the Rigden air voids.

Based on the concept of fixed and free asphalt binder, it is evident that mastic with binder volume that is less than the fixed volume of asphalt would be dry, stiff, and prone to degradation. Similarly, mastic with a relatively much higher volume of free asphalt would be more workable and provide more binder for coating the larger aggregate particles in the mix. Of course, an increase in the free volume of the

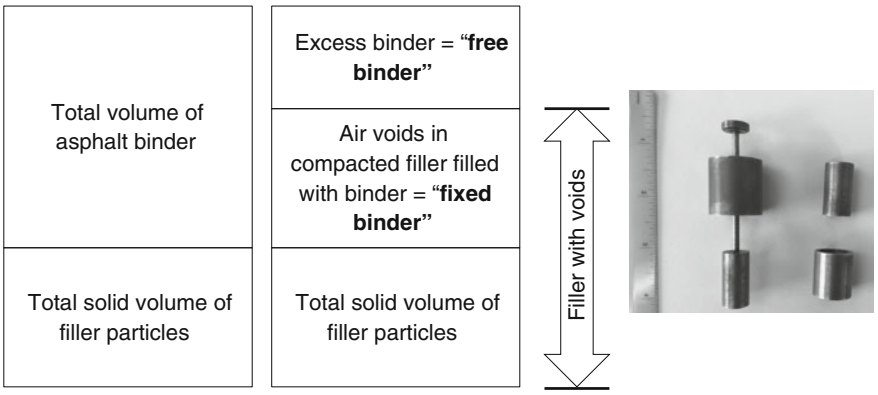


Fig. 6.1 A phase diagram showing the concept of fixed and free volume (*left*); a typical brass mold and a drop hammer used to determine Rigden air voids (*right*)

binder beyond an optimum will result in an overall reduction in the composite stiffness. Conversely, one could also visualize the importance of the physical properties of the filler on the fixed and free asphalt binder. For the same overall asphalt content, filler particles that result in higher Rigden voids will reduce the free binder and adversely affect the workability and durability of the mix and vice versa. In fact, earlier studies have shown that when finer filler particles are used in a mix, the mixture properties are very sensitive to changes in the binder content (Anderson et al. 1982).

One example of a case study where Rigden air voids were used to explain improved workability was in the case of a particulate additive used to produce warm mix asphalt. In this case, the particulate additive was added to the mix with notable improvements in mixture workability and compaction at reduced temperatures (Arega and Bhasin 2012). A more detailed investigation revealed that the fine particles of the additive that were added to the mix resulted in reduced Rigden voids. As a result, the proportion of free asphalt available for this mix was greater than the control mix and consequently resulted in improved workability even when the production temperatures were slightly lowered (Arega and Bhasin 2012). In summary, Rigden voids (and concomitant fixed and free asphalt) are a simple concept that can provide significant insights into mixture proportioning and understanding the filler–binder interaction in a mix.

Several studies over the past few decades have employed the concept of Rigden voids or something similar to better understand the role of filler particles in mastic and mixture performance. The goal of most of these studies was to develop an effective method to measure the stiffening effect of the filler particles and incorporate limits on the increase in stiffness during the mixture design process. A summary of findings and recommendations from some of these studies is discussed here.

Common methods to measure the increase in stiffness or stiffening ratio of the binder due to addition of filler are based on measurement of Rigden air voids in fillers, increase in viscosity, increase in temperature of the ring and ball test, and decrease in the penetration index. Other rheological tests (e.g., DSR and BBR as discussed in Chap. 2) can also be used to quantify the stiffening ratio of asphalt binders due to the addition of fillers (Cooley et al. 1998; Rayner and Rowe 2004). Irrespective of the method of measurement, almost all studies have consistently demonstrated that the stiffening effect and other mechanical properties of the mastic are dependent on the type and concentration of fillers. Particularly, the performance of asphalt mastics at high temperatures is influenced by both the concentration and type of mineral filler, whereas the performance of asphalt mastics at low temperatures is generally influenced by the concentration of the mineral filler (Brown et al. 1997).

Another interesting observation from some of the previous studies is that while the stiffness of the mastic increases with an increase in the concentration of the filler, in most cases the phase angle of the mastic was relatively less sensitive to the concentration of filler particles (Cooley et al. 1998). This has been demonstrated based on experimental measurements conducted by the Federal Highway

Administration (Shashidhar and Shenoy 2002) as well as analytical and computational models of asphalt mastics (Abbas et al. 2005; Buttlar et al. 1999; Shashidhar and Shenoy 2002; Dukatz and Anderson 1980). However, it must also be noted that these conclusions are typically true for fillers that are mostly inert and not active fillers. Active fillers, such as hydrated lime, that interact chemically or have strong physicochemical interactions with the binder can show behaviors that do not follow this trend. For example, one study has shown that the phase angle of mastics was very sensitive to the filler concentration when hydrated lime was used as partial replacement for some of the inert filler (Kim et al. 2004). The influence of active fillers on the properties of the mastic will be discussed in subsequent subsections with a more detailed discussion of mechanisms in Chap. 5.

There are a number of different models that are available for use to theoretically estimate the stiffening effect of filler particles in the mastic. A number of these models are derivatives of the Einstein's model for particles in a diluted solution. One particular example is the Marion-Pierce approach that has been used by several researchers (Heukelom 1965; Lesueur and Little 1999; Faheem and Bahia 2010). According to this model, the property of the mastic K_{mastic} can be estimated based on the property of the binder K_{binder}, volume fraction of the filler ϕ, and the maximum packing fraction of the dry filler based on the Rigden voids ϕ_m as follows:

$$\mathrm{K_{mastic}} = K_{binder}\left(1 - \frac{\phi}{\phi_m}\right)^{-2} \tag{6.1}$$

In Eq. 6.1, the property of interest K_{mastic} could be the viscosity of the mastic or the complex modulus. However, it must be noted that the above model and other similar models are only accurate at lower volume concentrations of the filler (Lesueur and Little 1999). At higher volume concentrations, the interaction between the filler particles is more dominant and not accounted for by these models resulting in an underestimated prediction.

In addition to the stiffening effect of fillers, several studies have evaluated performance of the mastic in terms of its fatigue cracking or fracture resistance and correlated it to the performance of the mixture. Examples include the use of dynamic mechanical analysis (DMA) to measure fatigue cracking and plastic deformation (Y. Kim et al. 2003; Airey et al. 2007) and the use of double edge notched test or DENT to measure essential work of fracture (Andriescu et al. 2006). These studies have shown that the tensile strength of the mastic typically increases with an increase in volumetric concentration of the filler up to approximately 55% (Cooley et al. 1998; Huschek and Angst 1980; Kandhal 1981). The primary mechanism for this increase in the tensile strength and fracture resistance has been shown to be crack pinning by the filler particles in the asphalt mastic (Smith and Hesp 2000). In simple terms, mechanisms such as crack deflection or crack pinning cause a growing crack front or tip to be deflected to either follow a more tortuous or energy intensive path or to be arrested, respectively.

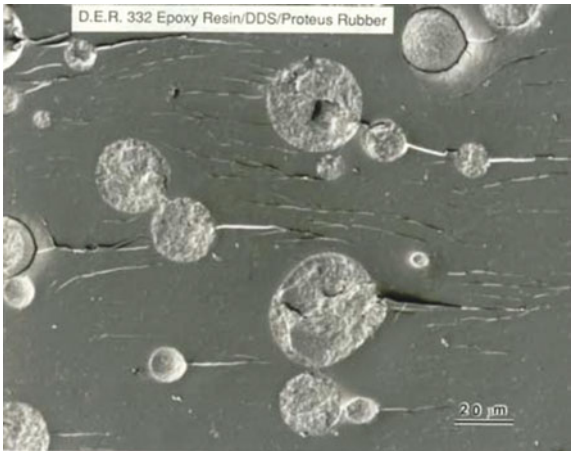


Fig. 6.2 A schematic showing the pinning and deflection of crack fronts in a soft matrix such as asphalt binder as it approaches the rigid inclusion particles (Reproduced from Polymer Science and Engineering, Sue (1991), with permission from John Wiley and Sons)

Figure 6.2 illustrates a photograph of the crack pinning mechanism in a soft and weaker matrix with rigid and stronger particle inclusions analogous to the mastic. Figure 6.3 illustrates that in most cases, the measured fracture toughness of various mastics lies within the theoretical boundaries predicted by the crack pinning mechanism for fillers that interact with the binder and fillers that do not interact with the binder (Smith and Hesp 2000).

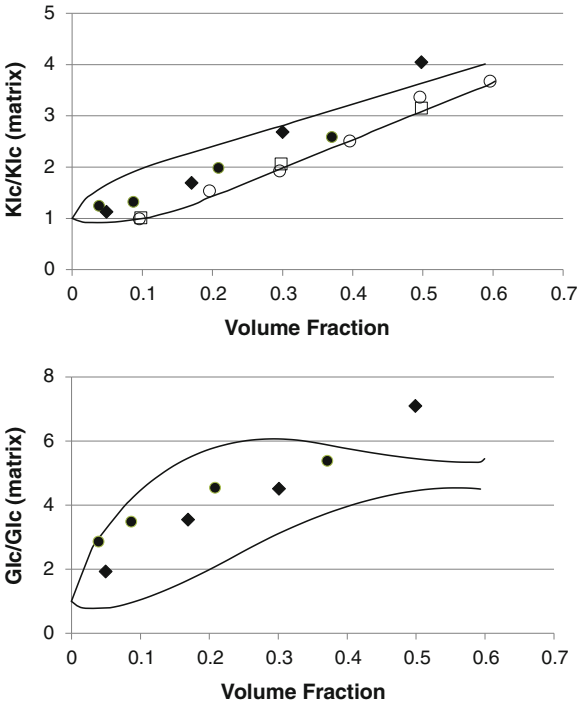


Fig. 6.3 Effect of the filler volume fraction on fracture toughness and energy. *Solid lines* are theoretical limits for non interacting fillers and interacting fillers. Glass spheres used-4 mm (*solid circles*); 114 mm (*solid diamonds*); limestone fillers used-coarse (*open circles*), fine (*open squares*) (adapted based on data from Rodriguez et al. 1996)

In summary, the increase in stiffness and fracture properties of the mastic due to the addition of filler particles can largely (but not completely) be explained based on the concepts of Rigden voids (fixed and free asphalt) and fracture toughening mechanisms such as crack deflection and crack pinning. From the discussion above, it is also clear that there is an optimum binder to filler ratio that will result in a cohesive mix with good workability and resistance to both rutting or permanent deformation and cracking. Binder proportions with low volume fractions of the free asphalt will result in workability and durability issues. Binder proportions with higher volume fractions of the free asphalt (low volume fraction of the particles) will compromise the stiffness of the composite resulting in issues related to permanent deformation and reduced fracture resistance.

6.2.2 *Physicochemical Interactions of Filler Particles in Mastic*

In the previous section, we discussed the role of filler particles to mechanically increase the stiffness and fracture resistance of the mastic. Several studies were able to demonstrate a good correlation between the physical properties such as Rigden voids and volume concentration of the filler particles to the stiffness of the mastic (Anderson et al. 1982; Kandhal 1981; Harris and Stuart 1995). Other studies further refined these correlations to include factors such as coefficient of uniformity, fineness modulus, and specific surface area of the filler particles (Harris and Stuart 1995). However, a few studies have also demonstrated that in some cases the physical properties of the fillers were not adequate to explain the stiffening effect of the filler in the mastic (Craus et al. 1978) or in some cases the improved fracture or fatigue cracking resistance of the mastic (Y. Kim et al. 2003). This was particularly true for active fillers (e.g., hydrated lime, certain types of slag, fly ash) that chemically or physico-chemically interacted with the asphalt binder. Such interactions are dependent on the nature of the filler particles as well as the chemical composition of the asphalt binder.

As an example, one study quantified the effect of such interactions using the Marion-Pierce model as shown in Eq. 6.1 (Lesueur and Little 1999). In this study, the maximum packing fraction measured on the dry filler ϕ_m was not determined based on Rigden voids but was rather back-calculated based on measured properties to quantify the effect of the filler. For example, for hydrated lime with asphalt binder AAM it was reported that $\phi_m = 0.63$. This value was typical for most asphalt binders. However, with another asphalt binder AAD this value was reported as $\phi_m = 0.20$ indicating that hydrated lime acts as a chemically active filler with AAD and essentially is three times as effective as a filler compared to AAM. This study also found that at lower temperatures, the stiffening effect of hydrated lime in asphalt binders AAD and AAM was similar. In other words, although hydrated lime behaved as active filler in asphalt AAD at high temperatures, it behaved as a

traditional inert filler in this binder at low temperatures. This filler-temperature effect was verified using a number of asphalt filler systems. Note that the binders AAM and AAD used in the above-mentioned study were also used in several other studies that helped the development of the Superpave performance grade (PG) binder specification. Readers interested in finding out more about these and other binders used in the SHRP can refer to the catalog report by Jones et al. (1993).

Physicochemical and chemical interactions between the filler particles (particularly active filler particles) and the binder have also been shown to influence the aging characteristics of the asphalt binder and moisture damage resistance of the mastic. Studies have evaluated the influence of different binder-filler pairs on the short-term and long-term aging characteristics of using unmodified and polymer modified asphalt binders (Plancher et al. 1976; Verhasselt 2004). These studies have demonstrated that the aging susceptibility was highly dependent on the nature of the filler and the asphalt binder. To illustrate the impact of these interactions consider the results from a study conducted by Plancher et al. (1976). Plancher et al. evaluated the aging susceptibility of different binders after allowing them to interact with fine particles of different mineral aggregates as well as active fillers (in their case hydrated lime). They quantified aging susceptibility in terms of a hardening index, which was the ratio of the viscosity of the binder after aging to the viscosity of the binder before aging. Results for 16 different pairs of binders and mineral fines with and without an active filler (hydrated lime) are shown in Table 6.1. Results from this table clearly show that different binders have different aging characteristics depending on the type of filler. For example, Binder C was extremely sensitive to interactions with the minerals and demonstrated aging indices that varied from 69 to 195. This variation is attributed to the differences in the physicochemical and chemical interactions between the filler particles and the binder. Such interactions are significantly exaggerated with active fillers such as hydrated lime. In this example, addition of hydrated lime reduced the aging index by 29–75% depending on the type of binder and mineral fines in the mix. The hypothesized mechanism for this reduction was that hydrated lime interacts with some of the polar functional groups that form during aging and disrupt the viscosity-building associations between molecules (see previous Chap. 5 for more details).

In addition to aging, the interactions between the binder and the filler particles are also responsible for determining the resistance of the mastic and the mixture as a whole to moisture-induced damage. For example, the physicochemical interactions between the mineral filler and the binder are responsible for the interfacial adhesion between these two components as well as the resistance of these interfacial bonds to moisture-induced damage. Studies based on binder chemistry have identified specific functional groups in the asphalt binder that strongly adsorb onto mineral aggregate surfaces to promote adhesion. For example, nitrogen-based functional groups (e.g., amines) from the binder have been shown to adsorb strongly to mineral surfaces and form moisture-resistant bonds (Petersen and Plancher 1982). Some other functional groups from the binder such as carboxylic acids have also been shown to adsorb strongly and adhere to mineral surfaces; however, these bonds have also been shown to be less resistant to debonding in the presence of

Table 6.1 Influence of binder and mineral interaction on aging of the binder

Binder	Mineral	Hardening index		
		Untreated	With hydrated lime filler	Percent reduction
Binder A	Limestone A	37	17	54
	Limestone B	31	12	61
	Quartzite	42	19	55
	Granite	41	21	49
Binder B	Limestone A	28	9	68
	Limestone B	20	6	70
	Quartzite	28	10	64
	Granite	31	14	55
Binder C	Limestone A	129	43	67
	Limestone B	69	17	75
	Quartzite	135	44	67
	Granite	195	38	81
Binder D	Limestone A	39	21	46
	Limestone B	23	17	26
	Quartzite	25	17	32
	Granite	32	16	50

moisture (Petersen and Plancher 1982). Some other studies have evaluated the moisture damage resistance of the binder-mineral bond based on the principles of physical chemistry and thermodynamics of the interface. This approach typically entails the measurement of surface free energy (also referred to as surface tension) of different asphalt binders and minerals. The surface energy values of these materials, along with the surface free energy of water, are then used to compute the work of adhesion between the binder and the mineral as well as the energy to promote debonding between these two materials in the presence of water. This approach can be used to identify the binder-mineral pair that is inherently most resistant to moisture-induced damage. Additional details on this approach are discussed in Chap. 4.

Note that other than debonding, in certain cases moisture damage may be concentrated in the mastic due to inherent deficiencies in the filler itself. For example, presence of excess clay or plastic fines in the filler material can render the mastic inherently susceptible to moisture damage. In summary, the physicochemical and chemical interactions between the filler particles and the binder can significantly affect the properties of the mastic in terms of its stiffness, aging characteristics, and resistance to distresses such as cracking and moisture-induced damage. It must also be noted that although the mechanisms related to the interactions between the filler and binder that influence stiffness, aging, and moisture damage resistance cited in this subsection were based on binders and mineral fillers in a mastic, these same mechanisms are also applicable for interactions between the binder and the coarse aggregates in the mixture. The main difference is that the

mastic incorporates fine aggregate particles with very high specific surface areas that exacerbate the effects due to the filler-binder interactions. Such interactions are particularly important in the case of active fillers such as hydrated lime and were discussed earlier in Chap. 5 (Little and Petersen 2005; Lesueur and Little 1999).

6.2.3 Considerations During Mixture Design

While the mechanisms that influence mastic and mixture properties due to the mechanical and physicochemical interactions between the binder and filler particles are well understood in the literature, putting these ideas into practice in the form of standard methods to design asphalt mixtures has been a challenge. This is primarily because of two reasons.

First, there are a variety of fillers with different physical and chemical characteristics that are encountered in practice. As a result, it is sometimes cost prohibitive to evaluate these characteristics each time a mixture design is to be carried out. Second, different types of asphalt mixtures have different sensitivities to the properties of the asphalt mastic. For example, one study has reported that addition of 2% excess fines to an already fine-graded mixture without changing the optimum asphalt content resulted in increased rutting but no change in fatigue cracking as observed in the field. Another study has reported that coarse-graded mixtures are more sensitive to small variations in filler content compared to fine-graded mixtures (Sebaaly et al. 2004). For stone matrix asphalt (SMA), a stiff binder-filler matrix is generally required to support the high stress concentrations that occur at stone-to-stone contact points within the mixture. It is common to use higher filler to aggregate ratios for SMA mixtures. However, excess fines may reduce the workability and result in a crack susceptible mixture.

In terms of current practice, the Superpave mix design method (from SHRP) recommends specification limits based purely on volumetrics (use of 0.6 to 1.2 filler-aggregate ratio). Many European specifications are based on both volumetrics and measurement of the stiffening ratio of the mastic due to addition of the filler. For example, the current European standard requires measurement of Rigden voids (values to be between 28 and 55%) (EN 1097-4) and measurement of stiffening as the increase in ring and ball test temperature of the mastic with respect to the binder (values to be between 8 and 25 °C) (EN 13179-1). Another example is the German filler test based on the amount of filler required to form a noncohesive mass when thoroughly mixed with a standardized amount and type of oil.

A more recent study supported by the National Cooperative Highway Research Program (NCHRP) revisited the recommendations for incorporating mastic properties while designing asphalt mixtures (reference). Some of the key conclusions from this study are summarized below.

The NCHRP study demonstrated that Rigden voids and viscosity of the mastic have a strong influence on the mixture workability (when all other factors are constant). Consequently, this study also proposed tentative limits for mastic

viscosity to ensure that the mixture workability was not compromised. In order to enforce this requirement, one would have to directly measure the mastic viscosity as an additional test or measure the Rigden voids for the filler and use a regression model that relates mastic viscosity to the Rigden voids and binder viscosity. The aforementioned study also demonstrated that the mastic properties influenced performance of mixtures in terms of resistance to rutting, fatigue cracking, low-temperature cracking and moisture-induced damage. However, in the case of mixture performance, several other factors in addition to filler-binder interactions were found to be significant. For example, in the case of rutting resistance both the coarse aggregate gradation and the type of binder play a more important role in the mixture behavior.

In summary, mastics or fillers are not very often evaluated and used during the mix design process. This is due to the fact that (1) mastics are never used in isolation in pavement applications and (2) testing and screening mastics during the mixture design process would entail an additional resource consuming design step. However, it is equally important to remember that evaluating mastics in isolation are an extremely important tool when evaluating active fillers or other innovative materials that may be considered for incorporation in asphalt mixtures.

6.3 Mortars or Fine Aggregate Matrix

Asphalt mortars or fine aggregate matrix (FAM) are generally regarded as a mix of asphalt binder and fine aggregate particles, typically finer than 1.18 mm (or passing the #16 sieve size). As discussed earlier, FAM can be treated as a homogenous matrix (hence the name) that binds the coarse aggregate particles together in an asphalt mixture. As in the case of asphalt mastics, asphalt mortars do not exist in isolation in a pavement structure but only as a part of an asphalt mixture. The only exception to this is perhaps extremely fine mixes that approach the particle size and gradation similar to that of the mortar or FAM. Consequently, FAM mixes are designed and studied to better understand the mechanisms of damage evolution in asphalt mixtures. This section will discuss two specific aspects of FAM mixes: (1) The methods that are used to design a FAM mix for evaluation and testing and (2) examples of case studies, wherein FAM was used as an evaluation tool and possibilities for similar use in the future.

6.3.1 Applications of Fine Aggregate Matrix

Some of the reasons that make it more efficient to evaluate FAM specimens are as follows: First, *the resources (time and manpower) required to evaluate FAM mixes are significantly lower than the resources required to test full asphalt mixtures; as a*

result for specific problem types (discussed later,) it is more efficient to evaluate FAM compared to full asphalt mixtures.

To explain this further, consider the fact that FAM mixes are usually tested in the form of a cylindrical test specimen that is typically 50 mm in height and 10–12 mm in diameter. Note that the specimen diameter is typically 8–10 times the maximum particle size, which is typically 1.18 mm. This allows for the test specimen to serve as a representative volume of the FAM mix. In other words, the specimen to particle ratio is large enough to allow the treatment of the FAM as a homogenous medium or continuum, wherein the properties of the individual particles, interfaces, and interparticle layers are not of interest; rather the focus is on the behavior of the composite as a whole. The smaller test specimen requires much less time and effort to fabricate compared to a full asphalt mixture. Also, the smaller test specimen allows the FAM to be evaluated using a smaller loading frame or testing device compared to a full asphalt mixture. In fact, most studies involved with FAM have been conducted using a dynamic shear rheometer (DSR) very similar to the one used to evaluate asphalt binders. Figure 6.4 illustrates a typical FAM test specimen along with the setup used to evaluate such specimens using a dynamic shear rheometer.

The four major forms of distress in asphalt materials in a flexible pavement are rutting, fatigue cracking, low-temperature cracking, and moisture-induced damage. Of these, distresses such as fatigue cracking, low-temperature cracking, and moisture-induced damage are mostly concentrated in and dictated by the mortar portion of the asphalt mixture. For example, rutting or resistance to permanent deformation is strongly influenced by the coarse aggregate particles (e.g., particle-to-particle contact, interlocking, etc.). Distresses, such as cracking, primarily originate and grow in the FAM portion of the asphalt mixtures. In this case, the coarse aggregates play a role in dictating local stress distribution in the matrix or

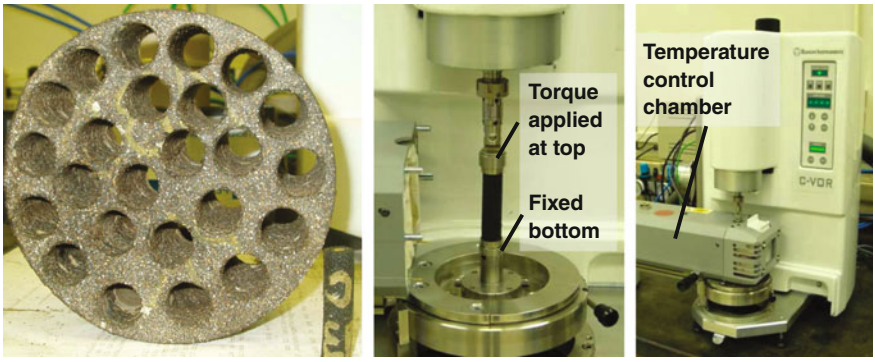


Fig. 6.4 A typical FAM test specimen (extreme *left corner*) cored out of a larger six inch diameter specimen compacted using a gyratory compactor; the specimen is typically evaluated using a dynamic mechanical analyzer (*middle*) with temperature controls (With permission from TRB through the National Academy of Sciences, (Little et al. 2006)

FAM, but for a given scenario or coarse aggregate type and distribution, it is ultimately the FAM portion of the mix that will resist crack nucleation and growth.

Similarly, owing to the high specific surface area of the smaller sized particles and the presence of the binder in the FAM, mechanisms related to moisture-induced damage such as softening of the binder and loss of interfacial adhesion (or stripping) are also concentrated in the FAM portion of the mix. Consequently, *the second advantage of using FAM as a test material is that it allows the investigation of distresses such as fatigue cracking and moisture-induced damage at a much higher resolution.* In other words, by investigating FAM it is possible to amplify and investigate the evolution of these distresses at higher sensitivity as opposed to using a full asphalt mixture.

There are several computational models that are focused on modeling and predicting the behavior of the asphalt mixture as a composite. In most of these models, the material geometry and property are required as an input (see Chap. 15). For a full asphalt mixture, geometry would entail details pertaining to the size and spatial distribution of individual aggregate particles, asphalt binder film between such particles and air voids. As discussed previously, aggregate particle dimensions can vary over orders of magnitude (75 μm–25 mm) in a typical mix. This makes it computationally untenable to recreate all the individual particles in a large representative volume of the mix. However, the problem can be alleviated by treating the asphalt mixture as a composite comprising of only coarse aggregate particles bound together by a homogenous matrix, i.e., the FAM. The third main advantage of evaluating FAM is that its properties can be used as an input to develop computational models to predict the behavior of full asphalt mixtures.

The following paragraphs discuss some of the case studies from the literature, where FAM was employed as a material of choice owing to one or more of the above reasons. The subsequent subsection discusses the methods that can be used to design FAM mixes and fabricate test specimens.

One of the earlier studies that involved the use of an asphalt mortar was to evaluate the fatigue cracking resistance of asphalt binders in the form of a composite (Y.R. Kim et al. 2003). In this case, the mortar was a mix of the asphalt binder and Ottawa sand. Although Ottawa sand is naturally mined, it is comprised of rounded particles of almost pure quartz that are very uniform in size and consistency. Owing to these properties, Ottawa sand is commonly used in laboratory experiments with hydraulic cement. In the study cited above, the goal was to evaluate the properties of the asphalt binder with this relatively standard form of natural sand.

In a subsequent study (Little et al. 2006), the focus was to evaluate the relationship between the surface free energy of asphalt binders and aggregates, and concomitant thermodynamic parameters such as work of adhesion and debonding to the moisture damage resistance, cracking, and self-healing properties of asphalt composites. In this study, it was necessary to experimentally measure the moisture damage resistance, fatigue cracking resistance, and self-healing properties of several different combinations of asphalt binders and aggregates along with their surface free energies. Since the goal was to isolate the role of material properties, specifically surface free energies, while all other factors remained constant (e.g.,

gradation, binder content, etc.), the researchers used mortars to evaluate the aforementioned performance characteristics of different mixes. Note that unlike the previous study (Y.R. Kim et al. 2003), in the latter study (Little et al. 2006) the interaction of the binder with the mineral aggregate was important and as such the Ottawa sand could not be used. Consequently, the latter modified the previous approach and produced mortars or FAM mixes using fine aggregates from typical full asphalt mixtures in lieu of the Ottawa sand. Owing to the efficiency of testing FAM specimens, researchers were able to conduct performance tests on several mixes to demonstrate that surface free energy played an important role in dictating the resistance of asphalt mixes to moisture-induced damage and also its self-healing properties.

FAM mixes have also been used as the material (and length scale) of choice to solve several other interesting problems, some of which are discussed below. Most of these applications can be classified into three broad categories:

1. As a tool by which to evaluate the influence of material properties and factors (e.g., additives, binder aging, mineral characteristics) on performance characteristics.
2. As the matrix to develop composite computational models for full asphalt mixtures.
3. As an efficient medium by which to develop analytical methods to evaluate composite performance, which can then be extended to full asphalt mixtures.

The following paragraphs present a few examples from each one of the above three categories of applications.

6.3.1.1 Use of FAM to Evaluate Influence of Material Properties and Other Factors on Performance

Several studies have used FAM as the length scale and material of choice to evaluate the influence of material properties or other factors on the performance of asphalt composites. Three examples are presented here to demonstrate the use of FAM to answer specific performance-related questions.

One of the studies involving FAM was to *quantify the impact of granite and limestone fillers on moisture-induced damage and also to find effective solutions to reduce the impact of granite fillers on moisture-induced damage* by either blending such fillers with limestone fillers and/or hydrated lime (Caro et al. 2008). In this study, researchers used FAM mixes with different combinations of fillers to quantify the impact of moisture-induced damage by conducting fatigue tests on dry and moisture conditioned FAM test specimens. Although FAM does not represent a full asphalt mixture, it does contain fine aggregates and fillers with the same mineralogy as the full asphalt mixture. Also fine aggregate particles and fillers have a much higher specific surface area compared to coarse aggregate particles and therefore represent a section of the mix, wherein the damage is highly concentrated.

More importantly, the use of FAM allowed the researchers to conduct tests on several replicates (with much less effort compared to testing full asphalt mixtures), which in turn improved confidence in the final test results.

In another similar study, researchers were *investigating the most efficient pairing of different asphalt binders (from different refineries and crude sources) with different mineral aggregates* (Bhasin et al. 2006). In this case, researchers were able to test FAM mixes using different combinations of mineral aggregates and asphalt binders to demonstrate the role of material properties such as surface free energy and aggregate angularity on the moisture sensitivity of asphalt composites. They were able to benchmark and validate their results against known field performance of several mixes in the field. As before, the use of FAM mixes allowed the researchers to efficiently evaluate a number of different mixes with significantly fewer resources.

As the third example, FAM was used to evaluate the impact of warm mix additives and long-term oxidative aging on the fatigue cracking resistance of asphalt composites (Arega et al. 2013). In this case, researchers were interested in evaluating two different binders with five different warm mix additives before and after long-term oxidative aging. This test matrix would entail running fatigue tests on 24 different mixes including control specimens. However, since the question of interest was the relative fatigue cracking resistance of these mixes before and after aging (all other factors such as aggregate gradation being the same), researchers were able to conduct these tests much more efficiently using FAM mixes in lieu of full asphalt mixtures. Recall that even in a full asphalt mixture, crack nucleation and growth primarily occur in the mortar or FAM portion of the mix. Therefore, the use of FAM mixes was a much more efficient path to answer this research question as opposed to running traditional fatigue tests on full asphalt mixtures. Interestingly, researchers demonstrated that the rank order for the fatigue cracking resistance of different mixes did not change considerably after long-term aging compared to the rank order before long-term aging.

6.3.1.2 Use of FAM to Obtain Properties Useful in Computational Modeling

The second category of applications for FAM mixes is in the area of multi-scale computational modeling. Computational models such as discrete element model or finite element model are used to predict the behavior of the material when subjected to any arbitrary loading condition. Depending on the problem of interest, such models may be developed at different length scales. For example, at the mixture level the problems of interest might include the following:

1. Conducting a parametric analysis on the influence of coarse aggregate particle gradation (or particle angularity) on the stiffness and damage resistance of asphalt mixtures,

2. Conducting a parametric analysis on the influence of air voids distribution on the stiffness of an asphalt mixture, and
3. Investigating the nature and rate of damage evolution in asphalt mixtures.

As discussed earlier, in all of the above studies, it is not feasible to model the precise geometry and spatial distribution of every single aggregate particle (ranging from finer than 75 μm to several millimeters in size) while recreating the structure of the asphalt mixture. Such an effort would require imaging or simulating the spatial distribution of all fine and coarse aggregate particles over a large volume using a very precise method that spans different length scales (micrometers to millimeters). Instead of recreating every single coarse and fine aggregate particle, it is significantly more advantageous to treat the mortar phase of the asphalt mixture (fine aggregate particles and the binder) as a single homogenous matrix that binds coarse aggregate particles together. This approach has been used in several studies to evaluate different issues of interest. For example, Freitas et al. (2006) were interested in modeling the evolution of microcracks within the asphalt mixture to simulate fracture or fatigue damage. To this end, they measured the fracture resistance of FAM using a well-defined geometry and then used that as a property input in a composite model of the full asphalt mixture comprising of coarse aggregates bound together using FAM.

As another example, Caro et al. (2010) conducted a parametric study using computational models to evaluate the influence of air voids distribution on the stiffness and moisture damage resistance of full asphalt mixtures. As before, they recreated the asphalt mixture as a composite of coarse aggregates and FAM, wherein the properties of FAM (e.g., viscoelastic behavior, damage resistance in dry and moisture conditioned states) were measured using simple geometries and used as a property input in the computational model.

The above and other similar studies are extremely useful because they provide valuable insight into mixture behavior using an efficient combination of computational and experimental methods. Figure 6.5 illustrates the typical decomposition of a full asphalt mixture into a matrix (FAM) and coarse aggregate particles used for computational models.

6.3.1.3 Use of FAM to Develop Constitutive Models for Material Behavior

The third and last category of applications for FAM mixes is to develop constitutive models for material behavior. Since FAM mixes present behavior that is very similar to full asphalt mixes (e.g., viscoelastic response and rate or frequency dependent fatigue cracking), they could be used as a more efficient alternative to develop constitutive models for material behavior. For example, Castelo-Branco et al. (2008) used FAM mixes to develop methods to quantify the fatigue cracking resistance of asphalt mixtures. More specifically, they developed a crack growth index that was dependent on the damage resistance of FAM specimens and surface

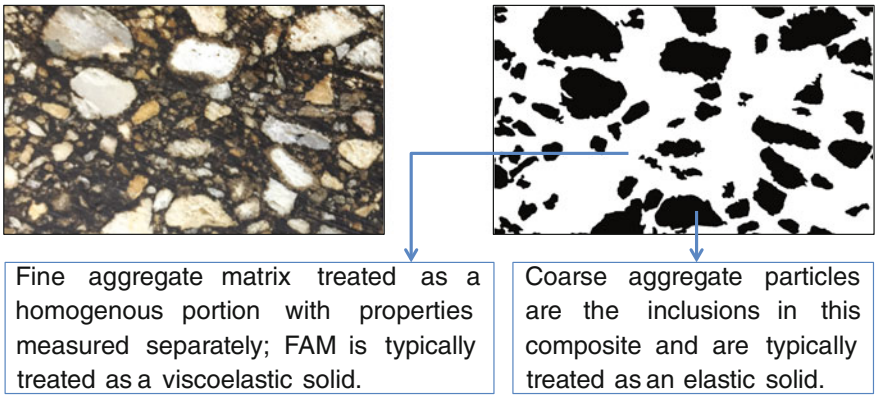


Fig. 6.5 A typical cross section of an asphalt mixture with coarse and fine aggregates (*left*) and a typical reduced geometry (*right*) used for computational modeling wherein the material is reduced to a composite of coarse aggregates (*dark regions*) and a matrix of FAM (*clear space*). Note that particles smaller than approximately 1 mm in size are not visible in the image on the left due limitations of image resolution and the fact that such particles are masked by a film of the asphalt binder during specimen sectioning

free energy of its components materials (i.e., mineral aggregates and binder). In another study, Karki et al. (2015) used FAM mixes to develop an experimental and analytical method to evaluate the fatigue cracking resistance and self-healing characteristics of asphalt composites. In this case, researchers were not only interested in fatigue damage and self-healing of asphalt mortars or FAM but were also interested in developing a method that could be applied to full asphalt mixtures. However, since the development of a method using full asphalt mixtures would require significant time and resources, it was more efficient to use FAM mixes instead and then propagate and apply the findings to full asphalt mixtures with any minor adjustments as needed.

6.3.2 Design of Fine Aggregate Matrix

The previous section demonstrated the different ways in which FAM mixes can be used as a diagnostic tool to evaluate the influence of material-related factors (e.g., additives and aging) on its behavior, or as a medium to develop computational or analytical models that describe the behavior of full asphalt mixtures. For the reader interested in using FAM in a similar or different application, the next natural question would be how to design and fabricate FAM mixes and test specimens, respectively. As far as the mix design is concerned, FAM mixes must be designed based on the application in mind. This section outlines a few of the most commonly used methods to design a FAM mix.

The first step to establish a mix design for FAM is to establish a maximum particle size or the sieve size through which 100% of the particles pass through. In most studies, this is selected to be 1.18 mm (or passing the #16 sieve size). Initially, the motivation for selecting this size was that particles smaller than this size are not typically detected without loss of detail using imaging techniques such as digital scanning of saw cut surfaces or X-ray computed tomography of full asphalt mixtures. Other studies (Underwood and Kim 2011) have suggested that a more rationale approach to select the maximum particle size is based on something similar to the Bailey method that was discussed in Chap. 3. Recall that in the Bailey method, the coarse and fine aggregate fractions are not defined based on an arbitrary cutoff size. Rather coarse aggregates are defined as aggregates that create voids when packed together, and fine aggregates are aggregates that fill these voids. Also, in this case a primary control sieve is defined to differentiate between coarse and fine aggregates, where the size of this sieve is 0.22 times the nominal maximum aggregate size. This approach can also be used to define the maximum aggregate size for the FAM mix.

Once the maximum aggregate size for the FAM mix is established, the gradation of the mix is simply reproduced to replicate the relative proportions of different aggregate sizes from the representative full asphalt mix (Fig. 6.6). The next important question has to do with the binder content.

There are two different approaches that are commonly used to determine the binder content in the FAM:

1. By assuming that all the binder in the mix is a part of the fine aggregate matrix, and the fine aggregate matrix holds coarse aggregate particles together as a composite, or

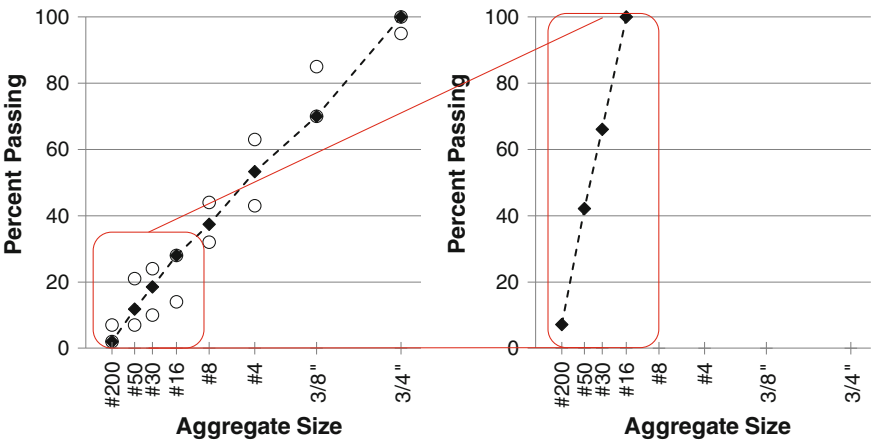


Fig. 6.6 Gradation of a typical full asphalt mixture (*left*) is used to determine the gradation of a FAM mix (*right*) after establishing a maximum aggregate size

2. By assuming that the coarse aggregate particles are coated by a thin film of the pure binder, and such coated coarse aggregate particles are held together by the fine aggregate matrix.

Determining the binder content for the first approach is fairly straightforward. The percentage of binder by weight of the aggregates in the full mix is simply expressed as the percentage of binder by weight of the fine aggregates in the mix. When following this approach, one important correction that must be considered is aggregate absorption, i.e., binder absorbed by coarse aggregates must be discounted in computing the overall binder content. As needed (particularly for absorptive aggregates), this correction can be determined by conducting a simple volumetric analysis of coarse aggregates mixed with sufficient amount of the binder.

Determining the binder content for the second approach requires that the binder film coating the coarse aggregate particles also be discounted from the binder content determined using the previous procedure. This can be done analytically or experimentally.

The analytical approach requires that the user makes some assumptions regarding the surface area of the coarse aggregates and the thickness of the binder film coating the aggregate particles (Branco 2008; Zollinger 2005). The binder content coating coarse aggregate particles can also be determined experimentally. This is typically done by producing a sample of the loose full asphalt mixture, separating the coarse aggregate particles from the rest of the mix, and finally determining the binder content in the coarse aggregate particles by using an ignition oven or other methods.

The choice between the two methods mentioned above is dictated by the objective of the exercise. For example, the former method can be used for studies involving the effect of asphalt binder type, chemical or polymer modifiers, aging, anti-strip agents, etc., on the performance of the composite. Either method can be used when the properties of FAM mixes are being measured for use as an input in a computational model that is specifically developed to comprise coarse aggregate particles held together by the fine aggregate matrix OR coarse aggregate particles with a binder film held together by the fine aggregate matrix.

Finally, the last variable in the production of FAM specimens is air voids. As before, the final objective of using the FAM will dictate the level of compaction used to produce FAM specimens. For studies involving investigation of fracture, fatigue, and healing characteristics as a function of material-related factors (e.g., binder type, modifier, aging), a fully compacted specimen with minimal or no air voids may be appropriate. If such studies were to include assessment of moisture-induced damage by conditioning the test specimens, then the FAM specimens can be compacted to 90–95% of their maximum density, allowing some room for air voids that will facilitate the moisture transport and conditioning process. Finally, if the objective is to measure FAM properties as an input for a computational model (similar to the one illustrated in Fig. 6.5), then the choice of whether or not to incorporate air voids in the FAM test specimens will be dictated by the approach used to create the model. In other words, if the computational

model explicitly includes air voids, FAM, and coarse aggregate particles, then the FAM specimens must be fabricated with minimal or no air voids and then used for measurement of material properties. On the other hand, if the computational model assumes that air voids are distributed within the FAM, then the FAM specimens must be fabricated accordingly for measurement of their properties.

6.4 Summary

In this chapter, we looked at two particular length scales of interest that lie between the component materials (asphalt binder and aggregate particles) and the full asphalt mixture: asphalt mastics and asphalt mortars or Fine Aggregate Matrix (FAM). Although mastics and FAM do not exist by themselves in an asphalt mixture, the evaluation of materials at these length scales offers a very cost- and time-effective pathway to:

1. investigate the effect of mineral aggregate interaction while designing liquid anti-strip agents,
2. design and determine optimal dosage of polymer and/or chemical modifiers, additives, extenders, softening agents etc. for use with different asphalt binders to enhance performance characteristics such as improved resistance to fatigue cracking, fracture, moisture damage resistance and improved healing,
3. determine the effect of long-term aging on the performance characteristics of different binder types while also being able to account for mineral-binder interactions,
4. conduct forensic analysis of material failures, and
5. develop computational mixture models that can simulate the behavior of full asphalt mixtures under a variety of different boundary conditions.

Finally, this chapter also discussed various ways in which such materials can be designed and produced in a laboratory to meet the objectives listed above.

6.5 Exercises

6.1. The table below gives the gradation of an asphalt mixture with an optimum binder content of 5.4% by weight of the mix. Determine the gradation and binder content of a mortar or Fine Aggregate Matrix that is representative of this mixture. Consider that no binder is absorbed in the coarse aggregate particles, and the FAM represents the matrix that binds the coarse aggregate particles (i.e., there is no additional binder between the FAM and the coarse aggregate surface).

Sieve size	Percentage passing
19.5 mm	0
9.5 mm	15.0
No. 4	40.0
No. 8	64.9
No. 30	84.3
No. 50	88.4
No. 200	96.4

6.2. Scenario A: The asphalt mixture can be modeled (e.g., in a computational finite element model) as a three component composite comprising coarse aggregate as inclusion, FAM as the matrix, and air voids as the third component dispersed within the FAM. Scenario B: The asphalt mixture can also be modeled as a two-component composite comprising coarse aggregate as inclusion and FAM as the matrix with air voids being an integral part of the FAM itself. You have been asked to determine the engineering properties of the FAM mix similar to the one described in question 6.1. using laboratory fabricated test specimens. Describe in detail how your specimen fabrication procedure would differ if you were modeling using Scenario A versus Scenario B.

6.3. Filler particles when added to an asphalt mix generally increase the stiffness of the mastic and hence the stiffness of the mix. In some rare cases, it is possible to add filler particles and observe that the stiffness of the mixture has reduced. Explain the possible reasons for this [Hint: Study the concept of Rigden voids in more detail].

References and Additional Reading

Abbas, A., et al. (2005). Modelling asphalt mastic stiffness using discrete element analysis and micromechanics based models. *International Journal of Pavement Engineering, 6*(2), 137–146.

Airey, G. D., et al. (2007). Asphalt mixture moisture damage assessment combined with surface energy characterization. In *Conference on Advanced Characterization of Pavement and Soil Engineering Materials, Athens, Greece*.

Anderson, D. A., Tarris, J. P., & Brock, J. D. (1982). Dust collector fines and their influence on mixture design. *Proceedings Association of Asphalt Paving Technologists, 51,* 363–397.

Andriescu, A., et al. (2006). Validation of the essential work of fracture approach to fatigue grading of asphalt binders. *Proceedings Association of Asphalt Paving Technologists*, 75.

Arega, Z., & Bhasin, A. (2012). *Binder rheology and performance in warm mix asphalt (Part-2)*. TX: Austin.

Arega, Z. A., Bhasin, A., & Kesel, T. De. (2013). Influence of extended aging on the properties of asphalt composites produced using hot and warm mix methods. *Construction and Building Materials, 44,* 168–174.

Bhasin, A., et al. (2006). Limits on adhesive bond energy for improved resistance of hot-mix asphalt to moisture damage. *Transportation Research Record: Journal of the Transportation Research Board, 1970*(1), 3–13.

Branco, V. C. (2008). *A unified method for the analysis of nonlinear viscoelasticity and fatigue damage of asphalt mixtures using the dynamic mechanical analyzer*.

Brown, E. R., et al. (1997). Development of a mixture design procedure for stone matrix asphalt. *Proceedings Association of Asphalt Paving Technologists, 66,* 1–24.

Buttlar, W. G., et al. (1999). Understanding asphalt mastic behavior through micromechanics. *Transportation Research Record: Journal of the Transportation Research Board, 1681,* 157–169.

Caro, S., et al. (2008). Probabilistic analysis of fracture in asphalt mixtures caused by moisture damage. *Transportation Research Record: Journal of the Transportation Research Board, 2057*(1), 28–36.

Caro, S., et al. (2010). Probabilistic modeling of the effect of air voids on the mechanical performance of asphalt mixtures subjected to moisture diffusion. *Journal of the Association of Asphalt Paving Technologists, 79,* 221–252.

Castelo Branco, V. T., et al. (2008). Fatigue analysis of asphalt mixtures independent of mode of loading. *Transportation Research Record: Journal of the Transportation Research Board, 2057*(1), 149–156.

Cooley, L. A., et al. (1998). Characterization of asphalt-filler mortars with superpave binder tests. *Journal of the Association of Asphalt Paving Technologists, 67,* 42–56.

Craus, J., Ishai, I., & Sides, A. (1978). Some physio-chemical aspects of the effect and role of filler in bituminous paving mixtures. *Proceedings Association of Asphalt Paving Technologists, 47,* 558–588.

Dukatz, E. L., & Anderson, D. A. (1980). The effect of various fillers on the mechanical behavior of asphalt and asphaltic concrete. *Journal of the Association of Asphalt Paving Technologists, 49,* 530–549.

Faheem, A., & Bahia, H. U. (2010). *Test methods and specification criteria for mineral filler used in HMA.* Washington, D.C.

Freitas, F. A. C. et al. (2006). *A Theoretical and Experimental Technique to Characterize Fracture in Asphalt Mixtures and Pavements*.

Harris, B. M., & Stuart, K. D. (1995). Analysis of mineral fillers and mastics used in stone matrix asphalt. *Proceedings Association of Asphalt Paving Technologists, 64,* 54–80.

Heukelom, W. (1965). The role of filler in bituminous mixes. *Proceedings Association of Asphalt Paving Technologists*.

Huschek, S., & Angst, C. (1980). Mechanical properties of filler-bitumen mixes at high and low serive temperatures. *Proceedings Association of Asphalt Paving Technologists, 49,* 440–475.

Jones, D. R. (1993). *SHRP materials reference library: Asphalt cements: A concise data compilation,* Report No. SHRP-A-645.

Kandhal, P. S. (1981). Evaluation of baghouse fines in bituminous paving mixtures. *Proceedings Association of Asphalt Paving Technologists, 50,* 150–210.

Karki, P., Li, R., & Bhasin, A. (2015). Quantifying overall damage and healing behaviour of asphalt materials using continuum damage approach. *International Journal of Pavement Engineering, 16*(4), 350–362.

Kim, Y., Little, D., & Song, I. (2003). Effect of mineral fillers on fatigue resistance and fundamental material characteristics mechanistic evaluation. *Transportation Research Record: Journal of the Transportation Research Board,* (03), 1–8. Retrieved May 22, 2014, from http://trb.metapress.com/index/8PP3381256555710.pdf.

Kim, Y. R., Little, D. N., & Lytton, R. L. (2004). Effect of moisture damage on material properties and fatigue resistance of asphalt mixtures. *Transportation Research Record: Journal of the Transportation Research Board, 1891,* 48–54.

Kim, Y. R., Little, D. N., & Lytton, R. L. (2003b). Fatigue and healing characterization of asphalt mixes. *Journal of Materials in Civil Engineering (ASCE), 15*(1), 75–83.

Lesueur, D., & Little, D. N. (1999). Effect of hydrated lime on rheology, fracture, and aging of bitumen. *Transportation Research Record: Journal of the Transportation Research Board, 1661,* 93–105.

Little, D. N., Bhasin, A., & Hefer, A. (2006). Final Report for NCHRP RRD 316: Using Surface Energy Measurements to Select Materials for Asphalt Pavement. Washington D.C.: Transportation Research Board of the National Academies.

Little, D. N., & Petersen, J. C. (2005). Unique effects of hydrated lime on the performance related properties of asphalt cements: Physical and chemical interactions revisited. *Journal of Materials in Civil Engineering (ASCE), 17*(2), 207–218.

Petersen, J., & Plancher, H. (1982). Chemistry of asphalt-aggregate interaction: Relationship with pavement moisture-damage prediction test. *Transportation Research Record: Journal of the Transportation Research Board, 843*, 96–104. Retreived May 5, 2014, from http://trid.trb.org/view.aspx?id=181667.

Plancher, H., Green, E. L., & Petersen, J. C. (1976). Reduction of oxidative hardening of asphalt by treatment with hydrated lime—A mechanistic study. *Proceedings Association of Asphalt Paving Technologists, 45,* 1–24.

Rayner, C., & Rowe, G. M. (2004). Properties of mastics using different fillers with both unmodified and eva-modified binders. *In Euroasphalt and eurobitume congress* (pp. 861–870). Vienna: Austria.

Rigden, P. J. (1947). The use of fillers in bituminous road surfacing mixtures. A study of filler-binder systems in relation to filler characteristics. *Journal of the Society of Chemical Industry,* 299–309.

Rodriguez, M. G., Morrison, G. R., vanLoon, J. R., & Hesp, S. A. (1996). LOW-TEMPERATURE FAILURE IN PARTICULATE-FILLED ASPHALT BINDERS AND ASPHALT CONCRETE MIXES (WITH DISCUSSION). *Journal of the Association of Asphalt Paving Technologists*, 65.

Sebaaly, P. E., et al. (2004). Field and laboratory performance of Superpave mixes in Nevada. *Transportation Research Record: Journal of the Transportation Research Board, 1891,* 76–84.

Shashidhar, N., & Shenoy, A. (2002). On using micromechanical models to describe dynamic mechanical behavior of asphalt mastics. *Mechanics of Materials, 34,* 657–669.

Smith, B. J., & Hesp, S. A. (2000). Crack pinning in asphalt mastic and concrete. *Transportation Research Record: Journal of the Transportation Research Board, 1728,* 75–81.

Sue, H.J. (1991). Study of Rubber-modified Brittle Epoxy Systems. Part II: Toughening Mechanisms Under Mode-I Fracture. 31(4).

Underwood, B. S., & Kim, Y. R. (2011). Experimental investigation into the multiscale behaviour of asphalt concrete. *International Journal of Pavement Engineering, 12*(4), 357–370.

Verhasselt, A. (2004). Short- and long-term aging with RCAT on bituminous mastics. *In Euroasphalt and eurobitume congress* (pp. 1429–1439). Vienna: Austria.

Zollinger, C. J. (2005). *Application of surface energy measurements to evaluate moisture susceptibility of asphalt and aggregate*. College Station, Texas A&M University.

Chapter 7
Asphalt Mixtures

7.1 Introduction

An asphalt mixture is a composite material that comprises mineral aggregates, asphalt binder, and air voids. Although, as discussed earlier, it is sometimes beneficial to treat an asphalt mixture as a composite of coarse aggregates bound together by an asphalt mortar (i.e., the binder, air voids, and fine aggregates are treated as a single component). Asphalt mixtures can be broadly classified into two groups: structural mixes and surface mixes. As the name suggests, structural mixes are mixes that provide structural capacity to the pavement, and their primary function is to effectively transfer loads from the vehicles to the pavement foundation (via the base and subbase). Surface mixes are mixes that are specifically designed to meet surface requirements such as adequate skid resistance, reduced vehicle–pavement noise, and effective water drainage. Generally, surface mixes are placed as a thin or very thin layer on top of an existing structural mix. In many cases, structural mixes also fulfill the role of a surface mix. Recall the discussion on the different types of aggregate gradations in Chap. 3. Aggregate structures that result in a dense-graded or gap-graded mix (the latter is also referred to as stone matrix asphalt or SMA) are typically used as structural mixes. Aggregate structures that result in an open-graded mix (also referred to as open-graded friction course or OGFC) are typically used as a surface mix.

Asphalt mixtures need to be tailored and designed for each specific geographic location and application. Different geographic locations have different climatic conditions, and the type of binder used should be selected such that it is appropriate for the climatic conditions at the construction site. More importantly, the mineral aggregates used in asphalt mixtures can differ significantly from one geographic region to another. As discussed in Chap. 3, since aggregates comprise about 95% by weight of the asphalt mixture, it is economically important to procure such aggregates locally. The two ingredients, i.e., the asphalt binder and the mineral aggregates, are produced by nature and processed by humans for engineering

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_7

applications. As a result, there are significant differences in the chemistry and performance of these materials over time, even when they are procured from the same source.

Even within the same geographic region, the intended use of the asphalt mixture will dictate the design of the asphalt mixture. For example, the requirements and economic constraints to design an asphalt mixture for a major interstate highway are obviously going to be very different from that of a rural road that runs in the vicinity of the highway. On account of these differences, there are no standard asphalt mixture designs that can be adopted universally. Instead, *the current engineering practice is to establish procedures for the design of different kinds of asphalt mixtures*. The expectation is that engineers can use such procedures to design mixtures for specific applications and geographic locations.

The end goal of a mixture design procedure is to determine:

1. the appropriate type of binder (Chap. 2),
2. the aggregate gradation or skeleton structure (Chap. 3), and
3. the optimum binder content (percent binder by weight of the mixture) such that the final mix demonstrates acceptable performance over its service life.

Chapters 2 and 3 have covered in detail the science and practice of selecting asphalt binders and aggregate structures, respectively. The focus of this chapter is on the third step, i.e., to understand the process of estimating the optimum binder content for an asphalt mixture for a given application.

The generic steps to determine the optimum binder content in a typical mixture design procedure are as follows:

1. The first step is to select the binder grade that is appropriate for the specific project. In some cases, it is possible that there is more than one producer that can supply the same binder grade to the specific location. In such cases, it may be of interest to run a comparative mix design for the binder of the same grade from different producers. Recall from Chap. 2 that although two binders may have the same grade, it does not imply that the two binders have a similar chemical makeup or will result in similar long-term performance characteristics. In fact, there is anecdotal evidence from highway agencies that certain mix designs performed much better merely by switching to a binder with a similar grade from a different producer.
2. The second step is to select the type of aggregate skeleton that is to be used for the specific application. Recall from Chap. 3 that several different types of aggregates structures can be created, and there are an infinite number of possibilities to arrive at a gradation that meets typical boundaries established as specification limits. However, techniques such as the Bailey method can help reduce this open-ended problem and provide a systematic way to arrive at a specific gradation (or two) for mixture design purposes.
3. The third step is to prepare aggregate and binder blends with varying mass fractions of the binder or simply binder contents. This step typically starts with an educated "guess" or estimate for the optimum binder content, expressed as

percent binder by weight of the mix. This is followed by preparation of blends that mix the mineral aggregate with the binder at the estimated value as well as 0.5 and 1% points above and below the estimate value. For example, if the estimated value for the optimum binder content (say from prior experience) is 4% binder by weight of the mix, then trial blends are prepared using 3, 3.5, 4, 4.5, and 5% of asphalt binder by weight of the mix.

4. The fourth step is to compact test specimens using the blends prepared in the previous step using some kind of a compaction device.
5. The fifth step is to evaluate the test specimens fabricated in the previous step to determine the optimum asphalt content that will result in a satisfactory mixture performance over the service life of the pavement.

The last two steps are the most crucial in ensuring that the final product used in pavement construction is durable and will result in satisfactory pavement performance. In fact, the last two steps are so important that they have been the subjects of several research studies over the past five decades all over the world. The last step, i.e., a method to evaluate the expected field performance of laboratory-compacted test specimens is still a subject of ongoing research across several institutions globally. Also, perhaps in the near future, computational modeling tools can further facilitate the process of finding an optimal gradation and binder content in steps 3 through 5. The following sections will discuss in detail aspects related to the last two steps.

7.2 Methods to Fabricate Laboratory Specimens

In most cases, an asphalt mixture is designed and validated for field performance in a laboratory environment. The final ratified mixture design or job mix formula (i.e., aggregate gradations and binder content) is then supplied to the contractor producing the asphalt. Therefore, it is absolutely important that the test specimens fabricated in the laboratory during the mixture design process replicate as closely as possible the mixture that is placed in the field not only in terms of the proportions of the ingredients *but also more importantly in terms of the internal structure of the asphalt mixture*. In other words, two mixture specimens that have the same component materials in the exact same proportions can still have significantly different internal structure of aggregate particles and distribution of air voids resulting in vastly different performance characteristics. Therefore, one of the cornerstones of establishing a sound methodology to design an asphalt mixture is to incorporate *a procedure to fabricate test specimens such that the internal structure of the test specimens is as similar as possible to the internal structure of the mixture compacted in the field during construction and service*.

It is also important to highlight here that typically laboratory compaction methods to fabricate test specimens are intended to replicate the internal structure of the asphalt mixture towards the end of its design life. In this sense, the laboratory compaction method not only simulates densification during construction, but also

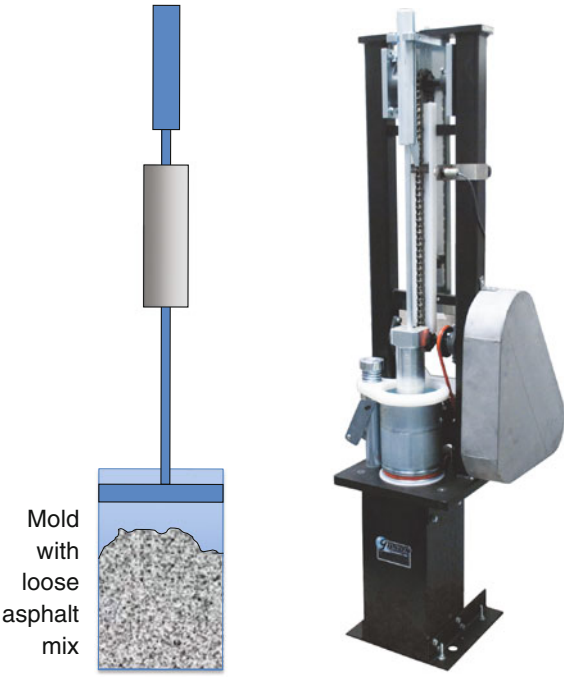


Fig. 7.1 A schematic showing the Marshall compaction method (*left*) and a photograph of automated Marshall compactors (*right*) (Image courtesy of Gilson Company Inc., USA)

densification that occurs during the service life of the pavement following construction. Several different laboratory-based devices have been developed over the decades to accomplish this. This section will present a brief background on some of these methods.

One of the earliest methods to compact asphalt mixture specimens for laboratory design was developed by Bruce Marshall for the Mississippi Highway Department in the USA in late 1930s. The US Army core of engineers began to study and use this device for the design of mixtures for airfield pavements during World War II. Figure 7.1 illustrates a schematic of the device along with a photograph of a Marshall compactor.

In summary, the Marshall compaction method involves the use of a hammer that drops from a fixed height into a mold with the loose asphalt mixture and compacts it. The hammer weighs about 10 lbs or 4.54 kg, and the cylindrical specimen is about 4 in. or 100 mm in diameter. The number of drops or blows typically ranges from 35 to 75 on each side of the specimen and is determined based on the volume of traffic expected during the design life of the pavement. The specimens fabricated using the Marshall compactor are used with the Marshall mix design method to determine the optimum binder content. This method will be briefly described in the following section. The biggest advantage of the Marshall compactor is that it is a relatively simple and a low-cost device. In fact, the manual compactor is nothing more than a guide rod with a weighted hammer designed to fall from a fixed height into a steel mold. On the other hand, the biggest criticism of the Marshall compactor

is that the compaction method is very unrealistic and does not accurately represent the compaction that occurs in the field using a steel wheel roller during construction or under traffic during service. In case of the latter, the material experiences a "kneading" action under the rolling wheel that results in a very different aggregate internal structure.

In the years that followed the introduction of the Marshall compactor, researchers and practitioners realized the need to develop a laboratory compaction method that produced test specimens that more accurately represent field-compacted mixes. At this point, one may argue that the best way to achieve this would be to use a rolling compactor and compact loose asphalt mixture into a slab to produce test specimens. However, such a procedure would be highly impractical to produce several test specimens in a short period of time for the purposes of routine mixture design.

One of the methods to produce laboratory-compacted samples that are similar to field-compacted samples is the kneading compactor. Francis Hveem in the 1920s developed a device to compact specimens that were about 4 in. or 100 mm in diameter using a kneading compactor. The kneading compactor applies a vertical compaction force to the loose mix in a mold via an approximately triangular-shaped foot that covers only a portion of the specimen face. Compaction is achieved by "tamping" the free face of specimen uniformly at different points. This process compacts only a part of the exposed surface of the mix in the mold, and the particles are allowed to move relative to each other creating a kneading action to densify the mix. Different versions of the kneading compactor have been developed based on this principle. One slightly different variation is the linear kneading compactor that comprises a loaded wheel or similar part that runs over a rectangular mold filled with loose mix and metal shims on top of the loose mix to transfer the load. The linear kneading compactor (LKC) and Arizona kneading compactor are examples of this approach. Figure 7.2 illustrates a schematic and photograph of the Hveem compactor.

Compared to the Marshall compactor, the specimens compacted using the kneading compactor are a much better representation of a field-compacted mix. However, despite this advantage, kneading compactors were not very common in practice. This is on account of the fact that such compactors are typically more expensive and require more space and resources to produce laboratory-compacted specimens.

The third category of compactors is the gyratory compactor. The advantage of the gyratory compactor is that it replicates the kneading action experienced by the mix using a device that has a much smaller footprint and can easily be used in a laboratory setting. An early form of the gyratory compactor was developed in Texas in the 1930s (Fig. 7.3). The Texas gyratory compaction method involves the use of a four-inch-diameter mold with two end plates. The mold itself is tilted at about 6°.

To prepare a test specimen, the mold is loaded with the loose mix, and a vertical pressure is applied to the specimen. The mold is then gyrated three times, thus inducing a kneading-type compaction of the sample. Since the vertical pressure is applied using a reaction frame, reduction in the specimen height due to compaction

Fig. 7.2 A schematic showing a schematic of the kneading compaction method (*left*) and a photograph of automated Hveem compactor (*right*) (Image courtesy of James Cox and Sons, USA)

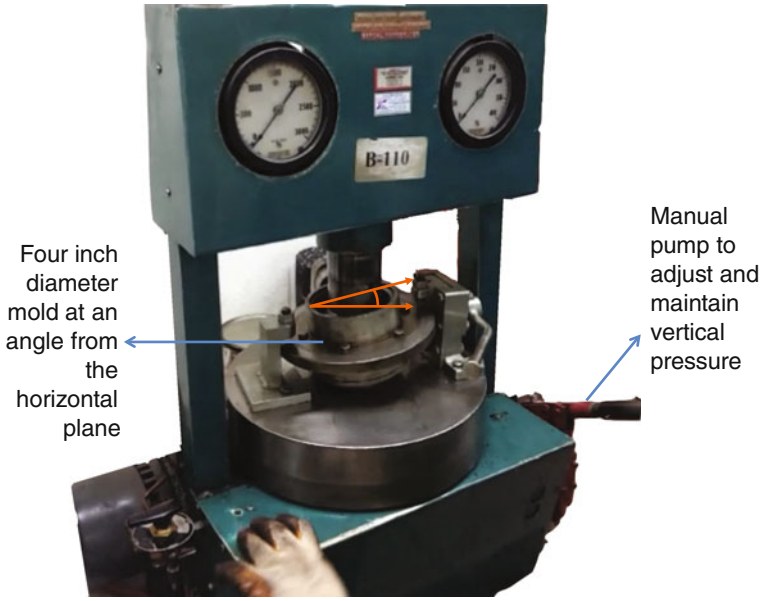


Fig. 7.3 A schematic showing the Kneading compaction method (*left*) and a photograph of automated Hveem compactor (*right*) (Photograph by A Bhasin)

relieves the vertical pressure. After the three gyrations, the vertical pressure is adjusted again to the target value, and three more gyrations are applied. This process is repeated until the specimen is compacted to a point where it would create a specific amount of resistance to the vertical pressure. Use of this method clearly requires a certain level of skill and practice before one can produce specimens with acceptable levels of repeatability.

The Texas gyratory compactor triggered the development of the French gyratory compactor in the 1950s. In the following decades, the French gyratory compactor was subjected to extensive research that involved evaluating the influence of factors such as aggregate gradation and filler content on the densification of the mixture. Subsequently, during the SHRP research, the gyratory compactor was finally adopted as the standard method to produce laboratory specimens to design and evaluate asphalt mixtures. This final form of the gyratory compactor was referred to as the Superpave gyratory compactor (SGC). The SGC allows the user to compact six-inch or four-inch-diameter specimens. The mold is typically operated at a constant vertical pressure of 600 kPa and 30 gyrations per minute. The angle of gyration is typically 1.25°, although the use of different angles depending on the type of the mixture will result in different compaction efforts (Button et al. 2006; Blankenship et al. 1994). The number of gyrations in the compactor is a function of the expected traffic on the pavement for which the mixture is being designed. Recall that during mixture design, the goal is to produce a test specimen that represents the mix density and state toward the end of its intended service life.

One of the most important features of the SGC is a displacement transducer that continuously records the height of the sample versus number of gyrations as the loose mix is being compacted (Fig. 7.4). This information, referred to as the densification curve, can provide meaningful insights into the stability of the mixture. For example, asphalt mixtures that densify very easily during the initial stages of compaction are tender mixes that tend to be unstable and also have a problem with segregation during construction. On the other hand, asphalt mixtures that continue to densify even after a significant number of gyrations (implying compaction effort) are also unstable and tend to have rutting problems during the service life of the pavement. The densification curves are therefore used as an integral part of the volumetric design of an asphalt mixture, which will be discussed in more detail later.

One obvious question at this stage is, how do we know that the specimens compacted using the SGC closely represent the internal structure and characteristics of the mix compacted in the field? Several studies have been conducted to answer this question; a few of these are briefly discussed here. In one study, Consuegra et al. (1989) evaluated the ability of five different compaction methods to produce mixtures with engineering properties nearest to those determined from field cores. The compaction methods that they compared were the mobile steel wheel simulator, Texas gyratory compactor, California kneading compactor, Marshall impact hammer, and Arizona vibratory kneading compactor. They used resilient moduli, indirect tensile strength and strain at failure, and tensile creep data as the basis to compare laboratory-compacted specimens (using different devices) to field cores. Their study demonstrated that the Texas gyratory compactor, which was also the

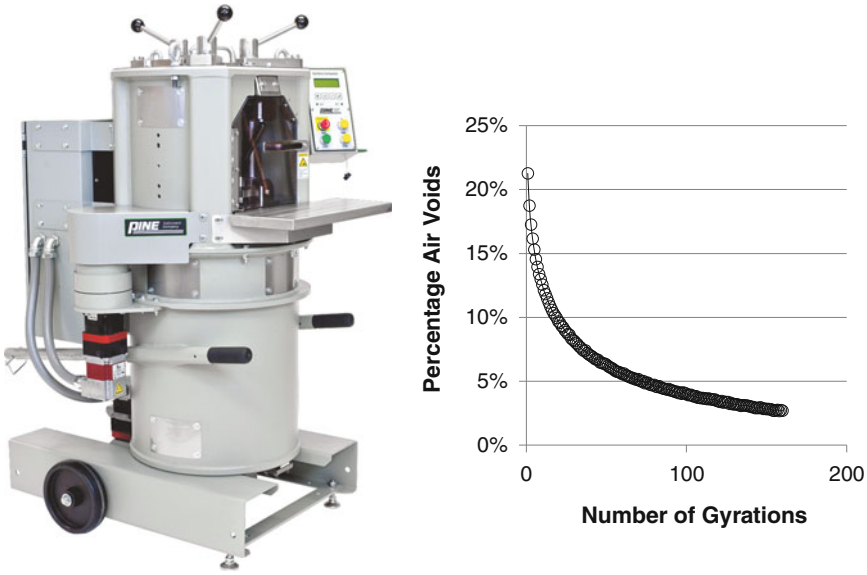


Fig. 7.4 A schematic showing the Superpave gyratory compactor (*left*) and a typical densification curve obtained during the compaction of a test specimen (*right*) (Compactor image courtesy of Gilson Company Inc., USA)

precursor to the SGC, provided the best match between laboratory-compacted specimens and field cores. Test specimens compacted using the Arizona vibratory kneading compactor and the Marshall impact hammer had the poorest match compared to the field cores. Despite these results, it must be noted that the gyratory compactor is not a perfect method that always produces the best results (in this case best implying the closest match with field compaction). Their study also indicated that there was no single laboratory compaction method that always provided the best match compared to the results of the field cores.

In a subsequent study, Button et al. (1994) also evaluated the ability of four types of compaction methods to produce specimens with engineering properties similar to field cores. Similar to the previous study, Button et al. (1994) compared test specimens produced using the Exxon rolling wheel compactor, Texas gyratory compactor, rotating base Marshall hammer, and Elf linear kneading compactor to field cores. They also used mechanical properties of mixtures measured using tensile and shear modes of loading as the evaluation criteria. Similar to the previous study, these authors also concluded that the gyratory method provided the best match when compared to the results from the field cores. Also, the Marshall rotating base compactor provided the poorest match compared to the results from the field cores. The third example is from another study that investigated gyratory compactor, kneading and rolling wheel compaction procedures (Harvey et al. 1994). In

this case, the gyratory compactor was found to place excessive emphasis on the asphalt binder and to inaccurately portray the role of asphalt–aggregate interaction in the performance of properly constructed pavements. These authors suggested that the rolling wheel compaction produces specimens that match the best with field-compacted specimens and recommended that such a device be used for the preparation of asphalt mixture specimens in the laboratory.

In summary, a large-scale rolling wheel compactor is perhaps the closest to replicating the compaction process that is experienced in the field. However, this approach and device naturally require a relatively much larger volume of the mixture to be compacted and are therefore highly resource intensive for use on a routine basis in a laboratory environment. The gyratory compactor, or the SGC in its most recent form, presents an optimal trade-off wherein the specimens compacted are similar to field samples and yet the device can be used to compact smaller test specimens that can be used for mixture design and testing purposes. Finally, on the subject of compaction, it is also worth pointing out that there have been studies undertaken to simulate the compaction of loose mixtures in an asphalt pavement. Reports and publications from such studies can be found under additional reading material (Masad et al. 2010).

7.3 Design for Optimal Binder Content

7.3.1 What is Optimal Binder Content?

The goal of the mixture design process is to determine the optimal binder content, i.e., optimal binder-to-aggregate or binder-to-mixture weight ratio that will result in optimum performance of the mixture in terms of its resistance to rutting and fatigue cracking. Recall the five steps in the generic mixture design procedure that were discussed earlier in this chapter. The first four steps involved selecting an aggregate gradation, binder type, preparing several different batches of mixtures with different binder–aggregate ratios, and finally compacting test specimens for each binder–aggregate ratio (discussed in the previous section).

This section will discuss, the last and the most important step, i.e., to determine the optimal mass percentage of the binder that will result in a mixture that is durable without any performance issues. A short discussion on the significance of accurately determining the optimal binder content is important here. Consider the following two extreme hypothetical scenarios. On the one extreme, it is possible to imagine a mix that has very little to no binder in it. Such a mix would obviously not be durable and would very easily fall apart under the action of traffic loads. From a simplistic viewpoint, one can say that such a mix would tend to crack and ravel easily. On the other extreme, one can imagine a mix that has extremely high binder content. For the purpose of this hypothetical example, consider a mix with 50%

binder by weight of the mix. Such a mix would obviously be prone to significant plastic deformations and stability issues (not to mention extremely expensive since binder is typically 20–30 times more expensive than aggregate). Therefore, it stands to reason that somewhere between these two extremes is a binder content that will result in a mixture that is stable, durable, and resistant to cracking and raveling. It is also important to highlight here that the optimal binder content is highly dependent on the aggregate gradation, binder type, and grade and compaction effort. The optimal binder content for a given mix is a very sensitive parameter. For example, a change in the binder content by just one percentage point from its optimum in a mixture (higher or lower) can result in a mixture that fails prematurely (due to rutting or durability issues). Simply put, the determination of optimum binder content entails using the test specimens fabricated at different mass percentages of binder by weight of the mix with some criterion or criteria to determine the optimum mass percentage of binder.

Several different mix design methods have emerged over the last few decades that specify different criteria to determine the optimum binder content. Most of these methods determine the optimum binder content based on mixture volumetrics and/or a combination of mixture volumetrics with one or more mechanical tests that measure indicators of mixture performance. A positive trend in the industry is to promote the use of and rely more on the performance indicators rather than mixture volumetrics. Before discussing a couple of examples on the mixture design method, it is important to briefly discuss mixture volumetrics.

7.3.2 Mixture Volumetrics

The term mixture volumetrics refers to a number of different volumetric properties of the mix. Although volumetric properties are not direct measures of engineering properties such as stiffness or resistance to plastic deformation or cracking, these properties serve as approximate indicators of expected performance. Also, given the fact that these properties can be measured very easily in a laboratory and require very little capital equipment, mixture volumetrics are often used as the sole basis for mixture design for many smaller scale projects. Finally, mixture volumetrics are also important because this information can be used for quality control purposes to ensure that the optimal "recipe" developed in the laboratory during mixture design is replicated in the field. In this section, we will limit our discussion to enlisting these properties and briefly describing what they reflect in terms of mixture behavior. Readers interested in learning more about the specific laboratory methods to measure mixture volumetrics and relevant formulas involved can find this information in documents listed in the additional reading material (references).

Mixture volumetrics refers to mass and volume relationships of the three components in an asphalt mixture, i.e., air, binder, and aggregate. The most commonly used parameters in this context (along with the typical notation in parenthesis) are as follows:

1. percentage of binder in the mix (P_b),
2. maximum specific gravity of the asphalt mixture (G_{mm}),
3. bulk specific gravity of the compacted mix or specimen (G_{mb}),
4. percentage air voids in the compacted mix,
5. percentage of absorbed binder (P_{ba}), and effective binder (P_{be}) in the mix,
6. voids in mineral aggregate (*VMA*), and
7. voids filled with asphalt (*VFA*).

The percentage of binder in a mix, (P_b), typically refers to the percentage of binder by weight of the mix (binder + aggregate). While this is the most common definition of the percentage of binder, some agencies may choose to define this as the percentage of binder by weight of the aggregate. As mentioned before, the ultimate goal of any mixture design process is to determine the optimum percentage of binder.

The maximum specific gravity of the mix refers to the mass per unit volume of the mix without any air voids. This is a very important property that must be measured as accurately as possible because all other properties are computed using G_{mm} as the basis. Note that during the mix design process, samples of loose asphalt mixtures at different binder contents are split into two parts. One part of the loose mixture (at each binder content) is used to measure the maximum specific gravity of the mixture, and the other part of the loose mixture (at each binder content) is used to compact test specimens. The maximum specific gravity of a mix will vary with the binder content (for a given binder type and aggregate type and gradation).

The bulk specific gravity and percentage of air voids are *properties of the compacted specimen* and in addition to the binder content depend on the extent of compaction effort applied to the specimen. The percentage of air voids is simply the ratio of bulk to maximum specific gravity subtracted from 100 (i.e.$100-(G_{mb}/G_{mm})$).

The percentage air voids is perhaps one of the most important parameters used in mixture design. Typically, it is expected that a dense-graded asphalt mixture in a newly constructed pavement will be compacted to approximately 93% of its maximum specific gravity (G_{mm}) or 7% air voids and would densify due to traffic to approximately 96% of G_{mm} or 4% air voids toward the end of its life. Consequently, typical mixture design procedures define *optimum asphalt content as the binder content in the compacted specimen that will result in 4% air voids when the specimen is compacted using a compaction effort that indirectly reflects the influence of expected traffic volume over the service life of the pavement*. This percentage of air voids is also referred to as the design air voids content. This 7–4% range is also considered as optimum because a higher upper limit would entail earlier densification of the mix and premature rutting. Similarly, a lower value of the design air voids content would result in a mix that has very little room for binder movement due to temperature changes and as a result can result in flushing and bleeding of the mix. Also, quality control procedures during construction and often payment bonuses and penalties toward the contractors are also tied to the ability of the contractor to place and compact the asphalt mixture to approximately 93% of G_{mm} or 7% air voids.

A couple of important things to note regarding the definition of optimum binder content at 4% air voids. First, there are obvious exceptions to the 4% design air void content approach. For example, porous or open-graded asphalt mixtures are designed and compacted at a much higher air void content. *Second, ideally the optimum binder content should be the binder content that optimizes the rutting and fatigue cracking characteristics of the mixture*. However, this would require running multiple mixture performance tests for each trial mix or binder content. While this may be worth the effort for larger scale projects, it may not be best suited for smaller projects. Consequently, although the volumetric definition of optimum binder content (i.e., binder content that results in 4% air voids at ultimate compaction levels) is less than ideal, it is less expensive to use and is therefore more commonly used particularly for smaller scale projects.

Returning back to the discussion on the volumetric properties, certain kinds of mineral aggregates are known to have some degree of porosity. In such cases, a small fraction of the asphalt binder tends to get "absorbed" into the aggregate close to the surface. This so-called absorbed binder does not effectively contribute to the function of the binder, i.e., to act as an interfacial adhesive layer binding aggregate particles together, and as such is not counted as effective binder. The term absorbed binder P_{ba} refers to this absorbed binder, and effective binder P_{be} is the total binder content after discounting for the absorbed binder.

The absorbed binder has a couple of important implications related to the design of asphalt mixtures. First, since binder contributes significantly to the cost of the mix, a higher percentage of absorbed binder would tend to increase cost of the mix without adding any significant benefit. Second, some studies suggest that depending on the composition of the asphalt binder and aggregate, only some chemical fractions of the binder are preferentially absorbed into the mineral aggregate leaving the residual binder in the mix with properties different from the bulk.

As an illustration, consider the saw-cut specimens from two different mixtures shown in Fig. 7.5. These two mixture specimens were prepared using the same aggregate gradation, aggregate type, and binder content and compacted to achieve similar densities. The only difference between the two specimens was the type of the binder. The specimen to the left used a stiff PG76-22 binder, whereas the specimen to the right used a softer PG64-22 binder. On careful observation, one can see the discoloration of the aggregate particles along the interface between the binder and mastic. This discoloration is more prominent for the softer binder, qualitatively indicating a higher degree of absorption compared to the stiffer binder. The discoloration is also more prominent at the center of the specimen than the edges, most likely because the center of the specimen cools more slowly, allowing the mix to remain at elevated temperatures and accelerating the absorption process. Also, when the complex moduli of the uncut specimens for these two mixtures were measured, the mixture with the stiffer binder (PG76-22) had slightly lower stiffness compared to the mixture with the softer binder (PG64-22). Given that the two mixtures were identical in every other respect and also that the bulk properties of

Fig. 7.5 Saw cut cross sections of two asphalt mixture specimens with same binder content, aggregate type and gradation but different binder types; the specimen on the left shows less absorption in the aggregates compared to the specimen on the right; the absorption can be seen in the form of discoloration of the aggregate particles at its interface with the binder (Photograph by A Bhasin)

the PG76-22 indicated that it was stiffer than the PG64-22 at the same test temperature, the only likely explanation was that the stiffness for the mix containing the PG64-22 binder was artificially being enhanced due to some fractions of the binder being absorbed by the aggregate particles (Arega and Bhasin 2012).

Finally, a short discussion about the VMA and VFA is important. Voids in mineral aggregate (VMA) reflects the percentage of volume occupied by air voids and asphalt binder in a compacted asphalt mixture specimen. In other words, the percentage VMA reflects the combined volume of air voids and asphalt binder in the mix. A higher percentage of VMA indicates that the mix contains a high percentage of air voids, or asphalt binder or both. Consequently, a mixture with a very high VMA indicates that the mixture may be susceptible to rutting and may also be unstable and prone to segregation. Similarly, a mixture with a very low VMA would indicate that the mixture has very low asphalt binder content or has been compacted to have very little room for air voids or both. Therefore, such a mixture may be prone to cracking due to lower durability and also to flushing and bleeding. The VMA for a mix can be computed using its other volumetric properties as $100 - (G_{mb}P_s/G_{sb})$, where, G_{mb} is the bulk specific gravity of the mix, P_s is the percentage of aggregate (s stands for stone) in the mix by weight of the mix, and G_{sb} is the bulk specific gravity of the aggregate.

Based on the above explanations, most agencies involved with specifying mixture designs have determined an optimum range of VMA that results in satisfactory performance for different types of asphalt mixtures. In this sense, VMA can be considered as the simplest form of performance indicator that can be incorporated into the mixture design procedure (although it is not necessarily the most accurate or robust). Note that the VMA indicates the volume percentage of binder and air voids put together in the asphalt mixture. In some cases, it is important to further determine the split of the VMA into air voids and asphalt binder. To this end, the parameter voids filled with asphalt or VFA is commonly used to indicate the fraction of the VMA that is made up of asphalt binder. VFA is typically

computed using the VMA and volume of air voids as $100\,(1-(V_{air}/VMA))$. Also, since VFA is dependent on two other volumetric properties (air voids and VMA) that are controlled or specified during the mixture design process, the use of the VFA to control the mix design is sometimes considered redundant.

7.3.3 Examples of Methods to Determine Optimum Binder Content

As discussed in the introductory section, the mixture design process typically involves five steps. The third and fourth steps involve preparing batches of loose mixes with different binder contents and compacting these mixes to produce representative specimens of the asphalt mixtures at these binder contents. The last step during the mixture design process is to use these representative specimens at different binder contents and determine the so-called "optimum" binder content. A few examples of methods to determine this optimum binder content are discussed in this subsection.

One of the oldest mixture design methods is the Marshall mix design method. Considering the fact that as of this writing, this method is still in use by several agencies across the world, it is important to at least briefly discuss this method here. In this method, batches of asphalt mixtures are prepared at different binder contents by weight of the mix. Each batch is separated into two parts; one part of the loose mix is used to determine the maximum specific gravity (G_{mm}), and the other part is used to compact test specimens using the Marshall compaction method as discussed earlier. Mixture volumetrics for the compacted specimens (VMA and percentage air voids) at each binder content are determined using mass and volume measurements. Readers are referred to additional reading material for more information on the detailed laboratory methods used to make these measurements and calculate the mixture volumetrics. In addition to these volumetric measurements, the Marshall mix design method also relies on some very simple performance indicators to make the final determination of the optimum binder content.

These performance indicators, Marshall stability and flow, are obtained by conducting a test referred to as the Marshall stability test. The Marshall stability test entails loading the sample diametrically by applying a displacement controlled load typically at a rate of 2 in. per minute using the Marshall stability apparatus (Fig. 7.6); the reaction load is continuously recorded during this process. Note that the load is applied using a collar that transmits the normal reaction force to a significant fraction of the circumference of the specimen. Marshall stability corresponds to the value of the maximum reaction load that the specimen experiences during the loading process. Marshall flow, measured in one-hundredths of an inch, is the total deformation of the mix when it reaches the maximum load. Together, the stability and flow serve as indicators of expected performance of the mix. For example, a mix with either a very low or very high binder content would either be

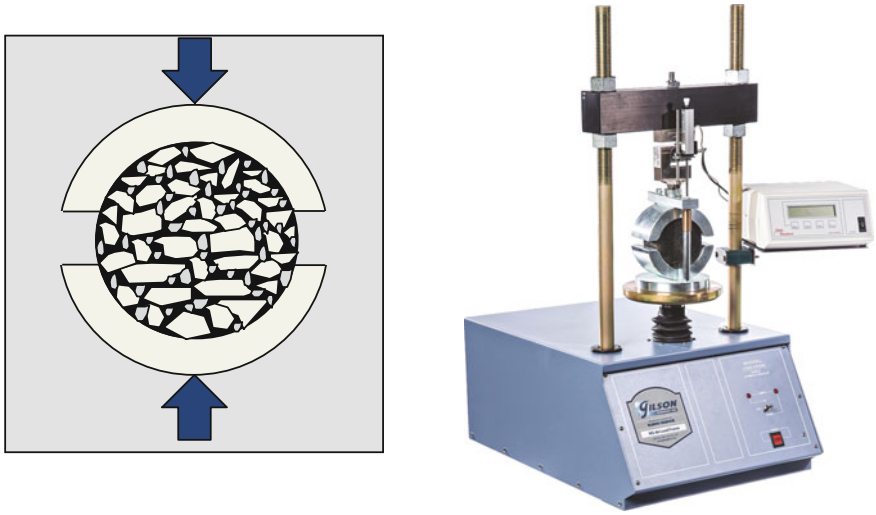


Fig. 7.6 Schematic (*left*) and the photograph (*right*) of the Marshall stability apparatus used with the mix design method (Image courtesy of Gilson Company Inc., USA)

prone to disintegration or rutting, respectively. Similarly, a very low flow value would indicate that the mix lacks ductility and could be susceptible to cracking and low durability, whereas a very high flow value would indicate that the mix is tender and is susceptible to flow and permanent deformation or rutting.

Mixture volumetrics as well as the Marshall stability and flow values are examined as a function of the different binder contents. This information is then used to determine the optimum binder content. A couple of different approaches can be conceived to achieve this. The first approach would be to define the optimum binder content based on one specific property and criterion and then verify whether this optimum binder content would fall within an established range for each of the other properties and criteria. Some of the generic steps based on this approach would be as follows:

1. Define the optimum binder content based on one criterion (a typical example is binder content at which the mixture has 4% air voids).
2. Determine the optimum by examining the relationship between percent air voids and binder content.
3. Determine the expected VMA, stability, and flow values at the optimum binder content using the respective relationships of these properties with respect to the binder content.
4. Finally, examine whether the VMA, stability, and flow at the optimum binder content are within certain acceptable limits. If these requirements are met, the mix design is complete. If not, the aggregate gradation will have to be changed and the mix will have to be redesigned.

The second approach is to define the optimum binder content as an average value of the binder contents that correspond to specific values for two or more properties. For example, in this case the optimum asphalt content can be defined as the average of the binder content that results in a certain percentage of air voids and the binder content that results in a specific value of stability.

The first approach is the most commonly used method to determine optimum asphalt content, i.e., by defining optimum (at typically 4% air voids) and then checking whether other properties at this binder content fall within a specified range.

In the remainder of this section, we will briefly discuss the Superpave mix design method that was developed as a result of the Strategic Highway Research Program (SHRP) in the USA during the 1990s. The Superpave mix design method was originally proposed to have three levels. The Level 1 design involves the use of volumetric methods to determine the optimum binder content combined with just one test to examine the moisture damage resistance of the asphalt mixture. The Level 2 design involves the use of volumetric methods to determine the optimum binder content supplemented by mixture performance tests conducted using the Superpave Shear Tester or SST for short. The Level 3 mix design is similar to Level 2 except that the performance tests are conducted at multiple temperatures using the Superpave Shear Tester.

In the years that followed the development of the Superpave mix design method, the Level 1 mixture design method gained the most momentum and was adopted by several different highway agencies in some form. On the other hand, the Level 2 and Level 3 mixture design procedures did not gain much traction primarily owing to the fact that these methods involved the use of the Superpave Shear Tester (SST) to measure mixture performance. The SST has two actuators that are perpendicular to each other and capable of applying a normal and a shear force to a cylindrical test specimen simultaneously. This allowed for conducting tests on specimens that would combine normal and direct shear stresses simultaneously. However, this hardware configuration also required testing equipment that had a much larger laboratory footprint as well as higher initial and operating costs. The cost of the equipment eventually became an impediment toward the adoption of the SST in the mixture design process. In the years since the development of the Superpave mix design method, several researchers have focused on the development of "Simple performance tests" with an intent to replace the SST. Our discussion here will be limited to the Superpave volumetric mix design method.

The basic steps of the mixture design process are the same as discussed in the introductory section of this chapter. Just to recap, the process involves selecting the appropriate binder grade for a specific location, selecting a trial aggregate gradation (or more than one trial aggregate gradation), and preparing batch mixes at different binder contents that typically vary by 0.5% by weight of binder. The batches of loose mixes are cured in a convection oven for typically 2h at the mixing temperature as determined by testing the binder. This 2-h curing period is intended to simulate short-term aging of the mix that occurs as the mix is produced in a hot-mix plant and transported to the construction site, placed, and compacted. Similar to the Marshall mix design method, after short-term aging, each batch is split into two

parts: One part is used to determine the maximum specific gravity G_{mm}, and the other part is used to compact two or more cylindrical specimens at each binder content.

As discussed in the previous section, one of the most distinguishing features of the Superpave mix design method is the use of the Superpave gyratory compactor (SGC) to produce compacted specimens to estimate the optimum binder content and the ability of the SGC to record the height of the specimen as a function of the compaction effort. Recall that by knowing the height and maximum specific gravity of the mix, it is possible to determine the percentage air voids as a function of the compaction effort. Specifically, the density or air voids content of the specimen are of interest at three different levels of compaction: initial number of gyrations (N_{ini}), design number of gyrations (N_{design}), and maximum number of gyrations (N_{max}). These three levels are defined based on the expected traffic volume for which the mixture is being designed. For example, a traffic volume of 0.3–3 million equivalent single axle loads (ESALs) would require that the loose mix be compacted using 115 gyrations of the SGC wherein the initial, design, and maximum number of gyrations are 7, 75, and 115. The specified values for these three parameters for different traffic volumes can be found in the additional reading material (Huber et al. 1994).

In the Superpave volumetric mix design method, *the optimum binder content is defined as the binder content at which the percentage air voids in the mix is 4% at the design number of gyrations* N_{design}. The optimum binder content is determined typically by examining a plot of the binder content versus number of air voids at N_{design}. Once the optimum binder content is determined based on this definition, it must then be verified for the following volumetric criteria before the mixture design can be fully established:

1. the percentage of air voids at the initial number of gyrations (N_{ini}) and at the optimum binder content must be above a minimum required value,
2. the percentage of air voids at the maximum number of gyrations (N_{max}) and at the optimum binder content must be above a minimum required value,
3. the voids in mineral aggregate (VMA) at the optimum binder content must be above a minimum required value,
4. the voids filled with asphalt (VFA) at the optimum binder content must be within a specified range of values, and
5. the filler-to-binder ratio must be within a specified range of values.

The rationale for checking the aforementioned volumetric properties is discussed below.

Typically, a good mix design should not result in a tender mix that is prone to segregation or early deformation even during construction. In other words, a good mix should offer some level of resistance to compaction. Mixes that are too easy to compact are also prone to segregation during construction and early deformation. A good indicator of this behavior is the percentage of air voids in the mix during the early stages of compaction in the SGC. A very low percentage of air voids very

early during compaction (specifically at N_{ini}) would indicate a tender mix. Therefore, the mix design procedure stipulates that at the optimum binder content the mix should have more than a certain percentage of air voids (or the density should be less than a certain percentage of the maximum density) at N_{ini}. Specifically, the requirement in the Superpave method is that the mix has more than 8.5%, 9.5%, and 11% air voids at N_{ini} for mixes designed for traffic volumes of less than 0.3, between 0.3 and 3 and more than 3 million ESALs.

On the other extreme, it is also desirable that mixes reach a certain percentage of their maximum density during their service life and do not continue to densify under extended traffic loading (which would suggest prolonged susceptibility to rutting). The density at the maximum number of gyrations, N_{max}, is a good indicator of this behavior. Specifically, the requirement in the Superpave method is that the mixture retains at least 2% air voids at N_{max} at its optimum binder content.

As discussed earlier, the voids in mineral aggregate or VMA is a volumetric indicator of the volume occupied by the air voids and asphalt binder together in the mix. As such, a low VMA would indicate that the mixture is either deficient in asphalt binder or air voids or both and therefore could result in problems such as poor durability, cracking, or bleeding. Therefore, the mix design method requires that the asphalt mixture should have a minimum value of the VMA at the optimum binder content. The required minimum value varies with the nominal maximum size of the aggregate. For example, a mixture with a 19 mm nominal maximum aggregate size must have a VMA that is above 13%. Since VMA is a measure of the volume occupied by the binder and air voids together within the mineral aggregate matrix, there is also a requirement for voids filled with asphalt or VFA to ensure that a certain fraction of the VMA corresponds to the asphalt binder. For example, a mixture that is designed for a traffic volume of 3–10 million ESALs is required to have a VFA between 65 and 78%. Finally, there is also a volumetric requirement that the filler-to-binder ratio in the mix be restricted to 0.6–1.2. A detailed description on the role of fillers and this requirement was presented in Chap. 6.

In summary, the volumetric mix design method is the simplest and most common method used to design asphalt mixtures. It cannot be overemphasized that although this method is the simplest, it is not the most robust. The biggest advantage of this method is that beyond mass and volume measurements, this method does not require any elaborate testing. The only exception to this is the requirement to measure moisture damage resistance of the asphalt mixtures at their optimum asphalt content using the tensile strength ratio (this will be discussed in a later chapter). Also, the most important capital equipments required in this process are the SGC and a convection oven to heat the asphalt mixtures.

The working assumption for this methodology is that if the mixture volumetrics are within a certain range, the mixture will not be susceptible to premature failure. While this may be an economically reasonable approach for several small-scale or low-value projects, in many cases materials and pavement engineers also require some form of a mechanical test to ensure that the mixture is not susceptible to premature distresses such as rutting, fatigue cracking, or moisture-induced damage. The lack of adoption of the Superpave Shear Tester (SST) led to the creation of an

industry of asphalt researchers who focused on the development and implementation of performance tests that could fulfill this requirement. The following chapters will discuss the various tools that are available for materials and pavement engineers to assess mixture performance.

As a final remark, the goal of performance tests is to evaluate possible mixture design combinations (different aggregate gradations, binder contents etc.) for distresses such as rutting and cracking and determining the mixture design that demonstrates acceptable resistance to both rutting and cracking at the lowest cost. Performance testing or laboratory evaluation to search for an optimal mix with such large number of variables can easily become a very expensive proposition. It is precisely for this reason that we believe that in the years to come, mechanics based and computational methods will emerge and serve as screening tools to facilitate performance based mixture design.

7.4 Summary

As of this writing, two of the many commonly used mix design methods are the Marshall mix design method and the Superpave Level 1 or volumetric mix design method (or some variation of these methods modified by road agencies for their local conditions). As discussed earlier, both methods require very little capital equipment and minimal material testing to establish the optimum binder content. *Also, both methods are based on the premise that if the volumetric requirements for the mixture are met, then the mixture will not be susceptible to premature damage due to rutting or cracking.* While this is a reasonable approach for small-value projects, material and pavement design engineers often require a higher degree of confidence on the durability of the proposed mixture designs for most high-value projects. This is even more important in light of the fact that during the last couple of decades, there have been several advances in additives and chemical modifiers that can be used to modify binder and mixture properties and production methods (e.g., chemical additives using Warm Mix Asphalt or WMA technology). The effect of such modifiers and processes on long-term mixture performance cannot be adequately captured through volumetric properties. In order to address this need, several mixture performance tests and concomitant parameters have been introduced in the past few years.

Chapter 8 of this book will discuss the mechanisms for the most common forms of distresses as well as the test and analytical methods that are used to quantitatively assess the susceptibility of different mixtures to these distresses.

7.5 Exercises

7.1. The maximum specific gravity of a mix is determined by taking a sample of the loose mix in a pycnometer, filling it with water and applying vacuum to displace any air voids trapped in the loose mix, and finally weighing the pycnometer with

the sample and water filled to the top with water. The weight of the pycnometer filled with water (without the sample) is also known. This allows the user to determine the volume of the mix without any air voids and determine the maximum specific gravity or G_{mm}. The bulk specific gravity is relevant for a compacted specimen and is measured by using the weight of the specimen and its volume. Given, the following information:

percent binder by weight of mix = 4.2%
specific gravity of binder G_b = 1.003 and aggregate G_{sb} = 2.61

Measurements with loose mix:
weight of loose mix in air = 1501.3 g
weight of pycnometer + water = 3002.4 g
weight of pycnometer + water (partially displaced by loose mix) + loose mix = 3895.2 g.

Measurements with compacted specimen:
weight of a compacted sample in air = 2970.1 g
weight of a compacted sample in water = 1680.9 g
weight of saturated surface dry (SSD) sample after removing from water = 2997.2 g.

Recall from basic physics that difference in the SSD weight and weight of sample in water are used to determine the weight of the water displaced and consequently volume using Archimedes' principle.

Use the above information to calculate:

1. maximum specific gravity of the mix (G_{mm})
2. bulk specific gravity of the compacted specimen (G_{mb})
3. percent volume of air voids in the compacted specimen.

7.2. Given the cost of binder at $500 per metric ton and aggregate at $30 per metric ton, what is the *total cost of the materials* that you estimate for the construction of an overlay 4-in. thick, 24′ feet (typical two lane with no median), 1 mile long to be constructed at 7% air void content.

7.3. Should the maximum specific gravity of the mix, G_{mm}, change with binder content? Why or why not?

7.4. The table below is information collected from preparation of specimens to determine optimum binder content using the Superpave volumetric mix design method:

Binder content	G_{mb}			G_{mm}
	N_{ini}	N_{des}	N_{max}	
4.5% Replicate 1	2.151	2.383	2.416	2.483
4.5% Replicate 2	2.158	2.388	2.421	
4.0% Replicate 1	2.152	2.379	2.410	2.505

(continued)

(continued)

Binder content	G_{mb}			G_{mm}
	N_{ini}	N_{des}	N_{max}	
4.0% Replicate 2	2.157	2.382	2.415	
3.5% Replicate 1	2.115	2.338	2.373	2.550
3.5% Replicate 2	2.117	2.341	2.376	

Also, the expected traffic for this mix is 3–30 million ESALs, and the maximum nominal size of the aggregate is 19 mm.

References

Arega, Z., & Bhasin, A. (2012). *Binder rheology and performance in Warm Mix Asphalt*. Austin, TX: University of Texas at Austin, Center for Transportation Research.

Blankenship, P. B., Mahboub, K. C., & Huber, G. A. (1994). Rational method for laboratory compaction of hot-mix asphalt. *Transportation Research Record*, 1454.

Button, J. W. et al. (1994). Correlation of selected laboratory compaction methods with field compaction. *Transportation Research Record*, (1454). Retrieved May 7, 2015, from http://trid.trb.org/view.aspx?id=424725.

Button, J. W., Chowdhury, A., & Bhasin, A. (2006). Transitioning from texas gyratory compactor to Superpave gyratory compactor. *Transportation Research Record: Journal of the Transportation Research Board, 1970*(1), 106–115.

Consuegra, A. et al. (1989). Comparative evaluation of laboratory compaction devices based on their ability to produce mixtures with engineering properties similar to those produced in the field. *Transportation Research Record*, (1228). Retrieved May 7, 2015, from http://trid.trb.org/view.aspx?id=308779.

Harvey, J. et al. (1994). Effects of laboratory specimen preparation on aggregate-asphalt structure, air void content measurement and repetitive shear test results. *Transportation Research Record*, (1454). Retrieved May 7, 2015, from http://trid.trb.org/view.aspx?id=424717.

Huber, G. A., Kennedy, T. W., & Anderson, M. (1994). *The superpave mix design manual for new construction and overlays*, Washington, D.C.

Masad, E. A. et al. (2010). *Modeling of hot-mix asphalt compaction : A thermodynamics-based compressible viscoelastic model*, Washington, D.C.

Chapter 8
Failure Mechanisms and Methods to Estimate Material Resistance to Failure

8.1 Introduction

In the previous section of this book, we discussed the most important elements of designing an asphalt mixture. These elements include (1) identifying and selecting component materials for the specific application (i.e., asphalt binder and mineral aggregate), (2) selecting an appropriate gradation for the mineral aggregate, and (3) identifying an optimal ratio of the binder and mineral aggregate such that the resulting mixture, when used in a specific pavement structure, will result in an acceptable performance. The next reasonable questions to ask are, how is performance measured? and what is an acceptable level of performance?

The overall performance or condition of a pavement is often quantified in terms of a pavement condition index (PCI). The PCI (or other similar metrics used by different highway agencies) reflects the overall condition of the pavement in terms of its ride quality and distresses such as rutting, cracking, raveling, flushing (or bleeding), and pothole formation. Some of the most common distresses in asphalt mixtures that influence the serviceability of a pavement are:

1. Rutting or permanent deformation
2. Fatigue cracking (and self-healing that reverses fatigue cracking)
3. Transverse, low-temperature, or more generally thermal cracking
4. Bleeding or flushing
5. Moisture-induced damage

In addition to the above distresses, asphalt binder in the mixture continuously undergoes aging (as discussed in Chap. 2). As a result, the mechanical properties of the mixture as well as the rates at which distresses evolve in the mixture also continuously change with time.

The focus of this chapter is to discuss the mechanisms of these distresses at the material level and also to introduce some of the common and more advanced methods that are used to quantify the resistance of a given material to these

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_8

distresses. Note that the mechanisms discussed in this chapter are from a phenomenological point of view. The mathematical representations of these phenomena using principles of mechanics are presented in second part of this book. Readers must recognize that it is important to carefully understand both of these aspects (phenomenon and mathematical modeling) to be able to fully and accurately model and predict the behavior of a material. Perhaps this is best reflected in an explanation provided by Malvern (1969), *"There may be exact mathematical solutions to the equations formulated in elasticity or in other branches of continuum mechanics, but the equations themselves are not exact descriptions of nature"*.

8.2 Understanding the Role of Pavement Versus Materials in Distress Evolution

Before we discuss the mechanisms that drive the different failure processes in an asphalt pavement, we need to briefly discuss the sources of different failures. *Understanding the differences in failure at different length scales is the most important step in being able to understand and design durable asphalt mixtures and pavements.* The ideas presented in this section are straightforward and simple but are very important. Readers are encouraged to review this section very carefully because it provides a basis to better understand the origins of different distresses.

Failure in a pavement structure can occur by design at the end of a pavement's intended service life or it could be premature because of improper design or construction. In a very broad sense, failure of a pavement can be attributed to failure of one or a combination of the following factors:

- component material selection (binder and aggregate),
- mixture design (aggregate gradation and binder content),
- pavement structure design (or thickness design of different layers), and
- construction.

Let us briefly discuss each one of the above factors. Component materials refer to the two main ingredients that constitute the asphalt mixture used in pavement construction, i.e., asphalt binder and mineral aggregate. In this context, chemical additives, polymers and other modifiers can be broadly considered to be a part of the binder. It is obvious that the use of poor quality raw materials will not result in a high-quality or durable asphalt mixture. Chapters 2 and 3 of this book discuss the attributes of asphalt binders and aggregates, respectively, that dictate the performance of the asphalt mixture as a whole. For example, consider a pavement intended to be built in a location that has generally warm weather for most of the year. An asphalt binder that has a very low stiffness at high temperatures would be a poor component choice for such a scenario. Such a binder would result in excessive deformations during the high-temperature summer months.

Selection of high-quality ingredients is necessary but not sufficient to ensure that the mix and pavement structure will perform well. In other words, selection of the

best ingredients does not guarantee the best end product. Mixture design is the next most important step to ensure that the ingredients are combined in optimal proportions such that the resulting composite has an optimal performance; this involves selecting the optimal gradation for the mineral aggregate and binder content by weight of the mix. Continuing the example from the previous paragraph, let us consider a scenario wherein one selects an asphalt binder that is appropriate for a pavement in a warm weather location. In this example, proper selection of the binder does not guarantee that the mixture and pavement will perform well. Using higher than optimal percentages of the binder in the mixture will result in excessive deformations and premature failure. Similarly, using lower than optimal percentage of the binder will result in durability issues.

The third source of failure is the design of the pavement structure. The design of a pavement structure involves determining the different pavement layers (subbase, base, HMA, etc.) and their thickness such that the stresses from the vehicles are transferred efficiently to the pavement foundation. Continuing with the previous example, let us now imagine that the component materials were selected and the mixture was designed such that the resulting asphalt mixture was of a very high quality and resistant to most distresses. However, the pavement structure could still fail if the thicknesses of the asphalt layer and/or the foundational layers beneath it are such that the stresses are not adequate to efficiently transfer the load from the vehicles. In other words, the performance of any material can ultimately only be interpreted in the context of the structure within which the material is used. Even the highest quality material will fail in a poorly designed structure. To further reinforce this idea, consider a simple-framed steel structure. Even when the highest quality steel is used in the structure, it can still fail if adequate cross sections are not provided to accommodate the applied loads. Summarily, even the best and superior performing asphalt mixtures can fail if used in an inadequate pavement structure. *A corollary to this is that with proper engineering and design it is possible to take advantage of superior performing mixtures in reducing the layer thickness and economizing the pavement design OR to use relatively poor performing but low cost mixtures by adjusting the layer thickness in the pavement design.*

The last and final source of distress evolution is construction. Even the best or most optimal component material selection practices, mixture design, and pavement thickness design will not result in satisfactory pavement performance if the pavement structure is not built properly. Sound construction practices are required to ensure that the asphalt mixtures are produced and pavement structures are built according to specifications.

8.3 Failure Mechanisms

In this section, we will discuss the mechanisms of the different kinds of failures that are most commonly encountered in flexible pavements. These mechanisms will be discussed from the point of view of the four different sources of distress discussed

in the previous section. In each of the following subsections, we will discuss the mechanism of failure, the role of pavement structure, mixture design, and component material selection in driving the failure mechanism. For the purposes of this section, we will not consider the role of construction practices in driving these forms of failure.

8.3.1 Rutting

Rutting is the result of accumulation of plastic strains or permanent deformation that occurs in the different layers of the pavement structure due to the action of repeated traffic loads. The component materials, mixture design, and structural design of the pavement dictate rutting or permanent deformation along wheel path.

In terms of the pavement structure, rutting that is visible on the pavement surface is due to the accumulation of plastic strains in all the layers of the pavement structure, i.e., plastic strains in the subgrade, subbase, base (stabilized or unstabilized), and the asphalt layers. Consider a scenario where rutting is the primary cause of premature pavement failure. In this case, rutting is evident from the examination of the pavement surface (Fig. 8.1). However, based on an examination of the surface alone, it is impossible to discern whether such failure is predominantly concentrated in the asphalt layer or whether it is a reflection of the excessive permanent deformation that has occurred in one or more of the foundation layers of the pavement structure or a combination of the two. In the example shown in Fig. 8.1, a trench was used to examine the pavement cross section. The examination revealed that the topmost layer of the asphalt pavement failed in rutting, whereas the other asphalt layers beneath it were almost free of any deformation. In this case, rutting is attributed to the top asphalt layer. For the remainder of this section, we

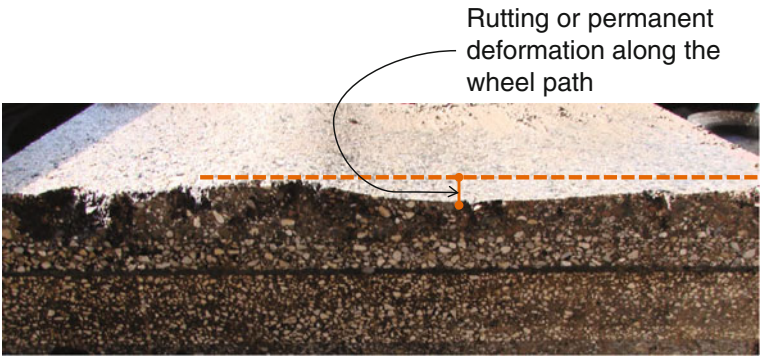


Fig. 8.1 Photograph of an asphalt section cut out of a pavement showing rutting in the asphalt layer along one of the wheel paths (Photograph by D Little)

will restrict our discussion to rutting that occurs in the asphalt layer and examine the mechanisms associated with the component materials and mixture design that dictate rutting.

There are two different mechanisms that drive permanent deformation in the asphalt mixture: densification and plastic or permanent shear deformation. As the name suggests, densification occurs due to an overall change in the volume of the material due to the repeated action of the wheel load. Asphalt mixtures are typically compacted to approximately 92–94% of their maximum density when placed in the field. In other words, after being placed in the field, asphalt mixtures have 6–8% air voids. These air voids typically get compacted due to the repeated action of the traffic wheel loads over the course of time. As such, the density of the asphalt mixture increases to approximately 96% with a corresponding change in the volume of the material. Since this change in volume occurs predominantly along the wheel paths, it will reflect in the form of rutting. However, the contribution of densification to the overall permanent deformation of an asphalt mixture is typically small.

The permanent deformation due to accumulation of plastic strain has a much more significant contribution to rutting observed in asphalt pavements. The accumulation of permanent deformation in the asphalt mixture is driven by several factors related to the component materials as well as the mixture design. The most significant factors are:

- shape, angularity, gradation, and to some extent toughness of the aggregate particles,
- the resistance of the asphalt binder to plastic deformation (as a function of temperature and age), and
- binder content in the asphalt mixture.

Note that the materials design engineer can manipulate most of the aforementioned variables to ensure optimal resistance to permanent deformation. We will briefly discuss the role of each of the above factors below.

The aggregate skeleton plays a very important load in transferring the loads from the asphalt layer to the pavement foundation. From a macroscopic viewpoint, the asphalt mixture can be treated as a homogenous material with certain mechanical properties that facilitates such load transfer. Assuming that the asphalt mixture is a homogenous mixture greatly facilitates the design of a pavement structure (i.e., to determine adequate layer thickness). However, in order to better understand the mechanism of distresses, it is important to examine and treat the asphalt mixture as a heterogeneous composite with a well-defined microstructure. In this context, the term microstructure is defined as the spatial distribution of the different component materials in the mixture. As discussed in chap. 6, an asphalt mixture can be treated as a composite of coarse mineral aggregate particles bound together by a continuous matrix of mastic or mortar (i.e., asphalt binder mixed with fine aggregate particles). Considering the internal structure of the mixture, the load applied to the surface of the mix is distributed via a series of stress transfers from one aggregate particle to its immediate neighboring aggregate particles either via direct contact or via a

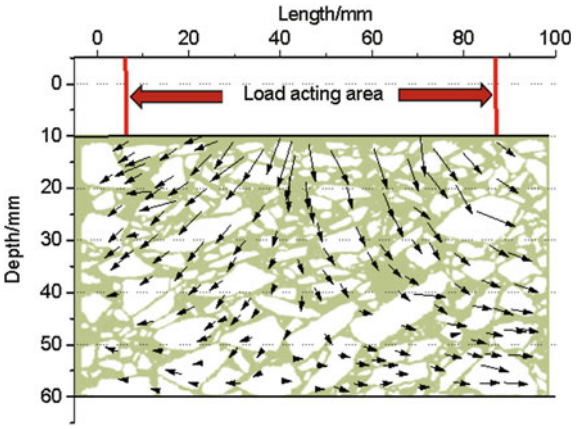


Fig. 8.2 Schematic showing transfer of loads within the asphalt mixture (Reprinted from Construction and Building Materials (Hu et al. 2015), with permission from Elsevier, copyright © 2015)

mastic or mortar interface between the particle (Fig. 8.2). With this general understanding of the load transfer mechanism, it is now possible to develop a better understanding of the factors that lead to rutting or permanent deformation.

In an asphalt mixture, permanent deformation is a manifestation of the permanent change in its internal structure due to external loads. Properties of the aggregate such as shape, angularity, and gradation dictate the propensity of the mixture to withstand such changes under the action of the external load. For example, an aggregate skeleton that is made up of cubical and angular aggregate particles will have a much higher internal friction and resistance to deformation as compared to a mix that is made up of rounded and smooth aggregate particles (Fig. 8.3). As a result, aggregate particles that are cubical and angular will in general provide higher resistance to both recoverable and permanent deformation. Chapter 3 of this book discusses in more detail the methods and metrics that are used to quantitatively assess and compare the angularity of aggregate particles.

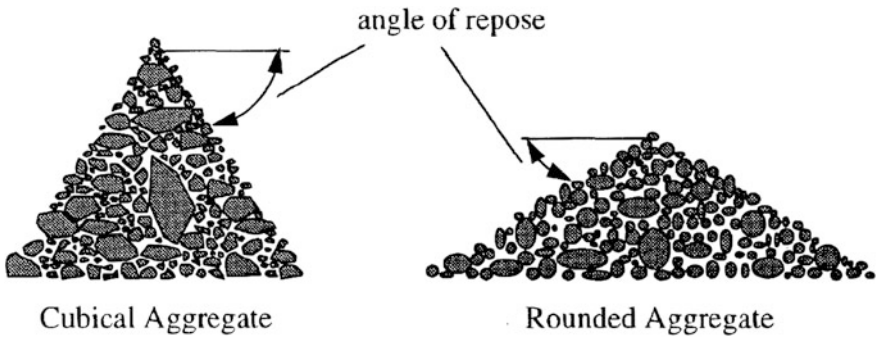


Fig. 8.3 Schematic showing the differences in angle of repose or internal friction for rounded and angular aggregate particles (McGennis et al. 1994)

In addition to the shape and angularity of individual aggregate particles, the distribution of these particles, i.e., the aggregate gradation is also important in dictating the efficiency of load transfer from one particle to another. Chapter 3 of this book has a discussion on the different types of aggregate gradations; of these, the dense and gap-graded aggregate skeletons are the most commonly used in structural mixes. The dense-graded mix relies on creating an aggregate skeleton that has close to the maximum possible density. This is achieved by filling in voids created by larger sized aggregate particles with successively smaller sized aggregate particles. Such a gradation would transfer loads from one large aggregate particle to another via a densely packed structure of successively smaller aggregate particles. The gap-graded mix relies on direct stone-to-stone load transfer between coarse aggregate particles creating a more efficient and direct pathway with least deformation through the mix. Recall from Chap. 3 that due to the high contact stresses in such a gradation, it is important that the gap-graded mix only be used with very high-quality and tough aggregates. Note that in the dense-graded mix, the mortar and the mastic incorporates the fine fraction of aggregate particles that play an important role in load transfer and resisting deformation. Whereas in the gap-graded mix, the asphalt mortar or mastic has less dominant role and is primarily responsible for holding the aggregate matrix together. In either case, the aggregate gradation is critical in dictating the efficiency of load transfer within the mixture, which in turn dictates both the elastic and permanent deformation in the mixture when subjected to an external load.

The properties of the asphalt binder also play an important role in resisting permanent deformation. As discussed in the previous paragraph, in a typical dense-graded mixture, load is transferred between coarse aggregate particles via the asphalt mastic or mortar that is made up of binder and finer aggregate particles. Therefore, in addition to the aggregate skeleton, the inherent ability of the binder to resist permanent or plastic deformation will also influence the overall rutting resistance of the asphalt mixture. Recall from Chap. 2 that asphalt binder is a time- and temperature-dependent viscoelastic material. When subjected to an external load the asphalt binder will deform with time. Depending on the magnitude and duration of the load, as well as the temperature of the pavement, a part of this deformation will recover but the remaining will be permanent. The composition of the binder strongly dictates its ability to resist plastic deformation. As a result, a binder that is typically more resistant to plastic deformation is desirable to produce a rut-resistant mixture. One notable difference with asphalt binders is that, unlike aggregate structures, the *total* deformation in the binder is not necessarily directly related to the overall deformation. Rather, only the permanent deformation of the binder contributes to rutting. For example, some highly polymer modified binders have relatively low stiffness and result in high total deformations for a given load, but most of this deformation is recoverable and not permanent.

Finally, the percentage of binder in the mix also governs the overall resistance of the mixture to permanent deformation. With the exception of very low temperatures (close to freezing or below), the stiffness of the binder is orders of magnitude lower than the stiffness of the aggregate particles. As a result, under general operating

conditions, the asphalt binder acts as the weak link in the internal load transfer process. A mix that has the ideal binder and aggregate, i.e., a binder with relatively very low propensity to accumulate permanent deformation and an aggregate that is cubical, angular with an efficient gradation, can still be prone to permanent deformation if the mass percentage of binder in the mix is higher than optimal.

In summary, a mix with rounded aggregates that does not provide efficient particle-to-particle and/or a mix with an aggregate gradation that does not provide an efficient pathway for internal distribution of stresses will be prone to higher recoverable or elastic deformations as well as permanent deformations. Also a mix with an asphalt binder that has a higher propensity to accumulate permanent deformation and/or with higher than optimal mass percentage of the binder will be more prone to permanent deformation.

8.3.2 Fatigue Cracking

Fatigue cracking is the result of nucleation of microcracks that grow and coalesce with each other to form larger macroscopic cracks due to the repeated action of wheel loads. Fatigue cracking is also referred to as alligator cracking on account of the fact that it manifests itself on the pavement surface in the form of a network of cracks along the wheel path (Fig. 8.4). As in the case of rutting, premature fatigue cracking of the pavement can also be attributed to issues related to component materials, mixture design, or pavement design.

In terms of the pavement structure, premature fatigue cracking can occur if the thickness of the asphalt and underlying layers is such that it induces unacceptable levels of tensile stresses in the asphalt mixture. Also, unlike rutting, fatigue cracking occurs only in the asphalt layer while the base and subbase layers of the pavement structure are assumed to be incapable of handling any tensile stresses or developing fatigue cracks. The only exception to this is in the case of cement or asphalt stabilized base. A couple of important things to note are as follows. First, an "unacceptable level of tensile stresses in the asphalt mixture" is a highly material-dependent property as will be discussed in more detail in the following paragraphs. Second, although fatigue cracks occur in the asphalt layer, the thickness and properties of all the pavement layers (base, subbase, etc.) dictate the state of the stress and resulting strain in the asphalt layer and ultimately the rate at which fatigue crack growth is likely to occur in the pavement. In other words, the magnitude of strain in the asphalt layer is a function of the applied load, stiffness of the materials used in different layers (including base and subbase), and the thickness of each layer. The fatigue life of the pavement is then dictated by the inherent fatigue cracking resistance of the asphalt mixture at that strain level. Therefore, *the growth of fatigue cracks is very intimately connected with both the inherent resistance of the asphalt mixture to cracking as well as the pavement structure within which such an asphalt mixture is used*.

Fig. 8.4 Typical fatigue cracks on a pavement section along the wheel path (Photograph by A Bhasin)

As a side note, in most cases, pavement design methods do not establish or take advantage of this connection and evaluate the fatigue cracking resistance of component materials, mixtures, and pavement structures in isolation from one another. Also this is mostly due to logistical issues rather than technical limitations. For example, the thickness design of the pavement is often times carried well in advance of component material and mixture design for bidding and letting purposes. In this case, the designer assumes typical values for the fatigue cracking resistance for the type of asphalt mixtures that are expected to be used. This approach (of designing pavement thickness not tailored to properties of a specific mix) inherently precludes the designer to take advantage of materials and mixtures with very high tolerance to fatigue cracking and reduce the pavement thickness and associated cost for a given design life.

For the sake of simplicity, and consistent with common practice as of this writing, let us isolate pavement structure from the component material selection and mixture design process. In other words, we now focus on the role of component materials and mixture design (i.e., percentage of binder and aggregate gradation) on the fatigue cracking resistance of an asphalt mixture for a given pavement structure. Recall that in the case of rutting, both aggregate and asphalt binder played an important role in resisting permanent deformation; perhaps it is not inaccurate to say that aggregate

properties and aggregate skeleton play a more dominant role when it comes to rutting. However, in the case of fatigue cracking, asphalt binder is the primary component that resists crack growth. For a given load, the aggregate structure surrounding the binder dictates the stresses experienced by the binder, but it is the ultimately the fatigue resistance of the binder that will inhibit the nucleation and growth of fatigue cracks. Therefore, in terms of component material properties, the properties of the binder are very important in dictating the ability of the mixture to resist fatigue crack growth. A comprehensive review on the measurement of fatigue resistance of binder can be found in the literature (Hajj 2017). The mechanism of fatigue crack growth and the role of component material properties and mixture are discussed in the following paragraphs.

The main characteristic of fatigue cracking is that the crack nucleation and growth does not occur precipitously but rather incrementally over several load cycles (typically hundred thousands or millions). In order to understand the mechanism of fatigue crack growth in asphalt pavements, it is more convenient to start with a review and a simplified understanding of the fatigue crack process in elastic materials. One of the basic properties of an elastic material is its tensile strength. A material fails if the applied stress exceeds the tensile strength or capacity of the material. For example, consider a simple cylindrical rod made up of a material with a tensile strength of 100 psi. For the purposes of this example, we will also ignore the variability associated with the measurements of such properties. Now, if we were to design the cross section of this bar for a given load such that the tensile stress is no more than 50 psi, we would expect that the bar will not fail because the applied stresses are well below the strength of the material. It is important to emphasize here the two different terms that we are using in this discussion: stresses experienced by the material and tensile strength of the material. The former is dictated by what we do with the material and is dependent on the geometry of the material in a given structure and the applied load. Whereas, the latter, i.e., the tensile strength is a material property and treated as a constant (for a given set of conditions). Therefore in this example, our hypothetical bar will not fail when used with a size and subjected to a load that results in a tensile stress of 50 psi the first time such a load is applied. However, when the load is cyclically applied and removed several times (perhaps a few hundred thousand or few million depending on the type of material), the bar eventually cracks and fails even though none of the load repetitions ever exceeded the original tensile strength of the material. This form of failure is referred to as fatigue cracking, and it is well recognized in several areas of engineering for over a century and is an area of specialization in itself.

In our simple example, the next logical question would then be: Why does the material fail over several cycles when the applied load in any given cycle does not exceed the tensile load of the material? To answer this question, we must consider the microscopic nature of the material and the concept of localized versus far field stresses. First, materials are heterogeneous. This is either by design (e.g., composites) or in the case of homogeneous materials heterogeneity is due to microscopic defects or flaws (Griffith 1921). As a result of the heterogeneity and/or the presence of defects,

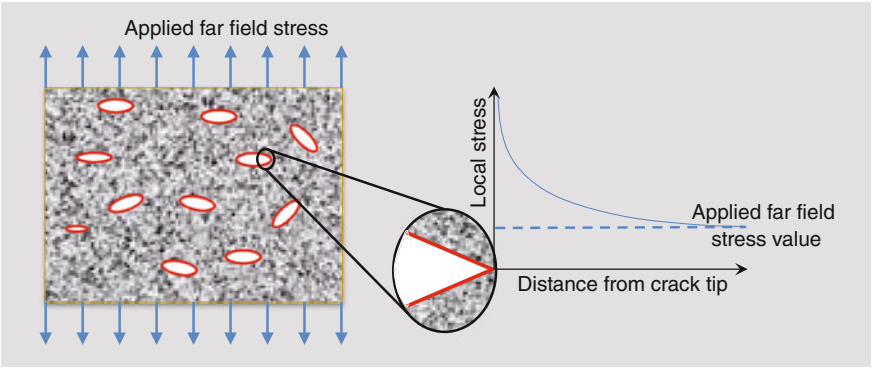


Fig. 8.5 Defects and local stresses in a material

when a material specimen is subjected to a certain level of external stress, the localized stresses at vulnerable locations within the material can be significantly higher than the average stresses experienced by the material far away from the defect. Figure 8.5 illustrates the localized amplification of stresses in the immediate vicinity of a microscopic defect. Also revisit Fig. 2.13 from Chap. 2 that shows the localized stress amplification in an asphalt binder sample at a microscopic length scale. Such localized increases in stresses enable the nucleation of microscopic cracks or new defects or cause the existing defects to grow in size. This process gradually occurs over multiple load cycles until macroscopic cracks form and coalesce leading to eventual material failure. A more mathematical description of this process can be found in the second part of this book in Chap. 14. Some basic methods to quantify fatigue cracking will be discussed in the subsequent subsections.

8.3.3 Transverse Cracking

Transverse cracking, as the name suggests, manifests in the form of cracks that occur transversely across the pavement section. In most cases, such cracks also occur periodically. Typically, transverse cracks in asphalt surfaces occur due to one of two reasons. First, the asphalt layer may be an overlay on top of an existing jointed or concrete pavement. In this case, cracks in the underlying pavement structure can create stress localization on the surface immediately above the crack. This in turn results in the creation and propagation of a crack on the overlaid surface, reflecting the crack beneath it. For this reason, such cracks are also commonly referred to as reflective cracking. Reflective cracks can be mitigated by a variety of different methods such as the use of interfacial layers or mixes that are specifically designed to resist the growth of such cracks. Also, reflective cracks are primarily driven by deficiencies in the structure. The second reason or form of transverse cracking is more relevant to this context of this book and chapter and is

Fig. 8.6 Typical transverse cracking (Photograph taken by A Bhasin)

referred to as low-temperature cracking. In the remainder of this section, we will restrict ourselves to this particular form of cracking Fig. 8.6.

Transverse, or low-temperature cracking is primarily dictated by the material properties. Unlike fatigue cracking, this form of cracking can occur more precipitously. To understand the mechanism that drives transverse cracking, let us first start by examining the behavior of an elastic material in a pavement structure. A segment of a continuous pavement without any joints can be treated as a geometry with fixed ends. In this case, when the temperature drops, the material in this segment would shrink. The thermal strain in this case would be $\alpha\Delta T_1$, where α is the coefficient of thermal expansion and ΔT_1 is the change in temperature. However, due to the fixed ends, this deformation is not allowed to occur and instead the segment develops a tensile stress, $E\alpha\Delta T_1$, that is dependent on the material stiffness, E. For an elastic material, this tensile stress will remain until the segment reverts back to its original neutral stress temperature. Also, if the temperature were to drop further by ΔT_2 after some time, say an hour, then the tensile stress would increase by $E\alpha\Delta T_2$. As the temperature drops, the tensile stress would continue to increase until (Fig. 8.7) it gets a point where it exceeds the tensile strength of the material; at this temperature, the material will fracture creating a transverse crack to relieve the thermal stress. This same process occurs in a viscoelastic material such as an asphalt mixture with one significant difference, i.e., a viscoelastic material tends to relax over time relieving and preventing the stresses from exceeding the strength of the material. Figure 8.7 shows a schematic of these stresses being relieved as the temperature decreases. However, an asphalt material that has a very high stiffness at low temperatures and/or a slow rate of relaxation will not relax effectively resulting in a build up of stresses that ultimately exceed the strength of the material and cause it to crack.

This mechanism can now be used to explain the role of mixture and component materials in driving low-temperature cracking. In terms of mixture properties, the binder content and aggregate gradation will dictate the internal state of stress of the binder and govern the growth of thermal cracks. Properties of the component materials, particularly the binder, are very important in dictating the propensity

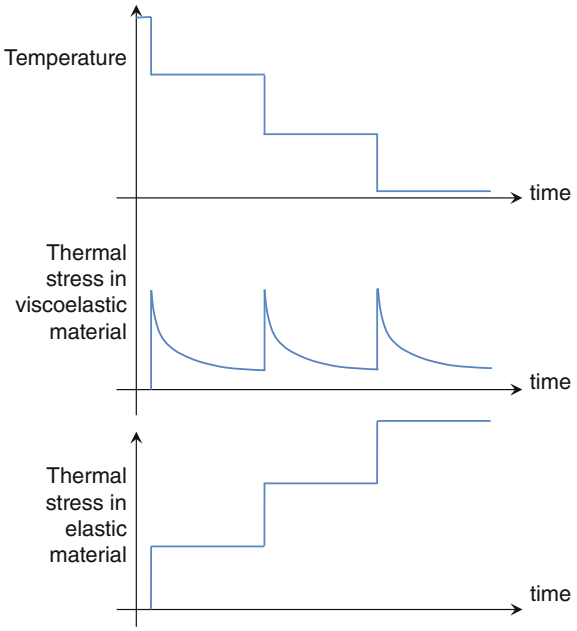


Fig. 8.7 Typical thermal stresses in elastic versus viscoelastic material; the schematic shows a step-by-step decrease in temperature for clarity and simplicity whereas realistically this process is continuous

toward low-temperature or transverse cracking. For example, the coefficient of thermal expansion of both these materials will dictate the magnitude of thermal strain that is developed in these materials per degree of drop in temperature. More importantly, the stiffness, rate of relaxation, and strength of the binder are critical in determining the resistance of the mix to thermal cracking. Recall that it is the asphalt binder in the mixture that will ultimately resist the thermally induced tensile stresses (unbound aggregate particles do not have the ability to withstand any tensile stresses) and also relax to relieve these stresses. This also explains why an ideal binder would have low stiffness, high rate of relaxation and high tensile strength at low temperatures. A low stiffness would imply that for a given thermal strain, the tensile stresses are low. A high rate of relaxation would result in faster relaxation of the thermally induced stresses and a high tensile strength would ensure that the thermal stresses are well below the strength of the material reducing the chances of thermally induced fatigue.

One final note in the context of low-temperature-induced transverse cracking is that it is not just the lowest pavement temperature that dictates this distress but the *rate at which temperature drops* in the pavement that is equally, if not more, important. In other words, low-temperature cracking can also occur if the temperature in a particular geographic region drops precipitously due to fast moving weather patterns. In such cases, the build up of stresses in the pavement can occur at a much faster rate than the ability of the material to relax ultimately leading to a scenario where the stresses exceed the strength of the material and result in a transverse crack. Finally, significant diurnal temperature changes over several months or years can also induce a fatigue type failure following a similar

mechanism. Therefore the term thermally induced cracking may be more appropriate compared to low-temperature cracking.

8.3.4 Moisture-Induced Damage

In addition to rutting and cracking, moisture-induced damage is also one of the most common forms of pavement distress. The most common visible form of moisture-induced damage is evident in the form of potholes. By some estimates, the repair of potholes costs agencies billions of dollars in maintenance; users of road ways pay even more money toward the maintenance of their vehicles due to damage caused by potholes (Cohen et al. 1997). While potholes are the most evident form of moisture-induced damage, moisture can act as a catalyst to exacerbate other forms of distress such as rutting and cracking. All three factors, the pavement structure, mixture design, and selection of component materials, play a very important role in the ability of the pavement to withstand the impact of moisture. Before we address the role of each of these factors, we must first try to understand the mechanism that drives moisture damage. Note that while moisture can also influence the properties of the pavement foundation (e.g., granular base or fine subgrade materials), for the purposes of this discussion, we will focus on the influence of moisture on the asphalt layer of the pavement structure (Fig. 8.8).

The moisture damage process begins with the presence of moisture gradients across the pavement section. For example, moisture may be available to the pavement due to naturally occurring rainwater on the surface. In this case, gravity, capillary action, and the stresses exerted by the vehicle tires on the pavement surface can force the moisture into the asphalt layer. Alternatively, moisture may also make its way to the asphalt layer through the foundation or embankment that may have higher moisture contents compared to the surface. In this case, the moisture gradients and capillary action can cause moisture to make its way into the asphalt layer.

Once the moisture makes it to the asphalt layer, it is transported within the asphalt mixture via the capillaries or interconnected air voids in the mixture. Moisture present in the voids and capillaries of the asphalt mixture can deteriorate the mixture by "erosion" of these capillaries. This phenomenon is particularly more exaggerated in open graded mixes that have much more open pathways and capillaries for the moisture to travel. In this case, the movement of vehicles can create a pumping action of the moisture within these capillaries causing the material around the capillaries to erode and deteriorate. Readers can refer to additional literature for a more detailed description of this phenomenon (Kringos et al. 2008).

Moisture trapped in the capillaries and voids of the mixture diffuses through the bulk of the mastic, albeit at a much slower rate, eventually making its way to the mastic- or binder-aggregate interface. During this process, the presence of the moisture in the mastic can soften and weaken the material. This process is referred as loss of cohesion in the mastic or the binder. Finally, once the moisture reaches the

Fig. 8.8 Typical example of a pavement section with severe moisture induced damage (Photograph by A Bhasin)

interface, the interfacial forces between the water, binder, and aggregate surface cause the water to strip the aggregate surface of the binder. This process is also referred to as stripping or loss of adhesion between the binder and the aggregate surface. At this point, the binder is not effective at holding the aggregate particles together. As a result, the aggregate particles break loose and are dislodged from the mixture eventually creating a pothole. An analogy to understand this process is that of a grease covered frying pan. In an attempt to clean such a pan, it is often convenient to leave it soaked under water for a while. Water (being a highly polar liquid) diffuses through the grease and eventually separates it from the substrate or the surface of the pan. Adding hot water can accelerate this process and so does addition of soap to the process. In the case of an asphalt mixture, the surface of the pan is somewhat similar to the surface of the aggregate and the grease is similar to the asphalt binder. A more detailed description of this interfacial process is discussed in Chap. 4 of this book. Also, readers are referred to additional literature for a more rigorous breakdown of the different physical, mechanical, and chemical processes that result in cohesive and adhesive failure (Hefer et al. 2005; Caro et al. 2008) Fig. 8.9.

Based on the above mechanism, we can now discuss the role of the pavement structure, mixture design and component materials in dictating moisture-induced damage. In terms of the pavement structure, it is absolutely critical that the

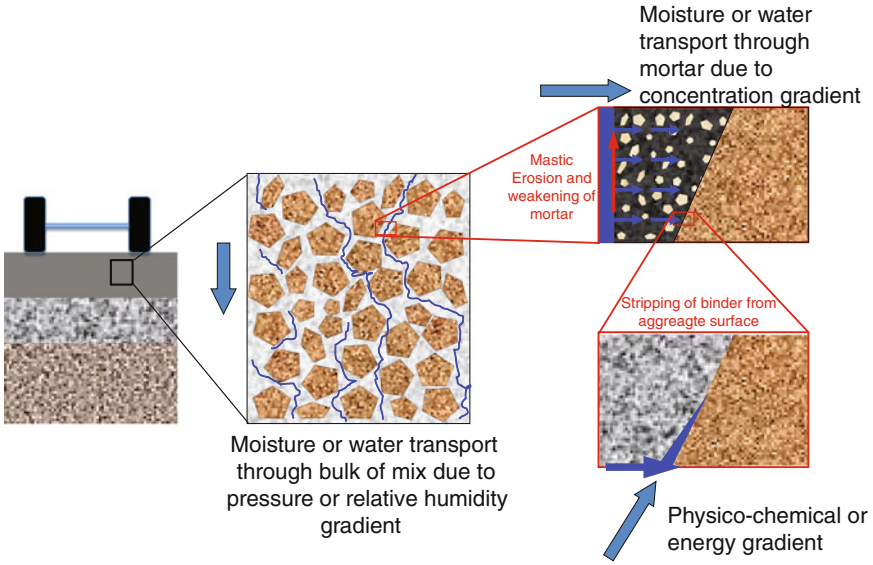


Fig. 8.9 A schematic showing the different stages involved in the moisture damage process

pavement is designed in a way that allows all moisture to be transported and drained away from the pavement structure. The pavement geometry and design of drains along the pavement are absolutely critical to ensure the durability and longevity of the pavement structure. As of this writing a significant research study was underway to identify and document best practices to achieve this. Readers interested in more details on this can find additional information in NCHRP Project 01-54.

In terms of the mixture design, the role of air voids and interconnected air voids is critical in dictating the susceptibility of the mixture to moisture-induced damage. Studies have shown that, all other factors being the same, a mixture with very large or open and interconnected air voids is relatively less prone to moisture damage. This is because the mixture will allow moisture to not only enter the matrix easily, but also exit the matrix easily. Similarly, a mix with very small air voids that are not well connected is also relatively less prone to moisture damage. In this case, it is very difficult for the moisture to enter the bulk of the mixture and cause serious damage. However, an intermediate level of air voids may be the least desirable in terms of resistance to moisture damage. In this case, the air voids may be just sufficient for moisture to slowly migrate into the bulk of the mix over time and possibly accelerated by the action of loads but not allow the moisture to escape easily. As a result, moisture is trapped in the mix and deteriorates it over time. Figure 8.10 is based on the results from a study (Arambula et al. 2007) investigating the influence of air void distribution on moisture-induced damage. As can be seen from the figure, there is a pessimum air void size at which the extent of moisture damage is highest. Readers must keep in mind that the relationship

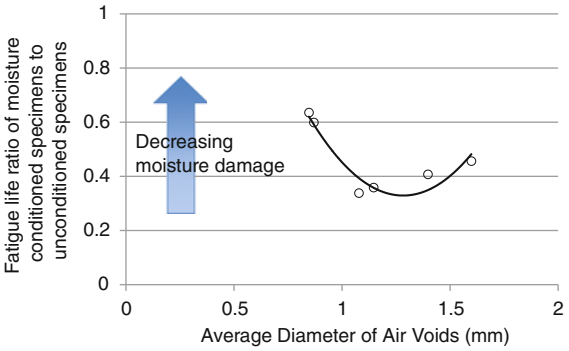


Fig. 8.10 Results from a study showing the impact of air void on moisture-induced damage (Masad et al. 2006)

between the size of air voids and total volume percent of air voids is not straightforward. Many factors such as particle size and gradation will dictate the size distribution of a specific volume of air voids. For example, a mixture with 4% air voids can have a few large air voids that are spatially dispersed or depending on the aggregate size and gradation, it could also have several smaller dispersed air voids. Tools such as X-ray tomography are often used to better understand such correlations and establish the connection between percentage volume of air voids, void size distribution, and ultimately propensity to moisture damage.

8.3.5 Aging

The properties of an asphalt mixture inherit the changes that occur in the properties of the asphalt binder due to oxidative aging over time. Aging of asphalt binders and consequently that of asphalt mixtures is not a distress by itself but it does have a significant influence on the rate of evolution of all the distresses discussed above. In some cases, such as permanent deformation or rutting, aging can help improve the ability of the asphalt mixture to resist the distress. In other cases, aging can amplify or have a mixed influence on the distress evolution. Since aging is not a form of distress by itself but rather affects the rate of evolution of other forms of distress; in this section, we will not discuss the role of pavement, mixture, or component materials, but restrict our discussion to the impact of aging on rutting, fatigue cracking, low-temperature cracking, and moisture-induced damage.

As discussed in Chap. 2 of this book, asphalt binder undergoes oxidative aging during the service life of the pavement. Aging results in the formation of more polar molecules that form agglomerates or more associated structures which in turn increases the stiffness of the binder and consecutively the stiffness of the asphalt mixture. An increase in stiffness also tends to increase the resistance of the material to plastic deformation. Therefore, over time aging helps improve the resistance of the asphalt mixture to permanent deformation or rutting. In fact, this is also the reason why rutting is more of a concern during the early years of a pavement life.

Also, mixture production technologies such as warm mix asphalt (WMA) that use lower production temperatures tend to reduce the extent of oxidative aging in asphalt binders and increase the susceptibility of asphalt mixtures to rutting during the early years of the pavement. On the other hand, use of reclaimed asphalt pavement (RAP) that contains not only aggregates, but also significantly aged asphalt binder in combination with virgin materials results in a mixture that is relatively more stiff and resistant to rutting. In many cases, both WMA and RAP are used in combination with each other to balance the effects of reduced oxidation in the virgin binder due to reduced mixing temperature and effects of the aged binder from the RAP.

While aging may increase resistance to rutting, it tends to have a mixed impact on fatigue cracking and low-temperature cracking. Let us first discuss the impact of aging on fatigue cracking. Studies on asphalt binders and mastics have shown that although aging tends to increase their stiffness and reduce their ductility, aging also tends to increase the strength and resistance to cracking, up to a certain extent (Arega et al. 2013; Sultana and Bhasin 2014). However, since the ductility of the binders reduces with aging, the nature of crack formation and growth also changes. The binder becomes increasingly more susceptible to brittle failure, albeit at a relatively higher stress level. The overall impact of aging on the rate of crack growth is therefore governed by the relative change in these properties and failure modes.

As discussed before, in order to resist low-temperature or thermal cracking the binder must have (1) a low stiffness, which would result in reduced thermal stresses, (2) high rate of relaxation, which would prevent the stresses from building up as the temperature drops, and (3) high tensile strength, which would ensure that the accumulated stresses remain below the strength of the material to prevent cracking. With aging, the stiffness of the binder increases and it reduces its ability to relax, which makes the binder more susceptible to low-temperature cracking. At the same time the tensile strength of the binder may also increase. This would partially offset the negative impact of increase in stiffness and reduction in relaxation rate. It is the relative change in the magnitude of these properties that will ultimately dictate the ability of the binder and the mixture to resist low-temperature cracking.

8.3.6 Bleeding or Flushing

In some cases, a very thin layer of excess asphalt binder oozes to the surface of the pavement along the wheel paths. This type of distress is often clearly visible to the naked eye in the form of a slick reflective surface that forms along the wheel path (Fig. 8.11). Bleeding typically occurs when the asphalt mixture has insufficient air voids and/or higher than optimal binder content. These factors in combination with consolidation due to traffic loads cause some of the binder to rise to the top of the pavement surface along the wheel paths. The excess binder significantly reduces the pavement texture and creates a very unsafe riding surface. Although bleeding is not

Fig. 8.11 Typical example of a pavement section with bleeding (Photograph taken by A Bhasin)

as common as other forms of distress, it can create a significant safety hazard in terms of reduced skid resistance. Finally, based on this description it is easy to recognize that bleeding is more of a mixture design problem rather than a component material or pavement structure related problem.

8.4 Terminology and Typical Approaches to Characterize Distresses

In the previous section, we developed a basic understanding of the most common forms of distresses that ultimately cause pavement failure. In this and the following sections, we will discuss some general approaches and commonly used test methods to measure the resistance of any given asphalt mixture to these distresses. We will first start by introducing some broad level ideas and terminology related to the characterization of asphalt materials.

8.4.1 Measuring Performance Indicators and Material Properties

In the asphalt industry, the business of evaluating component materials or asphalt mixtures for their resistance to one or more forms of distress is based on two broad approaches with some overlap: (1) indicator type or torture tests and (2) mechanistic

tests that measure material properties. Each of these approaches has certain advantages and disadvantages. Understanding the philosophical differences between these two approaches is critical not only for remainder of this chapter and book, but also more importantly for the future evolution of the asphalt materials and pavement industry.

We will first start with a very simple definition of material property, because we will use this term several times in the remainder of this chapter. *In the simplest sense, material property is a property that belongs to the material and is not influenced by the geometry or other extraneous conditions used to measure the property*. Consider a scenario where we measure the tensile strength of a steel bar. We can take such as steel bar and affix it to a universal testing machine (UTM) and apply a tensile load while measuring its deformation. The maximum load that the steel bar can withstand before it breaks is a measure of its tensile strength. But the load itself is only an indicator and not a material property. The material property would be the maximum tensile stress at failure, which is calculated by measuring the diameter of the rod and dividing the load by the area of cross section. We can use indicators to compare the strengths of different grades of steel. For example, a grade of steel with higher failure load than the other can be considered to be stronger in tension. We can also use indicators as a quality control tool, i.e., we can sample steel rods as they are being manufactured, run the test, record the maximum load, and use it as an indicator of quality and consistency. In fact, if an indicator is all that is needed for a specific purpose, then one can envision several different surrogates for an indicator test that do not necessarily require a sophisticated universal testing machine.

There are several advantages of using an indicator compared to a material property. For example, we only need to measure the maximum load before failure and we need not measure the diameter of the steel rod. This saves time because there is one less measurement to make and also some hardware cost because we do not need an extra device to measure the rod diameter. We also save time and expertise by not having to conduct calculations of stress. On the other hand, there are also several disadvantages of using an indicator compared to a material property. First, any variability in the cross-sectional area will add to the variability in the measured maximum load. If we were using the maximum tensile stress at failure, this variability would be accounted thereby increasing our confidence on the test results. Second, the indicator has no practical value from a design or prediction point of view. In contrast, the maximum tensile stress can be used to assess the possibility of tensile failure of an arbitrary cross section subjected to external loads. Also, this information can used to economize the design of a cross section. For example, if a specific grade of steel has higher than typical maximum tensile stress it is possible that you would need a smaller cross section for this specific grade thus economizing the cost of the project.

The above example is a simplistic illustration of the differences between measuring an indicator for simple comparison, material screening, or quality control purposes versus measuring a material property that can also achieve the same benefits and much more (e.g., predicting performance, optimizing design, and being able to accommodate changes in material types). While the latter has several

advantages, it also requires additional time, resources and expertise to accomplish. *These two broad approaches of measuring a simplistic indicator and measuring material properties define the backdrop against which most of the tests are developed to assess component material performance and asphalt mixture performance*. Note that these two approaches are not distinct and there is some overlap. Finally, it is important to recognize that similar to simplistic material indicators, fundamental material properties can also be used for screening materials, quality control and other purposes. However, the opposite is not true, i.e. material indicators cannot be used to design materials, structures, or predict pavement performance or life. We hope that over time the industry evolves towards being more focused on the measurement and use of fundamental material properties for materials and pavement design. Researchers and industry experts have developed hundreds of tests, if not more, to evaluate susceptibility of materials and mixtures to different distresses based on one or a combination of these approaches. While a comprehensive list of all of these tests (assuming it were feasible to produce one) is beyond the scope of this chapter and book, a few examples are discussed in the following section. But before we visit a few examples of these tests, we need to discuss a couple of terms that are also often used in this context.

8.4.2 Concept of Continuum

The term continuum is often times used in the context of measuring material properties and evaluating their resistance to specific forms of distress. Readers can refer to Part II of this book for a more detailed description. This section briefly discusses this idea here because of its importance in the different test methods that are used to evaluate material performance.

The term continuum is very commonly used in mechanics to describe how a material is being treated for the purposes of measuring and modeling its behavior. It basically means that the material is being treated as a continuous entity without any discrete boundaries or discontinuities. If one were to cut out a infinitesimally small portion of such a material from a larger specimen, this small portion would be exactly the same as every other infinitesimally small portion of the material. For example, if we were to treat an asphalt mixture as a continuum, we would not consider that the mixture has discrete aggregate particles floating around in a matrix of asphalt binder or mortar. Rather we could treat the mix as a homogenous continuous medium without any discontinuities. Even when the material experiences distresses such as fatigue crack growth, we would not be interested in the specific location and size of the different cracks but rather we would be interested in the overall impact of these cracks on the material specimen. In other words, the continuum approach smears the influence of aggregate particles, cracks, etc. on the entire specimen and it is the overall response averaged over the entire specimen that is of interest to us. Similarly, in the case of evaluating an asphalt binder, the binder

can also be treated as a continuum without regards to the spatial distribution of individual molecules.

It is very important to note that the treatment of a material as a continuum does not imply that the internal structure of the material is not important. It only implies that for the specific purpose, the overall response of the material is of interest rather than the specific mechanisms of the components within the material that result in the response. For example, in the treatment of an asphalt mixture as a continuum, it does not imply that distribution of cracks, aggregate particles, or aggregate size is not important. In fact, it is precisely these factors that will dictate the behavior and performance of the entire mixture as a continuum. However, for the purposes of characterizing the overall performance of the mixture, it can be treated as a homogenous continuum. Similarly, the microstructure and molecular fabric of the binder dictate the properties of the binder but for the purposes of evaluating the behavior of a specific binder, it can treated as a continuum. One important note in this context is the idea of a representative volume element. The specimen for any material that is evaluated using mechanical tests and treated as a continuum must be much larger than the largest discrete entity in the material. Using the example of an asphalt mixture, the smallest dimension of a test specimen must be at least several times larger than the largest aggregate particle in order to serve as a representative for measurement of material properties.

8.5 Examples of Test and Analytical Methods to Characterize Properties and Distresses

In this section, we will discuss a few typical test methods that are used to characterize the properties of an asphalt mixture as well as its resistance to different forms of distresses. As introduced before, some of these tests are merely torture tests or tests that only provide simple indicators of expected performance. Other tests are more fundamental in nature and are focused on measuring material properties that are related to performance rather than simplistic indicators. The intent of this section is to introduce readers to the philosophy and broad approaches to characterize asphalt mixtures through some common examples. A comprehensive list of all test methods or a detailed step-by-step description is beyond the scope of this book, but adequate references are provided along with the descriptions for readers interested in more details.

8.5.1 Complex Modulus

Before we discuss tests that are focused on measuring different forms of distress, we will first discuss the measurement of the most fundamental material property, that is the stiffness of the material in an undamaged state. Let us first consider a material such as steel that can be approximated as a linearly elastic isotropic material. The

elastic modulus of steel, along with Poisson's ratio, are the most basic properties that are required to compute the stresses and strains in any given steel structure. Without this very fundamental piece of information, there is practically no engineering design or prediction exercise that can be meaningfully carried out. The same is true for asphalt mixtures; a measure of stiffness is the most basic piece of information that is required to estimate the deformation of a pavement structure under external loads and to rationally design a pavement structure. The only difference is that the asphalt mixture is not linearly elastic. Recall from Chap. 2 that the stiffness of the asphalt binder is dependent on the time history of loading, temperature, and age of the binder. The asphalt mixture inherits all of these characteristics. As a result, the modulus of the asphalt mixture is also dependent on the rate of loading, temperature, and age.

An asphalt mixture does not have a single number that can be used to describe its modulus. The time or rate dependency of the mixture is accommodated by characterizing the stiffness of the mixture as a function of time using either creep compliance $D(t)$ or relaxation modulus $E(t)$ or as a function of loading frequency using complex modulus $|E^*(f)|$. The term $|E^*|$ denotes the magnitude of the complex modulus, also referred to as dynamic modulus. The creep compliance, relaxation modulus, and complex or dynamic modulus are the same as discussed in Chap. 2 for asphalt binders. The only difference in this case is that these properties are measured for an asphalt mixture, typically subjected to small compressive strains or deformations. Also, as in the case of asphalt binders from Chap. 2, temperature dependency of the material is accommodated using the concept of time–temperature superposition and developing a master curve. The complex modulus of an asphalt mixture is the most important mechanical property of any asphalt mixture that is required for any rational method to design the pavement structure.

The measurement itself is typically carried out using a cylindrical specimen that is 100 mm in diameter and 150 mm in height (typically obtained by coring and sawing a superpave gyratory-compacted specimen that is 150 mm in diameter and about 175–200 mm in height). The device typically used is a loading frame that is capable of applying sinusoidal loads to the test specimen at different frequencies. The load is applied to achieve 50–70 microstrains in the specimen at varying frequencies (e.g., 25, 10, 5, 1, and 0.1 Hz) and different temperatures (e.g., −10, 4, 25, 38, and 54 °C). At any given frequency, a fixed number of load cycles are applied and the complex modulus is determined using the data from the last few cycles to avoid the influence of initial transients on the data collected. Also the test is typically conducted by using the same test specimen starting with the lowest temperature and testing at all frequencies (from highest to lowest) at each temperature. This particular progression of applied loads ensures that the test specimen is evaluated at its highest stiffness first before proceeding to conditions where the stiffness of the specimen reduces (i.e., higher temperatures and/or lower frequencies). This progression ensures that the test specimen accumulates minimal permanent deformation and change to its internal structure during the test process. For the same reason, the applied loads are such that the strain amplitude in the specimen

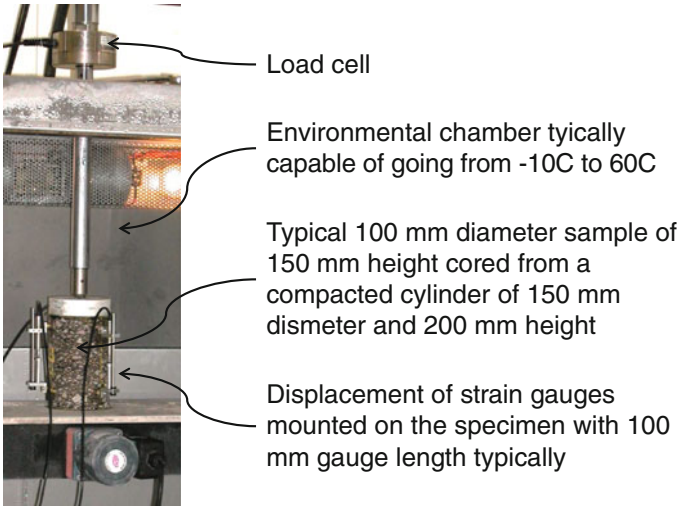


Fig. 8.12 Schematic of the set up and loads applied to measure complex or dynamic modulus of asphalt mixtures (Photograph by A Bhasin)

typically does not exceed 50–70 microstrains. Higher microstrains would cause the material to behave nonlinearly and induce damage. The choice of this range of microstrains is important because higher deformations could induce damage and smaller deformations would not generate a clear measurable response.

Figure 8.12 shows a schematic of the test set up used to measure complex or dynamic modulus of asphalt mixtures. Readers are referred to additional literature for a more detailed description on the research associated with the development of this test and a step-by-step description of the test procedure (American Association of State Highway and Transportation Officials 2005; Witczak et al. 2001). One interesting variation of this test method is to measure the dynamic modulus of a cylindrical test specimen that is about 150 mm in diameter and 50 mm in height by applying a compressive load diametrically across the specimen. This specific method was developed to be able to measure the modulus of specimens cored from the field, where it was not possible to have a height of 150 mm in all cases. Additional information about measuring dynamic modulus in this mode of loading can be found in the literature (Kim et al. 2004).

Finally, we should briefly discuss the typical results from measuring the complex modulus of an asphalt mixture and developing a master curve. Figure 8.13 shows the typical raw data collected from the measurement of complex modulus at different frequencies or loading times and temperatures. The abscissa on Fig. 8.13 shows the loading time (inverse of test frequency) on a log scale. The ordinate shows the modulus of the complex modulus on a log scale.

The asphalt mixture inherits the thermorheologically simple behavior of the asphalt binder as discussed in Chap. 2. Recall that a thermorheologically simple material is one that demonstrates a simple scalable relationship between the time-

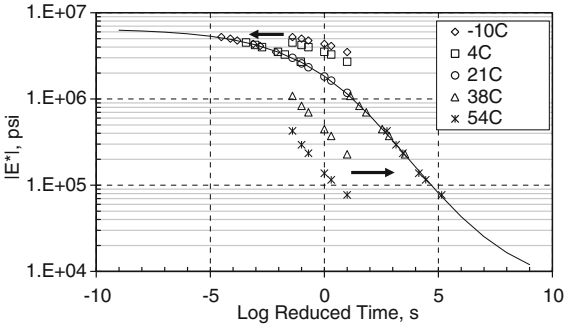


Fig. 8.13 Typical raw data collected from the measurement of complex or dynamic modulus and resulting master curve for an asphalt mixture

and temperature-dependent properties of the material. For the purposes of this example, let us define the reference temperature as 21.1 °C, which is the intermediate temperature at which all the data were collected. Now, one can observe from Fig. 8.13 that all the data from temperatures higher than the reference can be shifted to the right and all the data from temperatures lower than the reference temperature can be shifted to the left to form a single contiguous or smooth curve. In fact, as an exercise, readers can graphically conduct this exercise as follows. Each point from the temperature immediately below the reference temperature (4.4 °C in this case) can be shifted to the left to overlap with the existing data at the reference temperature. The "simple" part of the thermorheologically simple behavior is that the same horizontal shift is used for all the points at 4.4 °C to merge with (and also extend) the data at the reference temperature. This process of horizontal shifting blends data from the 4.4 °C temperature to the reference temperature and also extends the span of the data at the reference temperature. Next, the process is repeated with data from the next lower temperature, i.e. −10 °C in this case, extending the data at the reference temperature even more. Finally, the process is repeated for the two highest temperatures, except this time the horizontal shift is in the opposite direction. The final step is to plot the magnitudes of the four horizontal shifts (one each for −10, 4.4, 37.8, and 54.4 °C) against the temperature to produce the shift factor versus temperature relationship as shown in Fig. 8.14.

In practice, the process described above is not conducted manually but rather programmatically. Similar to the case of the asphalt binder, the master curve for the complex modulus of the mixture, $|E^*|$, is defined as shown in Eq. 8.1.

$$\log|E^*| = \delta + \frac{\alpha}{1 + \exp(\beta - \gamma \log \xi)} \tag{8.1}$$

In Eq. 8.1, α, β, γ, and δ are parameters that define the shape of the master curve. The term ξ refers to the reduced time, which is defined as the ratio of the true loading time (inverse of the frequency) to the shift factor as shown in Eq. 8.2.

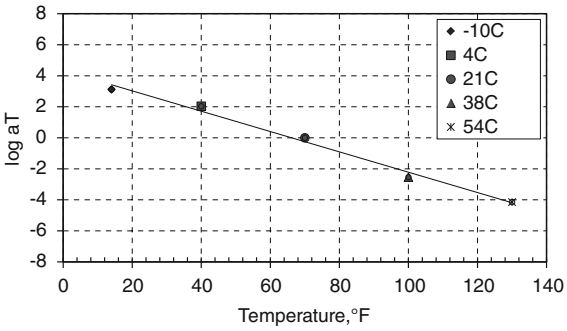


Fig. 8.14 Typical relationship between temperature and shift factors for the master curve shown in Fig. 13.13

$$\xi = \frac{t}{a(T)} \tag{8.2}$$

In Eq. 8.2, the term t is the loading of time of interest and $a(T)$ is the shift factor at the temperature of interest T. The concept of reduced time is central to the idea of a master curve. Simply put, the modulus of the material at any temperature and any duration of loading can be transformed to the modulus at the *reference temperature* at a *reduced time* of loading. Note that the term reduced time may be slightly misleading because depending on whether the temperature of interest is higher or lower than the reference temperature, the reduced time may in fact be greater than or less than the true time of loading of interest, respectively.

Once the complex or dynamic modulus data at multiple temperatures and frequencies are collected (five frequencies and five temperatures in this example), it is then used to determine the four shape parameters in Eq. 8.1 and the values for the four shift factors corresponding to each temperature: $a(-10)$, $a(4.4)$, $a(37.8)$, and $a(54.4)$. Note that the shift factor at the reference temperature is 1.0 by definition. These eight parameters (four shape parameters and four shift factors) are determined simultaneously by fitting the measured data to Eq. 8.1 to create a smooth master curve as shown in Fig. 8.13.

A couple of other observations regarding the master curve as shown in Fig. 8.13. First, the modulus of the mixture from the master curve at very small loading times (or high frequencies) at the reference temperature is an indicator of the behavior at very low temperatures and typical intermediate rates of loading. Similarly, the modulus at long loading times (or very low frequencies) at the reference temperature is an indicator of the behavior of the mixture at very high temperatures and typical intermediate rates of loading. Second, the curve has a distinct shape with two asymptotes at very low loading times (or high frequencies) and very long loading times (or low frequencies). These asymptotes are of course a reflection of the choice of sigmoidal function (Eq. 8.1) that was used to develop the master curve. However, readers must note that the choice of the sigmoidal function is based on the experimental observation of such asymptotes from testing. The phenomenological origins of these two asymptotes can be explained as follows. The

asymptote at very small loading times (or high frequencies) is because at very high loading rates and/or low temperatures, the behavior of the binder approaches that of an elastic material with little or no change in its modulus with a further increase in rate of loading or decrease in temperature. In fact, as discussed before, below a certain temperature referred to as the glass transition temperature, the asphalt binder behaves as a "glassy" elastic solid. A direct consequence of this is that below a certain temperature and/or above a specific rate of loading the modulus of the mixture is no longer affected by the temperature and rate of loading and its value reaches an asymptote. This behavior is also observed in the binder master curve. On the other extreme, at very slow rates of loading and/or high temperatures, the binder behaves as a fluid. Therefore, beyond a certain point an additional decrease in the rate of loading or an increase in the temperature does not change the contribution of the binder to the stiffness of the mixture. In fact, at this lower asymptotic value the binder makes very little to no contribution to the stiffness of the mixture and the modulus is mostly due to the particle-to-particle interaction in the aggregate skeleton. This asymptote is distinctly absent in the case of asphalt binders without any aggregates. Readers are encouraged to go through the accompanying example and carry out the exercise of developing and using a master curve to make predictions for the modulus of the material at different combinations of temperatures and frequencies.

8.5.2 Rutting

As discussed previously, rutting is one of the most common forms of distress in asphalt pavements. In this section, we restrict our discussion to some examples of methods that can be used to characterize the rutting resistance of asphalt mixtures. We will briefly discuss (1) simplistic methods that are based on measuring some kind of a performance indicator for materials screening purposes, and (2) methods that are based on measuring fundamental properties that can be used for materials screening purposes and performance prediction.

8.5.2.1 Simplistic Torture Tests to Compare Rutting Resistance of Materials

One example of a simple performance indicator or torture test is the wheel tracking test. Several variations of the wheel tracking test exist. In most cases, the test involves the use of a loaded wheel that cyclically tracks over a laboratory compacted test specimen at high temperatures. A couple of examples are the Hamburg wheel tracking device (HWTD) and the asphalt pavement analyzer (APA). The former has a loaded steel wheel that is in direct contact with the asphalt specimen typically submerged under water at high temperatures and the latter has a loaded

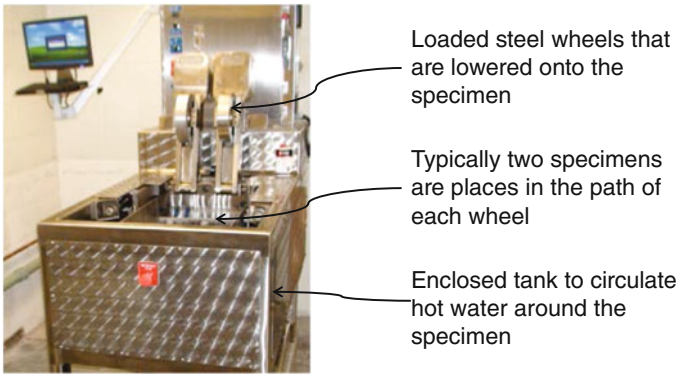


Fig. 8.15 An image of a typical Hamburg wheel tracking device (Photograph by A Bhasin)

grooved wheel that runs over a pressurized hose on top of the test specimen either in dry or wet conditions at multiple temperatures.

The HWTD is discussed here in some more detail and is one of many different variations of a wheel tracking test. The HWTD is commonly used by some state agencies in the USA as a go-no go or pass/fail test to screen asphalt mixtures susceptible to rutting (Fig. 8.15). The test is performed by oscillating an approximately 200 mm diameter and 47 mm wide steel wheel loaded with about 72 kg weight over a laboratory compacted asphalt mixture specimen that is about 65 mm in height submerged in water at high temperatures (typically at 50–60 °C). The permanent deformation of the specimen is recorded with reference to the number of passes of the loaded wheel. The test is typically conducted for 20,000 passes or until the specimen reaches a specified maximum deformation (e.g., 10 mm). Mixtures showing excessive susceptibility to moisture damage tend to undergo stripping and usually exhibit a sudden second change in the slope of the curve for rut depth versus number of passes after a certain number of cycles. Figure 8.16 shows typical deformation curves for samples that do and do not exhibit significant moisture damage in a HWTD test. The rutting resistance of a mixture is expressed

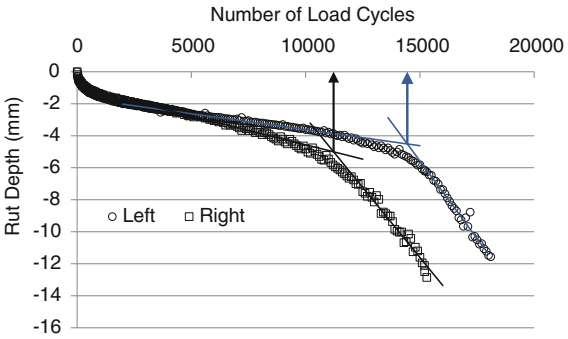


Fig. 8.16 Typical results obtained from the HWTD test

in terms of the observed rut depth at a specific number of load cycles or in terms of the number of load cycles required to reach a specific rut depth. Either parameter can be used as a pass/fail specification for mixtures to meet a design requirement. The presence of an inflection point can indicate the susceptibility of mixtures to moisture-induced damage and can also be used as a pass/fail requirement.

Torture or indicator type tests such as the HWTD have the same advantages and drawbacks as discussed earlier. To recap, the advantages are evident, i.e., these tests are easy to conduct, the results are straightforward to interpret, and the results can easily be incorporated into a specification requirement. On the other hand, such tests do not really measure a material property, the loading and torture conditions are unrealistic and most importantly the results from these tests do not have any predictive value. To further explain the last point, consider the results from the HWTD test. The results cannot be used to predict the actual rutting in the asphalt mixture in a pavement under any arbitrary set of conditions (traffic loads, traffic speeds, and temperatures). Also the results cannot be used in any meaningful way to guide pavement design for efficient materials. For example, if a mixture had extremely high resistance to rutting, there would be no effective method to use this information from a HWTD to optimize the thickness of the asphalt layer. A possible solution to overcome these shortcomings is to evaluate the damage evolution in the material using a more rigorous approach.

8.5.2.2 Robust Methods to Measure and Quantify Rutting Resistance of Materials

Before discussing a few methods to do so, it is important to briefly discuss how we quantify permanent deformation or the evolution of permanent deformation. We will only discuss this idea of quantifying permanent deformation in more simplistic terms here as an introduction to more mechanistic methods of measuring rutting in asphalt mixtures. Readers interested in more in-depth modeling can refer to Chap. 14 of this book for more details.

In simplistic terms, the total strain in the material can be decomposed as a sum of the recoverable strain and plastic strain (similar to Eq. 14.3) as shown in Eq. 8.3. Note that for a time-dependent material such as asphalt mixture (or binder), these strains are functions of time and the term recoverable strain also includes strain that ultimately recovers over a period of time after the load is removed. Similarly, the plastic strain is also time dependent, i.e., in addition to the magnitude of the load, the time history of the load also affects the magnitude of permanent deformation. Therefore, for a time-dependent material, such as an asphalt mixture, the total strain is decomposed as a sum of viscoelastic or recoverable strain (ε_{ve}) and viscoplastic or time-dependent plastic strain (ε_{vp}).

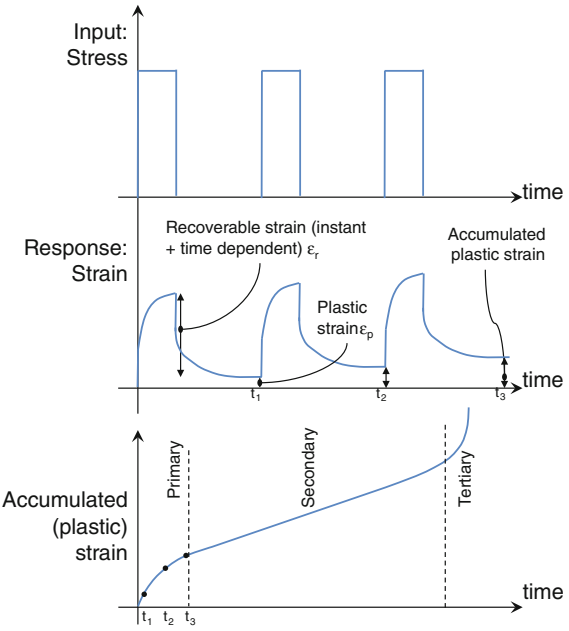


Fig. 8.17 A typical repeated load test and response showing total strain as well as recoverable and plastic strain

$$\varepsilon_{Total} = \varepsilon_{ve} + \varepsilon_{vp} \tag{8.3}$$

Experimentally, such an accumulation of plastic strain can be observed from a simple repeated load test. Consider a cylindrical specimen of an asphalt mixture subjected to repeated cycles of a compressive load. For the purposes of this example, each load cycle can comprise of a step load lasting for a fixed duration of time followed by a dwell time for which the load is entirely removed and the specimen is allowed to recover.

Figure 8.17 illustrates such a sequence of input loads as well as the typical deformation or strain measured on the specimen. The strain from any given cycle has a recoverable component (instantaneous as well as time dependent) and a permanent component (plastic strain). The bottom part of Fig. 8.17 also illustrates the total accumulated plastic strain over time. Such curves are commonly used in the flow time and flow number tests to evaluate the resistance of asphalt mixtures to permanent deformation (discussed in more detail in a later section).

Two important things to consider in this simplistic schematic are as follows. First, it is possible that in some cases the time-dependent strain might not have recovered completely at the end of the dwell time. As a result, strictly speaking, the residual strain at the end of the dwell time is not plastic strain but rather plastic strain with some residual recoverable strain that is still in the process of recovering. Second, the plastic strain measured here is for the specific duration and magnitude of loading. Also plastic deformation is a highly history-dependent phenomenon, i.e., the rate of accumulation of plastic strain depends on the history of loading and

previously accumulated plastic strain. This is due to strain hardening of the mixture. Due to these factors, the evolution of plastic strain as shown in the bottom part of Fig. 8.17 is not a material property that represents plastic strain evolution in the mixture but rather results from one specific loading magnitude and history. As such, these results (bottom part of Fig. 8.17) cannot be used to predict permanent strains with any arbitrary loading history. However, the results from such a loading history can serve as an indicator for rutting susceptibility; this will be discussed soon.

The two aforementioned gaps can be corrected by the use of appropriate mathematical models. For the first part, by knowing the linear viscoelastic properties of the mixture (such as creep compliance, relaxation or complex modulus), and the use of proper convolution integrals, it is possible mathematically separate the time-dependent recoverable strain from the irrecoverable or plastic strain at the end of the dwell time. This can be used to get a more accurate estimate of the accumulated plastic strain. The detailed mathematical description of this process can be found in Chap. 14.

The second gap was that the results as shown in Fig. 8.17 cannot be generalized to any arbitrary loading history (different rates or durations and temperatures). It is possible to address this shortcoming by using an appropriate and comprehensive model for viscoplasticity that captures the behavior of the material while accounting for different rates of loading. Such a model typically incorporates three components:

- an overstress function or a stress-based criterion that dictates the conditions under which the material exceeds its plastic limit and viscoplastic flow occurs accruing permanent deformation,
- a flow rule that defines the rate at which viscoplastic strain is accumulated, and
- a hardening or softening function that captures the history dependence of the material.

A more detailed description of these models can be found in Chap. 14 of this book. Readers are also referred to a chapter by Scwhartz and Kaloush on their treatment of this subject for more details (Schwartz and Kaloush 2009).

The parameters required to calibrate a comprehensive model, such as the one described above, are usually obtained from a test similar to the one illustrated in Fig. 8.17 but by subjecting the test specimens to different confining stresses and measuring both axial and radial deformations. In addition, linear viscoelastic properties and temperature-dependent behavior are obtained from a dynamic modulus test similar to the one described in the previous section.

A couple of simpler variations of the test shown in Fig. 8.17 are worth briefly mentioning here. These variations are referred to as the flow number and flow time tests. These involve testing a specimen at one temperature in unconfined conditions. Typically, the test specimen is 150 mm in height and 100 mm in diameter and is obtained from cutting and coring a 175–200 mm high and 150 mm diameter laboratory compacted specimen. In the flow number test, the specimen is subjected to

cyclic loads of a fixed amplitude following a haversine form (instead of the step load shown in the figure). The time period for the haversine load form is 0.1 s and dwell time is 0.9 s. The accumulated strain response from the test is recorded with respect to time or number of load cycles and is similar to the bottom schematic in Fig. 8.17. The response is divided into three zones: primary, secondary, and tertiary. The number of cycles at which the tertiary zone begins, identified by the change in the accumulated strain rate, is defined as the flow number value for the material (at the specific loading conditions). The flow time test is very similar to the flow number test, except that the applied load is not cyclic, but rather remains constant over time. The response from the flow time test is very similar to the flow number test and the time at which the flow of the specimen occurs (beginning of the tertiary zone) is referred to as the flow time. Readers interested in more details can refer to additional literature (Witczak et al. 2001; Bhasin et al. 2003).

An interesting feature of the flow time and flow number test is that when used in the simplistic manner described above, these tests can serve as simple indicator tests. In other words, flow number or flow time (obtained for a specific magnitude of load and temperature conditions) can be used as an indicator to compare the rutting resistance of different asphalt mixtures or as a screening tool to qualify asphalt mixtures for construction. A mixture with a relatively high flow number or flow time value would have a relatively higher resistance to rutting. However, these tests can also be modified to incorporate the effects of confinement and multiple loading conditions (magnitude and duration) to obtain parameters for a more fundamental viscoplastic model, which can then be used to predict the rutting behavior of asphalt mixtures under a variety of different conditions. *From this point of view, a transition from torture tests to these more basic tests is an important step forward in advancing the state of knowledge and practice in this area. The use of flow time and flow number tests would allow the agencies responsible for building road ways to use a simple test with indicators to screen and qualify materials and at the same time it would also create a stepping stone to incorporate advanced modeling and mechanistic tools into future practice.*

8.5.2.3 Summary

To summarize, we have looked at three broad categories of tests that can be used to assess rutting susceptibility of asphalt mixtures. The first category was that of torture tests, such as the wheel tracking tests. Such tests are repeatable, easy to perform and provide results that are straightforward to interpret. However, results from torture tests do not measure material properties, can sometimes be misleading (if the test conditions and stress states are very different from that of the pavement structure), and have very limited predictive value. The second category is that of tests such as the flow time and flow number tests. Such tests are simple and can be used to provide a performance indicator that can be used by road way agencies to screen and qualify materials for use on a routine basis. More importantly, with a few minor modifications and additional data collection these tests lead us into the

third category, i.e., tests and resulting data that can be used with viscoplasticity models to extract a fundamental material property related to the rutting characteristics of the material. Similar to a performance indicator, a basic material property can be used to compare the relative rutting resistance of different types of mixtures when used in the same structure and subjected to the same load and temperature history. However, unlike performance indicators, basic material properties can also be used to predict the performance of the material in the context of any given pavement structure and for any arbitrary loading and temperature history.

8.5.3 Fatigue Cracking

Fatigue cracking is also one of the most common forms of distress in asphalt pavements. Similar to the previous section on rutting, we will briefly discuss some simplistic methods to produce a performance indicator that can be used to rank the fatigue cracking resistance of different asphalt mixtures or qualify mixtures for construction purposes. We will also discuss some more fundamental approaches to characterize fatigue cracking resistance that can be used to predict the cracking resistance of asphalt mixtures in any given pavement structure subject to any arbitrary loading history.

Analogous to the torture tests used to evaluate rutting of asphalt mixtures, there are some simple tests that can be used to produce indicators that compare fatigue cracking or fracture resistance of different asphalt mixtures. Also as in the case of rutting, there are some tests and concomitant analytical methods that can provide both indicators of fatigue cracking resistance as well as a fundamental material property for this distress evolution with predictive capabilities. Before we discuss a few examples, it is important to recognize that there are several different types of tests that can be found in standards or literature. Most of these tests are indicator type tests and can be described using the following four main parameters:

(i) test specimen geometry,
(ii) loading input type,
(iii) loading form type, and
(iv) resulting parameter.

We will briefly discuss the variations in each of the above parameters that are typically used in developing and implementing a test method. We will then discuss a few specific examples in a bit more detail.

8.5.3.1 Typical Test Geometries Used for Fatigue and Fracture Tests

Regarding the first parameter, i.e., geometry, in order to subject a test specimen to tensile loads, the specimen must be typically gripped or glued to the ends with

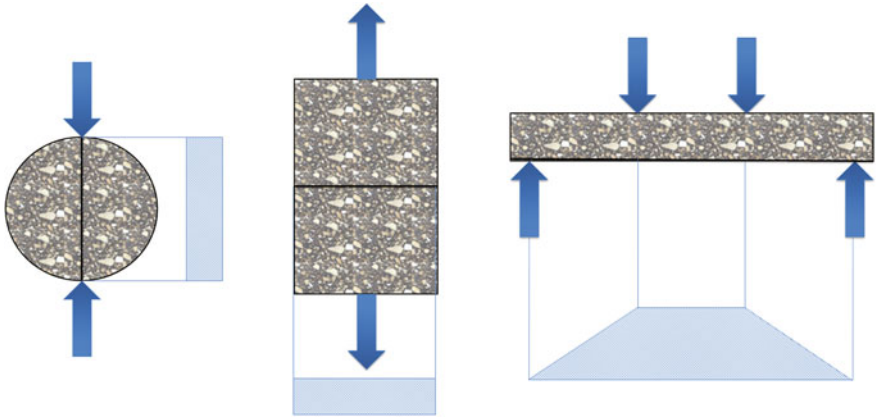


Fig. 8.18 Schematic showing typical geometries used to evaluate fatigue and fracture of asphalt mixtures

appropriate fixtures and attached to a loading frame (try to imagine applying a uniaxial tensile load to a specimen without gluing or gripping it in some form). The process of gluing or gripping the ends requires some additional work to accomplish the test and also creates some unwanted stresses near the ends of the specimen. Not withstanding these limitations, test methods such as the direct tension test and the overlay test utilize test specimens that are glued to fixtures used in loading frames to conduct fatigue or fracture tests. The direct tension test typically involves the use of a cylindrical laboratory compacted test specimen with both ends of the cylinder glued to metal plates (Fig. 8.18). The overlay test involves the use of a prismatic specimen cut out of a cylindrical compacted specimen or core. Note that the overlay test was intended to measure the rate of crack growth emanating from the bottom of an asphalt layer placed on top of a rigid jointed pavement. However, this test is also used as an indicator or torture test to measure fatigue cracking resistance of asphalt mixtures for conventional applications.

A few geometries that allow users to create a tensile stress without the need for gluing or gripping the specimen ends are by loading a beam in flexure (consider a simple three point or four point beam specimen) or by loading a cylinder diametrically to create an indirect tensile stress along the diameter (Fig. 8.18). In fact, owing to this convenience in sample preparation and handling, the beam fatigue test and indirect tensile test (IDT) are still used to measure fatigue and fracture properties of asphalt mixtures. The latter is more preferred on account of the fact that asphalt mixture specimens are typically prepared in the laboratory in the form of cylinders (about 150 mm in diameter or smaller) or cored from the field (again in the form of cylinders). A variation of the beam fatigue is the semicircular bending test that creates tensile stresses in a semicircular specimen that is simply supported on two ends of its flat side (for more details see Ozer et al. 2016).

8.5.3.2 Typical Loading Modes and Forms Used for Fatigue and Fracture Tests

The second and third parameters are the loading input type and form. Fatigue or fracture tests are typically conducted in one of the following two modes of loading: displacement or strain controlled mode and load or stress controlled mode. In case of the former, the test equipment is commanded to change the displacement or strain following a predefined load form and the reaction load or stresses in the specimen are monitored over time. The latter is just the opposite, i.e., the test equipment is commanded to change the load or stress following a predefined load form and the resulting deformation or strains are recorded over time. The loading forms that are typically used in fatigue or fracture tests are monotonically increasing at a certain displacement or loading rate, cyclic following a sinusoidal waveform, or cyclic following a triangular wave form. In some cases, users may also use a cyclic loading following a haversine form with a rest or dwell time after each cycle. Figure 8.19 is a schematic that shows the different loading input types and forms. The specific values must also be predefined for any given loading mode and form. For example, if a monotonically increasing displacement-controlled test is to be performed, then the rate at which the displacement is applied must be predefined as a test parameter. Similarly, if a sinusoidal cyclic displacement-controlled fatigue test is performed, then the frequency of the cyclic load as well as the magnitude of loading in each cycle must be predefined as a test parameter.

8.5.3.3 Typical Simplistic Metrics to Quantify Fatigue Resistance and Their Limitations

The fourth parameter in the fatigue or fracture tests is the metric that is extracted out of the test results to serve as an indicator for the fatigue or fracture resistance of the material. Several different metrics can be extracted depending on the type of test. For example, in the case of a simple monotonically increasing load or displacement-controlled test, the maximum load or displacement at failure can serve as an indicator of the fracture resistance of the material. For a cyclic test, the number of load cycles to failure can be used as an indicator. In these cases, failure itself is defined when the specimen is physically ruptured or in some cases when the specimen reaches 50% of its initial stiffness. Note that the failure criterion itself can be a subject of debate. One can argue that if a material is only 50% of its original stiffness it can be considered as "failed" because of the significant change in its properties. On the other hand, several binders have a significant fatigue life until physical macro cracks are formed much after the modulus has reduced by 50%. In short, the failure criterion itself must be considered carefully when evaluating material resistance to fatigue cracking. *Although, it is tempting to use these simple indicators (number of cycles to failure) as a measure of the fatigue cracking resistance of asphalt mixtures, these metrics can easily be misinterpreted or misleading*. The major drawback of using simple indicators (e.g., load cycles to failure)

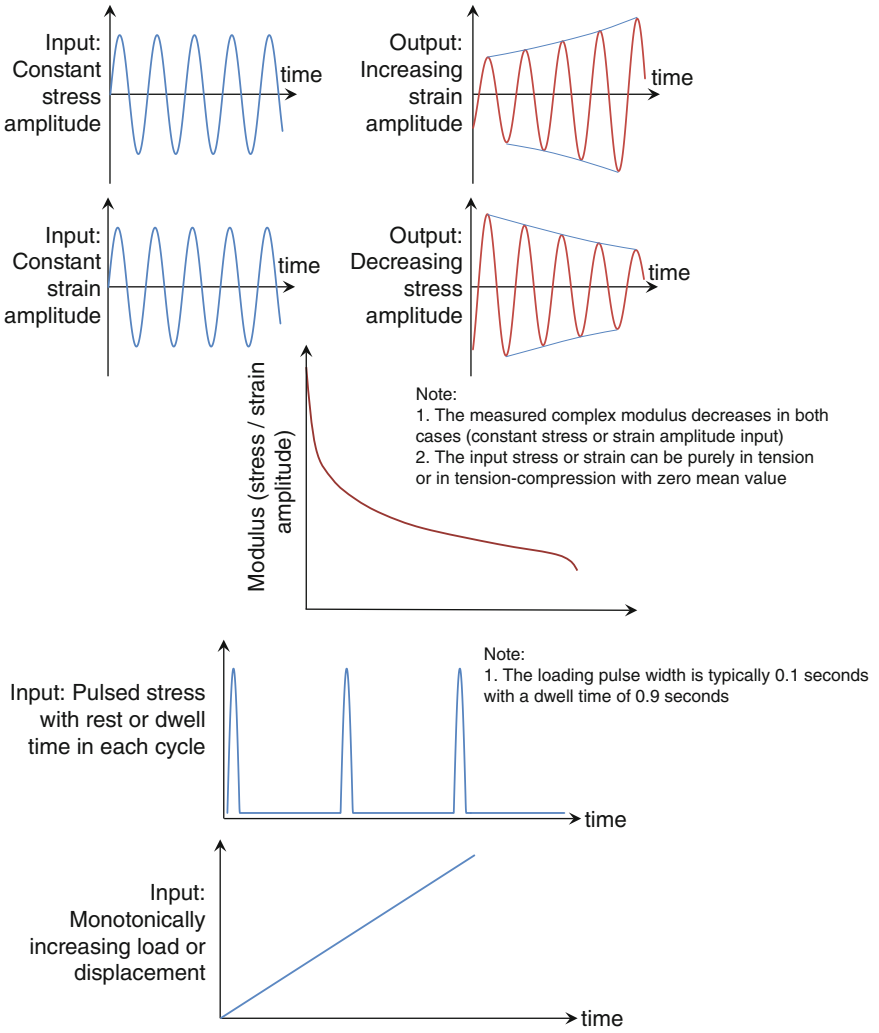


Fig. 8.19 Schematic showing typical loading modes and forms used in fatigue and fracture testing of asphalt mixtures

to rank order the fatigue cracking resistance of different mixtures is that the rank order may change if fatigue cracking resistance of different mixtures are compared using: (i) using different amplitudes of loading, (ii) different frequencies or rates of loading, and (iii) different modes of loading.

Let us consider the first two drawbacks, i.e., the rank orders for the fatigue cracking resistance of asphalt mixtures can change by simply using different stress (or strain) amplitudes or frequencies while conducting the test. For example,

consider two mixtures A and B that are evaluated for their fatigue cracking resistance by applying cyclic loads following a sinusoidal waveform at a frequency of 10 Hz and an amplitude of 50 kPa (the numbers here are just hypothetical for the example). Based on a simplistic indicator for the number of load cycles until failure, material A has higher number of load cycles to failure compared to material B and as a result A ranks better than B. It is now entirely possible, that this rank order will reverse if the test were to be conducted using different load amplitude and/or a different loading frequency. This apparent contradiction is neither unexpected nor unusual but does shed light on the fact that the use a simple fatigue life indicator can be misleading at times and most certainly cannot be used to assess the fatigue life of the mixture in the context of a pavement structure.

Similar apparent contradictions also arise from using different modes of loading. For example, typically a stiff mixture when tested in a controlled strain or displacement mode of loading will fail at lower number of load cycles as compared to a softer mixture. This is because for the same displacement or strain amplitude, the stiffer mixture will result in higher stresses. In contrast, if these mixtures were tested in a stress or load controlled mode, then typically the softer mix will fail at lower number of load cycles to failure because for a given load level it experiences much larger deformations. Therefore, the rank order of fatigue cracking resistance of two different mixtures can easily be reversed if evaluated using two different modes of loading. Of course, this example is only an illustration and not universally true, but it does highlight one of the most critical shortcomings of using simple indicators to evaluate fatigue cracking resistance.

In summary, as we have discussed before, simple indicators are not material properties but rather reflections of the material response under one specific set of loading conditions. These shortcomings can be overcome by using a correct combination of material testing protocol with appropriate damage model-based parameter. For example, continuum damage mechanics or energy-based parameters can be used to characterize fatigue damage resistance of asphalt mixtures as a material property that is independent of the loading mode, amplitude or rate.

8.5.3.4 Robust Methods to Characterize Fatigue Cracking Resistance

In this section, we will only visit some of the basic ideas to quantify fatigue cracking resistance of materials. Advanced and more mechanistic methods are covered in Chap. 14 of this book. As before, we will start developing our understanding based on a simple elastic material in which the rate or frequency of loading does not make a difference to the results.

The classical method to characterize fatigue cracking resistance of elastic materials is by means of a Wohler curve or simply referred to as the S-N curve. Figure 8.20 shows a schematic of a Wohler curve with two hypothetical materials A and B. The ordinate of this graph illustrates a stress ratio, i.e., the ratio of the applied stress to the ultimate stress or strength of the material and the abscissa in log scale shows the number of load cycles to failure. It is expected that if the applied

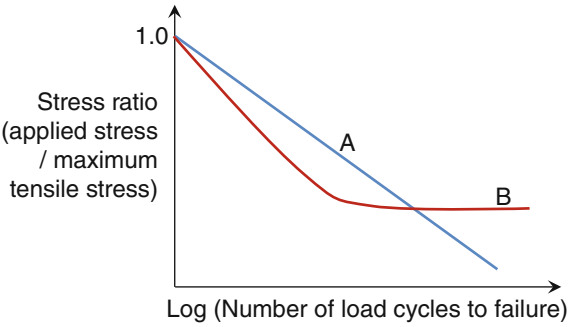


Fig. 8.20 Typical Wohler curve used to characterize the relationship between the tensile strength and fatigue life of a material

stress were the same as the strength of the material, the material would fail in the very first cycle. However, as this ratio reduces, the number of load cycles to failure increases until we reach a point where the relationship becomes asymptotic. This point is referred to as the endurance limit, i.e., the limiting strain below which the material is not expected to fail due to fatigue. This limit may or may not exist for all materials. This limit is known to exist in asphalt mixtures and is used in the design of "perpetual pavements"; this will be briefly revisited at the end of this section. Another point to note in the context of asphalt mixtures is that, tests such as the beam fatigue test are often times conducted using multiple strain amplitudes to generate a full S-N curve for each mixture type. Developing the S-N curves overcomes the first limitation discussed earlier and allows the user to examine the fatigue response of different materials at several different amplitudes as opposed to using the fatigue life at a single stress amplitude. The schematic shown in Fig. 8.20 also illustrates a scenario in which the rank order of the two different mixtures changes depending on the stress amplitude.

Another commonly used metric to quantify the fatigue cracking resistance of a material is based on the stress intensity factor, typically denoted as K_I in mode-I cracking (for a more detailed description readers are referred to Chap. 14). This approach is particularly beneficial in predicting crack growth if the geometry of the flaw or defect is known or can be approximated. The stress intensity factor is related to the applied stress and geometry of the flaw. Paris and his co-workers experimentally demonstrated that the rate of crack growth denoted as da/dN, where a is the crack size and N is the number of load cycles, follows a power function with respect to the change in the stress intensity factor ΔK (Paris and Erdogan 1963; Broek 1982). This relationship is shown in Eq. 8.4; the parameters of this power law function, C and n, are constants that are unique to the material. This form, typically referred to as Paris' law, is a phenomenological or empirical observation that is used as the basis for several forms of mechanistic models (in some sense analogous to the use of an elastic modulus to describe the stress–strain relationship of a material Fig. 8.21).

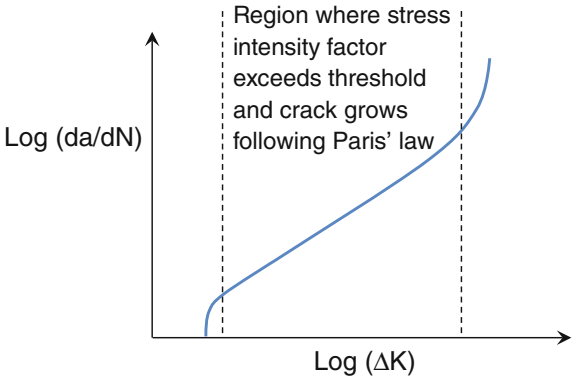


Fig. 8.21 Typical relationship between stress intensity factor and rate of crack growth

$$\frac{da}{dN} = C(\Delta K)^n \tag{8.4}$$

Note that the above discussion was primarily based on idealized elastic materials. We will try to use this as a basis to appreciate the challenges in being able to characterize the fatigue cracking resistance of asphalt mixtures that are time (or rate), temperature and age-dependent. First, let us consider time or rate dependency of asphalt binders and mixtures. Rate dependency would imply that the material response or fatigue cracking resistance of the material would not only depend on the magnitude of stress or strain, but also on the rate at which the magnitude is reached. As discussed before, fatigue tests are typically conducted by applying loads with a predetermined stress (or strain) amplitude following a sinusoidal wave form and monitoring the strain (or stress) response of the material until failure. In the case of an ideal elastic material, the frequency of this sinusoidal waveform would make very little difference (as long as the rate is not too high to create excess thermal changes in the test specimen). However, in case of asphalt mixtures, by changing the frequency of loading with the same stress (or strain) amplitude, one would get very different results. Similarly, asphalt materials are highly temperature dependent. Therefore, the fatigue life of the material would not only depend on the magnitude and rate of loading, but also on the temperature! Finally, all of these properties would change with oxidative aging of the material. In summary, even a simple two-dimensional S-N relationship would evolve into a multi-dimensional relationship to account for rate, temperature, and aging. Fortunately, there are mechanistic-based models and methods that can be used to simplify this complexity.

The variations in the fatigue cracking behavior due to frequency, amplitude, and mode of loading can be filtered to extract a single unique material property using different mechanistic approaches. Chapter 14 of this book presents a more detailed description of the mechanistic methods to characterize fatigue and fracture. A few highlights are briefly discussed here.

Let us first discuss reconciling the differences in fatigue test results for the same material that are due to the use of different rates of loading or frequencies. Recall that we are reconciling these differences in order to get to a material property that is independent of such external factors. In the case of asphalt materials, the same material results in different fatigue lives for different modes of loading because unlike elastic materials, asphaltic materials have a time-dependent response. The effect of this factor can be eliminated by transforming the response of the material into a "pseudoelastic" domain. Simply put, the stress–strain–time behavior of the time-dependent viscoelastic material subject to loading is mathematically transformed to a pseudostress–pseudostrain behavior of a hypothetical elastic solid. These transformations are made using correspondence principles and the only input required to carry out these transformations is the linear viscoelastic property of the material (e.g., any one of creep compliance, relaxation modulus, or complex modulus). The transformation (and subsequent inverse transformation of the solution) allows the user to apply the solutions and methods applicable for elastic materials to viscoelastic or time-dependent materials thereby filtering out the effects of rate or frequency of loading.

For the purposes of illustration, Eqs. 8.5 and 8.6 show one of the ways to transform stress and strain to pseudostress and pseudostrain denoted with a superscript R. Equation 8.7 shows the inverse transformation from pseudostrain to strain (note that the stress and pseudostress are the same in this form of transformation). In Eqs. 8.5 and 8.6, $E(t)$ and $D(t)$ are the relaxation modulus and creep compliance of the material, respectively, and E^R is the elastic modulus of the hypothetical elastic material (typically treated as 1). In the transformed pseudodomain, although the pseudostress and pseudostrain are functions of time, the ratio of these two parameters is the reference elastic modulus. Simply put, pseudostress and pseudostrain can be treated as stress and strain of an elastic material. A more detailed description of this transformation can also be found in the literature (Schapery 1984; Bhasin and Motamed 2011).

$$\sigma^R \overset{\text{def}}{=} \sigma \tag{8.5}$$

$$\varepsilon^R \overset{\text{def}}{=} \frac{1}{E^R} \int_0^t E(t-\tau) \frac{\partial \varepsilon}{\partial \tau} d\tau \tag{8.6}$$

$$\varepsilon \overset{\text{def}}{=} E^R \int_0^t D(t-\tau) \frac{\partial \varepsilon^R}{\partial \tau} d\tau \tag{8.7}$$

The variations in test results due to the use of different loading magnitudes or modes of loading can be reconciled using energy-based methods or continuum damage mechanics. In a very broad sense, the energy-based methods are based on the premise that failure of a specimen due to fatigue crack growth is based on the

total energy consumed to create microcracks that nucleate and coalesce to form macro cracks. For a given material, the total energy consumed in this process would be the same (after discounting for time-dependent effects) irrespective of the path taken, i.e., irrespective of whether it takes relatively more number of cycles at a low stress magnitude or fewer number of cycles at a higher stress magnitude. Several different methods based on this dissipated energy concept have been proposed. A detailed discussion of these approaches is beyond the scope of this chapter. However, readers interested in a comparative discussion of some of these methods can refer to Bhasin et al. (2009). Another excellent example on the use of the energy-based methods in conjunction with the indirect tensile test to assess fatigue cracking resistance of asphalt mixtures can be found in Roque et al. (2004).

The continuum damage mechanics-based approach can also be used to reconcile the differences in fatigue life of the same material due to different modes or magnitudes of loading. This approach was initially developed by Schapery (1981) for solid rocket propellants, which also have a time-dependent behavior similar to asphalt materials. This was then refined and applied to asphalt materials by Kim and co-workers (Park et al. 1996, Lee and Kim 1998, Daniel and Kim 2002, Underwood et al. 2010). In this sense, characterizing damage growth in asphalt mixtures is to some extent, literally rocket science. The continuum damage approach is based on three basic ideas.

First, it is assumed that the nucleation and growth of several microcracks or damage due to cyclic loading is well distributed and occurs uniformly throughout the test specimen being used to evaluate the fatigue life of the material. Therefore, damage to the specimen is not treated as a collection of microcracks but rather treated as a parameter that is distributed or smeared uniformly throughout the specimen volume. This parameter, based on its origins in thermodynamics, is also referred to as a state variable of damage and often denoted by the letter S in the literature. One can perhaps think of another common state variable, temperature, that is used to describe the state of the material. In this case, external work or heat supplied to a body will change its temperature state. Similarly, one can envision that the extent of damage to a body due to external work is quantified by a state variable S.

Second, the evolution of damage is quantified in terms of a reduction in the stiffness of the specimen. For example, if the stiffness were denoted as C, then it is reasonable to consider that this stiffness is a function of the state of damage, $C(S)$. Higher the state of damage, lower the overall stiffness of the specimen.

Third, based on the observation of several different materials behavior, it was proposed that the rate of evolution of damage has a unique power law relationship with the rate of change of strain energy density, W, as shown in Eq. 8.8. Of course, the strain energy density is related to the stiffness of the material as shown in Eq. 8.9. In Eq. 8.8, the term α in the power is a unique material property and in Eq. 8.9, the term ε is the strain. One important note is that Eq. 8.9 is applicable for elastic materials and can only be used with asphalt materials after the measured responses have been transformed into pseudoelastic domain.

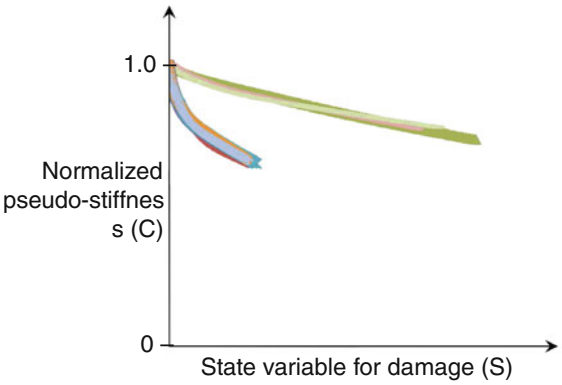


Fig. 8.22 Stiffness (C) versus damage (S) relationship for two different materials; the curves show that the relationship is unique irrespective of the mode or amplitude of loading and different for each material

$$\frac{dS}{dt} = \left(-\frac{\partial W}{\partial S}\right)^{\alpha} \tag{8.8}$$

$$W = \frac{1}{2}C(S)\varepsilon^2 \tag{8.9}$$

By replacing W from Eq. 8.9 in Eq. 8.8, it is possible to establish a unique relationship between the change in stiffness and the change in state of damage of the material. The premise (and the most powerful feature) of the continuum damage approach is based on this unique relationship. In practice, if one can determine this relationship using one experimental condition (rate and mode of loading), it would then be possible to use this relationship to predict damage evolution for any arbitrary mode and rate of loading. Figure 8.22 illustrates the results from an experimental study using two different materials. Several specimens of each material were tested under a variety of different loading amplitudes and rates. The figure shows that the relationship between the relationship between stiffness and damage is unique for a given material (and different from one material to another) for different fatigue testing and loading conditions. Readers are directed to additional reading to better understand and apply this methodology (Park et al. 1996; Lee and Kim 1998; Daniel and Kim 2002; Underwood et al. 2010).

8.5.3.5 Endurance Limit, Perpetual Pavements and Self-healing

Before, we conclude our discussion on fatigue damage and typical methods to quantify fatigue, we will briefly discuss three additional topics: (1) endurance limit, (2) perpetual pavements and (3) self-healing. Recall the schematic shown in Fig. 8.20. Material B in this schematic shows an asymptotic behavior in the S-N relationship at very low stress (or strain) amplitudes. This asymptote suggests that if the material experiences stresses (or strains) that are lower than this asymptotic value then the theoretical expected fatigue life would be infinite. The asymptotic value is

referred to as the endurance limit. Simply put, fatigue crack growth is not expected to occur as long as the material is within its endurance limit. The concept of endurance limit is not unique to asphalt materials; in fact, endurance limit is commonly used in designing metal parts subjected to cyclic loading to resist fatigue cracking. The concept of endurance limit takes us to the next idea, that is of a perpetual pavement. A perpetual pavement or pavement with limiting strain is a pavement structure that is designed so that the stresses or strains at the bottom of the asphalt layer remain below the endurance limit. As a result, there is no cracking that grows from the bottom of the pavement making its way upwards to the surface and eventual pavement failure (referred to as bottom-up cracking). Readers must note that in some cases the term "perpetual pavement" only suggests that the pavement is designed not to fail due to bottom-up cracking but may experience failure due to other forms of cracking or plastic deformation. For example, fatigue cracking can still occur near the surface of the pavement and close to the tires; this type of cracking is referred to as near surface cracking (more details in Wang and Al-Qadi 2009). However, in the case of near surface cracking, the distress easily manifests itself on the pavement surface and is relatively easier to repair (compared to bottom-up cracking).

The third topic of self-healing needs a bit more background. Most conventional materials have a well-defined temperature at which they change phase from solid to liquid (simplest example is that of water). However, at typical service temperatures, asphalt binders are viscoelastic materials that do not have a well-defined temperature above which they behave as liquids and below which they behave as elastic solids. In fact, asphalt binders, and by inheritance of properties asphalt mixtures, behave as viscoelastic solids when they are in a broad range of temperature. It is only at very low temperatures (of the order of −10 °C or below) that asphalt binders approach a behavior similar to that of elastic solids. Above such temperatures, the binder is typically a viscoelastic solid with molecules that have more mobility compared to stiff elastic solids; in fact, higher the temperature, greater the mobility. This mobility provides the asphalt binder the ability to self-heal microcracks at typical service temperatures. This essentially reverses some of the fatigue crack growth and increases the overall fatigue cracking life of the mix.

In the context of asphalt mixtures, self-healing can be defined as the process by which the crack growth in asphalt binders, which occurs due to repeated loading, is partially or completely reversed. For example, cracks may develop in an asphalt mixture due to repeated traffic loads. However, if the pavement is allowed to remain without the application of external load (or even when the applied load is very small) for a duration of time, crack-based damage is often partially recoverable due to the self-healing nature of the asphalt binders. The self-healing ability of asphalt binders has been demonstrated through several laboratory and field studies. The most important findings from these studies is that different asphalt binders have different abilities to self-heal and in some cases this ability is not necessarily tied to the binder stiffness. One implication of this finding is that it is possible to chemically engineer an asphalt binder that has excellent self-healing characteristics and can therefore resist or reverse fatigue crack growth at a material level. The second implication of this finding is that fatigue lives of asphalt mixtures predicted using typical laboratory

tests (that do not provide the opportunity for the material to self-heal) can significantly underestimate the actual fatigue cracking life of the mix and potentially incorrectly rank the realistic fatigue cracking resistance of different mixtures.

The self-healing of asphalt binders and mixtures is still a subject of ongoing research with several research studies focused on different aspects related to self-healing. For example, some studies have focused on understanding the mechanisms that drive nucleation and growth of microcracks at a molecular to a micrometer length scale as well as the relationship between the chemical makeup of the asphalt binder and its ability to resist cracking and self-heal. The current understanding from these studies is that the microstructure of the binder creates localized hot spots where fatigue cracking is likely to nucleate and grow. Further, a microcrack can self-heal in a two-step process that occur continually and in succession over time: crack surface wetting, i.e., the closing of the two surfaces of a microcrack, and interfacial strength gain, i.e., the process of gaining strength across a closed crack surface. Other studies have focused on the measurement of the intrinsic healing properties of the asphalt binder or the healing properties of an asphalt composite. For example, Fig. 8.23 shows a protocol to measure both fatigue life and healing (quantified as the percent reversal in damage) for an asphalt composite as a function of the duration of the rest period as well as a function of the extent of damage (quantified as the reduction in modulus compared to undamaged material) preceding the rest period. These results are typical representations of the self-healing phenomenon that illustrate what is intuitively expected, i.e., the amount of self-healing increases with an increase in the duration of the rest period and also that the amount of self-healing for a given rest period decreases if the extent of damage is higher preceding the rest period. The findings from these studies have further reinforced the idea that the self-healing characteristics of the asphalt mixture significantly influence its overall fatigue life and that such characteristics are dependent on the properties of the binder and its chemical makeup. (Bazin et al. 1967, Bhasin et al. 2008, Karki et al. 2016, Kringos et al. 2011).

8.5.3.6 Summary

To summarize, fatigue tests can be conducted using different geometries (e.g., three or four point bending, direct tension, indirect tension) and testing conditions (e.g., stress or strain controlled, frequencies, load amplitudes). The simplest parameter that can be obtained from such tests is the number of cycles it takes for a specimen to fail (where failure is defined as 50% of initial modulus or formation of macro cracks). It is tempting to use this simple and intuitive metric to compare performance of different mixes or qualify mixes for field use, however, it can also be misleading. This is because the number of load cycles to failure is not a material property but rather an indicator of a material property that is also dependent on the conditions (mode of loading, amplitude of loading, temperature, etc.) at which the indicator was obtained. Ranking mixtures for their resistance to fatigue cracking resistance based on such a simple indicator can be misleading because the rank

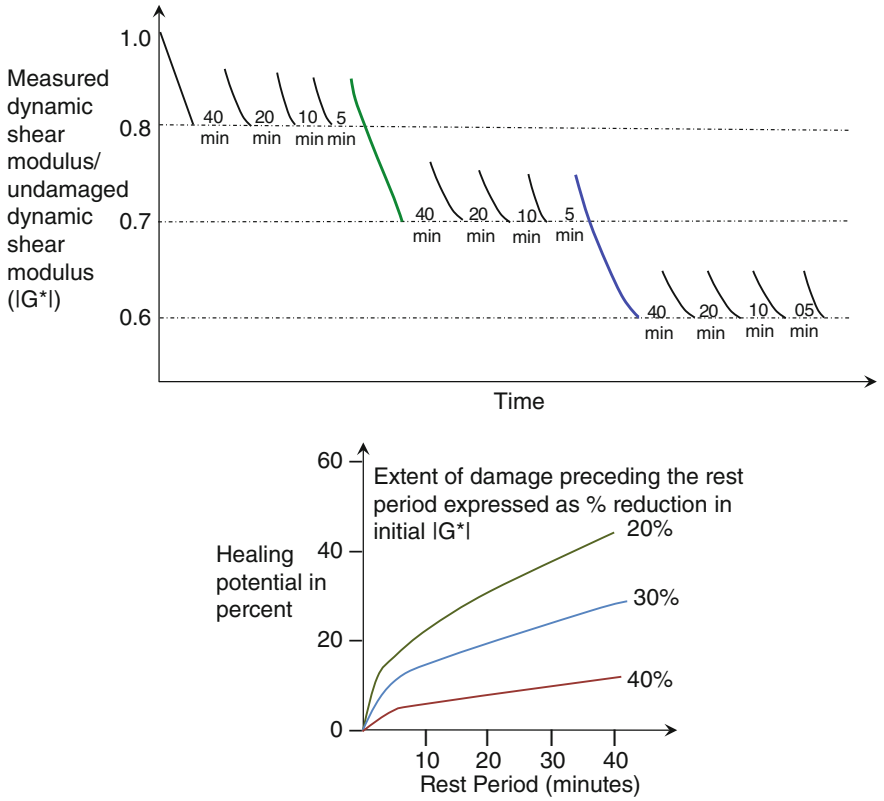


Fig. 8.23 Typical loading protocol with rest periods to measure both fatigue cracking and self-healing of asphalt composites as a function of the extent of damage preceding the rest period and duration of the rest period (*top*) and typical results for a composite (*bottom*) (Karki et al. 2015)

order may change if the materials were evaluated using a different set of test conditions. This shortcoming highlights the importance of using a more robust means to quantify fatigue cracking resistance. For elastic materials, one example for such a relationship is the Wohler curve that describes the fatigue life as a function of stress or strain amplitude. However, for viscoelastic materials such as asphalt mixtures, a robust measure must also be able to account for the influence of loading rate on the fatigue life and possibly also the influence of self-healing. To this end, approaches such as the continuum damage model can be extremely useful in using results from typical test methods to extract a fundamental material property that is related to fatigue cracking. The most significant advantage of extracting fatigue properties in this manner is that it can be used to predict the performance of the material (1) when it is subjected to any loading condition, (2) used in pavement structures with different layer thickness, and (3) when the influence of self-healing or endurance limit needs to be incorporated and used in the design of a pavement.

The use of such approaches should clearly be the cornerstone to evaluate the performance of materials in the near future.

8.5.4 Low Temperature Cracking

We have discussed the mechanism of low-temperature cracking in an earlier section. Recall that as the pavement temperature drops, the asphalt mixture tries to contract, but since there are no shrinkage/expansion joints in a flexible pavement, this results in tensile stresses that are developed in the asphalt mixture. At the same time, the viscoelastic nature of the asphalt binder allows it to relax and relieve these stresses. However, broadly speaking, an improper material selection for a climatic condition can result in a scenario in which the buildup of tensile stresses due to drop in pavement temperatures can occur faster than the rate at which the binder can relax or relieve these tensile stresses. In such cases it is possible that the tensile stresses exceed the strength of the material and result in transverse cracking. This is simply natures' way of creating an expansion/shrinkage joint if the engineer fails to properly account for the material properties. As before, in this section, we will very briefly discuss a simplistic indicator type or torture test to evaluate the resistance of a mixture to low-temperature cracking as well as a more robust way based on material properties to do the same. As a side note, this same process can also occur over multiple cycles due to daily temperature changes resulting in thermal fatigue.

8.5.4.1 Simplistic Torture Tests to Compare Low-Temperature Cracking Resistance of Mixtures

One example of a simple performance indicator or torture test is the thermal stress restrained specimen test (TSRST). Figure 8.24 shows a schematic of this particular test method. In this test method a prismatic asphalt mixture specimen is restrained from two ends in an environmental chamber. The temperature in the chamber is then

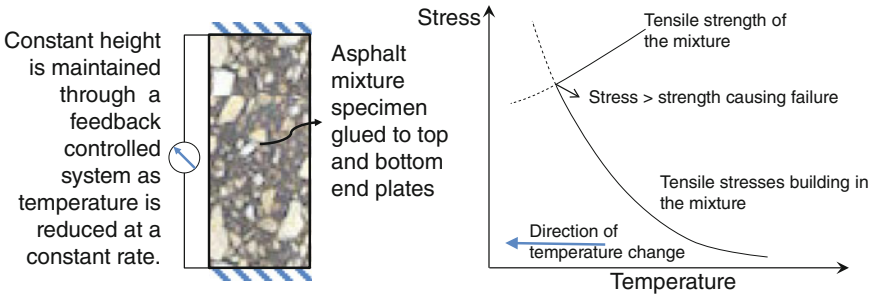


Fig. 8.24 A schematic showing the TSRST and typical results

gradually reduced. A feedback system is used to ensure that the length of the specimen does not change over time due to shrinkage in the loading frame or fixtures. The tensile load developed in the specimen is measured over time as the temperature drops at a predetermined rate. The test is continued until the tensile stresses in the specimen exceed its capacity or strength and the specimen fails. The temperature at which this failure occurs is referred to as the critical temperature; a lower critical temperature is an indicator of improved resistance to low-temperature cracking.

As in the previous cases with rutting and cracking, a simple indicator such as the critical temperature of failure is appealing to compare the performance of different materials or use as a material qualification parameter for pavement construction. However, as before, a simple indicator does not really measure material properties nor does it have the capability of predicting what would happen under any arbitrary set of external conditions. A standard simulation type test such as the TSRST can simulate only one condition, which may or may not represent the realistic conditions for any given job. For example, how would different materials react if the rate of cooling were slightly different from the rate of cooling that was used in the TSRST test? Such questions cannot be answered without using the basic material properties that drive a particular form of distress.

8.5.4.2 Robust Methods to Quantify Low-Temperature Cracking

As of this writing a significant amount of ongoing research is being conducted to measure properties related to low-temperature cracking of asphalt mixtures. A significant amount of work has been done to very precisely model crack growth in asphalt layers induced by low-temperature events, particularly when a sharp initial crack exists at the bottom of the asphalt layer from an existing jointed concrete pavement structure. Readers are referred to additional reading material for more details on work in this area (Dave and Buttlar 2010; Ahmed et al. 2013). In this subsection we will only briefly discuss the properties and phenomenological approach to model and predict the growth of low-temperature cracks. A more detailed discussion on the mathematical modeling would require the background from Part 2 of this book.

The rate at which tensile stresses develop in the asphalt pavement are dependent on four factors: (1) the rate at which the air temperature drops on the pavement surface (2) the rate at which the temperature in the pavement structure drops, (3) the coefficient of thermal expansion of the material, and (4) stiffness and the rate of relaxation of the material. First, the rate at which the air temperature drops is an extraneous factor that is dictated by the geographic location in which the pavement is constructed. Second, the rate at which the temperature in the pavement structure drops as well as the temperature of the asphalt layer at different depths depends on the thermal properties of the pavement material (e.g., thermal conductivity) as well as the temperature profile above and below the asphalt layer. In other words, once the thermal properties of the asphalt mixture and the temperature boundary conditions above and below the asphalt layer are known, it is possible to determine the

temperature of the asphalt layer at different depths. In fact, empirical models for such determinations have already been developed using the long term pavement performance (LTPP) database in the USA. Equation 8.10 shows one such equation, where T is the pavement temperature, H is the depth to surface in mm, Lat is the latitude of the section in degrees, and T_{air} is the low air temperature (Mohseni 1998).

$$T = -1.56 + 0.72\,T_{air} - 0.004\,Lat^2 + 6.26\log_{10}(H + 25) \tag{8.10}$$

Once the change in temperature in the asphalt layer is known, it can be combined with the coefficient of thermal expansion of the material to compute the thermal stress. Recall that for a simple elastic material with a coefficient of thermal expansion α and elastic modulus E experiencing a temperature drop of ΔT, the thermal strain is $\Delta T\alpha$ and if the material is restrained, then the tensile stress is $E\Delta T\alpha$. However, for a viscoelastic material, the rate at which the temperature drops is very important because it influences the strain rate, which in turn influences the stress evolution in the material. In addition, as we have discussed before the properties of the binder and by inheritance the properties of the asphalt mixture are temperature dependent. Therefore, the properties of the material that dictate the basic stress–strain–time relationship are also continuously changing with temperature.

Despite the multilayered complexity of this problem, it is possible to compute the stress evolution over time by knowing the basic time- and temperature-dependent material properties, coefficient of thermal expansion and the external rate at which the temperature changes over a period of time. For a more detailed description of this procedure, readers are referred to additional literature (Morland and Lee 1960; Muki and Sternberg 1961; Li et al. 2015), but a brief description of the basic approach is described here. Basically, the thermal stress at any point in time can be calculated using the same convolution form discussed before. However, in order to accommodate for the fact that properties of the material are also changing with temperature, this operation is carried out using reduced time. The thermal stress in reduced time ξ is given by the convolution integral:

$$\sigma(\xi) = \int_0^{\xi} E(\xi - \xi')\frac{\partial \varepsilon(\xi')}{\partial \xi'}d\xi' \tag{8.11}$$

where, ξ' is the variable of integration and $\varepsilon(\xi')$ is the thermal strain history in reduced time. The reduced time is defined as before and obtained by using the transformation:

$$\xi(t) = \int_0^{t} a[T(t')]dt' \tag{8.12}$$

where, $T(t)$ is the change in temperature T with time t. Equations (8.11) and (8.12) can be used to calculate the accumulated thermal stress as the temperature

T changes with time t following any form given by $T(t)$. The elegance of this approach is that it only requires the user to know the basic material properties, i.e., the relaxation modulus, reduced time versus temperature relationship (or time-temperature shift factor), and coefficient of thermal expansion, in order to compute the evolution of stresses in the material for any given arbitrary change in temperature over time. Finally, it is important to note that in order to predict failure, one needs to map the change in stress over time (or temperature) to the tensile strength of the material (which is also rate and temperature dependent) to determine the point of failure at which the stress exceeds the strength of the material.

8.5.4.3 Summary

To summarize, thermal cracking typically occurs when the material is trying to shrink due to a drop or reduction in the temperature but is unable to do so because of the absence of any expansion or contraction joints. The amount of stress that builds up in the material is a function of its coefficient of thermal expansion, stiffness, and change in temperature. Strictly speaking since viscoelastic materials are sensitive to the rate of loading or more generally the loading history, it is the rate at which temperature changes that is just as important as the magnitude of change in temperature. Since asphalt mixes are viscoelastic materials there are two competing time-dependent responses: (i) the build up of stresses over time that occurs due to a reduction in the temperature with time and the tendency of the material to contract and (ii) the relaxation of stresses over time due to the inherent ability of the material to relieve these stresses. A thermally restrained specimen test can be used to experimentally and empirically determine the rate of evolution of these stresses at a specific rate of cooling as well as the fracture point, i.e., the temperature at which such stresses exceed the strength of the material resulting in tensile failure. Although such a test is easy to conduct and results are straightforward to interpret, the results cannot be generalized and extended to predict failure at any arbitrary thermal history. To this end, basic material properties along with time-temperature superposition can be used to estimate the evolution of stresses within the asphalt for any thermal history.

8.5.5 Moisture-Induced Damage

In the pervious section, we had discussed that moisture damage is not only a mechanism that results in pavement failure (e.g., formation of potholes), but it also reduces material integrity and amplifies other forms of failure. We also discussed the mechanism of moisture damage in terms of moisture transport into the asphalt mixture, softening of the matrix (binder, mastic or mortar) and ultimately stripping of the binder from the aggregate surface. As we have done with other distresses, in this section, we will discuss some empirical ways to quantify the susceptibility of

materials and mixtures to moisture damage as well as some more fundamental methods to do the same.

8.5.5.1 Simplistic Torture Tests to Evaluate Sensitivity of Asphalt Materials and Mixture to Moisture-Induced Damage

One of the most direct methods to empirically estimate the moisture sensitivity of a given combination of asphalt binder and aggregate is referred to as the boil test. As the name suggests, the boil test involves the use of coarse aggregate particles that are coated with the asphalt binder and then subjected to accelerated moisture damage by placing these coated particles in boiling water for a specified period of time. The aggregates are then qualitatively examined to identify whether or not there was stripping of the binder from the aggregate surface. Such a test is clearly very easy to perform and requires very little capital equipment. At the same time, it is also easy to see that the results from this test are subjective indicators at best and not entirely robust to false negatives or positives. For example, the stiffness of the binder under such harsh conditions can influence the results from the test. Recall that the Hamburg wheel tracking test discussed earlier in this chapter is also used as an indicator of moisture damage by subjecting the mix to such harsh conditions (although in this case it is the compacted mixture specimen that is tested and the temperatures are far below boiling).

A more common approach or indicator to evaluate moisture sensitivity of different mixes is the ratio or percentage of a mechanical property in dry condition to the same property after moisture conditioning. Typically, a number of specimens of the mixture are compacted in the laboratory. Half the specimens are evaluated as it is using a mechanical test and the other half are subjected to a moisture conditioning process and then evaluated using the same mechanical test. The indicator for moisture sensitivity is then obtained as = 100 * (Mechanical property of moisture conditioned specimens/Mechanical property of dry specimens). In this sense, if the mixture is completely unaffected by moisture, then the moisture conditioned specimens would have the same property (value) as the dry or unconditioned specimens and result in a ratio of 100%. If the mixture is extremely sensitive to moisture, then the moisture conditioning process would deteriorate its mechanical property resulting in a very small ratio.

The generic method described above has two main variables: (1) the process used to achieve moisture conditioning and (2) the mechanical property that is measured using the test specimens before and after moisture conditioning. There are several different processes that can be used to moisture condition the test specimens. For example, test specimens may be subjected to a fixed number of freeze–thaw cycles. Another method to moisture condition the specimens is by first vacuum saturating the specimen under water and then leaving it under water at room or at an elevated temperature for a certain period of time. There are a number of different mechanical properties that can be measured using the dry and moisture conditioned specimens to assess the extent of moisture-induced damage. Indirect

tensile strength, complex modulus in tension or compression, fatigue life are just a few examples. By using different combinations of moisture conditioning procedures and mechanical tests, it is possible to conceive a multitude of different ways by which the moisture sensitivity of the mix can be quantified in terms of a ratio.

The two main advantages of this approach are that: (1) it is simple to conduct and easy to interpret, and (2) this methodology measures the net impact of moisture on the mixtures' mechanical properties including the moisture transport process within the mixture, cohesive softening of the binder or mastic, and adhesive stripping of the binder from the coarse and fine aggregate particles. In fact, owing to these advantages, the use of moisture sensitivity ratio is very commonly used as a screening and mixture design verification tool by several highway agencies and also to some extent by researchers.

Despite its advantages, there are some drawbacks associated with the use of moisture sensitivity ratio as a measure of moisture damage resistance for asphalt mixtures. First, this methodology implicitly assumes that the accelerated moisture conditioning process (e.g., freeze thaw cycles or vacuum saturation followed by conditioning) is a realistic representation of what happens in the field, which may or may not be true in all cases. Second, the biggest advantage of this methodology, i.e., capturing the net affect of all the mechanisms that drive moisture-induced damage is also its biggest disadvantage. Simply put, if a particular mixture is deemed to be moisture sensitive, it is nearly impossible to discern using this methodology the reason that makes the mix moisture sensitive, which in turn makes it nearly impossible to figure out a solution to fix the problem. Based on the several different mechanisms, discussed earlier in this chapter, one can speculate that a mix that is identified as being moisture sensitive could have an air void structure that allows water to be trapped within the bulk of the material, or a binder that allows faster diffusion of water through it and is more prone to softening in the presence of moisture, or an aggregate-binder pair that is thermodynamically less compatible and therefore prone to stripping, or an aggregate that has extremely low specific surface area providing poor mechanical bonding with the binder or a combination of any of the above. Unless the predominant cause for the moisture sensitivity of the material is known, it would not be possible to recommend a solution to address the issue. From this point of view, the methodology based on the moisture sensitivity ratio leaves the materials engineer simply guessing for solutions upon encountering a problematic mix.

The solution to the above shortcoming is that the components of a mixture and the mixture itself can be evaluated for material properties that drive moisture damage based on the mechanisms that were discussed earlier. This is discussed in the following subsection. While this approach may be apparently time consuming, it can result in significant cost savings for regions that frequently encounter issues related to moisture-induced damage.

8.5.5.2 Fundamental Methods to Evaluate Sensitivity of Asphalt Materials and Mixture to Moisture-Induced Damage

Moisture damage is a process that involves several different mechanisms including moisture transport through macroscopic interconnected voids in the mixture, pumping and scouring of the mastic on the pavement surface, diffusion of the moisture through the asphalt binder or mastic, softening or increased susceptibility of the binder or mastic to failure in the presence of moisture, migration of moisture to the binder-aggregate interface, and subsequent stripping. Each of these mechanisms can be modeled using specific material properties. For example, moisture transport through the mix can be modeled using geometry of the interconnected air voids, surface tension of water, and moisture gradients at the boundaries of the pavement or the macroscopic diffusivity of the moisture through the mix. Similarly, moisture transport through the binder or the mastic can be modeled by knowing the diffusivity of moisture through these materials. The propensity to cohesive and adhesive failure can be modeled by knowing the traction–separation behavior of the binder and the binder-aggregate surface in the absence and presence of moisture. Ultimately, in order to predict the overall moisture sensitivity of a mixture, all of these properties can be integrated into a comprehensive mathematical or computational model that predicts the behavior of the mixture in the presence of moisture. Readers interested in examples of different computational models and approaches that have been used to evaluate the moisture sensitivity of different mixtures are referred to additional literature (Kringos et al. 2008; Kringos 2007; Caro et al. 2010; Copeland 2007). Figure 8.25 illustrates the results from a typical computational simulation of a small section of an asphalt mixture; the results compare the results of the mix in dry state to two different mixes with different moisture diffusion rates.

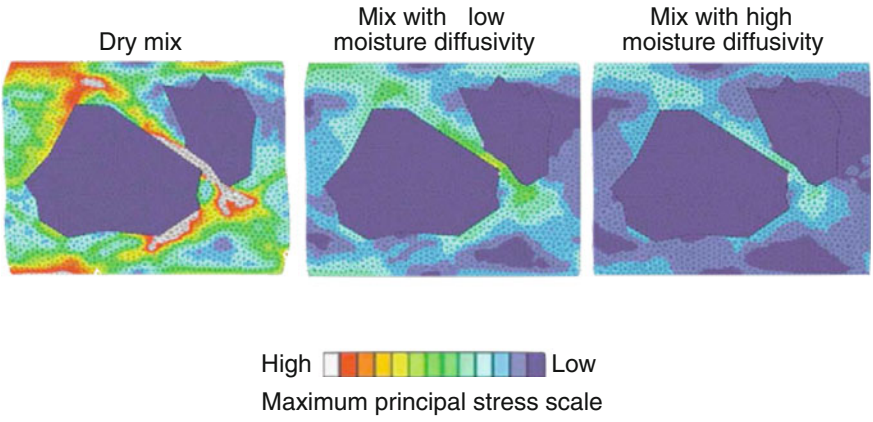


Fig. 8.25 Typical results from a computational model showing the influence of different durations of moisture exposure on the properties of the mix; the results show the maximum stress before crack initiation under dry condition and after moisture conditioning for 20 days for two different mixes with different rates of moisture diffusion (With permission from ASCE, Caro et al. 2010)

An alternative to a comprehensive mathematical model is to focus on one or more of the specific mechanisms and measurement methods that drive moisture damage, if such issues are identified a priori through experience or a process of trial and error. For example, in some cases, a highway agency may have identified certain binders or aggregates or binder-aggregate pairs to be problematic in terms of their high susceptibility to stripping. In this case, potential solutions would involve screening binder-aggregate pairs that have optimal resistance to moisture damage in terms of stripping. Such screening can be accomplished through the use of a more fundamental approach such as by measuring the surface free energies of the different binders (with and without any modifiers that could potentially be used) and aggregates and using this information to compute the potential for moisture sensitivity for all the different possible combinations of the aggregate-binder pairs. As another example, if certain mixes are more prone to moisture damage, despite the high inherent resistance of the binder-aggregate pairs to moisture-induced damage, then the moisture transport properties of the mix may be more closely examined (e.g., in terms of the void structure). Readers are directed to additional literature on the methods to measure and model different process that drive moisture-induced damage in asphalt mixtures.

8.5.5.3 Summary

To summarize, moisture-induced damage is a process that involves several different mechanisms including moisture transport at multiple length scales, scouring, cohesive damage, and adhesive damage. In this sense, moisture-induced damage can act as an amplifier to other forms of distress such as fatigue cracking. One holistic approach to quantitatively estimate the moisture sensitivity of any given mix is in terms of the moisture sensitivity ratio. This ratio can be determined in several different ways but it typically involves preparing a batch of mixture specimens, dividing the batch into two groups, using one group as a control, moisture conditioning the second group, measuring one or more mechanical properties of the control and conditioned group, and expressing moisture sensitivity of the mix as the ratio of the property for the moisture conditioned specimens to the control specimens. This holistic approach is simple to execute but incapable of identifying the dominant mechanisms of failure, if any. The alternative approach is focused on individual mechanisms. For example, one can measure rate of moisture transport through the mixture and binder and the degradation of adhesive and cohesive properties as a function of the moisture content. This information can be used in isolation to evaluate and compare the individual mechanisms of moisture-induced damage OR combined using a numerical or computational approach to examine the combined effect of all mechanisms in a specific mixture.

8.6 Exercises

8.1. Locate and image an asphalt pavement with distress (rutting, fatigue cracking, thermal cracking or surface damage). Discuss the failure observed and specifically address the potential causes for the distress in terms of materials (component materials and mixture design), structure, and construction as applicable. Exercise caution and beware of traffic and avoid stepping onto the road way.

8.2. Compute the minimum pavement temperature for your geographic location using air temperature data from the nearest available weather station at different pavement depths. To be conservative for the design and selection of materials resistant to low-temperature cracking, would you use the air temperature as the basis or the pavement temperature you have computed above?

8.3. For your local area (county, city, state), research and determine whether any mixture performance qualification tests are typically required. What are these tests? Would you categorize these tests as indicator type tests or fundamental?

References

Ahmed, S., et al. (2013). Cracking resistance of thin-bonded overlays using fracture tests, numerical simulations and early field performance. *International Journal of Pavement Engineering, 16*(6), 540–552.

American Association of State Highway and Transportation Officials. (2005). *Standard method of test for determining dynamic modulus of hot-mix asphalt concrete mixtures*, TP62 AASHTO.

Arambula, E., Masad, E., & Martin, A. E. (2007). Influence of air void distribution on the moisture susceptibility of asphalt mixes. *Journal of Materials in Civil Engineering, 19*(8), 655–664.

Arega, Z. A., Bhasin, A., & De Kesel, T. (2013). Influence of extended aging on the properties of asphalt composites produced using hot and warm mix methods. *Construction and Building Materials, 44,* 168–174.

Bazin, P., & Saunier, J. (1967, January). Deformability, fatigue and healing properties of asphalt mixes. In *Intl Conf Struct Design Asphalt Pvmts*.

Bhasin, A., et al. (2009). Quantitative comparison of energy methods to characterize fatigue in asphalt materials. *Journal of Materials in Civil Engineering (ASCE), 21*(2), 83–92.

Bhasin, A., Button, J., & Chowdhury, A. (2003). *Evaluation of simple performance tests on HMA mixtures from the South Central USA*. TX: College Station.

Bhasin, A., & Motamed, A. (2011). Analytical models to characterise crack growth in asphaltic materials and healing in asphalt binders. *International Journal of Pavement Engineering, 12* (4), 371–384.

Bhasin, A., Little, D. N., Bommavaram, R., & Vasconcelos, K. (2008). A framework to quantify the effect of healing in bituminous materials using material properties. *Road Materials and Pavement Design*, *9*(sup1), 219–242.

Broek, D. (1982). *Elementaty engineering fracture mechanics*.

Caro, S., et al. (2008). Moisture susceptibility of asphalt mixtures, Part 1: Mechanisms. *International Journal of Pavement Engineering, 9*(2), 81–98.

Caro, S., et al. (2010). Coupled micromechanical model of moisture-induced damage in asphalt mixtures. *Journal of Materials in Civil Engineering (ASCE), 22*(4), 380–388.

Cohen, B., et al. (1997). Potholes and politics: How congress can fix your roads.

Copeland, A.R. (2007). *Influence of moisture on bond strength of asphalt-aggregate systems*.

Daniel, J. S., & Kim, Y. R. (2002). Development of a simplified fatigue test and analysis procedure using a viscoelastic, continuum damage model (with discussion). *Journal of the Association of Asphalt Paving Technologists*, *71*.

Dave, E. V., & Buttlar, W. G. (2010). Thermal reflective cracking of asphalt concrete overlays. *International Journal of Pavement Engineering, 11*(6), 477–488.

Griffith, A. A. (1921). The phenomenon of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London, A221,* 163–198.

Hajj, R., Bhasin, A. (2017). The search for a measure of fatigue cracking in asphalt binders-a review of different approaches. *International Journal of Pavement Engineering*, 1–15.

Hefer, A., Little, D. N., & Lytton, R. L. (2005). A synthesis of theories and mechanisms of bitumen-aggregate adhesion including recent advances in quantifying the effects of water. *Proceedings of the Association of Asphalt Paving Technologists, 74,* 139–196.

Hu, J., Qian, Z., Liu, Y., & Zhang, M. (2015). High-temperature failure in asphalt mixtures using micro-structural investigation and image analysis, *84*, 136–145.

Karki, P., Li, R., & Bhasin, A. (2015). Quantifying overall damage and healing behavior of asphalt materials using continuum damage approach. *International Journal of Pavement Engineering, 16*(4), 350–362.

Karki, P., Bhasin, A., & Underwood, B. S. (2016). Fatigue Performance Prediction of Asphalt Composites Subjected to Cyclic Loading with Intermittent Rest Periods. *Transportation Research Record: Journal of the Transportation Research Board*, (2576), 72–82.

Kim, Y., et al. (2004). Dynamic modulus testing of asphalt concrete in indirect tension mode. *Transportation Research Record, 1891*(1), 163–173.

Kringos, N. (2007). *Modeling of combined physical-mechanical moisture induced damage in asphaltic mixes*.

Kringos, N., Scarpas, A., & De Bondt, A. (2008). Determination of moisture susceptibility of mastic-stone bond strength and comparison to thermodynamical properties. *Journal of the association of asphalt paving technologists 77*.

Kringos, N., Schmets, A., Scarpas, A., & Pauli, T. (2011). Towards an understanding of the self-healing capacity of asphaltic mixtures. *Heron*, *56*(1/2), 45–74.

Lee, H. J., & Kim, Y. R. (1998). Viscoelastic constitutive model for asphalt concrete under cyclic loading. *Journal of Engineering Mechanics*, *124*(1), 32–40.

Li, R., et al. (2015). Rheological and low temperature properties of asphalt composites containing rock asphalts. *Construction and Building Materials, 90,* 47–54.

Malvern, L. E. (1969). *Introduction to the mechanics of a continuous medium*. New Jersey: Princeton Hall.

Masad, E., Castelblanco, A., & Birgisson, B. (2006). Effects of Air Void Size Distribution, Pore Pressure, and Bond Energy on Moisture Damage. *Journal of Testing and Evaluation*, *34*(1), 1–9.

McGennis, R.B., Anderson, R.M., Kennedy, T.W., & Solaimanian, M. (1994). Background of Superpave Asphalt Mixture Design and Analysis, Report Number FHWA-SA-95-003, Performing Organization: Asphalt Institute, Lexington, KY, Sponsor: Federal Highway Administration, Washington, DC.

Mohseni, A. (1998). *LTPP seasonal asphalt concrete (AC) pavement temperature models*. VA: McLean.

Morland, L. W., & Lee, E. H. (1960). Stress analysis for linear viscoelastic materials with temperature variation. *Transactions of The Society of Rheology (1957–1977)*, *4*(1), 233–263. http://scitation.aip.org/content/sor/journal/tsor/4/1/10.1122/1.548856.

Muki, R., & Sternberg, E. (1961). On transient thermal stresses in viscoelastic materials with temperature-dependent properties. *Journal of Applied Mechanics*, *28*(2), 193–207. http://dx.doi.org/10.1115/1.3641651.

Ozer, H., Al-Qadi, I. L., Lambros, J., El-Khatib, A., Singhvi, P., & Doll, B. (2016). Development of the fracture-based flexibility index for asphalt concrete cracking potential using modified semi-circle bending test parameters. *Construction and Building Materials*, *115*, 390–401.

Paris, P. C., & Erdogan, F. (1963). A critical analysis of crack propagation laws. *Journal of Basic Engineering, 85,* 528–534.

Park, S. W., Kim, Y. R., & Schapery, R. A. (1996). A viscoelastic continuum damage model and its application to uniaxial behavior of asphalt concrete. *Mechanics of Materials*, *24*(4), 241–255.

Roque, R., et al. (2004). Development and field evaluation of energy-based criteria for top-down cracking performance of hot mix asphalt. *Journal of the Association of Asphalt Paving Technologist, 73,* 229–260.

Schapery, R. A. (1981). On viscoelastic deformation and failure behavior of composite materials with distributed flaws. *1981 advances in aerospace structures and materials*, 5–20.

Schapery, R. A. (1984). Correspondence principles and a generaiized J integral for large deformation and fracture analysis of viscoelastic media. *International Journal of Fracture, 25,* 195–223.

Schwartz, C. W., & Kaloush, K. E. (2009). Permanent deformation assessment for asphalt concrete pavement and mixture design. In Y. Richard Kim (Ed.), *Modeling of asphalt concrete*, (pp. 317–351). New York: McGraw-Hill Professional.

Sultana, S., & Bhasin, A. (2014). Effect of chemical composition on rheology and mechanical properties of asphalt binder. *Construction and Building Materials, 72,* 293–300.

Underwood, B. S., Kim, Y. R., & Guddati, M. N. (2010). Improved calculation method of damage parameter in viscoelastic continuum damage model. *International Journal of Pavement Engineering*, *11*(6), 459–476.

Wang, H., & Al-Qadi, I. (2009). Combined effect of moving wheel loading and three-dimensional contact stresses on perpetual pavement responses. *Transportation Research Record: Journal of the Transportation Research Board*, (2095), 53–61.

Witczak, M., Schwartz, C. & Von Quintus, H. (2001). *NCHRP Project 9-19: Superpave support and performance models management*. Washington, D.C.

Part II
Mechanics

Chapter 9
Mechanics of Continuous Solids

9.1 Introduction

The subject of this text is the study of the mechanics, thermodynamics, and chemistry of flexible road ways. In most cases, these concepts are deployed herein within the context of an idealization termed *continuum mechanics*, wherein it is assumed that there exists mass at every point in the object of interest. This assumption is approximate, but is nevertheless shown to be accurate for the purpose of modeling the thermomechanical response of many objects, especially where deformations and fracture are to be predicted, as in the case of road ways. An object that is modeled using this methodology is called a *continuum* or a *continuous body*. Augustin-Louis Cauchy (1789–1857) is regarded as the father of modern continuum mechanics. In a memoir presented to the Paris Académie on September 30, 1822, he introduced the concept of stress based on the assumption that a body is everywhere continuous (Maugin 2014). This may be regarded to be the fundamental underlying assumption common to the design and failure analysis of all macroscopic structural components today, including flexible pavements.

9.2 Mathematical Preliminaries

In order to construct scientifically based models for predicting the performance of flexible road ways, it is necessary to employ certain advanced mathematical concepts. This section introduces those concepts and the notation necessary to utilize those tools. Toward this end, Table 9.1 lists some mathematical terms that will be employed at various times within Part II of this text.

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_9

Table 9.1 Mathematical terminology

Mathematical symbol	Meaning
$\forall$ (right)	all (right)
(left) $\in$ (right)	(left) in (right)
(left) $\Rightarrow$ (right)	(left) implies that (right)
(left) $\Leftrightarrow$ (right)	$\Rightarrow$ and (right) implies that (left)
(left) iff (right)	(left) is true if and only if (right)
(left) $\equiv$ (right)	(left) is defined to be equivalent to (right)
(left) $\|$ (right)	(left) is parallel to (right)
$\emptyset$	a null set
(left) $\cap$ (right)	(left) intersected with (right)
(left) • (right)	vector dot product of (left) with (right)
$\|f\|$	magnitude of f
$\|\|f\|\|$	determinant of f
const	constant
$G_{\tau=0}^{\tau=t}\{f(\tau)\}$	function of the entire history, a functional

9.2.1 Index Notation

The study of flexible pavements utilizes the concepts developed within the field of continuum mechanics. A necessary consequence of the assumption that a body is continuous is that the body is modeled via differential and integral calculus, invented more or less simultaneously by Isaac Newton (1642–1727) and Gottfried Leibniz (1646–1716). Furthermore, the predictive methodologies are developed for three-dimensional objects that undergo changes in space and time.

Toward this end, consider an arbitrary three-dimensional continuous body with interior, V, and boundary, S, as shown in Fig. 9.1. For purposes of convenience, the entire body is denoted as $V + S$, and the spatial description of the body is constructed in terms of the position vector, $\vec{x}$, which in many practical circumstances may be written in Cartesian (mutually perpendicular, rectilinear, and right handed)

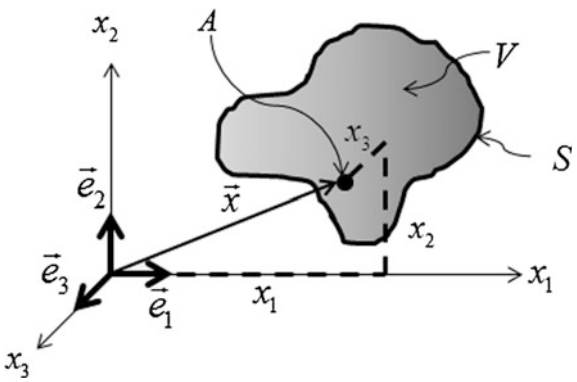


Fig. 9.1 Depiction of a three-dimensional continuum

coordinates (after René Descartes (1596–1650)) in this text, and is described in terms of components, x_i, and unit base vectors in the three coordinate directions, $\vec{e}_i$.

Thus, an arbitrary position vector, $\vec{x}$, may be written as follows, as shown in the figure:

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3 = x_i\vec{e}_i \tag{9.1}$$

Note that in the above *Einstein notation* has been utilized (after Albert Einstein (1879–1955)), commonly called *index notation*, and the following range and summation conventions have been employed:

(1) *Range Convention*: a subscript that is not repeated in a product is termed a free index, and it is interpreted to represent the entire range (3, unless otherwise noted) of its possible indices; and
(2) *Summation Convention*: a subscript that is repeated within a product is implied to be summed over its entire range (also 3, unless otherwise noted).

The above notation will be utilized extensively in this text. Thus, for example, according to the summation and range conventions:

$$a_i b_i \equiv a_1 b_1 + a_2 b_2 + a_3 b_3 \tag{9.2}$$

Furthermore,

$$c_{jj} \equiv c_{11} + c_{22} + c_{33} \tag{9.3}$$

In addition, the following simplified notation is utilized throughout this text:

$$f_{i,j} \equiv \frac{\partial f_i}{\partial x_j} \tag{9.4}$$

Thus, utilizing (9.3) and (9.4) together implies that

$$f_{k,k} \equiv \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3} \tag{9.5}$$

A special symbol that will be used throughout this text is called *the Kronecker delta* (after Leopold Kronecker (1823–91)), defined as follows:

$$\delta_{ij} \equiv \left\{ \begin{array}{lll} 0, & if & i \neq j \\ 1, & if & i = j \end{array} \right\} \tag{9.6}$$

The Kronecker delta has some interesting properties, such as the following:

$$\vec{e}_i \cdot \vec{e}_j = \delta_{ij} \tag{9.7}$$

In addition, the Kronecker delta is sometimes called the substitution operator (with respect to free indices) due to the following property:

$$a_i \delta_{ij} = a_j \tag{9.8}$$

9.2.2 Scalars, Vectors, and Tensors

One of the reasons that index notation is used in this book is due to the fact that coordinate transformations are easily performed with this notation. To see how this works, consider the position vector, $\vec{x}$, shown in Fig. 9.1. Suppose that this same vector is written with respect to another set of Cartesian coordinates originating at the same point but rotated with respect to the coordinates shown in Fig. 9.1, and this coordinate system is distinguished from the first by the use of primed superscripts, as shown in Fig. 9.2.

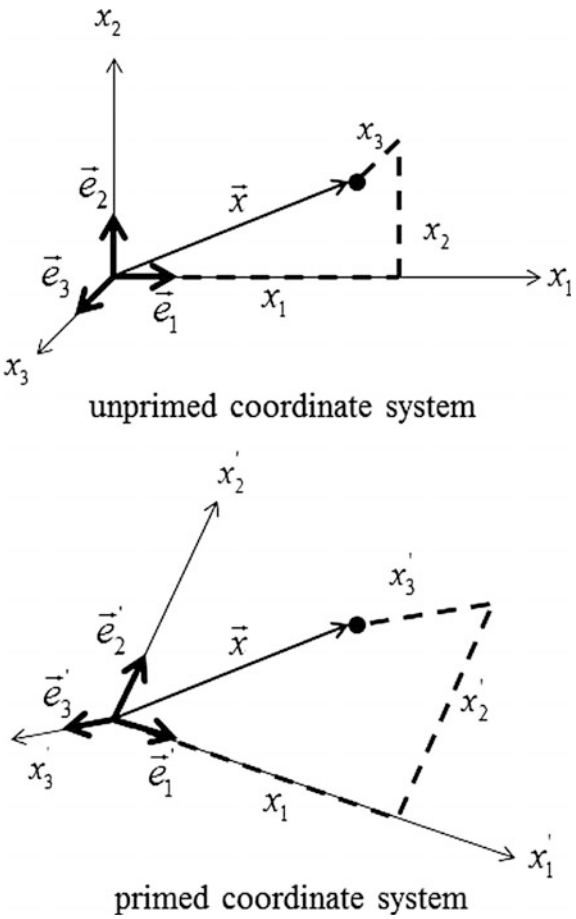


Fig. 9.2 Two different Cartesian coordinate systems used to define the position vector $\vec{x}$

Thus, it can be seen that

$$\vec{x} = x_i'\vec{e}_i' = x_i\vec{e}_i \tag{9.9}$$

The above equation may be used to determine the relationship between the components of $\vec{x}$ in the unprimed and primed coordinate systems by taking the dot product with a generic base vector in the primed coordinate system:

$$x_i'\vec{e}_i' \cdot \vec{e}_j' = x_i\vec{e}_i \cdot \vec{e}_j' \tag{9.10}$$

Note that, due to Eqs. (9.7) and (9.8), the above may be simplified to the following:

$$x_j' = a_{ij'}x_i \Leftrightarrow x_i = a_{j'i}x_j' \tag{9.11}$$

where *the direction cosine tensor,* $a_{ij'}$, is defined as follows:

$$a_{ij'} \equiv \vec{e}_i \cdot \vec{e}_j' = \cos(\vec{e}_i, \vec{e}_j') \tag{9.12}$$

It is apparent that Eq. (9.12) relates the two coordinate systems via the angles between the base vectors, assumed to be known a priori. Thus, if the components of the vector $\vec{x}$ are known in the unprimed coordinate system, Eq. (9.11) can be used to calculate the components of $\vec{x}$ in the primed coordinate system.

The direction cosine tensor should be utilized carefully. For example, due to definition (9.12), the direction cosine tensor is symmetric, that is:

$$a_{ij'} = a_{j'i} \tag{9.13}$$

On the other hand, the following can be verified from definition (9.12):

$$a_{ij'} \neq a_{ji'} \tag{9.14}$$

A variable that transforms from one Cartesian coordinate system to another (rotated) one via a single multiplication with the direction cosine tensor, $a_{ij'}$, is thus called a *vector*, or equivalently, a *first-order tensor*. An example is the position vector, $\vec{x}$, as is proven by Eq. (9.11). A variable that transforms without multiplication by the direction cosine tensor is called a *scalar*, or a *zero-order tensor*. An example is temperature.

Alternatively, a variable that transforms from one Cartesian coordinate system to another via two multiplications by the direction cosine tensor is called a *second-order tensor*. An example is the Kronecker delta, which can be shown to be a second-order tensor by using Eqs. (9.7) and (9.11) as follows:

$$\delta_{ij} = \vec{e}_i \cdot \vec{e}_j = (a_{ip'}\vec{e}'_p) \cdot (a_{jq'}\vec{e}'_q) = a_{ip'}a_{jq'}\delta_{p'q'} \tag{9.15}$$

The reader may by now have become suspicious that the number of subscripts attached to a variable determines the tensor order of that variable. Unfortunately, this is not always the case, as will be demonstrated in due course. Therefore, the reader is cautioned to always ensure that the order of a variable is determined by the number of multiplications by the direction cosine tensor required to transform from one Cartesian coordinate system to another.

There are also tensors of higher order than two, as will also be shown in Chap. 11.

9.2.3 Linearity

When one variable, $\mathcal{R}$, called the response, depends on another, $\mathcal{I} = \mathcal{I}(t)$, called the input, at a given time, t, it is called a function, and the function is denoted as $\mathcal{R}(t) = \mathcal{R}(\mathcal{I}(t))$. Conversely, when the response variable depends on the entire time history of the input variable, it is termed a functional, and it is denoted as $\mathcal{R}\{t\} = \mathcal{R}\{\mathcal{I}(\tau)\}$, where $-\infty \leq \tau \leq t$.

Given an input functional, $\mathcal{I}$, and a response functional, $\mathcal{R}$, a linear equation is defined to be one that satisfies the following conditions:

Superposition:

$$\mathcal{R}\{\mathcal{I}_1 + \mathcal{I}_2\} = \mathcal{R}\{\mathcal{I}_1\} + \mathcal{R}\{\mathcal{I}_2\} \tag{9.16a}$$

Homogeneity:

$$\mathcal{R}\{c\mathcal{I}\} = c\mathcal{R}\{\mathcal{I}\} \quad c = const \tag{9.16b}$$

where $\mathcal{I}$ is the input, and $\mathcal{R}$ is the output. Note that, since functions are special cases of functionals, Eqs. 9.16a and 9.16b also apply to functions. Accordingly, ***an equation, whether of algebraic, differential, or integral type is linear iff the above two conditions are satisfied.*** The concept of linearity will play an important role in the development of models for predicting road way performance.

9.2.4 Laplace Transforms

Laplace transforms (after Pierre-Simon Laplace (1749–1827)) can often be used to perform mathematical operations involving differential and integral equations such as those to be encountered in this text. Thus, the Laplace transform, $\mathcal{L}\{f(t)\}$, of a function, $f(t)$, is defined as follows:

$$\mathcal{L}\{f(t)\} = \bar{f}(s) \equiv \int_0^\infty e^{-st} f(t)dt \tag{9.17}$$

It can be seen from the above definition that the Laplace transform changes the independent variable from t (normally time in this text) on the left-hand side of Eq. (9.17) to s on the right-hand side. Furthermore, the properties of the Laplace transform are such that ordinary differential equations in t are converted to algebraic equations in s. The inverse Laplace transform, $\mathcal{L}^{-1}\{\bar{f}(s)\}$, of a function, $\bar{f}(s)$, can oftentimes be obtained from the following formula:

$$\mathcal{L}^{-1}\{\bar{f}(s)\} = f(t) = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{\gamma - iT}^{\gamma + iT} e^{st}\bar{f}(s)ds \tag{9.18}$$

Laplace transforms of commonly used functions can be found in reference books on the subject, and inversion of a typical function $\bar{f}(s)$ to the time domain can be accomplished by using these tables in reverse, or by employing a math software package such as MATLAB (Gilat 2008). Many of the Laplace transform pairs used in this text are listed in Table 9.2.

Table 9.2 Laplace transforms pairs

Equation no.	$f(t)$	$\bar{f}(s)$
1	$\delta(t)$	1
2	$H(t)$	$\frac{1}{s}$
3	t	$\frac{1}{s^2}$
4	t^2	$\frac{2}{s^3}$
5	$t^n \quad (n = 0, 1, 2, \ldots)$	$\frac{n!}{s^{n+1}}$
6	$t^n \quad (\forall n > -1)$	$\frac{\Gamma(n+1)}{s^{n+1}}$
7	$e^{-at} \quad (a = const)$	$\frac{1}{s+a}$
8	$\frac{1}{a}(1 - e^{-at}) \quad (a = const)$	$\frac{1}{s(s+a)}$
9	$\frac{t}{a} - \frac{1}{a^2}(1 - e^{-at}) \quad (a = const)$	$\frac{1}{s^2(s+a)}$
10	$e^{-at}\cos(bt) \quad (a, b = const)$	$\frac{s+a}{(s+a)^2 + b^2}$
11	$e^{-at}\sin(bt) \quad (a, b = const)$	$\frac{b}{(s+a)^2 + b^2}$
12	$e^{-at}t^n \quad (a = const, \forall n > -1)$	$\frac{\Gamma(n+1)}{(s+a)^{n+1}}$
13	$f(t - t_0)H(t - t_0) \quad (t_0 = const)$	$e^{-st_0}\bar{f}$
14	$\int_0^t f(t - \tau)g(\tau)d\tau$	$\bar{f} \cdot \bar{g}$

9.2.5 Carson Transforms

Carson transforms (after John R. Crenshaw (1886–1940)) are sometimes preferable to Laplace transforms for mathematical operations involving the materials to be discussed herein. The Carson transform is, therefore, defined as follows:

$$\tilde{f}(s) \equiv s\tilde{f}(s) \tag{9.19}$$

9.2.6 The Heaviside Step Function

The *Heaviside step function* (after Oliver Heaviside (1850–1925)), $H(t - t_1)$, is defined as follows:

$$H(t - t_1) \equiv \left\{ \begin{array}{ll} 0 & if \quad t - t_1 < 0 \\ 1 & if \quad t - t_1 \geq 0 \end{array} \right\} \tag{9.20}$$

where t is the generic variable of interest (normally time in the present text), and t_1 is a particular value of the variable, t. Thus, the Heaviside step function may be plotted graphically as shown in Fig. 9.3.

The Heaviside step function will be utilized extensively in this text.

9.2.7 The Convolution Integral

Sometimes it is convenient to employ hereditary integrals in time to model the behavior of continuous media. To see how this works, consider an object subjected to an kinetic input, $I(t)$, of the following form:

$$I(t) = I_0 H(t) \quad I_0 = const \tag{9.21}$$

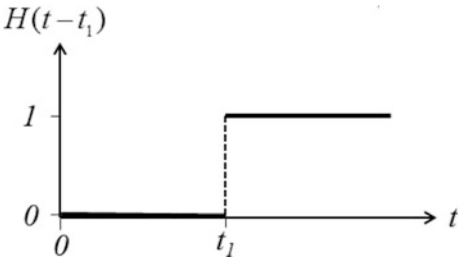


Fig. 9.3 Graphical depiction of the Heaviside step function

The object is observed to produce a kinematic output denoted, $R(t)$, such that *the compliance function*, *C(t),* may be defined as follows:

$$C(t) \equiv \frac{R(t)}{I_0} \tag{9.22}$$

Now, suppose that Eq. (9.22) is subsequently shown to satisfy linearity conditions (9.16a and 9.16b), thus implying that the response, $R(t)$, is linear with respect to the input, $I(t)$. Suppose furthermore that the object is now subjected to a series of stepwise inputs given by

$$I(t) = \sum_{i=1}^{n} I_i H(t - t_i) \quad I_i = const \tag{9.23}$$

as shown graphically in Fig. 9.4.

It then follows that since the response is linear, superposition holds and

$$\begin{aligned} R(t) &= \sum_{i=1}^{n} C(t - t_i) \frac{\Delta I_i}{\Delta t_i} \Delta t_i \Rightarrow \\ R(t) &= \lim_{\Delta t_i \to 0, n \to \infty} \sum_{i=1}^{n} C(t - t_i) \frac{\Delta \varepsilon_i}{\Delta t_i} \Delta t_i \Rightarrow R(t) = \int_0^t C(t - \tau) \frac{\partial I}{\partial \tau} d\tau \end{aligned} \tag{9.24}$$

where the generic timescale has been changed to τ.

The above hereditary integral is called a *convolution integral* and is also a variation of a *Boltzmann integral* (after Ludwig Boltzmann (1844–1906)). Thus, it can be seen that if the object is linear, the response to a complicated function of time can be calculated by time convoluting the compliance obtained from the simple input represented by Eq. (9.20) against the rate of change of the input over time. This important fact will become important when viscoelastic media are introduced in Chaps. 10 and 12.

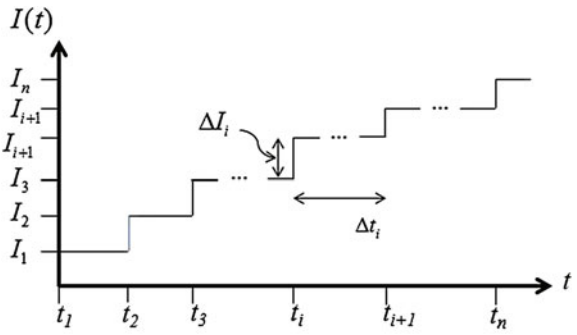


Fig. 9.4 Depiction of multiple steps of the input, $I(t)$

Similarly, it can be shown that an inverse convolution integral may be constructed as follows:

$$I(t) = \int_0^t D(t-\tau)\frac{\partial R}{\partial \tau}d\tau \tag{9.25}$$

where *D(t)*, called the modulus, is convoluted against the input kinematic history, $R(t)$, to produce the output kinetic history, $I(t)$.

It will be shown later that the convolution integral is often useful for modeling the material behavior of road way media.

9.2.8 The Dirac Delta Function

Suppose that the input similar to Eq. (9.21) is applied. According to Eq. (9.24), the resulting response will be given by

$$R(t) = \int_0^t C(t-\tau)\frac{\partial [I_0 H(\tau - t_1)]}{\partial \tau}d\tau \tag{9.26}$$

It can be seen that in the above integral the derivative of the Heaviside step function is undefined at the time $\tau = t_1$, and zero for all other times. Despite this discomforting revelation, it turns out that the integral of the function still exists and is (usually) single-valued. This fact will not be proven herein for the sake of time, but the theorem will be simply stated herein. To do this, first define the *Dirac Delta Function* (after Paul Dirac (1902–84)), $\delta(\tau - t_1)$, as follows:

$$\delta(\tau - t_1) \equiv \frac{\partial H(\tau - t_1)}{\partial \tau} \tag{9.27}$$

The sifting property is now stated as follows:

$$\int_{t-\Delta t}^{t+\Delta t} f(\tau)\delta(\tau - t)d\tau = f(t) \tag{9.28}$$

It therefore follows from Eq. (9.28) that Eq. (9.26) simplifies to the following:

$$R(t) = C(t - t_1)I_0 \tag{9.29}$$

thereby recovering the form described in Eq. (9.22).

9.2.9 The Divergence Theorem

The fundamental theorem of calculus states that, given a function, $f = f(x)$, that is continuous on the range $x_1 \leq x \leq x_2$, then

$$\int_{x_1}^{x_2} \frac{df}{dx} dx = f(x)|_{x_1}^{x_2} \tag{9.30}$$

It can thus be seen that the change in the function, $f(x)$, between two arbitrary points can be evaluated without recourse to integral calculus if the value of $f(x)$ is known at the limits x_1 and x_2. The above theorem can be used via careful manipulation of each coordinate direction to produce a three-dimensional version of the fundamental theorem of calculus. This result, called *The Divergence Theorem* or *Gauss' Theorem* (after Johann Carl Friedrich Gauss (1777–1855)), is stated without proof herein as follows (see Problem 9.6). Given a vector-valued bounded function of three-dimensional coordinates, $\vec{f}(\vec{x})$, the following identity must hold for any volume, V, with boundary, S, in a continuous body:

$$\int_V f_{i,i} dV = \int_S f_i n_i dS \tag{9.31}$$

where n_i are the components of the unit outer normal on the boundary, S. The above theorem will be used repeatedly throughout this text.

9.2.10 The Reynolds Transport Theorem

The Reynolds transport theorem states that, given a scalar-, vector-, or tensor-valued function, $F(\vec{x})$, (after Osborne Reynolds (1842–1912) (Malvern 1969):

$$\frac{d}{dt} \int_V \rho F dV = \int_V \rho \frac{dF}{dt} dV \tag{9.32}$$

The above form of the Reynolds transport theorem may not be familiar to some readers (such as fluid mechanists), and that is because it is written herein in a Lagrangian description (the approach to be employed herein), as opposed to an Eulerian description, to be discussed in detail in the following section.

This completes the prerequisite mathematical tools necessary to assimilate the materials in this text.

9.3 Kinematics and Strain

Consider once again a three-dimensional object with interior, V, and boundary, S, as shown in Fig. 9.1. The primary objective of mechanics is to predict the *motions* at all points both within a continuum and on its boundary due to externally applied loads as a function of the spatial coordinates, x_i, and time, t. The variables x_i and t are called the *independent variables*. *Kinematics* deals with the study of the motion of objects without regard to the external loads that produce these motions.

The objective is therefore to produce a model that is capable of predicting the set of *output variables* that are necessary to predict the motions of an arbitrary continuum as functions of these independent variables, as well as any *input variables*, including *loads, geometry, and material properties*. In order to accomplish this objective, consider the set of output variables (also sometimes called *dependent* or *field* variables) that are necessary to predict the motions of a continuous body. When mechanics problems involving road ways are considered, *deformations* come to mind first and foremost, where *deformable motions* are defined to be any motions that result in a change in the initial shape of the continuous body.

Toward this end, first define the *displacement*, u_i, of an arbitrary material point A in the body to be the difference between the location of point $\bar{A}$ in the deformed configuration $(t = t_1, \forall t_1 > t_0)$ and the location of that same point in the initial (undeformed) configuration $(t = t_0)$, denoted A, as shown in Fig. 9.5. For reasons that will become clear below, a second arbitrary point B (denoted $\bar{B}$ in the deformed configuration) is also shown in the figure.

It is self-evident that the continuous object is *deformed* iff the distance between any two arbitrary points A and B is a function of time, t. Otherwise, the object is termed a *rigid* body. Toward this end, consider Fig. 9.6, wherein a nomenclature is constructed pictorially describing the kinematics of the arbitrary points A and B.

It can be seen from the figure that the components of the coordinates of material points in the initial configuration are denoted by a_i (as distinguished from the direction cosine tensor, $a_{ij'}$, which contains two subscripts). Furthermore, the displacement of material points from their initial positions is described in terms of their initial locations, that is,

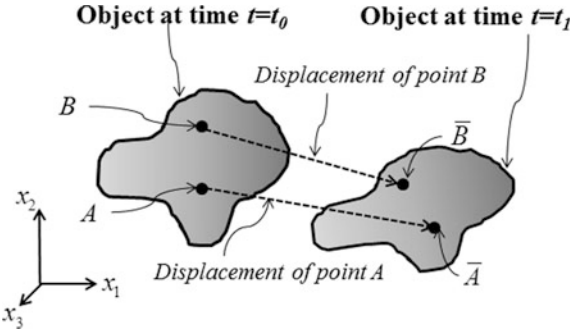


Fig. 9.5 Displacements of two arbitrary points A and B in a three-dimensional continuum

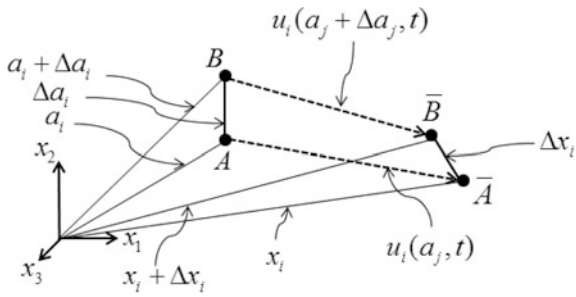


Fig. 9.6 Kinematics of the points A and B

$$x_i \equiv x_i(a_j, t) \tag{9.33}$$

Such a description of the displacement field is termed *a Lagrangian description* (after Joseph-Louis Lagrange (1736–1813), or Giuseppe Luigi Lagrangia in his native Italy), as opposed to *an Eulerian description* (after Leonhard Euler (1707–83)), described as follows:

$$a_i \equiv a_i(x_j, t) \tag{9.34}$$

These two different ways of viewing the displacement field may be distinguished by considering a practical example. Suppose that a vehicle is traveling down the highway. If the velocity of the vehicle is measured as a function of time, it is termed a Lagrangian description. On the other hand, if a highway patrolman is sitting in a vehicle on the side of the road and is measuring vehicular speeds with a radar device, the device is measuring velocities of passing vehicles at a fixed coordinate location on the highway. This perspective is termed an Eulerian description. In the case of asphaltic binder, each material point possesses memory, as we will see in Chap. 12. For this reason, the Lagrangian (or material) description will be utilized exclusively in this text.

The challenge here is to determine whether the object has undergone deformation in time, implying that the distance between any two arbitrary points changes as a function of time. To determine whether this is the case, first take the square of the difference (squaring obviates the possibility of negative numbers) between the line connecting points A and B in the displaced configuration and subtract the initial square of this length as follows:

$$\left|\overline{AB}\right|^2 - |AB|^2 = \Delta x_i \Delta x_i - \Delta a_i \Delta a_i \tag{9.35}$$

By utilizing Eq. (9.33), together with the chain rule of differentiation, and taking the limit as $\Delta x_k \to 0$ in the above equation, it can be seen that

$$\left|\overline{AB}\right|^2 - |AB|^2 = \lim_{\Delta x_m \to 0} \left[\Delta x_i(a_j) \Delta x_i(a_j) - \Delta u_i \Delta u_i\right] = \left[\frac{\partial x_i}{\partial a_j}\frac{\partial x_i}{\partial a_k} - \delta_{jk}\right] da_j da_k \tag{9.36}$$

where it should be noted from Fig. 9.6 that

$$x_i = u_i + a_i \tag{9.37}$$

Thus, substituting Eq. (9.37) into Eq. (9.36) results in:

$$\left|\overline{AB}\right|^2 - |AB|^2 = 2E_{ij}da_i da_j \tag{9.38}$$

where E_{ij} is the *Lagrangian strain tensor* (once again, after Joseph-Louis Lagrange), defined as follows:

$$E_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial a_j} + \frac{\partial u_j}{\partial a_i} + \frac{\partial u_k}{\partial a_i}\frac{\partial u_k}{\partial a_j}\right) \tag{9.39}$$

where the first-order terms are components of the *displacement gradient*. Note that it can be shown that the above is in fact a second-order tensor (see Problem 9.7), and that while the Lagrangian strain tensor is symmetric, it is necessarily nonlinear due to the quadratic term in the above equation. Nonetheless, it can be seen that if all components of this tensor are zero everywhere in a continuous body, the body is clearly moving (including both translation and rotation) as a rigid body. Thus, evaluating the Lagrangian strain tensor at every point in a continuous body may be seen to be a means of determining whether a continuum is deforming. This will become important when the process of constructing constitutive equations is developed in Chaps. 10–13.

While road ways can and often do undergo very large deformations, there are practical circumstances wherein all of the components of the displacement gradient are small compared to unity. Under this restriction, Eq. (9.39) simplifies to the following, called the *small strain tensor*:

$$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{9.40}$$

due to the fact that the deformed and undeformed coordinates become indistinguishable. Equations (9.40) are called t*he strain-displacement equations.* Note that they are linear equations and that the small strain tensor is symmetric.

Although some components of the displacement gradient may become large compared to unity in flexible pavements, it will be assumed in the remainder of this text that the small strain tensor is sufficiently accurate to model the response of road ways. This assumption is made for the sake of mathematical simplicity rather than accuracy, since including the second-order terms introduces considerable mathematical complexity into the problem.

The displacement field, u_i, and the strain field, ε_{ij}, are both output-dependent (or field) variables, and they are also termed *kinematic variables* because they are

associated with motions of the body of interest. It will be shown later that both of these variables must be known at all points and for all times within and on the boundary of a continuous body in order to predict the motions of the body.

9.4 Kinetics and Stress

Kinetics deals with the effects of mechanical loads acting on bodies. In order to construct a complete model capable of describing the motion of the continuum shown in Fig. 9.1, it is necessary to describe the kinetics associated with these motions. In order to do this, first consider an arbitrary point B within a continuous body. Suppose that an imaginary and arbitrarily oriented plane P is passed through the body at point B, as shown in Fig. 9.7.

Note that the orientation of the plane P is defined by *the unit outer normal vector* to this plane, $\vec{n}$, described by

$$\overrightarrow{n} = n_1\vec{e}_1 + n_2\vec{e}_2 + n_3\vec{e}_3 = n_i\vec{e}_i \tag{9.41}$$

9.4.1 The Traction Vector

Now suppose that a small area, ΔA, that is coplanar with plane P is chosen to surround the point B, as shown in Fig. 9.7. Suppose furthermore that there is a force, $\Delta\vec{F}$, acting on this area. Note that the force is not necessarily normal to plane P. *The traction vector*, $\vec{t}(\vec{n})$, is now defined as follows:

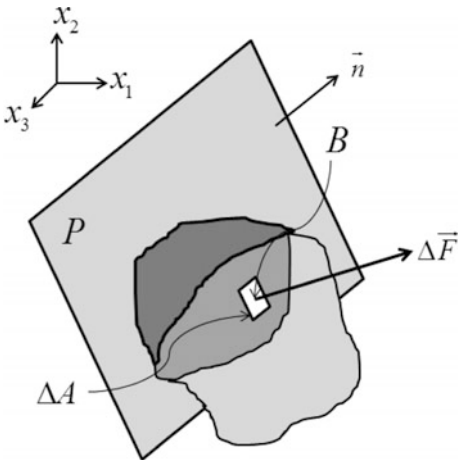


Fig. 9.7 Continuum with an imaginary plane P passed through it at an arbitrary point B

$$\vec{t}(\vec{n}) \equiv \lim_{\Delta A \to 0} \frac{\Delta \vec{F}}{\Delta A} \tag{9.42}$$

Note that the traction vector may be written in component form as follows:

$$\vec{t}(\vec{n}) = t_1 \vec{e}_1 + t_2 \vec{e}_2 + t_3 \vec{e}_3 = t_i \vec{e}_i \tag{9.43}$$

It can be seen that the traction is described in units of force per unit area.

Note also from Fig. 9.7 that a force is necessarily applied to the opposing face of the continuum on the plane P at the point B, and that the traction on this plane at point B is denoted by $\vec{t}(-\vec{n})$. It can be shown by deploying the conservation of linear momentum (see below) that:

$$\vec{t}(-\vec{n}) = -\vec{t}(\vec{n}) \tag{9.44}$$

In other words, the *tractions* acting on opposite faces at any point in a continuum are always of equal magnitude and opposite in sign. This theorem, called *Cauchy's Lemma*, was proven by Augustin-Louis Cauchy (1789–1857) (Cauchy 1822). It is a generalization of *Newton's Third Law* (after Isaac Newton (1642–1727)), which states that objects in contact exert equal and opposite *forces* on one another (Newton 1687). Cauchy's Lemma takes this law one step further, ensuring that not only are these forces equal and opposite, but the distribution of these forces, or tractions, are also of equal magnitude and opposite sign. Though this is a small theorem (thus termed a lemma), it is nevertheless an important theorem in continuum mechanics, as will be seen in the next section.

9.4.2 The Stress Tensor

Now suppose that the arbitrary point B in Fig. 9.7 is examined in further detail. As a thought experiment, suppose that three planes could be passed through the object, each normal to one of the coordinate axes, and the vertex of these three planes is just a tiny perpendicular distance removed from point B, denoted Δh. The result of this thought experiment is the depiction shown in Fig. 9.8.

It can be proven using Euclidean geometry (after Euclid (third century BCE)) that the areas of the three faces associated with the coordinate axes are related to the area of the exposed face, A, as follows:

$$A_i = n_i A \tag{9.45}$$

where n_i are the components of the unit outer normal to plane P, as described by Eq. (9.41). By employing conservation of linear momentum for the object shown in Fig. 9.8 and taking the limit as $\Delta h \to 0$, it can be shown that (together with Cauchy's Lemma!):

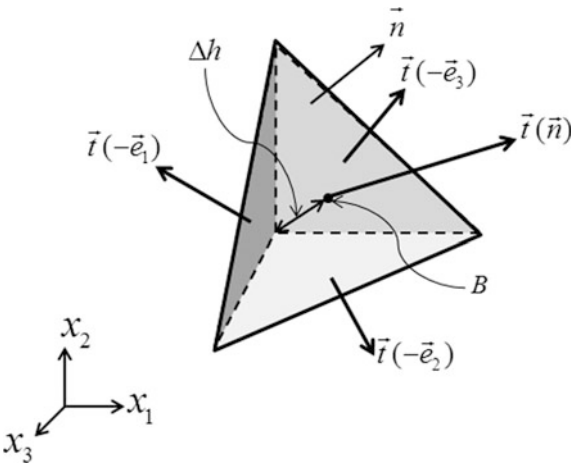


Fig. 9.8 Depiction of three additional planes passed near to point B and normal to the coordinate axes

$$\vec{t}(\vec{n}) = n_1\vec{t}(\vec{e}_1) + n_2\vec{t}(\vec{e}_2)n_3\vec{t}(\vec{e}_3) = n_i\vec{t}(\vec{e}_i) \tag{9.46}$$

Now define the components of the *Cauchy stress tensor* (Cauchy 1822), σ_{ij}, such that:

$$\vec{t}(\vec{e}_i) = \sigma_{ij}\vec{e}_j \tag{9.47}$$

The components of the stress tensor may, therefore be depicted on three orthonormal planes at the point B as shown in Fig. 9.9. Note that the stress is termed a tensor because it can in fact be shown that the stress transforms from one Cartesian coordinate to another via two multiplications by the direction cosine tensor, and this will be shown in the next section.

Note also that the components of the stress tensor with identical subscripts are normal to the planes on which they act, and for that reason they are called *normal stresses*. The remaining components of stress are termed *shear stresses*.

Substituting Eqs. (9.43) and (9.47) into Eq. (9.46) results in the following equation:

$$t_i = \sigma_{ji}n_j \tag{9.48}$$

The above all-important theorem was also first proven by Augustin-Louis Cauchy (Cauchy 1822), for which he is often called *The Father of Stress*. This is due to the fact that Eq. (9.48), called *Cauchy's formula*, proves that *in a continuum there are exactly nine components of the stress tensor that must be known at each and every point within that continuum in order to completely specify the state of kinetic loading within the continuum*. Thus, it can be seen that the stress tensor, a *kinetic variable*, will have to be included in the list of dependent output (or field) variables (along with the kinematic variables displacement and strain described above) necessary to describe the mechanics of a continuous body.

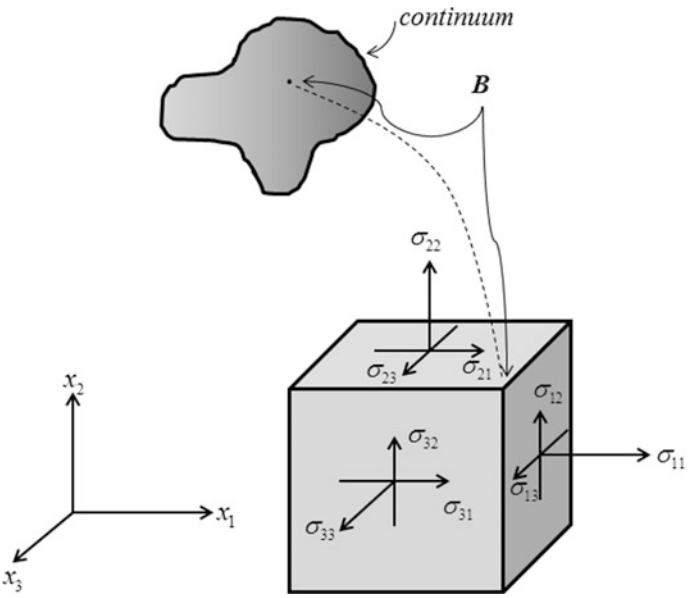


Fig. 9.9 Depiction of the components of the Cauchy stress tensor at point B in a continuum

It can be shown using conservation of angular momentum that the stress tensor is symmetric, that is, $\sigma_{ij} = \sigma_{ji}$, when there are no body moments in the continuum (see below). It will be assumed that there are no body moments in the materials modeled in this text, so that *there are six unique components of stress to be determined at every point in a continuum*.

There are quite a few important properties of the stress tensor that will become significant in the course of this text. Thus, before moving on to the important issues dealing with how to predict the stress at every point within a continuum, it is first propitious to discuss these properties in some detail.

9.4.3 Stress Transformations

Over the course of the succeeding two centuries since the definition of the Cauchy stress tensor, it has been found by experimental observation that the predicted stress in a continuous body can be used as a tool to predict failure of the body by fracture. Thus, *it may be possible to avoid failure by fracture if the stresses in a body can be predicted accurately*. Furthermore, the planes on which fracture occur seem to also be related to the stresses on these planes, so that it is desirable to determine the stress on the fracture planes. Since these planes are not usually known a priori, the coordinate system chosen by the structural analyst oftentimes does not align with the planes on which fracture occurs. It is, therefore, often necessary to transform the

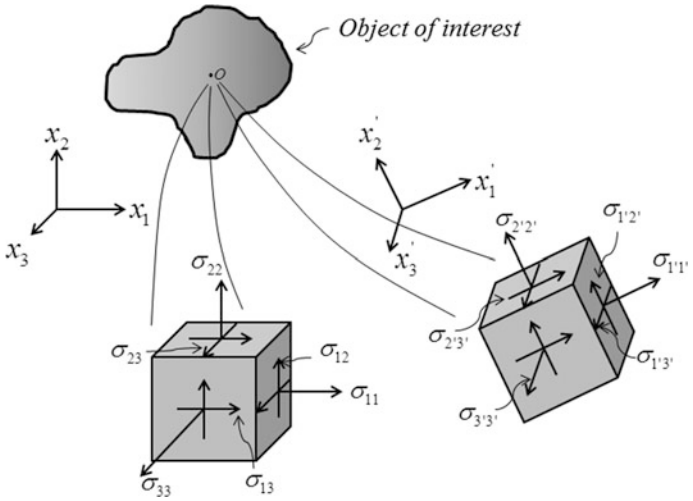


Fig. 9.10 Depiction of the components of the stress tensor in two different coordinate systems at point O in a continuous body

predicted stresses at a given point in the object from one Cartesian coordinate system to another that is rotated with respect to the one used to perform the predictions of the stresses in the body of interest, as shown in Fig. 9.10.

In order to construct a mathematical formula for carrying out a transformation of the stress at a given point from one Cartesian coordinate system to another, first recall Cauchy's formula, given by Eq. (9.48). Thus, in the rotated (primed) coordinate system, Cauchy's formula would be described as follows:

$$t_{p'} = \sigma_{q'p'} n_{q'} \tag{9.49}$$

where $n_{q'}$ are the components of the unit outer normal of the plane of interest in the primed coordinate system. In addition, since the traction, $\vec{t}$, and the unit outer normal, $\vec{n}$, are vectors, it follows that they both transform according to Eq. (9.11), that is,

$$t_{p'} = a_{ip'} t_i \tag{9.50}$$

and

$$n_j = a_{q'j} n_{q'} \tag{9.51}$$

Now, substitute Eq. (9.50) into Eq. (9.49), Eq. (9.48) into this result, Eq. (9.51) into that result and employ symmetry property (9.13), thus producing the following equation:

$$a_{ip'}\sigma_{ji}a_{q'j}n_{q'} = \sigma_{q'p'}n_{q'} \Rightarrow (a_{ip'}a_{q'j}\sigma_{ji} - \sigma_{q'p'})n_{q'} = 0 \tag{9.52}$$

Since the components of the unit outer normal vector need not be null, it follows from the above that

$$\sigma_{q'p'} = a_{q'j}a_{p'i}\sigma_{ji} \tag{9.53}$$

The above equation proves that stress is a second-order tensor. This is due to the fact that unlike the traction vector (a first-order tensor), which is composed of components that are not necessarily perpendicular and parallel to the plane on which it acts, the components of the stress tensor must by definition be perpendicular or parallel to the planes of interest, thereby requiring an additional multiplication by the direction cosine tensor. This is one of the primary properties of stress that make it a physically complicated variable.

9.4.4 Principal Stresses

One of the important properties of the stress tensor is that at each material point in a continuous body, there are planes on which there are no components of stress parallel to these planes (as pointed out by Cauchy in his 1822 memoir). These components of stress are termed *principal stresses*, and the planes on which they act are called *principal planes* (a terminology introduced by Cauchy himself).

It can be shown that the normal stress attains its maximum (and minimum) on one of these planes at each material point. Thus, it is not surprising that experimental evidence has supported the supposition that the normal stress components on these planes, as well as the orientations of these planes, may be related to both fracture and inelastic deformations.

It is, therefore, propitious to develop a mathematical technique for determining these stress components, as well as the orientations of the planes on which they act. In order to formulate this technique, consider a plane at a material point on which there are no components of shear stress. In this circumstance, it is also the case that the traction vector is perpendicular to this plane, that is,

$$\vec{t}\|\vec{n} \Rightarrow \vec{t} = \sigma|\vec{n}| \Rightarrow t_i = \sigma n_i \tag{9.54}$$

where σ is according to the definition above the value of the principal stress on that plane. On the other hand, Cauchy's formula, Eq. (9.48), applies at this selfsame material point. Thus, equating (9.48) and (9.54) and utilizing Eq. (9.8) results in the following:

$$(\sigma_{ji} - \sigma\delta_{ji})n_j = 0 \tag{9.55}$$

The above is a classical eigenvalue problem. There are three eigenvalues of σ, denoted σ_1, σ_2, and σ_3 that satisfy Eq. (9.55). Since the components of the stress tensor are all real and symmetric, it can be shown by using Cardano's method (after Gerolamo Cardano (1501–76), although the formula was apparently discovered due to the combined efforts of Scipione del Ferro (1465–1526) and Niccolò Tartaglia (1499/1500–1557)) that all three of these roots are real, thus resulting in three principal stresses at each point in a continuous body. The values of these three principal stresses can be determined by solving for the roots of the determinant of the parenthetical term in Eqs. (9.55). This cubic equation is given by

$$\left\|\sigma_{ij} - \sigma\delta_{ij}\right\| = 0 \Rightarrow \begin{Vmatrix} \sigma_{11} - \sigma & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} - \sigma & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} - \sigma \end{Vmatrix} = 0 \Rightarrow -\sigma^3 + I_1\sigma^2 - I_2\sigma + I_3 = 0 \tag{9.56}$$

where

$$I_1 \equiv \sigma_{11} + \sigma_{22} + \sigma_{33} = \sigma_{ii} \tag{9.57}$$

$$I_2 = \frac{1}{2}\left(I_1^2 - \sigma_{ij}\sigma_{ji}\right) \tag{9.58}$$

$$I_3 = \frac{1}{3}\left(3I_1I_2 - I_1^3 + \sigma_{ij}\sigma_{jk}\sigma_{ki}\right) \tag{9.59}$$

where I_1, I_2, and I_3 can be shown to be invariant with respect to coordinate rotation, and for this reason they are called *stress invariants*. For example, multiplying Eq. (9.53) by the Kronecker delta and employing Eq. (9.15) results in the following:

$$\delta_{p'q'}\sigma_{p'q'} = \delta_{p'q'}a_{p'i}a_{p'j}\sigma_{ij} = \delta_{ij}\sigma_{ij} \Rightarrow \sigma_{p'p'} = \sigma_{ii} \tag{9.60}$$

The above proves that I_1 is invariant with respect to coordinate rotation. A similar approach may be employed to prove that I_2 and I_3 are invariants (see Problem 9.13).

Thus, it can be seen that if the components of stress are known at a given point in a continuous body with respect to an arbitrary set of Cartesian coordinate axes, Eqs. (9.57)–(9.59) may be employed to calculate the stress invariants, and these may subsequently be substituted into Eq. (9.56), thereby providing a means of calculating the three principal stresses at that material point. Furthermore, once the principal stresses are calculated, the resulting values may be successively substituted back into Eq. (9.55) to determine the components of the unit outer normal vector for each principal stress. It can be shown that since the three principal stresses are real, the three unit outer normal vectors are mutually perpendicular, thus

implying that the principal planes at any material point in a continuous body lie on the faces of a cube passing through the material point.

A similar line of reasoning may be utilized to determine the maximum shear stresses at a material point, and the planes on which they occur with respect to the base Cartesian coordinate system.

9.4.5 Deviatoric Stresses

As mentioned above, Cardano's method may be utilized to solve the cubic Eq. (9.56) (see Problem 9.11). In order to employ this method, it is necessary to construct a transformation of the stress tensor. This is accomplished by first defining *the average hydrostatic stress* as follows:

$$\sigma_0 \equiv \frac{\sigma_{ii}}{3} = \frac{I_1}{3} \tag{9.61}$$

The above derives its name from the fact that if a block of material is subjected uniquely to hydrostatic pressure, p, the value of σ_0 will be equal to the negative of this pressure, which is defined to be positive in compression. The average hydrostatic pressure will turn out to be an important physical parameter in the course of developing constitutive models in Chap. 13. Note that due to its relation to the stress invariant I_1, it can be seen that σ_0 is also a stress invariant.

For reasons that will become clear below, a transformation of the stress tensor is now introduced. Define *the deviatoric stress tensor* as follows:

$$\sigma'_{ij} \equiv \sigma_{ij} - \sigma_0 \delta_{ij} \tag{9.62}$$

Note that the above implies that there must necessarily also exist three *deviatoric principal stresses*, defined by

$$\sigma'_i \equiv \sigma_i - \sigma_0 \tag{9.63}$$

The determination of these three deviatoric principal stresses proceeds by direct correspondence to the method for determining the principal stresses discussed in the previous section. Thus, Eq. (9.56) is transformed to the following:

$$\left\|\sigma'_{ij} - \sigma' \delta_{ij}\right\| = 0 \Rightarrow \begin{Vmatrix} \sigma'_{ij} - \sigma' & \sigma'_{12} & \sigma'_{13} \\ \sigma'_{21} & \sigma'_{22} - \sigma' & \sigma'_{23} \\ \sigma'_{31} & \sigma'_{32} & \sigma'_{33} - \sigma' \end{Vmatrix} = 0 \Rightarrow -(\sigma')^3 - J_2 \sigma' + J_3 = 0 \tag{9.64}$$

where due to the definition of the deviatoric stress, Eq. (9.62)

$$J_1 \equiv \sigma'_{ii} = (\sigma_{ij} - \sigma_0 \delta_{ij})\delta_{ij} = \sigma_{ii} - 3\sigma_0 = 0 \tag{9.65}$$

Furthermore, by similarity to Eqs. (9.58) and (9.59)

$$J_2 = \frac{1}{2}\sigma'_{ij}\sigma'_{ij} \tag{9.66}$$

and

$$J_3 = \frac{1}{3}\sigma'_{ij}\sigma'_{jk}\sigma'_{ki} \tag{9.67}$$

Note that due to the similarity between the above two equations and (9.58) and (9.59), respectively, J_2 and J_3 are also invariant with coordinate rotation (see Problem 9.13). They are, therefore, termed *deviatoric stress invariants*. It is also convenient to define an additional physically based quantity called *the octahedral shear stress, τ_0, as follows:*

$$\tau_0 \equiv \sqrt{\frac{2}{3}J_2} \tag{9.68}$$

Note that the above is by definition also an invariant.

Lest the reader forget, the primary objective at the moment is to calculate the principal stresses. However, a slight detour is warranted at this point. The deviatoric stress, σ'_{ij}, is a very important concept in the modeling of deformable fluids, wherein the deviatoric stress is a function of the rate of deformation. As such, deviatoric stress has great physical significance when dealing with fluids. Of course, the subject of this text is flexible pavements, and such structures are subjected to moisture (which is a fluid when in either the liquid or gaseous state) ubiquitously. Furthermore, the asphalt binder utilized in flexible pavements is a fluid in the unprocessed state. Thus, the concept of deviatoric stress will become very important from a physical standpoint later in this text.

For now, however, the deviatoric stress tensor may be used as a tool to calculate the principal stresses at a material point in a continuous body. To see how this may be accomplished, first assume that the state of stress has been predetermined at the material point of interest. Table 9.3 (Cardano's method) may then be used to calculate both the principal stresses and the components of the unit outer normals to the principal planes.

While the above technique provides a means of calculating the principal stresses and principal planes in terms of the stress components at any given material point, it is cumbersome, whereas the graphical technique to be described below is not only straightforward, it provides added physical insight.

Table 9.3 Procedure for calculating principal stresses

Step 1: Use the predetermined stress components σ_{ij} to calculate the deviatoric stress state:	
$\sigma'_{ij} \equiv \sigma_{ij} - \sigma_0 \delta_{ij}$	(9.62)
Step 2: Use σ'_{ij} to calculate the deviatoric stress invariants and τ_0:	
$J_2 = \frac{1}{2}\sigma'_{ij}\sigma'_{ij}$	(9.66)
$J_3 = \frac{1}{3}\sigma'_{ij}\sigma'_{jk}\sigma'_{ki}$	(9.67)
$\tau_0 \equiv \sqrt{\frac{2}{3}J_2}$	(9.68)
Step 3: Substitute J_2 and J_3 into Eq. (9.60). Define the following:	
$\varphi \equiv \cos^{-1}\left(\sqrt{2}\frac{J_3}{\tau_0^3}\right)$	(a)
Step 4: Calculate the three roots of Eq. (9.60) as follows:	
$\sigma'_1 = \sqrt{2}\tau_0 \cos\left(\frac{\varphi}{3}\right)$	(b)
$\sigma'_2 = \sqrt{2}\tau_0 \cos\left(\frac{\varphi}{3} + 120^\circ\right)$	(c)
$\sigma'_3 = \sqrt{2}\tau_0 \cos\left(\frac{\varphi}{3} + 240^\circ\right)$	(d)
Step 5: Use the deviatoric principal stresses to calculate the principal stresses:	
$\sigma_i \equiv \sigma'_i + \sigma_0$	(9.63)
Step 6: Substitute each of the principal stresses, σ_i, recursively to calculate the three components of the three unit outer normal vector for each of the principal stresses:	
$(\sigma_{ji} - \sigma\delta_{ji})n_j = 0$	(9.55)

9.4.6 *Stress Analysis Using Mohr's Circle*

In many practical circumstances either analysis or experimentation will reveal that the significant components of stress at a material point lie in a plane, called *plane stress*, as shown in Fig. 9.11a. If there is a component of normal stress perpendicular to the plane described in Fig. 9.11b, then the state of stress is called *generalized plane stress*, and for purposes of simplicity, we have defined that plane to be the $x_1 - x_2$ plane. For the case of generalized plane stress, there is a simplified technique for performing transformations of stress from one coordinate system to another that is rotated about the x_3-axis (Allen 2013). Note that generalized plane stress includes as a special case *plane strain* (in which the three components of strain on the plane normal to the x_3 coordinate direction are zero).

Now, suppose that we want to determine whether the state of stress in the primed coordinate system is equivalent to the stress in the unprimed coordinate system for the case of generalized plane stress (note that the x_3 coordinate direction is the same in both depictions), as shown in Fig. 9.12. In order to determine the relation between the stress components in the primed and unprimed coordinate systems, first pass a cutting plane through the stress block in the unprimed coordinate system rotated an angle θ about the x_3 axis, as shown in Fig. 9.13.

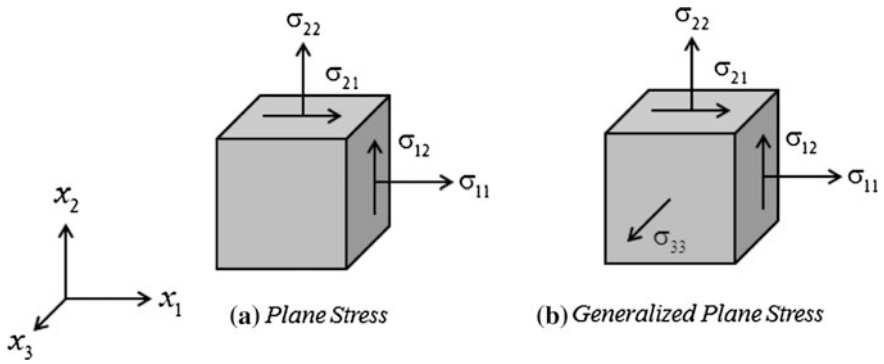


Fig. 9.11 Plane stress and generalized plane stress

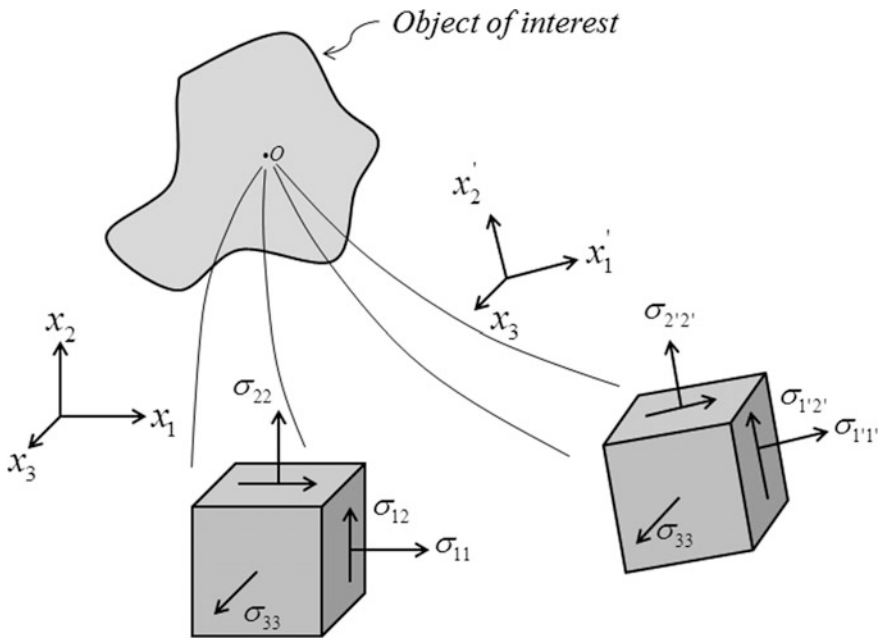


Fig. 9.12 Planar stress Transformations

Using the free body diagram resulting from this cutting plane, as shown in Fig. 9.13, we will now sum forces in the x_1' coordinate direction, recalling that the stresses must be multiplied by the area over which they act in order to produce forces. In order to do this, the area of the face normal to the x_1' axis is denoted as A. This results in the following (since $\sigma_{21} = \sigma_{12}$):

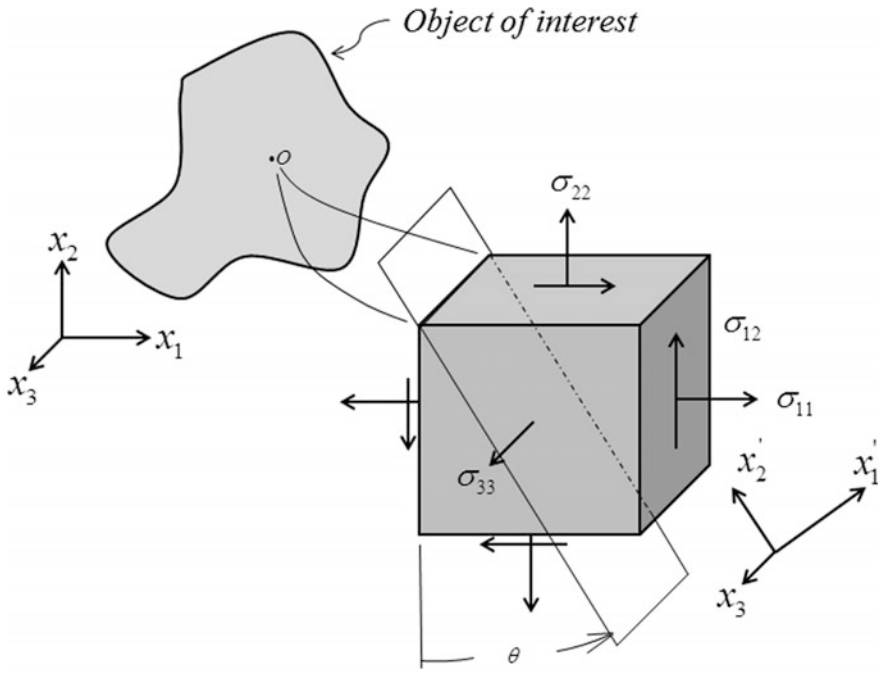


Fig. 9.13 Cutting plane in primed coordinate system passed through the stress block in the unprimed coordinate system

$$\sum F_{x_1'} = 0 = \sigma_{1'1'}A - \sigma_{11}\cos\theta(a\cos\theta) - \sigma_{22}\sin\theta(a\sin\theta) - \sigma_{12}\sin\theta(A\cos\theta) - \sigma_{12}\cos\theta(A\sin\theta) \tag{9.69}$$

Rearranging terms and dividing through by A results in

$$\sigma_{1'1'} = \sigma_{11}\cos^2\theta + \sigma_{22}\sin^2\theta + 2\sigma_{12}\sin\theta\cos\theta \tag{9.70}$$

Similarly, summing forces in the x_2' direction will result in

$$\sigma_{1'2'} = -(\sigma_{11} - \sigma_{22})\sin\theta\cos\theta + \sigma_{12}(\cos^2\theta - \sin^2\theta) \tag{9.71}$$

Equations (9.70) and (9.71) can be used to calculate the components of stress in the primed coordinate system as functions of the stress components in the unprimed coordinate system for any given value of θ (Fig. 9.14).

Unfortunately, Eqs. (9.70) and (9.71) are cumbersome to utilize for purposes of stress analysis. A more convenient graphical technique was introduced by Karl Culmann (Culmann 1866). This technique was later expanded and used to great effect for the prediction of failure by Otto Mohr (1835–1918), from whence comes

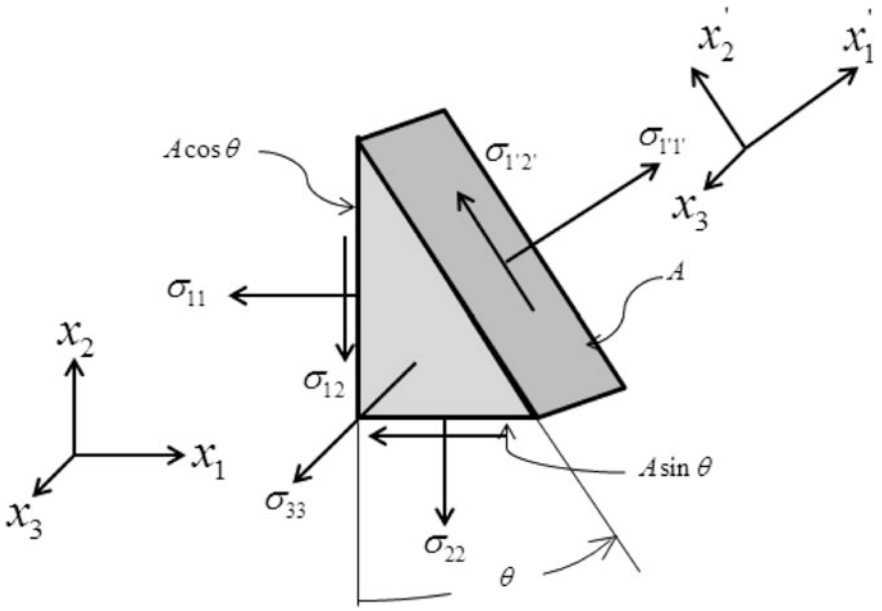


Fig. 9.14 Free body diagram of stress block with cutting plane

the name applied to this technique—*Mohr's circle* (Mohr 1868). In order to construct this graphical technique, first recall the following trigonometric identities:

$$\cos^2\theta = \frac{1+\cos 2\theta}{2} \quad \sin^2\theta = \frac{1-\cos 2\theta}{2} \quad 2\sin\theta\cos\theta = \sin 2\theta \tag{9.72}$$

Substituting the above identities into Eqs. (9.70) and (9.71) and rearranging will result in the following two equations:

$$\sigma_{1'1'} - \left(\frac{\sigma_{11}+\sigma_{22}}{2}\right) = \left(\frac{\sigma_{11}-\sigma_{22}}{2}\right)\cos 2\theta + \sigma_{12}\sin 2\theta \tag{9.73}$$

$$\sigma_{1'2'} = -\left(\frac{\sigma_{11}-\sigma_{22}}{2}\right)\sin 2\theta + \sigma_{12}\cos 2\theta \tag{9.74}$$

Squaring and adding the above two equations will result in the following:

$$\left[\sigma_{1'1'} - \left(\frac{\sigma_{11}+\sigma_{22}}{2}\right)\right]^2 + [\sigma_{1'2'} - 0]^2 = \left[\sqrt{\left(\frac{\sigma_{11}-\sigma_{22}}{2}\right)^2 + \sigma_{12}^2}\right]^2 \tag{9.75}$$

It can be seen that the above equation is equivalent to the general equation of a circle, shown in Fig. 9.15, and given by:

$$[x-a]^2+[y-b]^2=[c]^2 \tag{9.76}$$

where, by comparison of Eqs. (9.75) and (9.76), the following transformation of variables is apparent:

$$\begin{aligned} x &\rightarrow \sigma_{1'1'} \\ y &\rightarrow \sigma_{1'2'} \\ a &\rightarrow \frac{\sigma_{11}+\sigma_{22}}{2} \\ b &\rightarrow 0 \\ c &\rightarrow \sqrt{\left(\frac{\sigma_{11}-\sigma_{22}}{2}\right)^2+\sigma_{12}^2} \end{aligned} \tag{9.77}$$

Applying the above transformations to the circle shown in Fig. 9.15 results in the graphical representation shown in Fig. 9.16.

Note that all possible orientations for a plane stress state at a given material point in an object that are rotated about the z-axis lie on the circle shown in Fig. 9.16. Furthermore, because the trigonometric half angle formulas in Eq. (9.72) have been employed, angles shown in Fig. 9.16 are exactly twice as large as angles depicted in the real world.

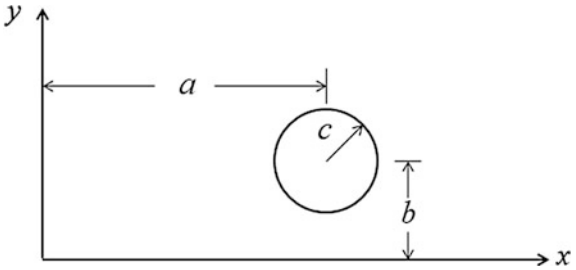


Fig. 9.15 General equation of a circle

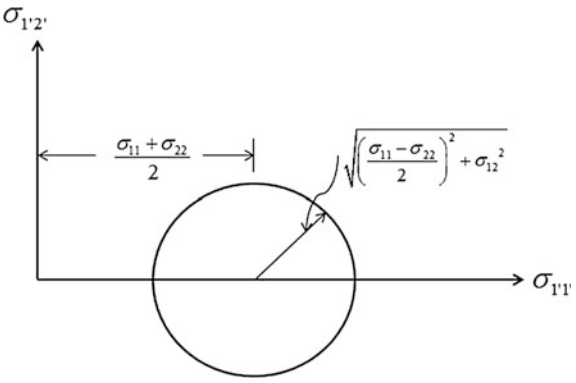


Fig. 9.16 Mohr's circle for plane stress

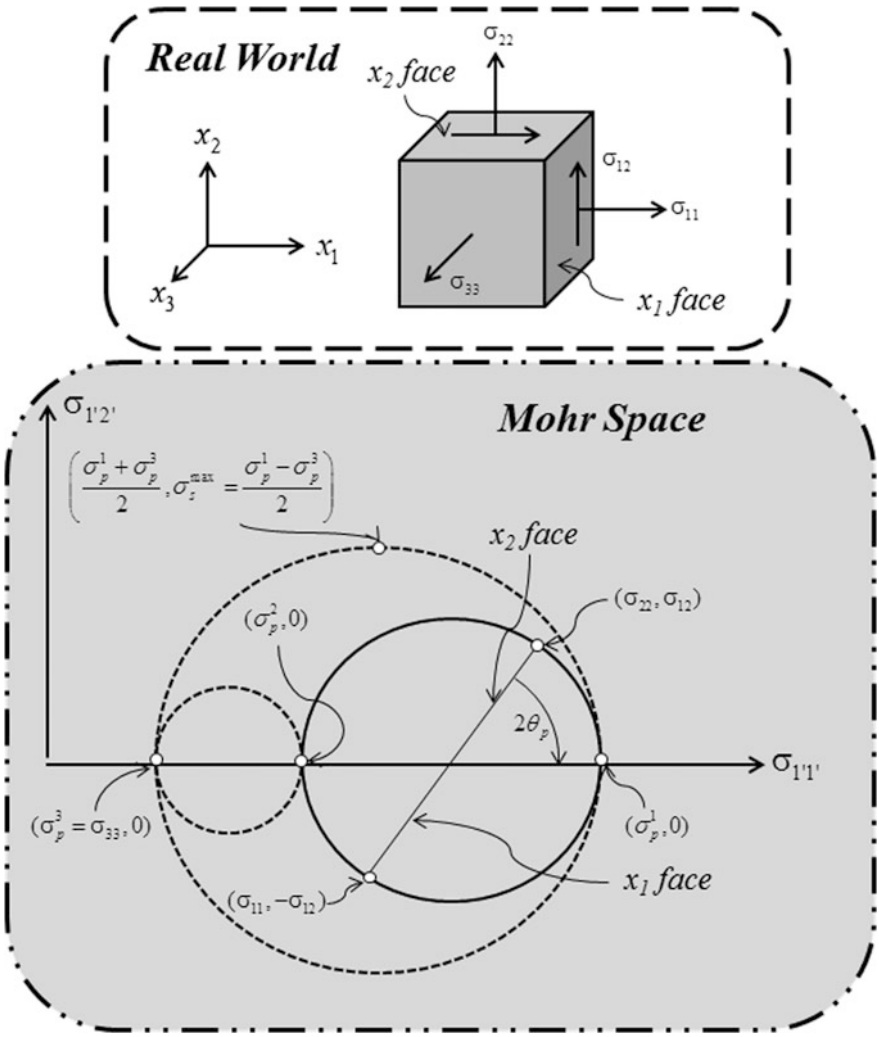


Fig. 9.17 Mohr's circles for generalized plane stress

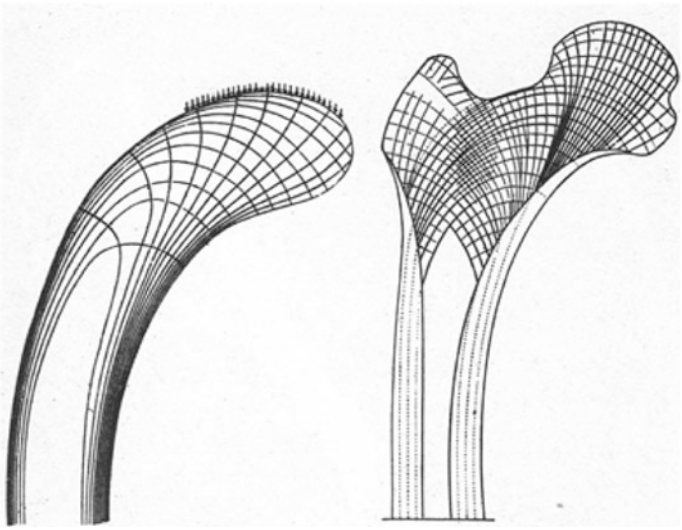


Fig. 9.18 Lines of principal planes (*Left*) in a curved crane compared to the trabecular alignment in a human proximal femur (*Right*) (Culmann 1866; Meyer 1867; Wolff 1870; Thompson 1917)

It can be shown that for the case of generalized plane stress, the depiction of Mohr's circle shown in Fig. 9.16 can be generalized to produce three circles, as shown in Fig. 9.17.

Mohr's circle was created nearly a century and a half ago, in a time when graphical techniques were the most powerful mathematical tools for many engineering applications. An early example (in 1867) of the physical power of this method is demonstrated in Fig. 9.18 (Culmann 1866; Meyer 1867; Wolff 1870; Thompson 1917). The figure shows the predicted lines of principal stresses in a crane analyzed by Culmann on the left. On the right is a depiction by Wolff of the experimentally observed trabecular alignment in the proximal femur of a human. This impressive demonstration of the importance of principal stresses is said to have occurred by coincidence when Professor Culmann visited the dissecting room of his colleague Hermann Meyer in Zurich. Upon seeing a section of bone, Culmann is said to have cried out, "That's my crane!" (Thompson 1917).

Today it is no longer necessary to utilize such antiquated graphical methods due to the power of computers. Nonetheless, there is much information to be extracted from a careful study of Mohr's circles that is physically significant with respect to the failure of solids. Several important deductions are given in Table 9.4.

Table 9.4 Properties of stress deduced from Mohr's circles

Important deductions from Mohr's circles
• *The three circles subtend the horizontal axis at points labeled* σ_p^1, σ_p^2 *and* σ_p^3
• *These normal stresses are called principal stresses because the shear stresses on the planes are zero*
• *A vertical diameter of the largest circle subtends the circle at the top and bottom, where the shear stress attains it maximum,* $\sigma_s^{\max}$
• *The angle (in Mohr space) between the* x_1 *(or* x_2*) face and a horizontal diameter of the circle is twice the angle from the* x_1 *(or* x_2*) face to a principal plane in the real world (denoted as* $2\theta_p$ *on Mohr's circle)*
• *Since principal planes and planes of maximum shear stress are always normal to one another in Mohr space, they are exactly 45 degrees apart in the real world*
• *For the case wherein* $\sigma_p^1 \geq \sigma_p^2 \geq \sigma_p^3$*, it follows that* $\sigma_s^{\max} = \left(\sigma_p^1 - \sigma_p^3\right)/2$ *is the radius of the largest circle*

Table 9.5 Procedure for drawing Mohr's circles

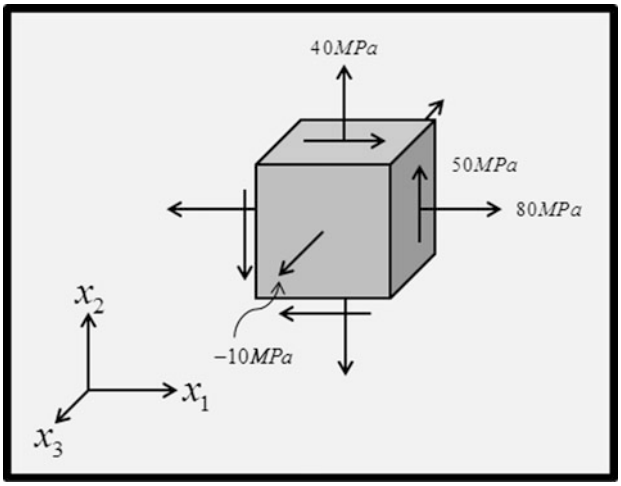
Procedure for Plotting Mohr's Circles
(1) *Plot the coordinates of the* x_1 *face* $(\sigma_{11}, \pm\sigma_{12})$ *and the* x_2 *face* $(\sigma_{22}, \mp\sigma_{12})$ *using the sign convention that shear stresses* ***that cause a clockwise couple are positive, and shear stresses that cause a counterclockwise couple are negative***
(2) *Draw a straight line connecting these two points*
(3) *The intersection of the line drawn in step 2 and the horizontal axis is the center of the circle. Use a compass to draw the circle that passes through the two points drawn in step 1*
(4) *Label the diameter of the circle that ends at the two points plotted in step 1 as the* x_1*-face and* x_2*-face as appropriate*
(5) *Label the three principal stresses, including the out-of-plane normal stress (even if it is zero!), and draw the two remaining circles, as depicted in Fig.* 9.17
(6) *Label the maximum shear stress at the top of the largest circle,* $\sigma_s^{\max}$*, and calculate the value of the angle* $2\theta_p$*, also depicted in Fig.* 9.17

Currently, there are several open-source software programs available on the Internet that may be utilized for the purpose of constructing Mohr's circles for a given state of stress (at no cost to the user). By importing the values of the stress components for a given material point in an object with respect to an arbitrary set of coordinate axes, Mohr's circles are produced by the software. In addition, there are a few software programs that will even construct diagrams of the material point, with the stress components labeled on the diagrams. It is therefore far less difficult to perform analysis of stress at a given material point than in former times, when computer algorithms were not available for this purpose.

For those who prefer to take the time to construct their own Mohr's circles graphically, Table 9.5 presents the procedure for doing so.

Example Problem 9.1

Given: Suppose that one or more of the models developed in this text have been used to predict the state of stress as a function of location in an object with known loads, geometry, and material properties, and this analysis has resulted in the stress state shown below at a point in the object identified by the designer as a critical point where failure of the object may be initiated.



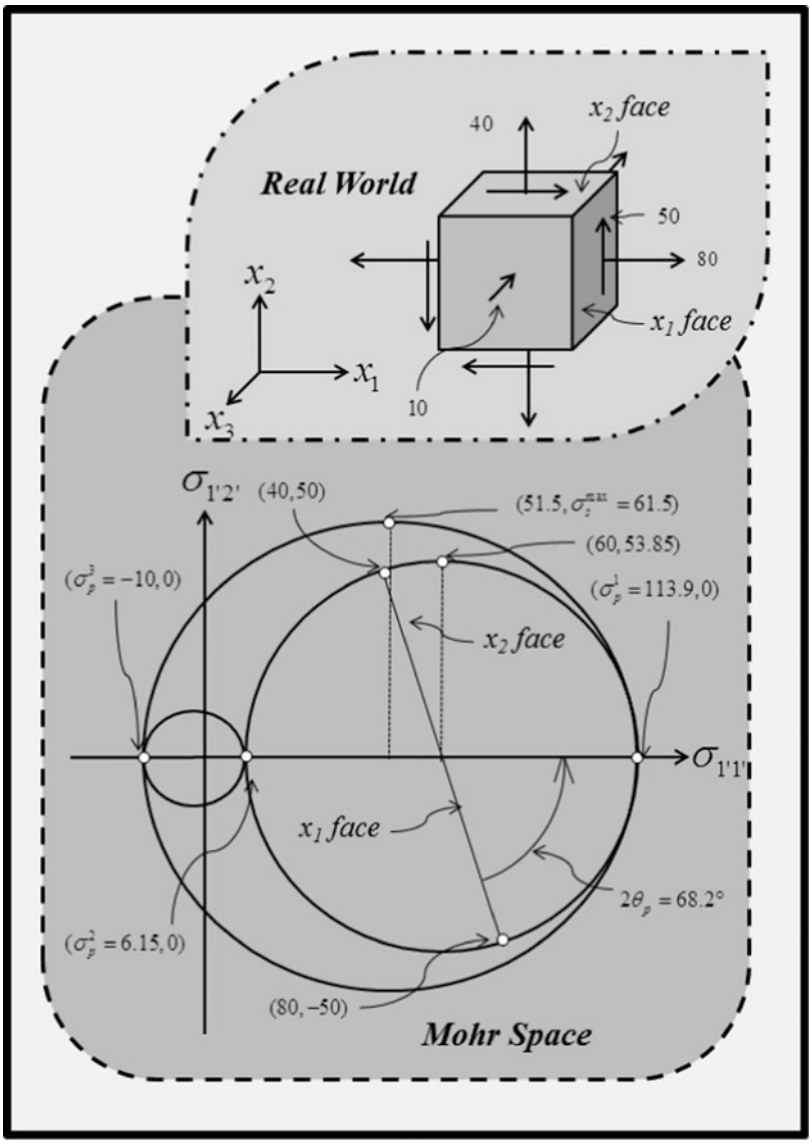
Required:

(a) Plot Mohr's circles and determine the principal stresses and the maximum shear stress

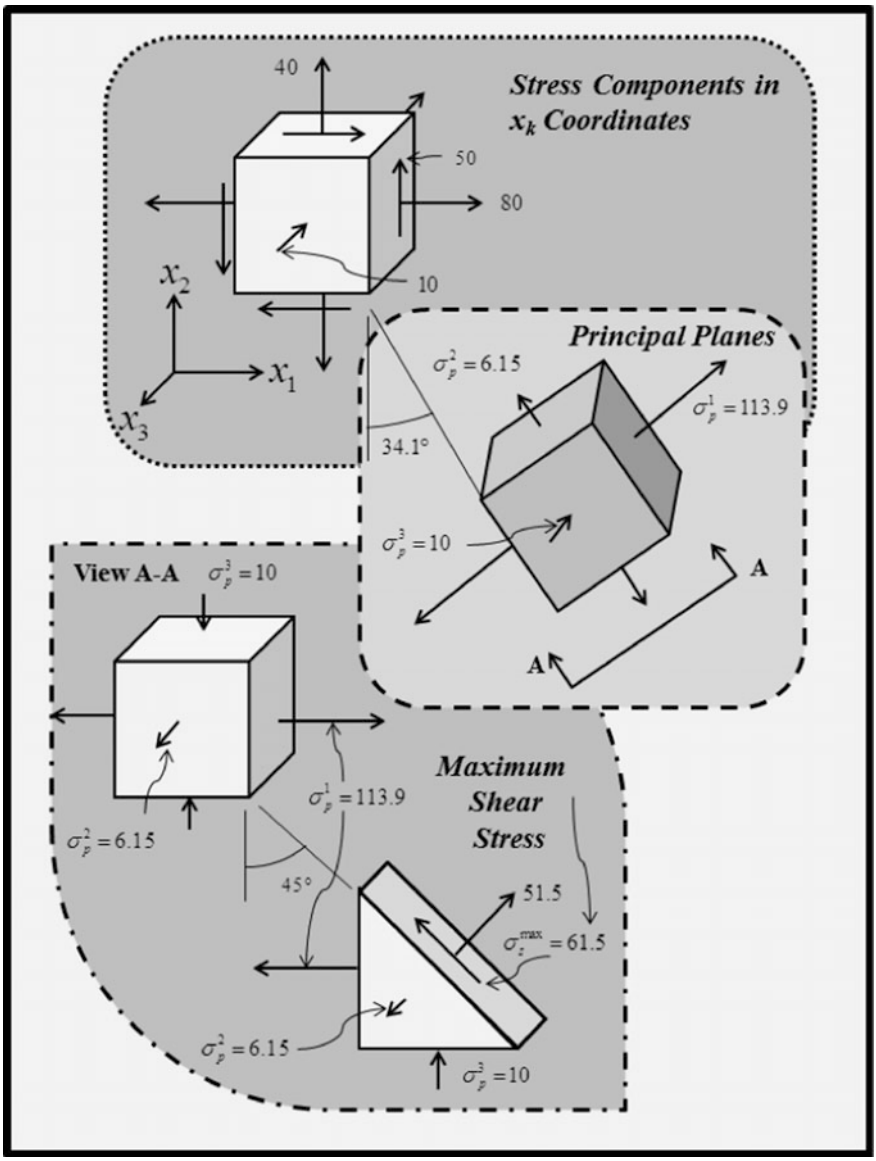
(b) Draw sketches of the principal planes and the plane of maximum shear stress, showing the stress components on these planes

Solution:

(a) Mohr's circles



(b) Sketches in the real world


Stress Components in
x_k Coordinates
40
50
80
10
x_2
x_1
x_3
Principal Planes
$\sigma_p^2 = 6.15$
$\sigma_p^1 = 113.9$
34.1°
$\sigma_p^3 = 10$
A
A
View A-A
$\sigma_p^3 = 10$
$\sigma_p^1 = 113.9$
$\sigma_p^2 = 6.15$
Maximum
Shear
Stress
45°
51.5
$\sigma_s^{max} = 61.5$
$\sigma_p^2 = 6.15$
$\sigma_p^3 = 10$

9.5 Conservation Laws

In the previous section, the concept of the stress tensor, $\sigma_{ij} = \sigma_{ij}(x_k, t)$, was introduced. Since this new variable is symmetric, when it is appended to the components of the deformation vector, $u_i = u_i(x_k, t)$, and the components of the strain tensor, $\varepsilon_{ij} = \varepsilon_{ij}(x_k, t)$, it becomes apparent that it is necessary to know fifteen field variables at every point in a continuum in order to characterize the mechanics of the object of interest.

It would seem self-evident that the successful prediction of these variables will necessarily require that a model be developed that contains no less than fifteen equations at every point in a continuous body. As will be shown later, this is indeed the case, and these equations are termed *field equations*. Where do these equations come from? The answer to this question is in reality contained within a history lesson. Our forefathers have developed a set of *universal laws* for this purpose. In modern terms, these are called *conservation laws,* and these are listed below:

(1) Conservation of Mass
(2) Conservation of Charge
(3) Conservation of Momentum
(4) Conservation of Energy
(5) Entropy Production Inequality

There are other laws of nature that exist, but they either do not apply to the mechanics of continua, or they are not universal in nature. An example is the profoundly important Darwin's law, which applies to evolution, but within the context of this text, it does not apply to mechanics (it is also not known whether Darwin's Law is a universal law, whereas scientists are relatively certain that those listed above do in fact apply throughout the universe). These conservation laws have been postulated in a variety of ways historically, including via the method of "fluxions" invented by Isaac Newton, the variational formulations proposed by the Bernoullis (principally John Bernoulli (1667–1748) and Daniel Bernoulli (1700–1782)) and expanded by Jean Le Rond d'Alembert (1717–1783) and Joseph-Louis Lagrange, not to mention via Noether's Theorem (Emmy Noether (1882–1935)).

Newton's method of fluxions, having been more or less replicated (and in some cases even expanded upon) simultaneously by Gottfried Leibniz, became one of the two competing approaches that are still in common practice today, the other being the variational approach. Following the method deployed by Cauchy (not to mention his monumental contributions to the development of the theory of fluxions into modern integro-differential calculus), we will develop the above list of conservation laws using Newton's (and Cauchy's more detailed) approach. Later, in Chap. 14, we will see that these may in fact be converted into the variational approach, which is preferable for the purpose of constructing computational solution procedures such as the finite element method.

9.5.1 Conservation of Mass

As it turns out, for the purpose of predicting the response of flexible pavements, the conservation of mass is normally trivially satisfied so long as there are no chemical reactions or ablation of mass. It will be assumed that this is the case in this text (except when dealing with the effects of moisture in Chaps. 10 and 14, as well as aging in Chap. 12). Thus, the conservation of mass is satisfied by assuming that the mass density of the road way, $\rho = \rho(x, t)$, does not change in time. This may be stated mathematically as follows:

$$\rho(x_k, t) = \rho(x_k, 0) \tag{9.78}$$

where $\rho(x_k, 0)$ is a necessary initial condition. It will be assumed that the initial mass density of the road way is known at all points in the road way, so that the conservation of mass is satisfied trivially herein. Thus, in this text (except when dealing with aging and/or moisture change), it will be assumed that the mass density is a *material property* that is known a priori. The case wherein moisture is a significant effect will be discussed in detail in Chaps. 10 and 14. In Chap. 14, it will be shown that the conservation of mass may be used to predict the temporal and spatial distribution of moisture in pavement, a necessary component of modeling the effects of moisture on pavement performance.

9.5.2 Conservation of Charge

The conservation of charge is also negligible assuming that there is no coupling between mechanical and electromagnetic phenomena. It will be assumed that this is the case for normal road way applications. Thus, for the applications considered in this text, the conservation of momentum, the conservation of energy, and the entropy production inequality are the only nontrivial conservation laws necessary for the purpose of modeling flexible pavements (except when considering the effects of moisture and/or aging, as noted in the previous section). These are constructed in the following sections.

9.5.3 Conservation of Momentum

Conservation of Momentum is a modern name for *Newton's Second Law of Motion* (after Isaac Newton (1642–1727)), which states the following (Newton 1687):

The change of motion is proportional to the motive force impressed: and is made in the direction of the right line in which that force is impressed.

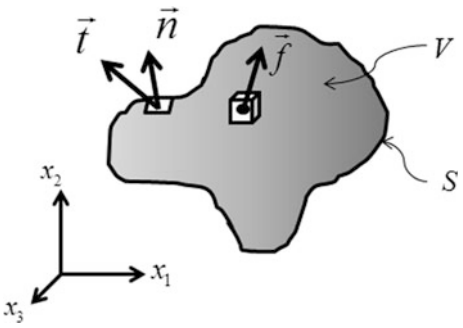


Fig. 9.19 Depiction of a continuous body with forces acting both on the interior and the exterior

Our modern interpretation of the above is as follows (Malvern 1969):

The time rate of change of the total momentum of a ***given set of particles*** *equals the vector sum of all the* ***external*** *forces acting on the particles of the set.*

Note that the above law must be applied for both linear and angular momentum. In order to construct a statement of conservation of momentum for a continuous body containing a fixed set of material points, consider once again Fig. 9.1. For convenience, the body is reconstructed in Fig. 9.19, and generic descriptors for the types of forces acting on the body are shown in the figure.

It is shown in the figure that the traction vector, $\vec{t}$, acts on an area, dS, on the boundary of the body, S. In addition, there are body forces per unit mass, $\vec{f}$, acting on a volume, dV, in the interior of the body, V. Conservation of linear momentum therefore requires that the following be true (Malvern 1969):

$$\int_S \vec{t} dS + \int_V \rho \vec{f} dV = \frac{d}{dt} \int_V \rho \vec{v} dV \tag{9.79}$$

where the derivative d/dt is the *material derivative* (Malvern 1969), and $\vec{v}$ is the velocity vector, given by

$$\vec{v} = \frac{d\vec{u}}{dt} \tag{9.80}$$

Dotting the above with the base vector $\vec{e}_i$ will thus result in

$$\int_S t_i dS + \int_V \rho f_i dV = \frac{d}{dt} \int_V \rho v_i dV \tag{9.81}$$

Substituting Cauchy's formula, Eq. (9.48), into the above and applying the Divergence Theorem, Eq. (9.31), as well as the conservation of mass, Eq. (9.78), and the Reynolds Transport Theorem (9.32) will result in the following:

$$\int_V \left[\sigma_{ji,j} + \rho f_i - \rho \frac{dv_i}{dt}\right] dV = 0 \tag{9.82}$$

The above is a global statement of the conservation of linear momentum in the body, $V + S$. Since the volume is arbitrary, the above also implies that

$$\sigma_{ji,j} + \rho f_i = \rho \frac{dv_i}{dt} \quad \forall \vec{x} \in V + S \tag{9.83}$$

The above is a local statement of conservation of linear momentum for all points in the body called *The Cauchy Equations of Motion* (Cauchy 1822). In cases wherein the loads are applied to the body sufficiently slowly that the rate of change of the velocity (acceleration), $d\vec{v}/dt$, is negligibly small, the above reduces to *the differential equations of equilibrium:*

$$\sigma_{ji,j} + \rho f_i = 0 \quad \forall \vec{x} \in V + S \tag{9.84}$$

For many commonly encountered applications in road ways, it can be assumed that the acceleration term in Eq. (9.83) may be neglected. This includes cyclic loadings applied to the road way by vehicular traffic, but does not include the effects of impact loading. Unless stated otherwise, equilibrium Eq. (9.84) will be utilized for the purpose of modeling flexible pavements in the remainder of this text.

Note that the above represents three equations in the nine unknown components of the stress tensor. As such, these equations are *statically indeterminate*, that is, they are insufficient to calculate the stress in a general body. In the few special cases wherein they are sufficient for this purpose, the problem is called *statically determinate*.

Without proving it herein, it can be said that conservation of angular momentum is satisfied at every point in the body *iff* (Malvern 1969) (see Problem 9.10):

$$\sigma_{ij} = \sigma_{ji} \quad \forall \vec{x} \in V + S \tag{9.85}$$

Thus, satisfying the conservation of momentum has supplied six equations that must be satisfied at every point in a continuous body for the purpose of determining the nine components of the stress tensor. Equivalently, one can view the stress tensor as symmetric due to Eq. (9.85), and in this case, there are three equilibrium Eqs. (9.84) and six unknown components of the stress tensor. This is the viewpoint that will be taken in the remainder of this text.

9.5.4 Conservation of Energy

Within the context of flexible pavements, the conservation of energy is a very important concept that is essential to a complete understanding of pavement

performance. This is due to the fact that pavement undergoes numerous forms of energy dissipation throughout the life of the road way. This can be due to the flux of heat, the creation of cracks, chemical changes, and material dissipation. The laws of thermodynamics must necessarily be utilized in order to understand and predict the effects of such phenomena.

The Conservation of Energy is a modern name used for *The First Law of Thermodynamics*, which states the following (Clausius 1850):

In a thermodynamic process involving a closed system, the increment in the internal energy is equal to the difference between the heat accumulated by the system and the work done by it.

Within the context of the above law, the modern interpretation for a collection of material points in a continuous body is as follows (Malvern 1969):

The rate of change of the internal work is balanced by the sum of the heat added to the body and work done on the body.

A mathematical statement of the above is given by:

$$\frac{dU}{dt} = \frac{dQ}{dt} + \frac{dW}{dt} \tag{9.86}$$

where U is the internal energy of the body, Q is the heat contained within the body, and W is the work performed on the body. In order to apply the above to an arbitrary continuous object, consider the object consisting of a fixed set of material points, as shown in Fig. 9.20, where $\vec{q}$ is the heat flux per unit area, and r is the internal heat source per unit mass.

The application of Eq. (9.86) results in the following equation for the object, $V + S$:

$$\frac{d}{dt}\int_V \rho u dV = -\int_S q_i n_i dS + \int_V \rho r dV + \int_S t_i v_i dS + \int_V \rho f_i v_i dV \tag{9.87}$$

where u is the internal energy per unit mass. The Divergence Theorem (9.31), the Cauchy equations of motion (9.83), the conservation of angular momentum, and the

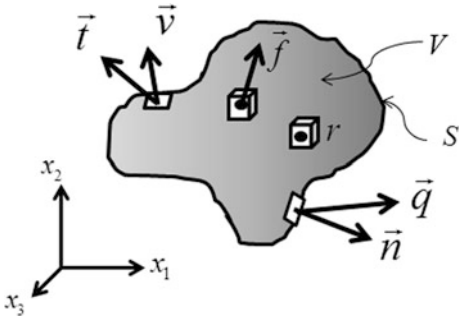


Fig. 9.20 Depiction of a continuous body subjected to various forms of energy

Reynolds transport theorem (9.32) may be employed to convert the above to the following form:

$$\int_V \left(\rho \frac{du}{dt} + q_{i,i} - \rho r - \sigma_{ij} \frac{d\varepsilon_{ij}}{dt}\right) dV = 0 \tag{9.88}$$

The above is a global statement of the conservation of energy in the body, $V + S$. Since the volume is arbitrary, the above also implies that

$$\rho \frac{du}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} - q_{i,i} + \rho r \quad \forall \vec{x} \in V + S \tag{9.89}$$

The above is a local statement of conservation of energy for all points in the body. Note that while it does indeed supply one additional equation, several new unknowns have been introduced, so that satisfaction of the conservation of energy necessarily adds significant complication to the problem of modeling flexible pavements.

9.5.5 The Entropy Production Inequality

The Entropy Production Inequality is a modern name used for *The Second Law of Thermodynamics* (also called *The Clausius-Duhem Inequality,* after Rudolph Clausius (1822–88) and Pierre Duhem (1861–1916)), which states the following (attributed to Sadi Carnot (1796–1832) (Clausius (1850):

The entropy in an isolated system can never decrease.

Within the context of the above law, the modern interpretation for a collection of material points in a continuous body is as follows:

The rate of change of the entropy within an isolated body is never negative.

A mathematical statement of the above is given by:

$$\frac{d\bar{S}}{dt} \geq 0 \tag{9.90}$$

where $\bar{S}$ is the entropy contained within the body. It should be noted that the total rate of change of entropy in the body is mitigated by the rate of entropy input to the body, where the entropy input to the body is changed by the addition of heat. Once again referring to Fig. 9.20, the application of Eq. (9.90) results in the following equation for the body, $V + S$:

$$\frac{d}{dt}\int_V \rho s dV + \int_S \frac{q_i n_i}{T} dS - \int_V \frac{\rho r}{T} dV \geq 0 \tag{9.91}$$

where s is the entropy per unit mass and T is the temperature, defined by *The Zeroth Law of Thermodynamics* (attributed to James Clerk Maxwell (1831–79)) (Maxwell 2001):

If two systems are each in thermal equilibrium with a third system, they are also in thermal equilibrium with each other.

Equation (9.91) is a global statement of the entropy production inequality. By employing the divergence theorem (9.31) and the Reynolds transport theorem, it may be restated as follows:

$$\rho \frac{ds}{dt} + \frac{q_{i,i}}{T} - \frac{\rho r}{T} \geq 0 \quad \forall \vec{x} \in V + S \tag{9.92}$$

The above is a local statement of the entropy production inequality for all points in the object. Note that while it does indeed supply one additional constraint, it also introduces two new variables (the entropy and the temperature), so that, as in the case of conservation of energy, the entropy production inequality necessarily adds significant complication to the problem of modeling flexible pavements.

The equations and inequality derived in Sect. 9.5 comprise all of the information available from universal laws for the purpose of modeling the response of flexible pavements.

9.6 Summary

In this chapter, notation has been introduced along with some mathematics that is necessary for the purpose of modeling flexible pavements. The independent variables $\vec{x}$ and t have been introduced, and field variables have been introduced for the purpose of modeling the performance of flexible pavements. These variables which depend on the independent variables $\vec{x}$ and t within the road way are shown in Table 9.6:

Table 9.6 List of field variables required to predict performance of flexible pavements

Field variable name	Variable symbol	Number of components
Displacement vector	u_i	3
Strain tensor	ε_{ij}	6
Stress tensor	σ_{ij}	6
Body force per unit mass	f_i	3
Internal energy per unit mass	u	1
Heat flux vector	q_i	3
Heat source per unit mass	r	1
Entropy per unit mass	s	1
Temperature	T	1
Total		**25**

Table 9.7 List of field equations necessary to model flexible pavements

Equation name	Formula	Equation number	No. of equations
Strain-displacement Equations	$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	(1.40)	6
Equations of motion	$\sigma_{ji,j} + \rho f_i = \rho \frac{dv_i}{dt} \quad \forall \vec{x} \in V + S$	(1.83)	3
Conservation of energy	$\rho \frac{du}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} - q_{i,i} + \rho r$	(1.89)	1
Entropy production inequality	$\rho \frac{ds}{dt} + \frac{q_{i,i}}{T} - \frac{\rho r}{T} \geq 0$	(1.92)	0 (inequality)
Total			**10 + 1 inequality**

Note that while certain additional variables have been introduced, such as the traction vector, $\vec{t}$, and the velocity vector, $\vec{v}$, they are not included in the above list because they can be obtained from the variables listed via the use of simple ancillary equations ((9.48) and (9.80), respectively). In addition, the mass density, ρ, is treated as a known material property, thereby satisfying the conservation of mass Eq. (9.78). Thus, in this chapter, twenty-five field variables have been introduced that are necessary to model the performance of flexible pavements.

This chapter has also introduced a number of field equations that must be satisfied in order to predict the performance of flexible pavements. These are listed in Table 9.7.

Note that at this point, the model consists of 10 equations (plus one inequality) in 25 unknowns. Thus, the introduction of the conservation laws has resulted in the problem being significantly underposed (implying that there are not enough equations to model the unknowns). Therefore, it is apparent that additional information will be required to construct a complete model for flexible pavements. Such equations are called *constitutive equations*. The construction of these equations will be the subject of Chaps. 10–13.

9.7 Problems

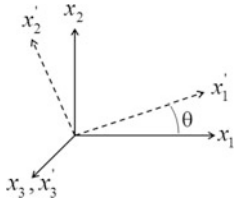
Problem 9.1
Given: The following matrix:

$$[b_{ij}] \equiv \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 5 \\ 3 & 1 & 9 \end{bmatrix}$$

Required:

(a) Evaluate $\delta_{ij}b_{ij}$ and b_{ii}
(b) Evaluate $b_{ij}b_{ij}$ and $b_{ij}b_{ji}$
(c) Evaluate $b_{ij}b_{jk}b_{ki}$
(d) Evaluate $c_i = e_{ijk}b_{jk}$ and $c_i = e_{jki}b_{kj}$
(e) Evaluate $d_{ij} = b_{ik}b_{kj}$ and $d_{ij} = b_{jk}b_{ki}$

Problem 9.2
Given: Two coordinate systems are constructed relative to one another as shown below.



Required:

(a) Evaluate the components of $a_{ij'}$ and $a_{ji'}$ in terms of θ
(b) Evaluate $a_{ip'}a_{jq'}\delta_{p'q'}$ using the results in a)

Problem 9.3
Given: The principles of superposition and linearity must be satisfied in order for a function or functional to be linear.
Required: Determine which of the following are linear.

(a) $$f_1(t) = a_0 + a_1x_1 + a_2(x_1)^2 + a_3x_1t$$

(b) $$f_2(x) = x + \frac{dx}{dt} + x\frac{d^2x}{dt^2}$$

(c) $$f_3\{\varepsilon(\tau)\} = \int_0^t E(t-\tau)\frac{\partial\varepsilon(\tau)}{\partial\tau}d\tau$$

(d) $$f_4\{g(t)\} = \int_0^\infty e^{-st}g(t)dt$$

Problem 9.4
Given: The definition of the Laplace transform, Eq. 9.17.

Required: Prove the following lines of Table 9.2.

(a) Line 1
(b) Line 2
(c) Line 3
(d) Line 4
(e) Line 7

Problem 9.5
Given: A kinetic variable, *R,* is a functional of a kinematic variable, *I,* as follows:

$$R(t) = \int_0^t E_1 e^{(t-\tau)} \frac{\partial I(\tau)}{\partial \tau} d\tau$$

Required: Evaluate *R(t)* for each of the following kinematic input histories (assuming I_1, t_1, and ω are constants):

(a) $$I(\tau) = I_1 H(\tau)$$

(b) $$I(\tau) = I_1 H(\tau - t_1)$$

(c) $$I(\tau) = I_1 \tau H(\tau)$$

(d) $$I(\tau) = I_1 \sin \omega\tau H(\tau)$$

Problem 9.6
Given: The fundamental theorem of calculus, Eq. (9.30).
Required: Prove the divergence theorem, Eq. (9.31).

Problem 9.7
Given: The Lagrangian strain tensor, Eq. (9.39).
Required:

(a) Prove that it is a second-order tensor.
(b) Prove that it is nonlinear

Problem 9.8
Given: The rotation tensor is defined as follows:

$$\omega_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right)$$

Required:

(a) Prove that

$$\left|\overline{AB}\right|^2 - |AB|^2 = 2\left[\varepsilon_{ij} + \frac{1}{2}(\varepsilon_{ki} - \omega_{ki})(\varepsilon_{kj} - \omega_{kj})\right] da_i da_j$$

(b) Prove that

$$\varepsilon_{ij} = 0 \rightarrow \left|\overline{AB}\right|^2 - |AB|^2 = 0$$

Problem 9.9
Given: Newton's First Law (summation of forces), Cauchy's Lemma, and Eq. (9.47).
Required:

(a) Prove Eq. (9.44).
(b) Prove Eq. (9.46).
(c) Prove Eq. (9.48)

Problem 9.10
Given: Newton's First Law (summation of moments)
Required: Prove that the stress tensor is symmetric in the absence of body moments

Problem 9.11
Given: The deviatoric stress tensor is defined by Eq. (9.62).
Required: Prove that deviatoric stress is a second-order tensor.

Problem 9.12
Given: The stress state at a given material point with respect to an arbitrary set of Cartesian coordinates (x_1, x_2, x_3) is found to be as follows (measured in MPa):

$$\sigma_{ij} = \begin{bmatrix} 0.10 & 0.20 & 0 \\ 0.20 & 0.30 & 0.10 \\ 0 & 0.10 & -0.20 \end{bmatrix}$$

Required:

(a) Find $\vec{t}(\vec{n})$ on a plane that does not intersect the x_3 axis and intersects the x_1 axis at the coordinate location $x_1 = 1.0$ and the x_2 coordinate axis at $x_2 = -1.0$.
(b) Rotate (normal to the x_3 axis) the coordinate axes to this new plane and find the components of the stress tensor with respect to the new axes

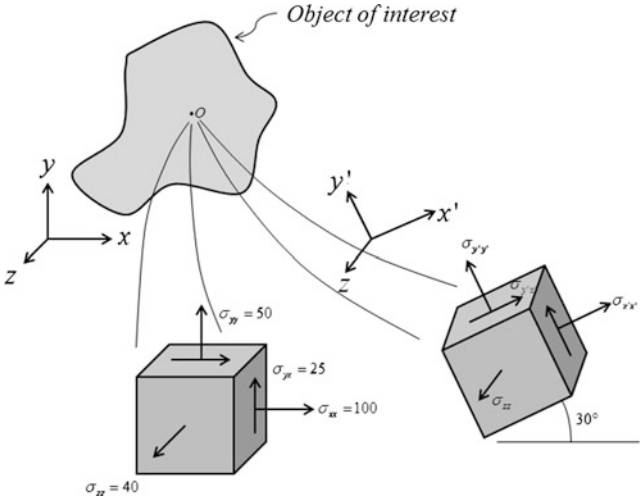
Problem 9.13
Given: The stress state described in Problem 9.12.
Required: Use Cardano's equation to find the following:

(a) Find the stress invariants, I_1, I_2, I_3, J_2 and J_3.
(b) Find the principal stresses, σ_1, σ_2, and σ_3.
(c) Find the unit normal vectors to the planes of the principal stresses, $\vec{n}_1, \vec{n}_2$, and $\vec{n}_3$, respectively, with respect to the Cartesian coordinate system (x_1, x_2, x_3) described in Problem 9.12.

Problem 9.14
Given: Eqs. (9.58), (9.59), (9.66), and (9.67).
Required:

a) Prove that I_2 and I_3 are invariants.
b) Prove that J_2 and J_3 are invariants.

Problem 9.15
Given: An engineer has performed a stress analysis of the object below using the unprimed coordinate axes shown. That analysis has produced the state of stress shown in the diagram at Point O. All stress components are in MPa.



Required:

(a) Calculate the stress components $\sigma_{x'x'}, \sigma_{x'y'}, \sigma_{y'y'}$ in the primed coordinate system assuming that the primed coordinate system is rotated 30° counterclockwise about the z-axis as shown in the diagram.
(b)

(1) Draw the three Mohr's circles at point O on a piece of graph paper using a compass.

(2) Label the points $(\sigma_{xx}, \sigma_{xy})$, $(\sigma_{yy}, \sigma_{xy})$, and $(\sigma_{zz}, 0)$ on the graph.
(3) Label the x- and y-planes on the circle.
(4) Label the three principal stresses on the graph and calculate their values.
(5) Label the maximum shear stress on the graph and calculate its value.
(6) Label the face rotated thirty degrees counterclockwise from the x-face and calculate the state of stress on this face.

(c)

(1) Draw sketches showing the principal planes.
(2) Draw a sketch showing the plane of maximum shear stress.

Problem 9.16
Given: The Lagrangian form of conservation of mass given by Eq. (9.78).
Required: Derive an Eulerian form of conservation of mass.

Problem 9.17
Given: Eq. (9.79).
Required: Convert to Eq. (9.82).

References

Allen, D. (2013). *Introduction to the mechanics of deformable solids—bars and beams*. Springer.
Cauchy, A. (1822). Lecture presented to the French Academy of Sciences, Paris.
Clausius, R. (1850). Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen. *Annalen der Physik und Chemie* (Poggendorff, Leipzig), 155(3), 368–394.
Culmann, R. (1866). Die Graphische Statik. Zurich.
Gilat, A. (2008). *MATLAB: An introduction with applications*. Wiley.
Malvern, L. (1969). *Introduction to the mechanics of a continuous medium*. Prentice-Hall.
Maugin, G. (2014). *Continuum mechanics through the eighteenth and nineteenth centuries*. Springer.
Maxwell, J. (2001). *Theory of heat* (9th ed.). Courier Dover Publications.
Meyer, H. (1867). Die arkitectur der spongiosa, Arch. f. anat. und phys (Vol. XLVII).
Mohr, O. (1868). Z architek u ing ver. Hannover.
Newton, I. (1687). *Principia—vol. 1—the motion of bodies*. University of California press.
Thompson, D. W. (1917). *Growth and form*. Cambridge University Press.
Wolff, J. (1870). Die innere architektur der knochen, Arch. f. anat. und phys. (Vol. L).

Chapter 10
One-Dimensional Constitutive Theory

10.1 Introduction

In Chap. 9, the mechanics of continuous solids was reviewed, and it was shown that the conservation laws are insufficient to model the mechanical response of solids. This implies that there must be additional information included in any cogent model for predicting the performance of flexible pavements. Because the conservation laws are independent of the material considered, it stands to reason that the additional information required should somehow be capable of distinguishing between the material constitution in different road ways. For that reason, the additional equations that are necessary to complete the model are called *constitutive equations*.

The development of constitutive theory may be said to have begun with *Da Vinci's experiment,* in which Leonardo Da Vinci (1452–1519) described the first material experiment known to us. His description, written around the year 1500, reads as follows (Da Vinci 1972; Parsons 1939):

> The object of this test is to find the load an iron wire can carry. Attach an iron wire 2 bracchia (about 1.3 m) long to something that will firmly support it, then attach a basket or any similar container to the wire and feed into the basket some fine sand through a small hole placed at the end of a hopper. A spring is fixed so that it will close the hole as soon as the wire breaks. The basket is not upset while falling, since it falls through a very short distance. The weight of sand and the location of the fracture of the wire are to be recorded. The test is repeated several times to check results. Then a wire of one-half the previous length is tested and the additional weight it carries is recorded, then a wire of one-fourth length is tested and so forth, noting each time the ultimate strength and the location of the fracture.

This experiment is still used today to characterize the mechanical response of many materials. Still later, around 1660, Robert Hooke (1635–1703) performed experiments on a variety of springs, as shown in Fig. 10.1 (Jardine 2005).

As a result of these experiments, he postulated *Hooke's Law: ut tensio sic vis* (as the load, so the displacement), as given by the following equation:

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_10

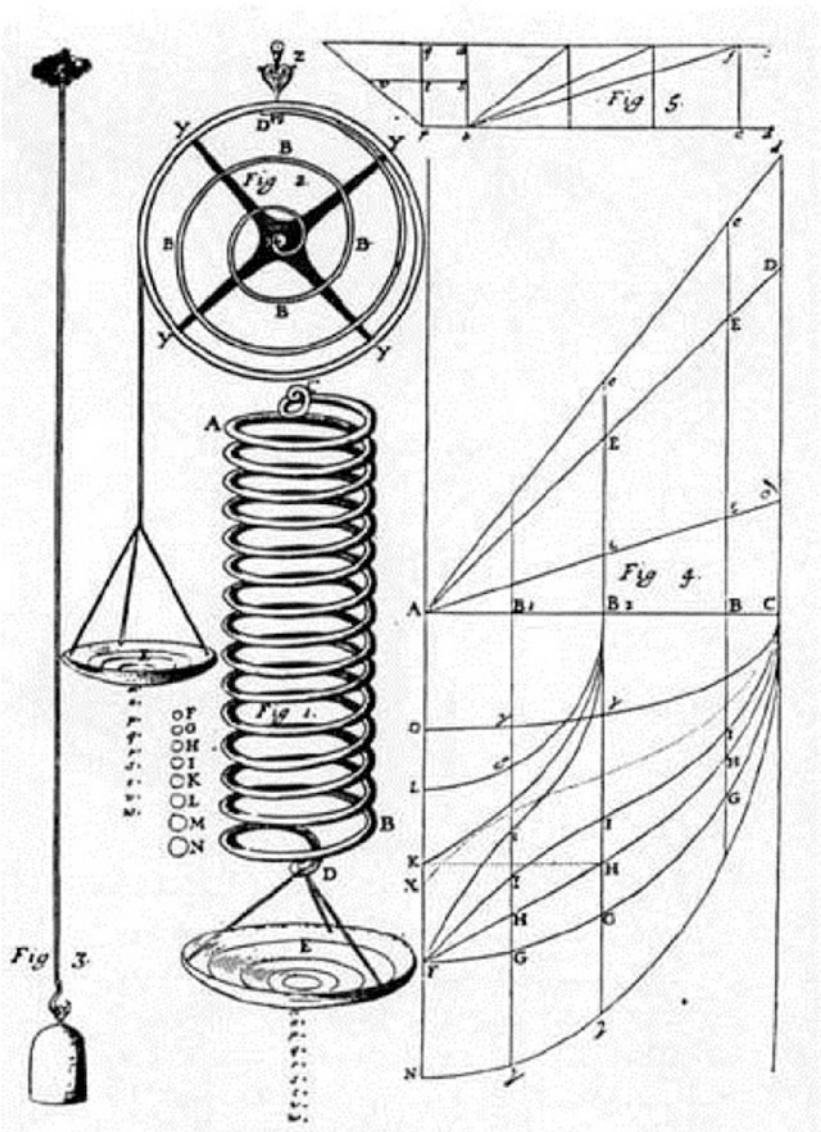


PLATE TO HOOKE'S LECTURE 'OF SPRING' 1678.

FIG. 1. Wire helical spring stretched to points *s*, *p*, *q*, *r*, *s*, *t*, *v*, *w*, by weights F, G, H, I, K, L, M, N.
FIG. 2. Watch spring similarly stretched by weights put in pan.
FIG. 3. The 'Springing of a string of Brass Wire 36 ft. long'.
FIG. 4. Diagram of velocities of springs.
FIG. 5. Diagram of law of ascent and descent of heavy bodies.

Fig. 10.1 Depiction of Robert Hooke's experiments

$$F = k\delta \tag{10.1}$$

where F is the magnitude of the force applied to the spring, δ is the measured displacement of the spring in the direction of the applied loading, and k is called *the spring constant*. This rather pithy Latin statement may be said to be the beginning of the science of constitutive theory, and just as in the case with many developments in science, the field of constitutive theory has taken a rather circuitous path right up to recent times.

Materials often have a way of behaving badly, that is, their constitutive behavior can be simple, or it can be quite complex. As a result, the mathematical complexity of the resulting constitutive model for a given class of materials can in turn be straightforward, or it can be quite challenging. Therefore, since the remaining field equations (described in Chap. 9) are identical for all materials, it has become customary to initiate the process of constructing the mechanics model by first assuming the general character of the constitutive equations, and subsequently subdividing the resulting solution methodology according to the mathematical complexity anticipated for the particular material model described by the chosen set of constitutive equations. Thus, for example, in the field of fluid mechanics, materials are subdivided into Newtonian and non-Newtonian (to be discussed below), and in the field of solid mechanics, materials are subdivided into linear elastic, nonlinear elastic, viscoelastic, elasto-plastic, viscoplastic, and so on. Several of these models will be introduced briefly in this chapter.

Fortunately, over the past half century, a fairly rigorous methodology has developed for constructing constitutive equations that is capable of aiding in the prediction of the mechanical response of objects. On the other hand, this methodology, as well as the mathematical form of the resulting equations, can be quite cumbersome and complex. As a result, it is instructive to begin the discussion of this subject on a somewhat simplistic level, thereby preparing the reader for the approaching complexity. In order to accomplish this simplification, this chapter will begin with a discussion of constitutive theories in one spatial dimension.

10.2 One-Dimensional Constitutive Experiments

A simple test not unlike those performed by Da Vinci and Hooke serves the purpose of introducing one-dimensional constitutive models. This test is described as follows. Consider a prismatic (constant shape of the cross section in the x_1 coordinate direction) test specimen subjected to slowly applied colinear and self-equilibrating forces, $\vec{F}$ and $-\vec{F}$, respectively, via a set of (relatively) rigid platens, as shown in Fig. 10.2.

For convenience, a Cartesian coordinate system is used, and the x_1 coordinate direction is selected to be collinear with the applied forces. The magnitude of the forces, F, is determined by measuring the displacement in an inline Hookean spring

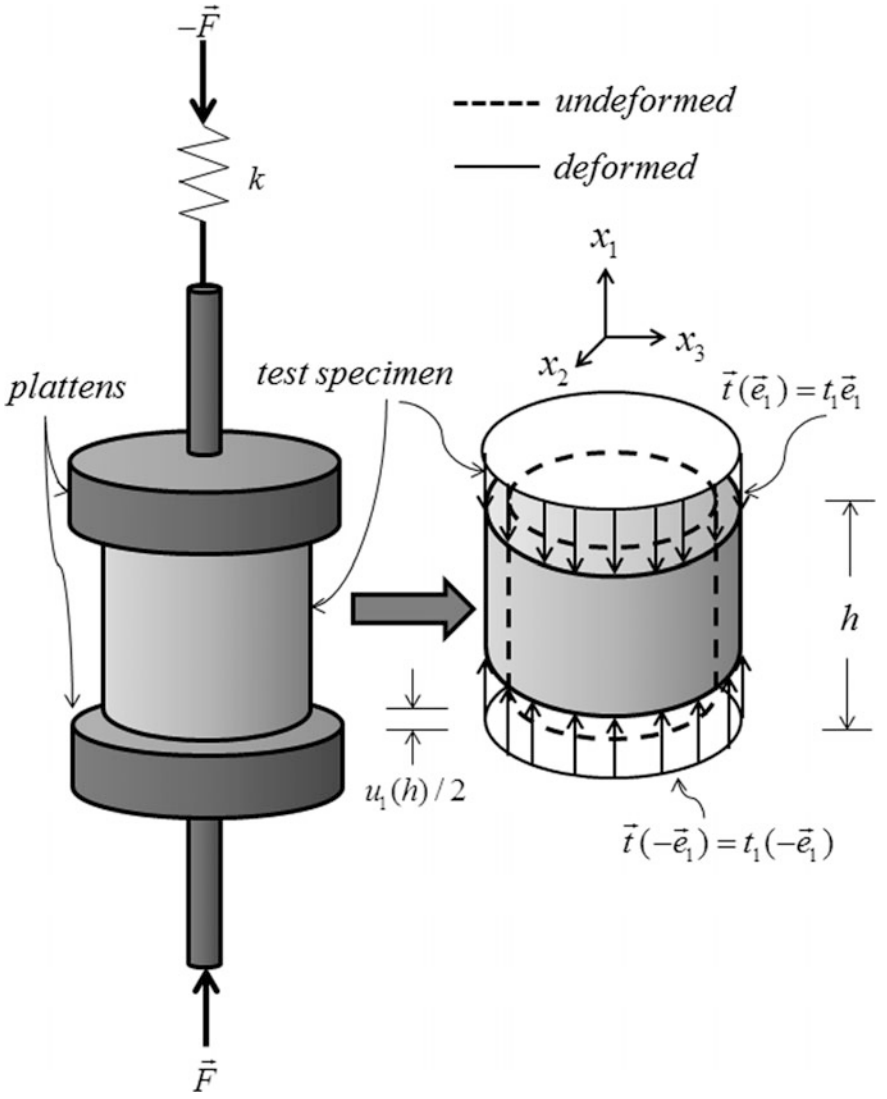


Fig. 10.2 Depiction of a uniaxial test

(called a load cell) and deploying Eq. (10.1). Due to conservation of linear momentum (quasi-static case), the magnitude of this force is equivalent to the magnitude of the force applied to the top and bottom surfaces of the specimen. In order to extract material properties directly, it is necessary that the material properties within the test specimen to be spatially homogeneous (at least macroscopically). For reasons that will become clear below, this test is called *a uniaxial test*.

Note that in the test, the force is assumed to be applied in compression. This is due to the fact that road way materials are normally subjected to compressive loading. Furthermore, road way materials typically do not perform well in tension, demonstrating much lower strength in tension than in compression. Indeed, in many base and sub-base materials, the specimen will fail immediately upon application of uniaxial tensile loading.

If the platens are lubricated so that no lateral forces are applied to the specimen during testing, then it can be shown (using conservation of momentum) that the forces applied to the specimen by the plattens will be (approximately) evenly distributed on the top and bottom of the test specimen, and it follows from Eq. (9.42) that the resulting boundary traction, $\vec{t}(\vec{e}_1)$, can be calculated using the following equation:

$$\vec{t}(\vec{e}_1) = t_1\vec{e}_1 = -\frac{F}{A}\vec{e}_1 \tag{10.2}$$

where A is the area of the top (or bottom) surface of the test specimen. Furthermore, using Cauchy's formula, Eq. (9.48), it can be shown that at every point on the top surface, the uniaxial stress is given by

$$\sigma \equiv \sigma_{11} = t_1 = -\frac{F}{A} \tag{10.3}$$

Thus, it can be seen that the uniaxial stress is spatially homogeneous on the top surface of the test specimen. A similar process will lead to the same value of the stress on the bottom surface of the specimen. Furthermore, if the body force is negligible compared to the applied force, F, and the load is applied sufficiently slowly for the momentum terms to be negligibly small, then it can be shown using the conservation of linear momentum, Eq. (9.83), that the stress component, σ_{11}, is independent of x_1, x_2 and x_3 at every point within the specimen, and that all of the remaining components of stress within the specimen are identically zero. It has therefore been proven that in this test, the stress is everywhere uniaxial, spatially homogeneous, and equivalent to the stress on the boundary, given by Eq. (10.3).

The above proof is significant since it can be seen from Eq. (10.3) that the stress at all points in the specimen can be calculated simply by knowing the applied force and the surface area of the test specimen. Note that the subscripts on the uniaxial stress have been dropped for notational simplicity. This simplification will be utilized for convenience throughout the current chapter.

As shown in Fig. 10.2, it is also possible to measure the relative displacement in the x_1 coordinate direction between the top and bottom surfaces of the specimen, $u_1(h)$, where h is the longitudinal dimension of the specimen. Assuming that the strain in the test specimen is also spatially homogeneous (to be elucidated below), it follows from strain-displacement Eq. (9.40) that the axial strain is spatially constant at every point within the specimen and can be determined from the boundary displacement as follows:

$$\varepsilon \equiv \varepsilon_{11} = \frac{u_1(h)}{h} \tag{10.4}$$

Note that although there will necessarily be other components of strain that are nonzero in this experiment, the determination of these components will not be necessary in order to determine the relationship between the uniaxial stress and the axial strain. Note also that because the stress and strain are spatially homogeneous in this test, it is sometimes called a spatially homogeneous boundary value problem (whereas the displacement field in fact varies linearly in coordinates).

Thus, in this experiment, it is possible to measure from boundary and load train displacements both the uniaxial stress, σ, and the uniaxial strain, ε, as functions of time *at all points within and on the boundary of the test specimen*. Such a circumstance is unusual and stems from the following special circumstances during the test:

(1) The loads are applied normal to two coplanar flat boundaries;
(2) the test specimen is prismatic in the loading direction;
(3) the loads are applied slowly enough that the momentum terms can be neglected;
(4) the loads are applied in such a way that the tractions on the boundary are spatially homogeneous;
(5) the material properties within the specimen are spatially homogeneous; and
(6) the loading is large compared to the gravitational forces applied to the test specimen.

If any of the above assumptions is violated, then the state of stress (as well as the strain) within the specimen will not necessarily be spatially homogeneous.

More importantly, from this experiment, *it is possible to determine the mathematical relationship between the uniaxial stress and strain*. Such an experiment is therefore called *a constitutive experiment*.

The above experiment is also mathematically one-dimensional because a single component of the stress, $\sigma = \sigma_{11}$, can be related mathematically to a single component of the strain, $\varepsilon = \varepsilon_{11}$. This test then serves as the setting for this chapter, entitled One-Dimensional Constitutive Theory. The purpose will be to develop mathematical models that accurately reproduce the relationship between the uniaxial stress and strain. It will be assumed throughout this chapter that the material models developed herein refer to this (or a similar) uniaxial test. The one-dimensional mathematical models discussed below are observed in many cases to be accurate when compared to actual experiments on road way materials.

10.3 Elastic Material Model

Perhaps the simplest realistic one-dimensional constitutive model capable of capturing the mechanical response of road way materials is *the elastic material model*. For this class of materials, *the stress at a given coordinate location, $\vec{x}$, and time, t, is*

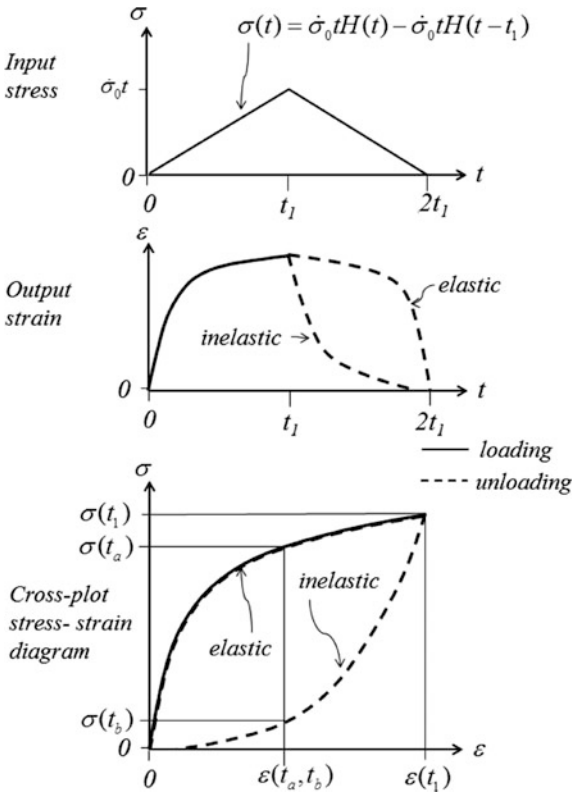


Fig. 10.3 Uniaxial test showing elastic and inelastic response

defined to be a single-valued function of the strain at that coordinate location and time. This class of material model is therefore described mathematically as follows:

$$\sigma(\vec{x}, t) \equiv \sigma(\varepsilon(\vec{x}, t)) \tag{10.5}$$

where the parentheses signify "*a function of.*" It is essential to recognize that a given material may or may not perform in accordance with the elastic material model. In order to determine if a material can be described by the elastic material model, it is necessary to perform a few experiments, although these will not necessarily guarantee that the material is elastic. A set of experiments to determine whether a material is elastic would typically include the following:

(1) A uniaxial test, normally performed in load control and at constant loading rate, $\dot{\sigma}_0$, as shown in Fig. 10.2;
(2) the specimen tested in (1) should also be unloaded at the same unloading rate; and
(3) the uniaxial test performed in (1) and (2) should be performed at several different loading rates.

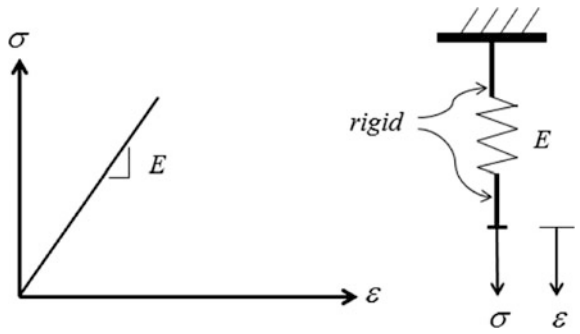


Fig. 10.4 One-dimensional Hookean material behavior

As shown in Fig. 10.3, the results of the above experiments, when output on a cross-plot stress–strain diagram, would necessarily have to be single valued in order to satisfy Eq. (10.5). Note that for the inelastic case, there are two different possible values of the stress for the same value of strain, thus violating Eq. (10.5). Note also that the results for various loading rates should fall on the same stress–strain curve both for loading and unloading in order to ensure that the material does not exhibit loading rate dependence, as this would also violate the elastic material model described by Eq. (10.5). Although the above experiments would not guarantee that the material is elastic, it is normally observed experimentally that materials that exhibit uniqueness of the relationship between stress and strain for this suite of experiments is at least approximately elastic for uniaxial stresses up to the value $\dot{\sigma}_0 t$.

In the special case wherein the above stress–strain curve is linear, the material is called *linear elastic* or *Hookean* (after Robert Hooke). In accordance with Hooke's law, Eq. (10.1), a mechanical analog—a spring—is used to depict the Hookean material model, as shown in Fig. 10.4.

Thus, uniaxial Hookean material behavior, similar to Eq. (10.1), is modeled by

$$\sigma = E\varepsilon \Leftrightarrow \varepsilon = D\sigma \tag{10.6}$$

where E is a material property called *Young's modulus* (after Thomas Young 1773–1829). Also, $D = 1/E$ is called *the uniaxial compliance*. Linear elastic material behavior is observed in most solids in which the primary cause of deformation is molecular bond stretching. In this case, the material dissipates little if any energy due to mechanical loading, so that this stored energy, called *strain energy,* is completely recovered when the specimen is unloaded (see below). Such mechanical behavior is typically observed in rock used as aggregate in road ways, as well as some types of soil, not to mention metals, such as the steel used in concrete reinforcement.

If the cross-plot stress–strain curve shown in Fig. 10.3 is not a straight line, but Eq. (10.5) is nevertheless satisfied, the material is called *nonlinear* elastic. Some materials (such as rubber) sometimes demonstrate nonlinear elastic material behavior. For these materials, a higher order algebraic expression is required for Eq. (10.5) in order to accurately reflect the experimentally observed relation between stress and strain. For example, one might utilize the following power law:

$$\sigma = A\varepsilon^n \tag{10.7}$$

where A and n are material properties that are curve fit to the experimental results. Indeed, where rubber is used as a toughener in asphalt binder, a nonlinear elastic material model may be appropriate.

Aggregates (see Chap. 4) utilized in asphalt concrete are observed to behave elastically up to the onset of slip and/or fracture. Unfortunately, the elastic material model does not reflect the actual material behavior of most road way materials for typical loadings that road ways are subjected to by vehicular traffic. This revelation should be self-evident due to the fact that the elastic material model disallows permanent deformation, whereas anyone who has ever driven a road way is well aware that they invariably undergo significant (sometimes even catastrophic) permanent deformations. For these materials, models that are more mathematically complicated are required in order to accurately reproduce the observed mechanical material behavior. Therefore, it is propitious to consider some of these so-called *inelastic material models*.

10.4 Viscous Material Model

A viscous material is defined to be one in which the stress is a single-valued function of the rate of strain (or velocity gradient). This class of materials is described mathematically as follows:

$$\sigma(\vec{x}, t) \equiv \sigma(\dot{\varepsilon}(\vec{x}, t)) \tag{10.8}$$

where the dot above the strain implies the time derivative of the strain.

Consider, for example, three viscous specimens subjected to constant strain rate, as shown in Fig. 10.5. As depicted in the figure, the resulting stress–strain curve will yield little information, since the absence of strain in the above definition implies that the stress is independent of the strain. Thus, it is more appropriate to cross-plot the stress versus the rate of strain, as shown in the figure.

In the special case wherein the resulting cross-plot of stress versus rate of strain curve is linear, as shown in Fig. 10.6, the material is called *Newtonian* after Isaac Newton, who was apparently the first person to derive this relationship for fluids, including both gases, such as air, and liquids, such as water. In this case, a mechanical analog called a dashpot (a simple shock absorber) is used to depict the material behavior, as shown in the figure.

Thus, one-dimensional (most commonly, pure shear) Newtonian material behavior is modeled by

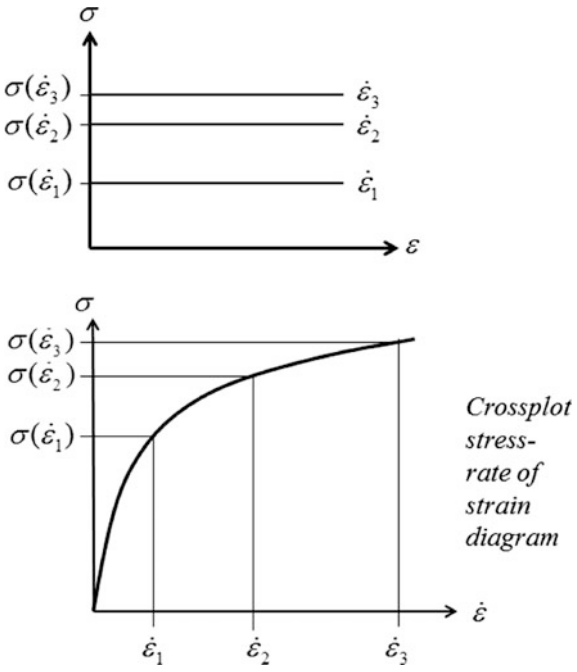


Fig. 10.5 One-dimensional testing of a viscous material

$$\sigma = \eta \dot{\varepsilon} \tag{10.9}$$

where η is a material property called the viscosity.

When the curve obtained by plotting the stress versus the rate of strain is nonlinear, the material is called *non-Newtonian*, and in this case, it is necessary to expand Eq. (10.8) in higher order terms in the rate of strain in order to accurately model the mechanical response of the material. For example, one might utilize the following power law:

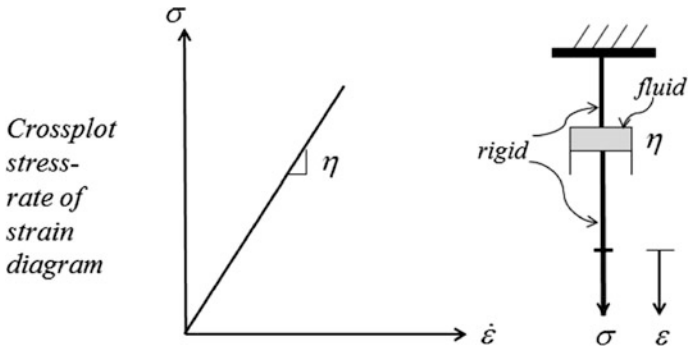


Fig. 10.6 One-dimensional Newtonian material behavior

$$\sigma = A\dot{\varepsilon}^n \tag{10.10}$$

where A and n are material properties that are curve fit to the experimental results. A typical non-Newtonian viscous road way material is bitumen at elevated temperatures.

The viscous material model is accurate only for modeling fluids, in which a significant portion of the molecular bonds within the material is broken (without load application) at the current temperature. Since definition (10.8) implies that a viscous material will deform continuously under constant loading, road way designers should avoid using viscous materials as structural components in road ways. This type of material model would, therefore, seem to be inappropriate for modeling road ways. However, asphalt that comes from the wellhead is usually a liquid (see Chap. 2), not to mention the water that is deployed with Portland cement to make it a useful binder. Furthermore, moisture due to climate is virtually unavoidable within road ways.

Thus, despite the rather dubious structural properties of viscous materials, it would seem that the viscous material model discussed in this section will play a necessary role in road way design. It will be shown in later chapters that this is indeed the case.

10.5 Viscoelastic Material Model

Sometimes materials exhibit constitutive behavior that is neither elastic nor viscous. Obvious examples occur when a solid is embedded in a liquid (or vice versa), such as partially saturated soils, Portland cement, asphalt (both the binder (see Chap. 2), embedded with fines (see Chap. 5), and the concrete embedded with aggregate (see Chaps. 6 and 7)), soft biologic tissue, and even such exotic materials as cooked pasta. One class of materials that responds in this way is termed *a viscoelastic material,* in which *it is assumed that the stress is a single-valued functional of the strain*, defined mathematically as follows:

$$\sigma(\vec{x}, t) \equiv E_{\tau=0}^{\tau=t}\{\varepsilon(\vec{x}, \tau)\} \tag{10.11}$$

where the generic timescale has been changed to τ in order to clarify the meaning of *the modulus functional*, *E{}*, which maps the entire history of strain into the value of the stress at the current time, t. The above mathematical formulation is called *a strain (or modulus) formulation* because the strain is the input and the stress is the output variable. *A stress (or compliance) formulation* is given by the following form:

$$\varepsilon(\vec{x},t) = D_{\tau=0}^{\tau=t}\{\sigma(\vec{x},\tau)\} \tag{10.12}$$

where $D\{\}$ is termed *the compliance functional*.

The decision as to which of the above two formulations the modeler chooses to use normally depends on the particular problem to be solved. Since the mathematics necessary to predict the mechanical response of viscoelastic materials is notoriously complicated, this complexity will often dominate the choice of formulation. Fortunately, in most practical circumstances, there exists a mapping between the functionals $E\{\}$ and $D\{\}$, so that only one needs to be formulated and characterized experimentally (see Chap. 12).

In order to introduce to the reader the nuances of viscoelastic material behavior, consider, for example, a uniaxial creep test, defined as follows:

$$\sigma(t) = \sigma_0 H(t - t_0) \quad \sigma_0 = const \tag{10.13}$$

As shown in Fig. 10.7, a typical viscoelastic material is observed to undergo time-dependent deformation under constant loading. Of course, this type of test is entirely unnecessary if the material is elastic, since materials that are elastic do not exhibit time dependence, therefore obviating creep with time. In addition, as shown from the figure, the material deforms instantaneously on application of the load, also obviating the viscous material model. Thus, it can be seen from this test that neither the elastic nor the viscous material model described above will accurately reproduce the constitutive response of this material.

Note that the curve of strain versus time in Fig. 10.7 is nonlinear, but does this imply that the material is nonlinear? The answer is—*not necessarily*—and this is confusing to most newcomers to viscoelasticity. In order to understand this issue, it

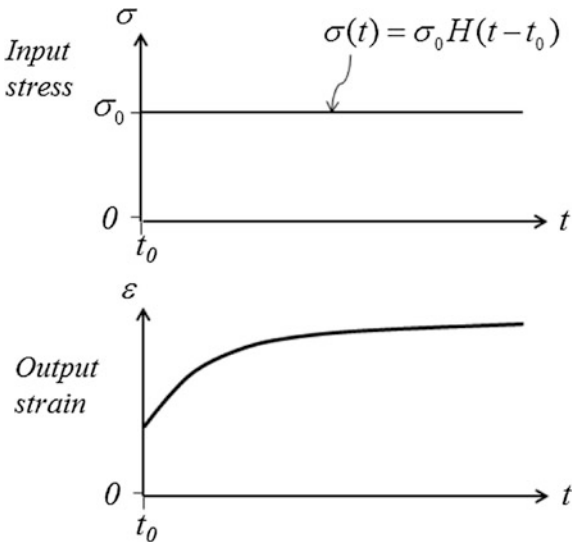


Fig. 10.7 Uniaxial creep test

is necessary to refer to the definition of linearity discussed in Chap. 9. Toward this end, Eqs. (9.16a) and (9.16b) are recast here in terms of the input variable, $\sigma(t)$, and the output variable, $\varepsilon(t)$, thus implying a viscoelastic stress formulation, as depicted in Eq. (10.12). Accordingly, the two conditions for linearity are written as follows:

$$\text{Superposition: } \varepsilon(t) = D_{\tau=0}^{\tau=t}\{\sigma_1(\tau) + \sigma_2(\tau)\} = D_{\tau=0}^{\tau=t}\{\sigma_1(\tau)\} + D_{\tau=0}^{\tau=t}\{\sigma_2(\tau)\} \tag{10.14a}$$

$$\text{Homogeneity: } \varepsilon(t) = D_{\tau=0}^{\tau=t}\{c\sigma(\tau)\} = cD_{\tau=0}^{\tau=t}\{\sigma(\tau)\} \tag{10.14b}$$

In order to check *experimentally* for linear viscoelastic material response, it is therefore necessary to choose a specific experiment, i.e., a particular form for the mapping, $D\{\}$. Accordingly, suppose that the simple creep test described is chosen, that is, let

$$c\sigma(\tau) = \sigma_0 H(\tau) \tag{10.15}$$

In order to check for linearity, both (10.14a) and (10.14b) must be experimentally observed to be satisfied. Suppose that homogeneity, Eq. (10.14b), is checked first. To do this, substitute Eq. (10.15) into Eq. (10.14b) as follows:

$$\varepsilon(t) = D_{\tau=0}^{\tau=t}\{\sigma_0 H(\tau)\} = \sigma_0 D_{\tau=0}^{\tau=t}\{H(\tau)\} \quad ? \tag{10.16}$$

The above must be satisfied in order for the homogeneity test to be passed. To perform this check, first define *the creep compliance* observed in the creep test as follows:

$$D(t) \equiv \frac{\varepsilon(t)}{\sigma^0} \tag{10.17}$$

The above implies that

$$D(t) = D_{\tau=0}^{\tau=t}\{H(\tau)\} \tag{10.18}$$

Thus, it can be seen from Eq. (10.16) that the creep compliance must be independent of the loading, σ_0, applied in the creep test in order for the homogeneity condition to be satisfied. Therefore, to check for satisfaction of the homogeneity condition, the usual set of tests is to subject several specimens to creep tests, each with an increase in the applied loading, as shown in Fig. 10.8.

As shown from Fig. 10.8, for this particular example, the creep compliance is approximately the same for the first three input stress levels, but the compliance increases dramatically at the fourth stress level. Therefore, the material can be said to have passed the homogeneity test up to the stress level σ_3, thus implying that the material *may* be linear viscoelastic up to this stress level, but is definitely nonlinear for stress levels greater than σ_4.

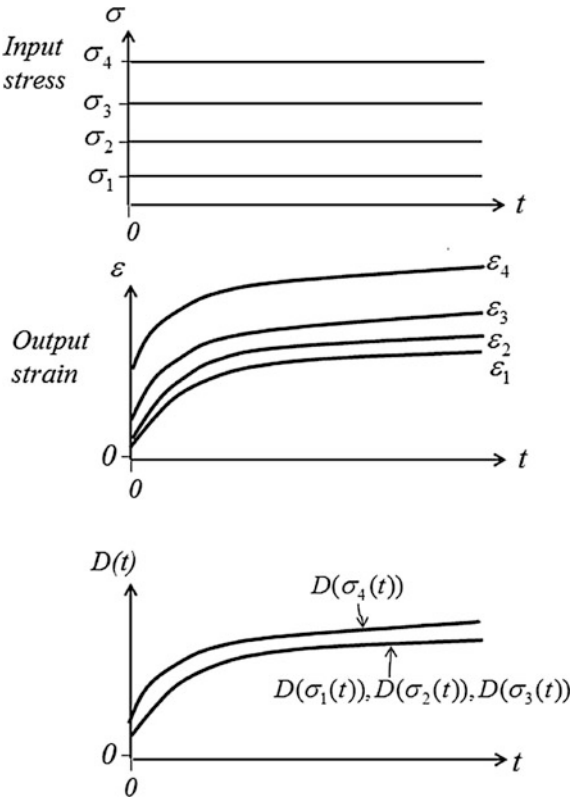


Fig. 10.8 Testing a viscoelastic material for homogeneity condition

Since the superposition principle has not yet been checked, it is premature to assume that the material displays linear response. Nonetheless, some modelers do not perform a superposition check, perhaps because they perceive that this experiment is not as simple to perform as the homogeneity check. However, there is one fairly straightforward experiment that can be performed as a check for superposition, and that is *a creep and recovery test*, as described by the following:

$$\sigma(t) = \sigma_0 H(t) - \sigma_0 H(t - t_1) \tag{10.19}$$

Substituting the above into the superposition condition, Eq. (10.14a), will thus result in the following (assuming that homogeneity has already been checked and verified):

$$\begin{aligned} &\varepsilon(t) = D_{\tau=0}^{\tau=t}\{\sigma_0 H(\tau) - \sigma_0 H(\tau - t_1)\} = D_{\tau=0}^{\tau=t}\{\sigma_0 H(\tau)\} - D_{\tau=0}^{\tau=t}\{\sigma_0 H(\tau - t_1)\}? \\ &\quad \Rightarrow \varepsilon(t) = \sigma_0 D(t) - \sigma_0 D(t - t_1) \quad \forall t > t_1\,? \end{aligned} \tag{10.20}$$

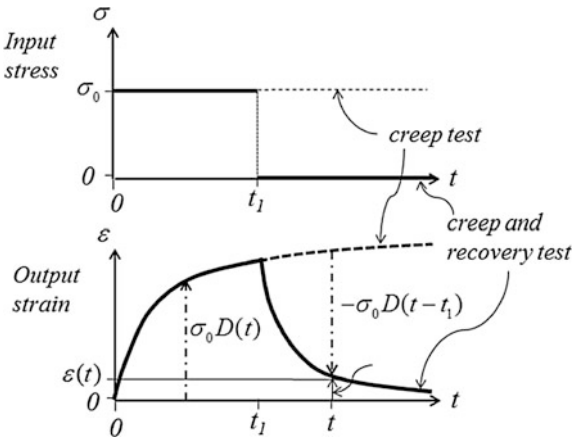


Fig. 10.9 Testing a viscoelastic material for superposition condition

The above implies that in order for superposition to hold in this experiment, for times greater than $t = t_1$, the time at which the loading is removed, the experimentally measured strain, $\varepsilon(t)$ must be equal to the strain that would have been observed in a creep test begun at time $t = 0$, denoted $\sigma_0 D(t)$ in accordance with Eq. (10.17), minus the strain observed *in the same creep test*, except that the value subtracted is for time $t - t_1$, denoted $\sigma_0 D(t - t_1)$. The procedure for checking whether superposition holds in this experiment is thus depicted in Fig. 10.9, wherein the two vertical dotted lines in the creep diagram denote the two terms on the right-hand side of Eq. (10.20). If the difference between these two values is not equivalent to the observed strain in the creep recovery test, then superposition is not satisfied and the material is necessarily nonlinear at this level of stress. In this case, a nonlinear viscoelastic material model must be utilized. This subject will be discussed further in Chap. 12.

For the moment, it will be assumed that the experiments on the material of interest pass both the superposition and the homogeneity tests, implying that the material is linear viscoelastic. In this case, it is still necessary to construct a precise mathematical representation of the functional model described by Eq. (10.11). Various possibilities have, in fact, been proposed by numerous researchers over the past 150 years.

Early attempts (beginning in the 1860s) took a quite simplistic approach with the thought in mind that the more complicated the mathematical expression, the more difficult would be the resulting mathematical technique necessary to obtain solutions with the model. For example, consider the case of placing a single dashpot in parallel with a single spring, called *the Voigt model* (originally introduced by Oskar E. Meyer (1834–1909), but named after Woldemar Voigt (1850–1919), who reintroduced it in 1892 (Leaderman 1943)), as depicted in Fig. 10.10, where $D_V \equiv 1/E_V$ is the spring compliance, which is assumed to be constant in time.

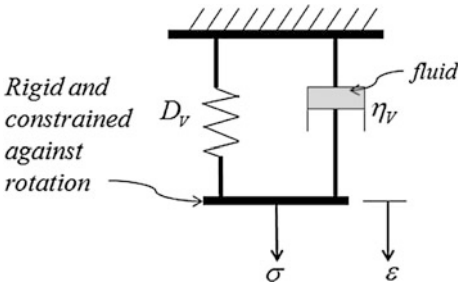


Fig. 10.10 Voigt model

In order to construct the constitutive equation governing the behavior of the above mechanical analog, it is treated as a heterogeneous continuum and the resulting mathematical problem is solved and homogenized. Toward this end, it is assumed that the load is applied slowly, so that the inertial effects can be neglected. Furthermore, it is assumed that the horizontal bar is constrained to translate without rotation. This results in the following governing equations for the *initial boundary value problem* (IBVP):

Conservation of linear momentum:

$$\sigma = \sigma^s + \sigma^d \tag{10.21}$$

Kinematics:

$$\varepsilon = \varepsilon^s = \varepsilon^d \tag{10.22}$$

Material constitution:

$$\text{Spring: } \sigma^s = E_V \varepsilon^s = \varepsilon^s / D_V \tag{10.23}$$

$$\text{Dashpot: } \sigma^d = \eta_V \dot{\varepsilon}^d \tag{10.24}$$

where the superscripts s imply "spring" and the superscripts d imply "dashpot."

The above IBVP can be solved by substituting Eqs. (10.22)–(10.24) into Eq. (10.21), thus resulting in the following ordinary differential equation:

$$\sigma = \frac{\varepsilon}{D_V} + \eta_V \dot{\varepsilon} \tag{10.25}$$

The above is called *the governing differential equation for a Voigt model*, thus demonstrating that viscoelastic materials can be represented by differential equations. Note also that it can be shown that the above equation is linear, so that both the response functionals $E\{\}$ and $D\{\}$ are linear for this material model. To see whether this material model accurately captures the material behavior of typical viscoelastic materials, consider once again a creep test, described by Eq. (10.13). Substituting this equation into governing differential equation (10.25) and solving for the output strain results in the following:

$$\sigma^0 = \frac{\varepsilon}{D_V} + \eta_V \frac{d\varepsilon}{dt} \Rightarrow \varepsilon(t) = D_V \sigma^0 \left(1 - e^{\frac{-t}{D_V \eta_V}}\right) \tag{10.26}$$

It then follows from Eq. (10.17) that for the Voigt model, the creep compliance is given by:

$$D_V(t) = D_V \left(1 - e^{\frac{-t}{D_V \eta_V}}\right) \tag{10.27}$$

where it should be noted that D_V represents the Voigt spring compliance (which is a constant), whereas $D_V(t)$ represents the Voigt creep compliance, which is a function of time.

The above equation is plotted in Fig. 10.11. Note that due to the character of the exponential term e^{-at}, the creep compliance exhibits observable change over only about two decades of time. When a creep test is performed on real materials such as asphalt, the compliance changes over a much larger span of time than two decades, as shown in the figure. Therefore, the Voigt model does not appear to be a very accurate representation of actual material behavior. Nonetheless, the Voigt model is clearly an improvement over the elastic material model, which predicts no creep at all.

It now follows from Eq. (10.25) that since the material is linear, a convolution integral may be utilized to predict the response to a stress history that is more complicated than a creep test by utilizing the Voigt creep compliance as follows:

$$\varepsilon(t) = \int_0^t D_V(t - \tau) \frac{\partial \sigma}{\partial \tau} d\tau \tag{10.28}$$

Thus, it can be seen that the Voigt analog can be represented by either the differential Eq. (10.25) or the convolution integral Eq. (10.28). Deployment of each of these approaches will reveal that these two mathematical formulations produce identical results. The answer as to which formulation is more propitious for predicting the response of flexible pavements will be discussed in Chaps. 12 and 13.

Next consider *a relaxation test*, given by:

$$\varepsilon(t) = \varepsilon_0 H(t) \quad \varepsilon_0 = const \tag{10.29}$$

The above test is typically quite a bit more difficult to perform in the laboratory than is a creep test, which only requires a load frame capable of supporting the applied loading. A relaxation test, on the other hand, requires the construction of a testing machine that can apply a deformation essentially instantaneously and hold that deformation constant in time. Although this type of test can be performed with the aid of modern electronic equipment, it is nonetheless challenging.

Supposing that this test can be performed, consider the prediction of the response that is made by the Voigt model. Substituting Eq. (10.29) into governing differential Eq. (10.25) results in the following:

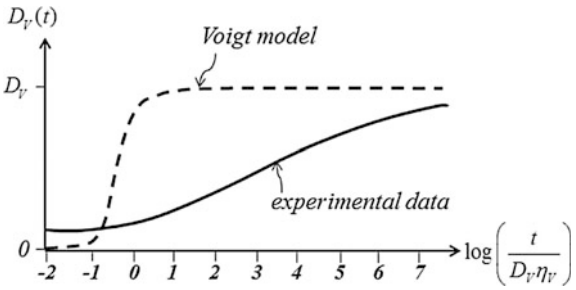


Fig. 10.11 Voigt model predicted creep compared to typical material response

$$\sigma = E_v \varepsilon_0 \tag{10.30}$$

Now define *the relaxation modulus* as follows:

$$E(t) \equiv \frac{\sigma(t)}{\varepsilon_0} \tag{10.31}$$

It follows that for the Voigt model

$$E_V(t) = E_V = const \tag{10.32}$$

Thus, the Voigt model predicts that during a stress relaxation test, the stress will not change in time. This does not model the material response observed in typical viscoelastic media well at all. Thus, the Voigt model is not accurate for modeling time-dependent response of road way materials (Fig. 10.11).

An alternative view was taken by James Clerk Maxwell, who instead placed a spring in series with a dashpot, thereby creating *the Maxwell Model* (Maxwell 1867), as shown in Fig. 10.12.

This results in the following governing equations for the *initial boundary value problem*:

Conservation of linear momentum:

$$\sigma = \sigma^s = \sigma^d \tag{10.33}$$

Kinematics:

$$\varepsilon = \varepsilon^s + \varepsilon^d \tag{10.34}$$

Material constitution:

$$\text{Spring: } \sigma^s = E_M \varepsilon^s \tag{10.35}$$

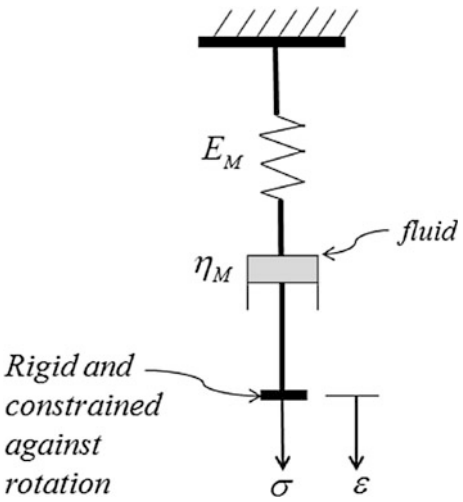


Fig. 10.12 Maxwell model

$$\text{Dashpot: } \sigma^d = \eta_M \varepsilon^d \tag{10.36}$$

The above IBVP can be solved by differentiating Eqs. (10.34) and (10.35) with time and substituting (10.33) and (10.36) into the result, thus leading to the following ordinary differential equation:

$$\dot{\varepsilon} = \frac{\dot{\sigma}}{E_M} + \frac{\sigma}{\eta_M} \tag{10.37}$$

The above is called *the governing differential equation for a Maxwell model*, thus once again demonstrating that viscoelastic materials can be represented by differential equations.

To see if this material model accurately captures the material behavior of typical viscoelastic materials, consider once again a creep test, described by Eq. (10.13). Substituting this equation into governing differential Eq. (10.37) and solving for the output strain results in the following:

$$\varepsilon(t) = \frac{\sigma_0}{E_M} + \frac{\sigma_0}{\eta_M} t \tag{10.38}$$

It then follows from Eq. (10.17) that for the Maxwell model, the creep compliance is given by:

$$D_M(t) = \frac{1}{E_M} + \frac{t}{\eta_M} \tag{10.39}$$

When a creep test is performed on solid materials such as asphalt, the compliance change is not linear in time. Therefore, the Maxwell model does not appear to be a very accurate representation of actual material behavior.

Next consider *a relaxation test*, given by Eq. (10.29). Substituting this into Eq. (10.37) and solving will result in the following:

$$\sigma(t) = E_M \varepsilon_0 e^{-\frac{E_M}{\eta_M} t} \tag{10.40}$$

The resulting relaxation modulus is thus:

$$E_M(t) = E_M e^{-\frac{E_M}{\eta_M} t} \tag{10.41}$$

As indicated in Fig. 10.11, the above equation will decay over about two decades of time, whereas the relaxation modulus in real materials usually decays over many decades in time. Thus, once again it is found that the Maxwell model is not accurate for real materials.

It now follows from Eq. (10.25) that since the material is linear, a convolution integral may be utilized to predict the response to a stress history that is more complicated than a creep test by utilizing the Maxwell creep compliance as follows:

$$\sigma(t) = \int_0^t E_M(t-\tau) \frac{\partial \varepsilon}{\partial \tau} d\tau \tag{10.42}$$

Thus, it can be seen that, just as in the case of the Voigt analog, the Maxwell analog can be represented by either the differential Eq. (10.37) or the convolution integral Eq. (10.42).

It has been indicated in this section that the Voigt and Maxwell models do not accurately capture the response of typical structural materials. It will therefore be necessary to construct even more complicated mathematical equations in order to accurately reproduce the response of realistic viscoelastic materials such as asphalt concrete. This is an advanced subject that will be covered in Chap. 12.

Nevertheless, the point to be taken from this section is that once a viscoelastic constitutive model has been chosen that is accurate for the material at hand, it can in principle be used to predict the mechanical response of an arbitrarily shaped continuous body subjected to essentially any external loading. This then is the concept of all continuum mechanics models. Everything else is just mathematical details. Unfortunately, the mathematical details necessary to construct a model that is accurate for viscoelastic materials can be very complicated, as will be shown in Chap. 12.

Since the history dependence exhibited in the viscoelastic material model is clearly a significantly more complicated mathematical model than that proposed for elastic and viscous materials, it is preferable to avoid using this type of model whenever possible. Unfortunately, it is not possible to predict the performance of

flexible pavements with simple elastic material models in most practical circumstances, especially those that involve predictions of long-term pavement performance. Fortunately, computational tools have been developed in the last few decades that make it possible today to solve problems using the viscoelastic material model, and these will be discussed in Chap. 14.

The reader may ask why even bother going down this path, that is, creating complicated mathematical models to predict the response of viscoelastic media. The answer is really quite simple: In order to accurately predict the failure of viscoelastic media such as asphalt binder (see Chap. 2) and asphaltic concrete (see Chap. 7), the physics of the energy dissipation resulting from mechanical loading must be incorporated within the model. Otherwise, road way designers will be forced to design road ways experimentally, which is comparable to the way historians believe the Egyptians built the great pyramids. Historians have also suggested that the approach taken by the Egyptians bankrupted their civilization (Allen 2014).

10.6 Elasto-Plastic Material Model

Many materials do, in fact, exhibit uniaxial elastic material behavior up to a measurable value of stress, termed the uniaxial yield stress, σ_Y, as shown in Fig. 10.13, and the elastic response up to the uniaxial yield stress is typically linear. Furthermore, although the material behavior after yielding is history dependent, it is nonetheless oftentimes rate independent.

This class of materials is called *elasto-plastic*, and the mathematical definition of *a uniaxial elasto-plastic material* is as follows (Allen and Haisler 1985):

$$\begin{aligned} &\sigma(\vec{x},t) = E\varepsilon(\vec{x},t) \quad \forall\sigma \le \sigma_Y \\ &\sigma(\vec{x},t) = E_{\tau=0}^{\tau=t}\{\varepsilon(\vec{x},\tau)\} = E_{\tau=0}^{\tau=t}\{\varepsilon(\vec{x},c\tau)\} \quad \tau \to c\tau \quad \forall\sigma > \sigma_Y \end{aligned} \tag{10.43}$$

The above model is demonstrated graphically in Fig. 10.14.

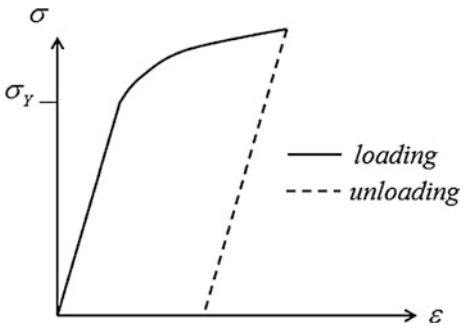


Fig. 10.13 Depiction of uniaxial elasto-plastic material behavior

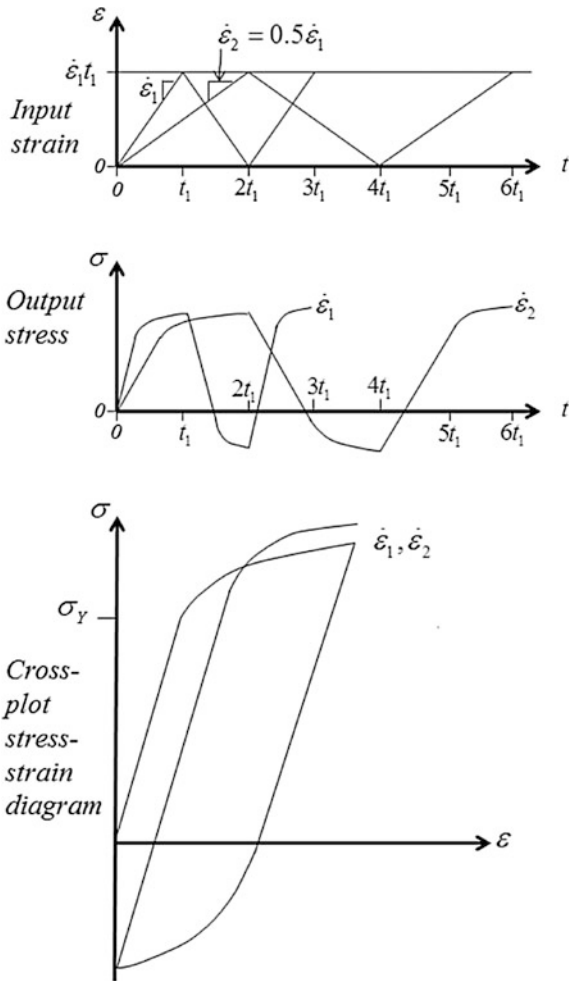


Fig. 10.14 Uniaxial tests to check for elasto-plastic rate independence

Note from the figure that the yield point, σ_Y, is not a material constant, but instead depends on the loading history. This complication will be discussed further in Chap. 13.

It should be clear that the elasto-plastic constitutive model is a special case of the (nonlinear) viscoelastic material model in which the material response is independent of strain rate. This definition is chosen because some materials (including soils with little or no moisture), such as those used in road way base and sub-base applications (after consolidation and drying), can exhibit this type of history dependence. Furthermore, aggregates (see Chap. 3) may exhibit elasto-plastic response due to intergranular slip and granular fracture (see Chap. 8). This type of model is also useful for predicting the response of metals below about 30% of their melting temperature.

As will be shown in Chap. 13, this type of material behavior can be quite complicated, but it is often nonetheless mathematically simpler as well as more accurate than that produced by a viscoelastic material model.

10.7 Viscoplastic Material Model

The terminology *viscoplastic material model* is not well defined in the research literature. What can be said is that there exists a class of materials that are modeled as a special case of the viscoelastic constitutive model discussed above in which the following commonality occurs:

(1) The material exhibits a yield point, as in the elasto-plastic material, and the material response is often linear elastic below this stress;
(2) the material response at stresses in excess of the yield point, σ_Y, is rate dependent;
(3) the material response is typically (but not always) modeled with a set of first order *nonlinear* differential equations; and
(4) the material exhibits permanent deformation on unloading.

Viscoplastic material models were initially developed in the latter half of the twentieth century for the purpose of modeling metals at elevated temperature (above about three-tenths of their melting temperature). However, in recent years, they have been deployed for a variety of other materials, including asphalt concrete (see Chap. 7), Portland cement concrete, and even moisture-laden road way base materials. One such model will be presented in three dimensions in Chap. 13.

10.8 Thermo- and Hygro-Type Material Models

Roadways are subjected to environmental conditions that affect their performance (see Chap. 8). It will be shown in subsequent chapters specifically how these effects can be incorporated into mechanical constitutive models. Within the context of the current section dealing with one-dimensional material models, the effect of both temperature and moisture must be included in the constitutive model if these effects are to be considered in the design process. This effect can readily be seen by performing a simple test on the specimen shown in Fig. 10.2. To do this, the specimen is placed within an environmental chamber and the temperature is slowly increased (or decreased) within the chamber. It will be observed that the specimen will change its size and shape if it is unconstrained from deformation. On the other hand, it will be observed that it will be required to apply loads to the boundary of the object in order to maintain the shape it took on before the temperature was

changed. In order to model this temperature dependence, it is necessary to generalize all of the models described above. This is accomplished by including temperature as an input variable (on the right-hand side), and when the temperature is included, the prefix thermo-is added to the name of the material model. Thus, the following generalizations are constructed to accommodate temperature effects:

(1) *Thermoelastic material model:*

$$\sigma(\vec{x},t) \equiv \sigma(\varepsilon(\vec{x},t), T(\vec{x},t)) \tag{10.44}$$

(2) *Thermoviscoelastic material model*:

$$\sigma(\vec{x},t) \equiv E_{\tau=0}^{\tau=t}\{\varepsilon(\vec{x},\tau), T(\vec{x},t)\} \tag{10.45}$$

where it should be understood that the temperature is an unknown field variable similar to the stress and strain. As will be shown in Chap. 11, the temperature is predicted by using the conservation of energy (see Chap. 9).

For example, consider the case of the thermoelastic material model given by Eq. (10.44). Suppose that the stress is expanded linearly in the strain and temperature, thereby resulting in the following:

$$\sigma = a_1\varepsilon + a_2\Delta T \tag{10.46}$$

where ΔT is defined to be change in temperature from the temperature in the undeformed state, and a_1 and a_2 are material constants. In order to measure the thermal coefficient, a_2, consider an experiment in which a uniaxial bar of length L is subjected to a spatially constant temperature change, as shown in Fig. 10.15.

It is usually observed that when the longitudinal strain resulting from the temperature change is plotted against this change, the resulting line is very nearly straight, and the slope of this line is called *the coefficient of thermal expansion,* α, defined by:

$$\alpha \equiv \varepsilon/\Delta T \tag{10.47}$$

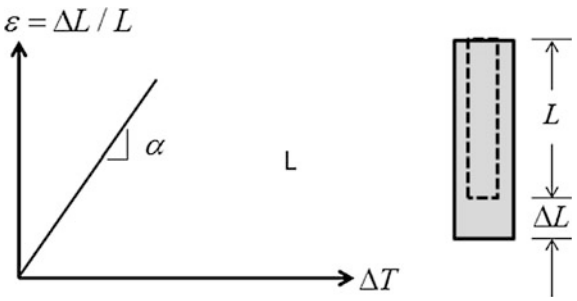


Fig. 10.15 Depiction of procedure for experimentally determining the coefficient of thermal expansion

Note that this result is only correct if the bar is unconstrained against expansion, implying that the uniaxial stress is zero. Now, consider the case wherein the stress is zero in Eq. (10.46). Equating the resulting equation to Eq. (10.47) will result in the following:

$$\alpha = -\frac{a_2}{a_1} \tag{10.48}$$

In as much as the coefficient a_1 has already been established in Eq. (10.6) to be equivalent to Young's modulus, E, substitution of this fact and Eq. (10.48) into Eq. (10.46) now results in the following:

$$\sigma = E(\varepsilon - \alpha \Delta T) \tag{10.49}$$

thereby leading to the result that if the bar is completely constrained ($\varepsilon = 0$), the resulting stress will be

$$\sigma = -E\alpha\Delta T \tag{10.50}$$

Very similar physical outcomes to those observed for temperature occur if the effect of moisture is taken into consideration. Once again, suppose that the specimen shown in Fig. 10.2 is subjected to a change in the moisture within an environmental chamber. The specimen will be observed to change its shape if unconstrained. Thus, moisture must be included in constitutive models.

In order to model this moisture dependence, first define the moisture per unit mass, $M = M(\vec{x}, t)$, as the mass of water contained in a unit mass of the host material. Using this field variable, it is necessary to once again generalize all of the models described above. This is accomplished by including moisture as an input variable (on the right-hand side), and when the moisture is included, the prefix hygro- is added to the name of the material model. Thus, the following generalizations are constructed to accommodate moisture effects:

(3) *Hygro-thermo-elastic material model:*

$$\sigma(\vec{x}, t) \equiv \sigma(\varepsilon(\vec{x}, t), T(\vec{x}, t), M(\vec{x}, t)) \tag{10.51}$$

(4) *Hygro-thermo-viscoelastic material model*:

$$\sigma(\vec{x}, t) \equiv E_{\tau=0}^{\tau=t}\{\varepsilon(\vec{x}, \tau), T(\vec{x}, t), M(\vec{x}, t)\} \tag{10.52}$$

As will be shown later, the moisture is predicted by using conservation of mass (see Chap. 14).

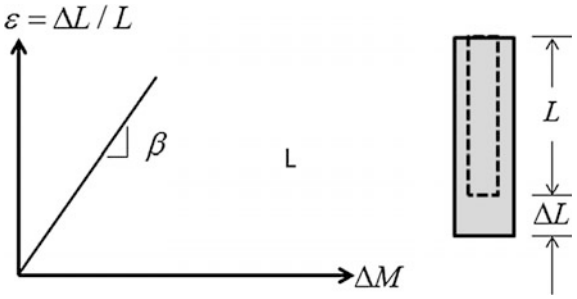


Fig. 10.16 Depiction of procedure for experimentally determining the coefficient of moisture expansion

For example, consider the case of a hygro-elastic material model given by Eq. (10.51) at constant temperature. Suppose that the stress is expanded linearly in the strain and moisture, thereby resulting in the following equation:

$$\sigma = b_1 \varepsilon + b_2 \Delta M \tag{10.53}$$

where ΔM is defined to be the change in moisture from the moisture in the undeformed state, and b_1 and b_2 are material constants. In order to measure the moisture coefficient, b_2, consider an experiment in which a uniaxial bar of length L is subjected to a spatially constant moisture change, as shown in Fig. 10.16.

It is usually observed that when the longitudinal strain resulting from the moisture change is plotted against this change, the resulting line is very nearly straight, and the slope of this line is called *the coefficient of moisture expansion,* β, defined by:

$$\beta \equiv \varepsilon / \Delta M \tag{10.54}$$

Note that this result is only correct if the bar is unconstrained against expansion, implying that the uniaxial stress is zero. Now, consider the case wherein the stress is zero in Eq. (10.53). Equating the resulting equation to Eq. (10.54) will result in the following:

$$\beta = -\frac{b_2}{b_1} \tag{10.55}$$

In as much as the coefficient b_1 has already been established in Eq. (10.6) to be equivalent to Young's modulus, E, substitution of this fact and Eq. (10.55) into Eq. (10.53) now results in the following:

$$\sigma = E(\varepsilon - \beta \Delta M) \tag{10.56}$$

Thereby leading to the result that if the bar is completely constrained ($\varepsilon = 0$), the resulting stress will be

$$\sigma = -E\beta\Delta M \tag{10.57}$$

The effects of moisture and temperature are observed in essentially all materials, including road way pavements and base materials. Perhaps the effects are more significant in road ways than in most other circumstances due to the fact that road ways are oftentimes constrained against motion in the direction of travel (called plane strain). Furthermore, temperature and moisture change are both ubiquitous and unavoidable in road ways. That of course is the bad news. The good news for the road way designer is that despite the ominous appearance of Eqs. (10.51) and (10.52), it is not nearly so difficult to model the effects of moisture and temperature as in former times, and this is due to the development of modern high-speed computers. As will be shown in due course, the prediction of the effects of these environmental changes on road way performance is no longer beyond the grasp of road way designers.

10.9 Summary

In this chapter, the concept of a constitutive model has been introduced. Such a model is utilized in this text to relate the kinematic variables to the kinetic variables in a solid subjected to thermomechanical loading. Because the constitutive behavior of real materials can be quite complicated, a series of constitutive models has been developed historically, and these models increase in mathematical complexity as more physically intricate behavior of materials is encompassed within the models. Because this mathematical complexity is quite challenging, the models to be developed in succeeding chapters in three dimensions have been presented in this chapter in one-dimensional form as a means of introducing this subject in as mathematically and physically simple a form as possible. Accordingly, one-dimensional models have been introduced for elastic, viscous, elasto-plastic, viscoelastic, and viscoplastic media, and these models are necessary to capture the observed material behavior of base materials, asphalt binders (see Chap. 2), aggregate (see Chap. 3), and asphalt mixtures (see Chaps. 7 and 8). In addition, the effects of moisture and temperature on one-dimensional constitutive behavior have been discussed (see Chap. 8). Finally, experimental techniques for obtaining material properties necessary for characterizing one-dimensional constitutive equations have been detailed in this chapter.

The concepts developed in this chapter will be expanded to three dimensions in Chaps. 11–13, wherein considerable mathematical complexity will become unavoidable. Nonetheless, the fundamental concepts introduced in this chapter in one dimension will in no way be altered in succeeding chapters of this text. Therefore, whenever the reader finds this oncoming complexity cumbersome, referral back to this chapter will oftentimes serve to clarify without loss of physical accuracy.

10.10 Problems

Problem 10.1
Given: Sufficient resources to buy a rectangular parallelepiped of rubber (an eraser).
Required: Using the concepts laid out in Sect. 10.2, devise and construct a uniaxial testing machine that is relatively stiff compared to the rubber, and:

(a) Devise a means of calculating the load applied to the eraser by your machine.
(b) Devise a means of determining the relative axial motion of the platens.
(c) Utilize your machine to apply a prescribed loading history to the eraser.
(d) Plot on three graphs the following: loading versus time, platen displacement versus time, and uniaxial stress versus axial strain.

Problem 10.2
Given: The apparatus you invented in Problem 10.1.
Required: Construct a solid material by embedding olive oil in cornstarch. Mold this material into a uniaxial constitutive specimen, as described in Sect. 10.2. Use it to do the following:

(a) Repeat the tests performed in Problem 10.1.
(b) Determine whether your new material is elastic or inelastic.

Problem 10.3
Given: The apparatus you invented in Problem 10.1.
Required: Construct a solid material by embedding water in clay soil and kneading it into a homogeneous solid. Use it to do the following:

(a) Plot a uniaxial stress versus strain diagram.
(b) Determine at what value of stress the material ceases to behave elastically.
(c) Determine whether the inelastic uniaxial material behavior is elasto-plastic or viscoelastic.

Problem 10.4
Given: The apparatus you constructed in Problem 10.1.
Required: In accordance with Sect. 10.5, perform creep and creep and recovery tests on the materials to discern whether the material is linear viscoelastic for the following:

(a) The material you fabricated in Problem 10.2.
(b) The material you fabricated in Problem 10.3.

Problem 10.5
Given: A Voigt model
Required: Use it to do the following:

(a) Derive Eq. (10.25).

(b) Show that Eq. (10.25) is linear.
(c) Derive Eq. (10.27).
(d) Using Eq. (10.25), find $\varepsilon(t)$ for an input stress history $\sigma = \sigma_0 \tau H(\tau)$, $\sigma_0 = const$
(e) Using Eqs. (10.27) and (10.28), find $\varepsilon(t)$ for the input stress history listed in part (e).

Problem 10.6
Given: A Maxwell model
Required: Use it to do the following:

(a) Derive Eq. (10.37).
(b) Show that Eq. (10.37) is linear.
(c) Derive Eq. (10.41).
(d) Using Eq. (10.37), find $\sigma(t)$ for an input strain history $\varepsilon = \varepsilon_0 \tau H(\tau)$, $\varepsilon_0 = const$
(e) Using Eqs. (10.41) and (10.42), find $\sigma(t)$ for the input stress history listed in part (e).

Problem 10.7
Given: Eqs. (10.46) and (10.47).
Required:

(a) Derive Eq. (10.49) for a thermoelastic material.
(b) Utilize Eq. (10.49) to predict the strain in a uniaxial bar composed of A36 steel subjected to an axial stress of $\sigma = 0.05$ MPa and a temperature change of $\Delta T = 50\,^{\circ}\text{C}$.
(c) Utilize Eq. (10.49) to predict the strain in a uniaxial bar composed of high strength concrete subjected to an axial stress of $\sigma = 0.05$ MPa and a temperature change of $\Delta T = 50\,^{\circ}\text{C}$.

References

Allen, D. (2014). *How mechanics shaped the modern world*. Springer.
Allen, D., & Haisler, W. (1985). *Introduction to aerospace structural analysis*. Wiley.
Da Vinci, L. (1972). I libri di meccanica. Reconstructed from the original notes by Arturo Uccelli, Kraus Reprint, Nendeln, Liechtenstein.
Jardine, L. (2005). *The curious life of Robert Hooke: The man who measured London*. Harper Perennial.
Leaderman, H. (1943). *Elastic and creep properties of filamentous materials and other high polymers*. Washington, D.C.: The Textile Foundation.
Maxwell, J. (1867). *On the dynamical theory of gases*. London: Philosophical Transactions of the Royal Society.
Parsons, W. (1939). *Engineers and Engineering in the Renaissance*. Cambridge, MA: MIT Press.

Chapter 11
Elasticity and Thermoelasticity

11.1 Introduction

The previous chapter introduced the one-dimensional thermomechanical constitutive behavior of several classes of materials that are necessary in order to capture the observed material behavior of the constituents utilized in flexible road ways (see Chaps. 2–8). In this chapter, the discussion focuses on fully three-dimensional thermomechanical material behavior of elastic and thermoelastic media.

11.2 Multidimensional Linear Elasticity

Within the context of continuum mechanics, the term *elasticity* implies the field of mechanics in which the objective is to predict the mechanical response of an arbitrarily shaped continuous body composed uniquely of elastic material to externally applied loads. This subject seems to have originated with the works of Joseph-Louis Lagrange (1736–1813), Sophie Germain (1776–1831), Claude-Louis Navier (1785–1836), and Augustin-Louis Cauchy (1789–1857) in Paris during the first quarter of the nineteenth century (Todhunter and Pearson 1893; Maugin 2014). As detailed in Chap. 9, the works of Cauchy are particularly noteworthy. As a result of these efforts, by the year 1822 the theory of elasticity was essentially reduced to a problem in applied mathematics.

Unfortunately, prior to the development of additional mathematics (not to mention the electronic computer), it was exceedingly difficult to obtain practical solutions to the elasticity problem. This is due to the fact that the model for elastic media is mathematically complicated. But this all changed in the latter half of the twentieth century, when computational mechanics afforded a methodology (called the finite element method) for solving essentially any linear elasticity problem. Therefore, this section should be considered by the reader to be essentially a history

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_11

lesson, albeit one that is necessary as a prerequisite to a complete understanding of the more advanced models to be discussed later in this text.

The subject of elasticity begins with the definition of a three-dimensional elastic material, that is, a three-dimensional generalization of one-dimensional Eq. (10.5). Thus, in three dimensions *an elastic material is defined to be one in which the stress tensor is a single-valued function of the strain tensor*. Mathematically, this is represented as follows:

$$\sigma_{ij}(\vec{x}, t) \equiv \sigma_{ij}(\varepsilon_{kl}(\vec{x}, t)) \tag{11.1}$$

In the linear case, assuming that the body possesses a stress-free state, the above simplifies to the following:

$$\sigma_{ij} = E_{ijkl}\varepsilon_{kl} \tag{11.2}$$

where E_{ijkl} is a set of 81 *material constants* called *the modulus tensor*. It can be shown that the modulus tensor is in fact a fourth-order tensor, i.e., the modulus tensor transforms from an unprimed Cartesian coordinate system to a rotated (primed) one via the following:

$$E_{p'q'r's'} = a_{p'i}a_{q'j}a_{r'k}a_{s'l}E_{ijkl} \tag{11.3}$$

The coefficients of the modulus tensor can in principle be determined from experiments, analogous to the one-dimensional case discussed in Chap. 10. Thus, utilizing the above constitutive model, together with the theory developed in Chap. 9, it is possible to predict the mechanical response of a three-dimensional linear elastic continuous body.

11.2.1 The Linear Elastic Boundary Value Problem

The model that is produced by adjoining the linear elastic constitutive Eq. (11.2) with the other parts of the model is called *the linear elastic boundary value problem.* The construction of this model begins with Tables 9.6 and 9.7, wherein the list of unknowns and available equations is described in detail. The reader will recall that Table 9.6 listed 25 unknown variables to be predicted as functions of position, $\vec{x}$, and time, t, in the continuous body $V + S$. These are as follows:

Displacement vector	u_i
Strain tensor	ε_{ij}
Stress tensor	σ_{ij}
Body force per unit mass	f_i
Internal energy per unit mass	u
Heat flux vector	q_i

Heat source per unit mass	r
Entropy per unit mass	s
Temperature	T

The prediction of these variables is afforded with the use of the 10 equations and one inequality shown in Table 9.7, plus the linear elastic constitutive Eq. (11.2).

This may seem daunting at first glance, but when dealing with an elastic body the problem simplifies considerably. This is due to the fact that the body is by definition *at constant temperature* in both time and space, thus obviating the necessity to include temperature, T, in the model. Since the temperature is constant, the heat flux vector, $\vec{q}$, and the heat source term, r, must necessarily be identically zero, and it follows that for the linear elastic case the internal energy, u, can be treated as an ancillary variable due to the fact that the conservation of energy simplifies to a means of calculating the internal energy once the stress field, $\sigma_{ij}(\vec{x}, t)$, and strain field, $\varepsilon_{ij}(\vec{x}, t)$, are predicted with the model.

In addition, the entropy, s, will be shown (see below) to be an ancillary variable. This reduces the above total of 25 variables to the following 18 variables: $u_i, \varepsilon_{ij}, \sigma_{ij}$ and f_i. The remaining field equations for a linear elastic body listed in Table 9.7 (along with Eq. (11.2)) are therefore (renumbered) as follows:

Linear Elastic Constitutive Equations:

$$\sigma_{ij} = E_{ijkl}\varepsilon_{kl} \tag{11.2}$$

Strain-Displacement Equations:

$$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{11.4}$$

Equations of Motion:

$$\sigma_{ji,j} + \rho f_i = \rho \frac{d^2 u_i}{dt^2} \tag{11.5}$$

where the rate of change of the displacement has been substituted for the velocity vector, $\vec{v}$.

The above set totals to 15 equations. This would appear at first glance to be three to few equations (15 equations in 18 unknowns). However, a careful examination of the unknown field variables will reveal that the body force vector, f_i, can be predetermined using Newton's gravitational law (Newton 1687) when this is the source of the body force, which is normally the case when dealing with road ways. Thus, the above 15 equations are sufficient to predict the 15 unknowns, $u_i, \varepsilon_{ij}, \sigma_{ij}$, as functions of space and time at all points within a linear elastic body.

In many circumstances, the momentum terms in Eq. (11.5) may be neglected, called *quasi-static*, and in this case the governing field equations need not be supplemented with initial conditions (for the case wherein the loading is constant

in time). Thus, the following information must be known on the boundary, S, called *boundary conditions*:

$$\text{Tractions}: \quad t_i = \hat{t}_i \quad on \quad S_1 \tag{11.6a}$$

$$\text{Displacements}: \quad u_i = \hat{u}_i \quad on \quad S_2 \tag{11.6b}$$

where the symbol $\hat{}$ implies that the quantity is known a priori, and

$$S = S_1 + S_2 \quad S_1 \cap S_2 = \emptyset \tag{11.7}$$

The above set of field equations, together with boundary conditions (11.6a, 11.6b), constitute a so-called *well-posed boundary value problem* (BVP), and this is due to the fact that there exists a solution for the unknown field variables in a linear elastic body for any given set of input *loads*, *geometry*, *and material properties*. Furthermore, because all of the governing equations and the boundary conditions are linear, it can be shown that any solution that satisfies all of the field equations and the boundary conditions is unique, called *the uniqueness theorem* (Timoshenko 1970). This BVP is therefore a very powerful tool for predicting the mechanical response of a linear elastic body, and for this reason the *quasi-static linear elastic boundary value problem* is restated in Table 11.1.

Methods for solving the elastic boundary value problem will be discussed in later sections. Before leaving this section, however, it is perhaps useful to elucidate the subject of boundary conditions. As described in Table 11.1, there are two types of boundary conditions: tractions and displacements. In reality, these are just mathematical approximations of the physics on the boundary of a continuous body. As shown in Fig. 11.1, there are several different possible physical idealizations that may present themselves when two bodies come into contact with one another. For purposes of this discussion, it will be assumed that an object comes in contact with a road way.

As shown in the figure, in case (b) the object (such as a steel road grader) is much stiffer than the road way. In this case, the road way conforms to the shape of the object over the area of the contact surface. The road way, therefore, conforms to the shape of the object, and the boundary conditions applied to the road way by the object are of displacement type.

In case (c), the road way is much stiffer than the object (such as a tire), and the object conforms to the shape of the road way. In this case, the elastic boundary value problem may be first solved for the object (assuming that it is indeed linear elastic!), and the resulting solution may be utilized to calculate the distribution of tractions on the boundary of the object (in this case, the road way). Using Cauchy's lemma, Eq. (9.44), equal and opposite tractions may now be applied by the object to the road way, and these boundary conditions may be used as input for the purpose of predicting the response of the road way.

Table 11.1 Quasi-static linear elastic boundary value problem

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S),\ f_i(\text{in } V+S)$

Geometry: $\vec{x}(\text{on } S),\ n_i(\text{on } S)$

Material Properties: ρ, E_{ijkl}

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}$ **= 15 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = E_{ijkl}\varepsilon_{kl}$	6
Total	**15**

Boundary Conditions:

Tractions: $t_i = \hat{t}_i \quad on \quad S_1$

Displacements: $u_i = \hat{u}_i \quad on \quad S_2$

In case (d), both the object and the road way have comparable moduli, so that it is not possible to solve one or the other problem separately. In this case, both objects must be modeled simultaneously using Table 11.1, with so-called matching conditions (either matching displacements or tractions) where the two objects come into contact with one another. Such a problem is called *an interaction problem*, and in this case the modeler has no choice but to apply *matching conditions* along the surface where the two objects are in contact and solve the two problems simultaneously. This type of problem is usually quite complicated mathematically, and should therefore be avoided if at all possible.

The type of boundary conditions that are physically observed in the application to be solved will necessarily have a significant impact on the mathematical solution method chosen. This topic will be taken up in a later section.

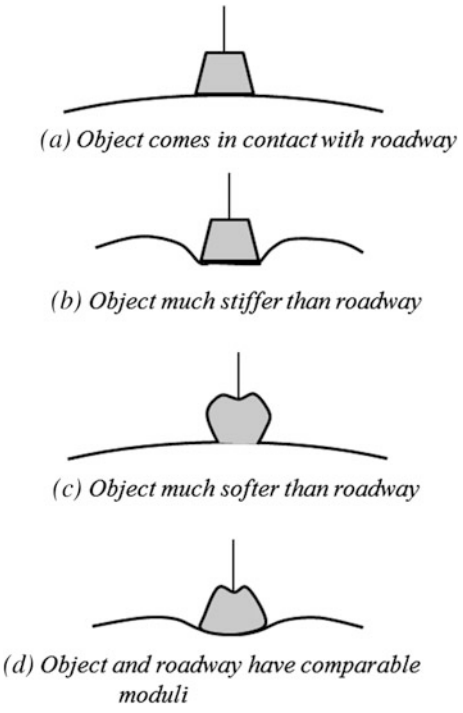


Fig. 11.1 Physics of mechanical boundary conditions

11.2.2 Thermodynamic Constraints on Elastic Material Behavior

In the previous section dealing with elastic material behavior, it was argued (without rigor or proof) that the thermodynamics of elastic media is trivial and therefore unnecessary. This is by and large true, but when dealing with more complex materials, the thermodynamics becomes both significantly more complicated and substantially more useful. Thus, although not much information is gained by studying the thermodynamics of elastic media, it is nonetheless useful to provide a short introduction to this subject.

Accordingly, consider a body that is at spatially and temporally constant temperature. In this case, the conservation of energy, Eq. (9.89), simplifies to the following:

$$\rho \frac{du}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} \tag{11.8}$$

Similarly, the entropy production law, inequality (9.92), simplifies to:

$$\rho \frac{ds}{dt} \geq 0 \tag{11.9}$$

Now, recall the definition of an elastic material, Eq. (11.1). On the basis of this model, it is also assumed that in an elastic material both the internal energy and the entropy are single-valued functions of the strain:

$$u(\vec{x}, t) = u\big(\varepsilon_{ij}(\vec{x}, t)\big) \tag{11.10}$$

and

$$s(\vec{x}, t) = s\big(\varepsilon_{ij}(\vec{x}, t)\big) \tag{11.11}$$

Substituting Eq. (11.10) into Eq. (11.8) and applying the chain rule of differentiation, thus results in the following equation:

$$\rho \frac{\partial u}{\partial \varepsilon_{ij}} \frac{d\varepsilon_{ij}}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} \tag{11.12}$$

The above equation may be recast in the following useful form:

$$\left[\sigma_{ij}(\varepsilon_{kl}) - \rho \frac{\partial u}{\partial \varepsilon_{ij}}(\varepsilon_{kl})\right] \frac{d\varepsilon_{ij}}{dt} = 0 \tag{11.13}$$

It is apparent that the terms inside the brackets in Eq. (11.13) depend on the components of the strain tensor, whereas the term outside the brackets depends only on the time derivative of the strain tensor, which is independent of the strain tensor at any given time. Equation (11.13) thus leads to the conclusion that the only way it can be true for all conditions is for the term in brackets to be identically zero, that is, for an elastic material:

$$\sigma_{ij} = \rho \frac{\partial u}{\partial \varepsilon_{ij}} \tag{11.14}$$

The above implies that there exists *a potential function* for the stress tensor in terms of the strain tensor, and that potential function is precisely the internal energy, $u = u(\varepsilon_{ij})$. As will be seen shortly, this revelation is a useful result.

As described in Eq. (11.14), the internal energy is a potential function for the stress tensor in an elastic material. In the case of a linear elastic material, this implies that the internal energy must be a second-order function of the strain tensor. Thus, suppose that the internal energy is expanded in the strain as follows:

$$u = A + B_{ij}\varepsilon_{ij} + \frac{1}{2}E_{ijkl}\varepsilon_{ij}\varepsilon_{kl} \tag{11.15}$$

Since energy may be assumed to be relative, the leading coefficient, A, is arbitrary and may therefore be set to zero. The above may be substituted into Eq. (11.14) with the following result:

$$\sigma_{ij} = \sigma_{ij}^{R} + E_{ijkl}\varepsilon_{kl} \tag{11.16}$$

where $\sigma_{ij}^{R} \equiv B_{ij}$ is called *the residual stress*, since it is the stress predicted when the input strain is zero. Equation (11.16) is therefore a generalization of Eq. (11.2) that accounts for residual stress.

Consider a similar procedure applied to the entropy production inequality. Substituting (11.11) into (11.9) and applying the chain rule of differentiation, thus results in the following inequality:

$$\left[\frac{\partial s}{\partial \varepsilon_{ij}}(\varepsilon_{kl})\right]\frac{d\varepsilon_{ij}}{dt} \geq 0 \tag{11.17}$$

Once again noting that the terms outside the brackets do not depend on the terms inside the brackets, the above inequality leads to the conclusion that the terms inside the brackets must be zero:

$$\frac{\partial s}{\partial \varepsilon_{ij}} = 0 \tag{11.18}$$

Equation (11.18) together with Eq. (11.11) now leads to the important conclusion that the entropy, s, in an elastic body is constant for all times. This is a rather profound conclusion.

It should be apparent to the reader that in order for a road way to fail due to any cause whatsoever, some entropy must be generated within the road way. And yet, the elastic material model denies the possibility that this can occur. Thus, it can be concluded that an elastic material model is not sufficient for the purpose of predicting failure of road ways. On the contrary, in order to predict failure of road ways, there must be some means of entropy generation included within the model. This important observation will be elaborated upon in later sections.

11.2.3 Material Symmetry

Linear elastic constitutive Eq. (11.2) implies that the relationship between the stress and strain contains 81 (3^4) coefficients since E_{ijkl} is a fourth-order tensor. Now, an interesting question comes to mind—Are there actually 81 *unique* coefficients in the

modulus tensor? If so, there may be quite a few experiments required to populate the modulus tensor for a given linear elastic material. This is indeed a very good question.

As will be shown below, the number of unique coefficients in the modulus tensor will depend on the physical symmetry observed within the material microstructure.

11.2.3.1 Anisotropic Material Behavior

When the material possesses no observable physical symmetry at the microscale, it is termed *anisotropic*. To determine the number of unique coefficients in the modulus tensor, first note that since the stress tensor is symmetric (Eq. (9.85)). It follows from Eq. (11.2) that:

$$\sigma_{ij} = \sigma_{ji} \Rightarrow E_{ijkl}\varepsilon_{kl} = E_{jikl}\varepsilon_{kl} \Rightarrow E_{ijkl} = E_{jikl} \Rightarrow 54 \text{ unique coefficients} \tag{11.19}$$

Next, note from Eq. (9.40) that the strain tensor is also symmetric, and it follows that:

$$\varepsilon_{kl} = \varepsilon_{lk} \Rightarrow E_{ijkl}\varepsilon_{kl} = E_{ijlk}\varepsilon_{kl} \Rightarrow E_{ijkl} = E_{ijlk} \Rightarrow 36 \text{ unique coefficients} \tag{11.20}$$

Now, consider Eq. (11.15) once again. The second derivative of this equation with respect to the strain, therefore, results in the following:

$$\frac{\partial^2 u}{\partial\varepsilon_{ij}\partial\varepsilon_{kl}} = E_{ijkl} \Rightarrow E_{ijkl} = E_{klij} \Rightarrow 21 \text{ unique coefficients} \tag{11.21}$$

It thus follows from Eqs. (11.19)–(11.21) that for the most anisotropic material possible there are only 21 unique coefficients in the modulus tensor, E_{ijkl}, thereby implying that the experimental determination of the material constants will be significantly less cumbersome than that presupposed as first glance.

For purposes of further mathematical simplification, it is useful to introduce a simplified notation herein. This notation, called Voigt notation, is defined as follows:

$$\begin{array}{llll} \sigma_1 \equiv \sigma_{11} & \sigma_4 \equiv \sigma_{23} & \varepsilon_1 \equiv \varepsilon_{11} & \varepsilon_4 \equiv 2\varepsilon_{23} \\ \sigma_2 \equiv \sigma_{22} & \sigma_5 \equiv \sigma_{13} & \varepsilon_2 \equiv \varepsilon_{22} & \varepsilon_5 \equiv 2\varepsilon_{13} \\ \sigma_3 \equiv \sigma_{33} & \sigma_6 \equiv \sigma_{12} & \varepsilon_3 \equiv \varepsilon_{33} & \varepsilon_6 \equiv 2\varepsilon_{12} \end{array} \tag{11.22}$$

where it should be noted that the tensorial (meaning two subscripts) shear strains have been doubled. While this notational change introduces mathematical simplification, improper usage can lead to erroneous model predictions. Thus, this notation should be used with caution.

Using the above notation, Eq. (11.2) may now be written as follows:

$$\sigma_\alpha = E_{\alpha\beta}\varepsilon_\beta \tag{11.23}$$

where, due to Eqs. (11.19)–(11.22), it follows that:

$$E_{\alpha\beta} = \begin{bmatrix} E_{1111} & E_{1122} & E_{1133} & E_{1123} & E_{1113} & E_{1112} \\ E_{1122} & E_{2222} & E_{2233} & E_{2223} & E_{2213} & E_{2212} \\ E_{1133} & E_{2233} & E_{3333} & E_{3323} & E_{3313} & E_{3312} \\ E_{1123} & E_{2223} & E_{3323} & E_{2323} & E_{2313} & E_{2312} \\ E_{1113} & E_{2213} & E_{3313} & E_{2313} & E_{1313} & E_{1312} \\ E_{1112} & E_{2212} & E_{3312} & E_{2312} & E_{1312} & E_{1212} \end{bmatrix} \tag{11.24}$$

thus, demonstrating in symmetric matrix form the 21 unique coefficients in the modulus tensor for the most anisotropic material possible.

There are very few practical circumstances wherein materials are completely anisotropic. At least at the continuum scale, nearly all materials exhibit some degree of symmetry that simplifies the number of unique coefficients contained within the modulus tensor still further.

11.2.3.2 Linear Elastic Orthotropic Material

To see how these simplifications can be obtained, it is instructive to consider an example. Suppose that a material (such as an asphalt mixture (see Chap. 7)) that is homogeneous to the naked eye is actually composed of tiny aligned ellipsoidal inclusions embedded within a spatially homogeneous and isotropic binder. When viewed within a microscope, the material then appears something like the cartoon depicted in Fig. 11.2. Suppose furthermore that a Cartesian coordinate system is constructed that coincides with the principle axes of the ellipsoids, as shown in the figure. Finally, suppose that there is equal probability of finding one of the inclusions everywhere within the object (called *statistically homogeneous*), so that the material is macroscopically spatially homogeneous. Note that if the dimension of the inclusions is extended to infinity in any of the coordinate directions, the inclusions become continuous fibers that are elliptic in cross section.

It should be apparent that at the macroscopic scale the material will possess three mutually perpendicular planes of reflective symmetry, and this is in fact the definition of an orthotropic material. As shown in Fig. 11.2, the coordinate axes have been constructed so that the planes of reflective symmetry are normal to the coordinate axes shown in the figure. As a result of this observed reflective symmetry, the macroscopic material properties will be unaffected by a rotation of 180° about any one of the coordinate axes. For example, consider a rotation of 180° about the x_3-axis, as also shown in the figure. The direction cosine tensor Eq. (9.12) relating these two coordinate systems is given by:

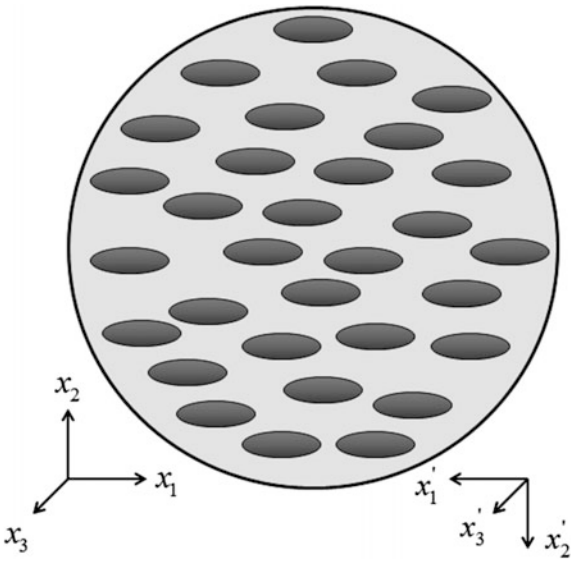


Fig. 11.2 Cartoon of a microscopic image of a continuum composed of aligned ellipsoids embedded within a homogeneous and isotropic binder

$$a_{p'i} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{11.25}$$

Substituting the above into the second-order tensor transformation for stress and strain (similar to Eq. (9.15)), and this result into Eq. (11.2) will thus result in the following:

$$E_{2311} = E_{2322} = E_{2333} = 0 \tag{11.26}$$

Similarly, rotations about the x_1 and x_2 coordinate axes will result in the following:

$$E_{1211} = E_{1222} = E_{1233} = E_{1311} = E_{1322} = E_{1333} = 0 \tag{11.27}$$

The above *symmetry constraints* thereby reduce the number of unknown coefficients in the modulus tensor from 21 to 11.

It can be seen that the above symmetries also preclude coupling of shear strains in orthotropic media, thus it also follows that:

$$E_{2313} = E_{2312} = E_{1312} = 0 \tag{11.28}$$

The above results thereby reduce the number of unknown coefficients from 12 to 9. Thus, substituting Eqs. (11.26) through (11.28) into Eq. (11.24) will result in the following modulus matrix for a linear elastic orthotropic material:

$$E_{\alpha\beta} = \begin{bmatrix} E_{1111} & E_{1122} & E_{1133} & 0 & 0 & 0 \\ E_{1122} & E_{2222} & E_{2233} & 0 & 0 & 0 \\ E_{1133} & E_{2233} & E_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & E_{2323} & 0 & 0 \\ 0 & 0 & 0 & 0 & E_{1313} & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{1212} \end{bmatrix} \tag{11.29}$$

It is possible to create a road way material that is macroscopically orthotropic by a variety of means. For example, rolling the road way during curing can align the aggregate in such a way as to produce macroscopically orthotropic material (see Chap. 8). While this is certainly possible, the effect is normally neglected in predictive models. Instead, the model is simplified still further, as described below.

11.2.3.3 Linear Elastic Transversely Isotropic Material

Consider Fig. 11.2 once again. Suppose that the aligned ellipsoids are circular in cross section normal to the x_1-axis. In this case, the material properties will be independent of coordinate direction in the $x_2 - x_3$-plane. Such a material is termed *transversely isotropic*, and in this case the modulus matrix will reduce to the following form containing five unknown coefficients:

$$E_{\alpha\beta} = \begin{bmatrix} E_{1111} & E_{1122} & E_{1122} & 0 & 0 & 0 \\ E_{1122} & E_{2222} & E_{2233} & 0 & 0 & 0 \\ E_{1122} & E_{2233} & E_{2222} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{E_{2222} - E_{2233}}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & E_{1212} & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{1212} \end{bmatrix} \tag{11.30}$$

Roadway materials that are reinforced with aligned bars of circular cross section will exhibit macroscopically transversely isotropic material behavior.

11.2.3.4 Linear Elastic Isotropic Material

A material in which *the material properties at a fixed material point are invariant with respect to coordinate rotation* is called *isotropic*. Such a circumstance occurs (macroscopically) in cementitious road way materials in which both the fines and the aggregate are particulates that are randomly oriented.

It follows that for linear elastic isotropic media:

$$E_{i'j'k'l'} = E_{ijkl} \tag{11.31}$$

Since the linear elastic modulus tensor is a fourth-order tensor, it transforms according to Eq. (11.3). Thus, substituting Eq. (11.31) into Eq. (11.3) will result in the following:

$$E_{ijkl} = \frac{1}{3}(G_2 - G_1)\delta_{ij}\delta_{kl} + \frac{1}{2}G_1\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \tag{11.32}$$

where G_1 and G_2 are the only two unknown coefficients for linear elastic isotropic media. Substitution of Eq. (11.32) into the linear elastic constitutive model (11.2) will result in the following form for an isotropic material:

$$\sigma_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + 2\mu\varepsilon_{ij} \tag{11.33a}$$

where

$$\lambda \equiv \frac{G_2 - G_1}{3}, \quad \mu \equiv \frac{G_1}{2} \tag{11.33b}$$

The above equations are called *the Lamé equations* (after Gabriel Lamé 1795–1870), and the two material properties are called *the Lamé coefficients*. In fact, in the nineteenth century it was thought by much of the scientific community that there was only one unique coefficient in the modulus tensor for the special case when the material is *isotropic* (Love 1906). This was due at least in part to the fact that prior to the invention of electronic equipment, and it was difficult to measure properties (such as Poisson's ratio) accurately.

Equation (11.33a, 11.33b) may be written in matrix form as follows:

$$E_{\alpha\beta} = \begin{bmatrix} E_{1111} & E_{1122} & E_{1122} & 0 & 0 & 0 \\ E_{1122} & E_{1111} & E_{1122} & 0 & 0 & 0 \\ E_{1122} & E_{1122} & E_{1111} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{E_{1111}-E_{1122}}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{E_{1111}-E_{1122}}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{E_{1111}-E_{1122}}{2} \end{bmatrix} \tag{11.34}$$

where

$$E_{1111} = \lambda + 2\mu \quad E_{1122} = \lambda \tag{11.35}$$

Linear elastic material behavior may just as easily be represented by the following so-called *stress formulation*:

$$\varepsilon_{ij} = D_{ijkl}\sigma_{kl} \tag{11.36}$$

where it should be apparent that material symmetry constraints will result in similar forms for the compliance tensor, D_{ijkl}, and associated matrix as those reported above for the modulus tensor and matrix. Since the material is linear and the stress–

strain relation is algebraic, there exists a unique mapping between the stress and strain formulations for linear elastic media.

The isotropic linear elastic constitutive equations are reported in the literature in a variety of ways, but it should be noted that *there are always two unique material coefficients* in linear elastic isotropic media. For example, consider the uniaxial test described in Fig. 10.2. If the lateral normal component of strain, ε_{22}, is measured in this test, then it is common to determine *Poisson's ratio* (after Siméon Denis Poisson 1781–1840) as follows:

$$v \equiv \frac{-\varepsilon_{22}}{\varepsilon_{11}} \tag{11.37}$$

The above coefficient, together with Young's modulus, E, obtained from the same experiment, results in two independent material properties that can be obtained from this simple experiment. They are often called *the engineering coefficients*. It can be shown by substitution into Eq. (11.33a, 11.33b) that the compliance matrix for a linear elastic isotropic material may be written in terms of the engineering coefficients as follows:

$$D_{\alpha\beta} = \frac{1}{E}\begin{bmatrix} 1 & -v & -v & 0 & 0 & 0 \\ -v & 1 & -v & 0 & 0 & 0 \\ -v & -v & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2(1+v) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(1+v) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(1+v) \end{bmatrix} \tag{11.38}$$

Inversion of the above matrix will result in the following alternate form of Eq. (11.34):

$$E_{\alpha\beta} = \frac{E}{(1+v)(1-2v)}\begin{bmatrix} 1-v & v & v & 0 & 0 & 0 \\ v & 1-v & v & 0 & 0 & 0 \\ v & v & 1-v & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2v}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2v}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2v}{2} \end{bmatrix} \tag{11.39}$$

Thus, there are several different ways of writing the modulus and compliance for isotropic linear elastic media and, just as a reminder, these are all equivalent to one another.

11.2.3.5 Material Property Characterization

In the case described above, the engineering coefficients are deployed for the case wherein the material is easily tested using a single uniaxial test, as described in

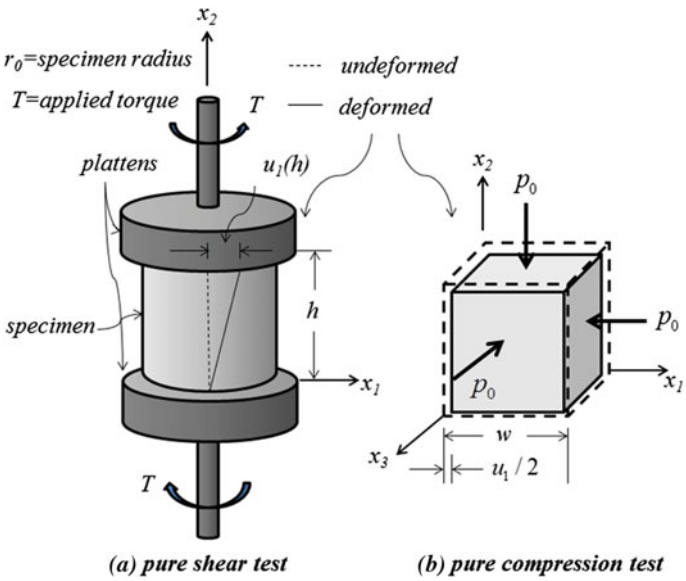


Fig. 11.3 Two tests for obtaining material properties for isotropic linear elastic material

Fig. 10.2. However, in some materials, such as asphalt and road way base materials, it is often impractical to perform a uniaxial tension or compression test because the material cannot sustain significant uniaxial tensile or compressive loading. In this case, one possibility is to perform two tests: a pure shear test and a pure compression test, as depicted in Fig. 11.3.

These two tests are described as follows:

(a) pure shear test:

$$\varepsilon_{12} = \dot{\varepsilon}^0_{12} t \quad \dot{\varepsilon}^0_{12} = const \quad \varepsilon_{ij} = 0 \quad \forall\, other i, j \tag{11.40}$$

where it can be shown that (Allen 2013)

$$\varepsilon_{12} = \frac{u_1(h)}{2h} \quad \sigma_{12} = \frac{2T}{\pi r_0^3} \tag{11.41}$$

where the quantities in Eq. (11.41) are defined in Fig. 11.3a. In this test, *the shear modulus*, G, is defined as follows:

$$G \equiv \frac{\sigma_{12}}{2\varepsilon_{12}} \tag{11.42}$$

(b) pure compression test:

$$\sigma_{11} = \sigma_{22} = \sigma_{33} = -p_0 \quad \sigma_{23} = \sigma_{13} = \sigma_{12} = 0 \tag{11.43}$$

where p_0 is called *the hydrostatic pressure*, with the result that (assuming that the specimen is a cube):

$$\varepsilon_{11} = \varepsilon_{22} = \varepsilon_{33} = \frac{u_1}{w} \tag{11.44}$$

where the quantities are defined in Fig. 11.3b. In this test, *the bulk modulus*, K, is defined as follows:

$$K \equiv -\frac{p_0}{\varepsilon_{kk}} \tag{11.45}$$

A comparison of Eqs. (11.34), (11.35), (11.36), and (11.42) will reveal the following:

$$\begin{array}{ll} G = \mu = \frac{G_1}{2} & \mu = G \\ \Leftrightarrow & \\ K = \frac{G_2}{3} = \frac{3\lambda + 2\mu}{3} & \lambda = K - \frac{2G}{3} \end{array} \tag{11.46}$$

Equations (11.46) may be substituted into Eq. (11.33a, 11.33b) to obtain the following equivalent tensorial representation of a linear elastic isotropic material:

$$\sigma_{ij} = \left(K - \frac{2G}{3}\right)\varepsilon_{kk}\delta_{ij} + 2G\varepsilon_{ij} \tag{11.47}$$

where K and G are sometimes called *the geologic coefficients*. Matrix Eqs. (11.39) may be substituted into Eqs. (11.23) and compared to Eq. (11.47), with the result that

$$\begin{array}{ll} G = \frac{E}{2(1+\nu)} & E = \frac{2G(3K+2G)}{3K} \\ \Leftrightarrow & \\ K = \frac{E}{3(1-2\nu)} & \nu = \frac{(3K-2G)}{2(3K-4G)} \end{array} \tag{11.48}$$

The above relations may be substituted into Eq. (11.39) to produce an equivalent representation of the linear elastic isotropic modulus matrix using the geologic coefficients.

Numerous other forms of the linear elastic isotropic constitutive equations exist in the open literature. These forms are normally based on the accessibility of the equations to readily available material properties for the material in question. Nevertheless, it is important to note that in all cases wherein the material is isotropic

and linear elastic, there are only two unique material properties. A few have been listed here. The choice as to which should be used is nothing more than a matter of expediency, as all are equivalent to one another.

11.2.4 Solution Techniques for the Linear Elastic Boundary Value Problem

As mentioned previously, the essential formulation of the linear elastic boundary value problem was completed by 1822 (Todhunter and Pearson 1893; Allen 2013, 2014; Maugin 2014). Thereafter, mathematicians and engineers turned their attention to the challenge of obtaining mathematical solutions to problems with specific loads, geometric shapes, and material properties. This turned out to be a challenging task, as the mathematical problem described in Table 11.1 was (and still is) a formidable one. Despite the fact that the governing equations for a linear elastic material are all mathematically linear (see Chap. 9), there are nonetheless 15 equations in 15 unknowns, and nine of the equations are differential equations. Thus, mathematicians quickly realized that these equations are quite complicated.

Perhaps the first mathematician to solve a variety of useful elasticity problems was Adhémar Jean Claude Barré de Saint-Venant (1797–1886). He recognized that under certain circumstances the mathematical model could be simplified to a more practical set of equations from whence analytic solutions were possible. Much of this early work focused on one of two simplifications: reducing the mathematical dimensionality of the equations; and reducing the number of primary equations via boundary conditions. Each of these simplifications will be discussed briefly below.

11.2.4.1 Reducing the Mathematical Dimensionality of the Equations

Solution of problems involving the full 15 equations in 15 unknowns poses an enormous challenge simply due to the number of equations involved. Thus, it is pragmatic to reduce this number of governing equations wherever possible. Perhaps the most obvious simplification results when the state variables depend on only one or two coordinate directions rather than that encountered in the fully three-dimensional case. For example, suppose that the unknowns depend only on the x_1 and x_2 coordinates. Such a circumstance occurs when the object of interest is homogeneous and prismatic in the x_3 coordinate direction, and the externally applied loads and boundary conditions are independent of x_3, so that the object can be idealized as two-dimensional, as shown in Fig. 11.4 (see Chap. 9).

In this type of problem, the field variables can be shown to reduce to the following set of eight primary variables: $u_1, u_2, \varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{xy}, \sigma_{xx}, \sigma_{yy}, \sigma_{xy}$.

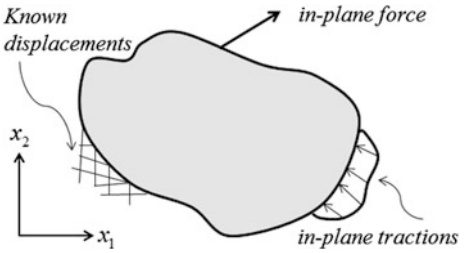


Fig. 11.4 A mathematically two-dimensional object

If the out-of-plane stress component, σ_{zz}, is zero, then the out-of-plane component of displacement, u_3, as well as the out-of-plane normal strain component, ε_{zz}, can be calculated a posteriori given this fact. Thus, the governing field equations shown in Fig. 11.1 will reduce to eight primary equations in the above eight unknowns. Such a problem is termed *a plane stress boundary value problem* because all of the nonzero components of stress in the object lie in the $x_1 - x_2$-plane.

A similar case, called *plane strain* occurs when motions in the out-of-plane direction are constrained to zero (or are constant in x_3), leading to the result that the out-of-plane displacement and strain are zero, but the out-of-plane component of stress is not zero. In this case, a similar set of eight primary equations results.

A slightly more complicated physical circumstance occurs when the object shown in Fig. 11.4 is subjected to an evenly distributed out-of-plane loading in addition to the in-plane loading, and this case is termed *generalized plane strain.* All three of these special cases lead to dramatic mathematical simplifications in the complexity of the resulting boundary value problem, and there are many practical circumstances wherein they may be used to great effect. For example, the plane of a road way normal to the direction of travel is constrained against motion in the direction of travel (assuming that the road way is macroscopically homogeneous and cracks do not occur), so that the plane strain assumption is often utilized to great effect in this circumstance.

Another means of reducing the mathematical dimensionality of a boundary value problem is to utilize non-Cartesian coordinates. Although this technique introduces certain mathematical complications, it may nonetheless simplify the problem. For example, if the object of interest is shaped like a cylinder, and the loads and material properties are *axisymmetric* (meaning constant along the curvilinear surface normal to the cylindrical axis), then the problem may be solved in *cylindrical coordinates*, as shown in Fig. 11.5. In this case, the solution would necessarily have to be three-dimensional in Cartesian coordinates, whereas it is two-dimensional in cylindrical coordinates.

In fact, a generalization of the problem shown in Fig. 11.5, the case of a prismatic rod subjected to self-equilibrating end torques, was one of the most important solutions obtained by Saint-Venant (1856). There are other curvilinear coordinate systems that are capable of reducing the mathematical dimensionality of a boundary

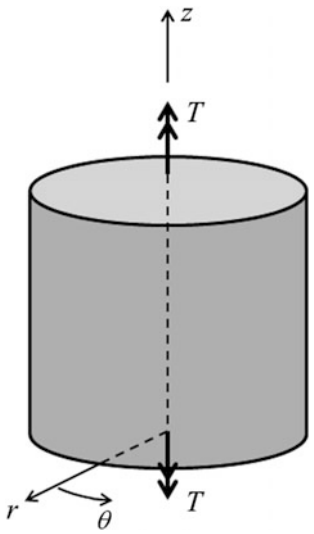


Fig. 11.5 An axisymmetric boundary value problem

value problem, and these can be used to great effect when attempting to solve elasticity problems.

11.2.4.2 Reducing the Complexity via Boundary Conditions

As mentioned earlier in this chapter, boundary conditions in elasticity problems occur in two different physical ways (see Fig. 11.1). These are either of displacement or traction type, and the solution method that will be propitious will depend on which type of boundary conditions occurs in the problem of interest.

Dirichlet Problem

The simplest mathematical case occurs when the entire boundary, S, is subjected to displacement boundary conditions. In this case, strain-displacement Eq. (11.4) may be substituted into isotropic constitutive Eqs. (11.33a, 11.33b), and this results into equations of motion (11.5) to produce *the Navier equations*, a version of which was first reported by Claude-Louis Navier (1785–1836) in 1821 (Malvern 1969):

$$(\lambda + \mu)u_{j,\,ji} + \mu u_{i,\,jj} + \rho f_i = \rho \frac{d^2 u_i}{dt^2} \tag{11.49}$$

The above equations, together with the displacement boundary conditions, comprise a set of three equations in the three unknown displacement components, $u_i = u_i(\vec{x}, t)$. As such, they comprise a well-posed boundary value problem. Once

they are solved, the strain may be computed using Eq. (11.4), and the stresses may be determined using Eq. (11.2). In the case wherein the motions may be assumed to be quasi-static, the right-hand side of Eq. (11.49) is negligible, thereby further simplifying the problem.

The above problem, called *the Dirichlet problem* (after Peter Gustav Lejeune Dirichlet 1805–1859), is perhaps the most direct formulation of the three-dimensional elasticity problem. Unfortunately, it is rarely of use in road way applications. This is due to the fact that road ways often come into contact with objects that are softer than the road way material itself (such as tires), thereby making it unlikely that the road way may be modeled accurately with displacement boundary conditions.

Neumann Problem

The second mathematical simplification due to boundary conditions occurs when the entire boundary of the object is subjected to traction boundary conditions, termed *the Neumann problem* (after John von Neumann 1903–57). This problem is significantly more complicated than the Dirichlet problem. In addition to the fact that there are six primary unknown stress components, due to the fact that the boundary conditions are all of traction type, there is a uniqueness issue that must be dealt with. To deal with this anomaly, an arbitrary line integral of the displacement field within the object is constructed (Cesaro 1906). In order for this line integral to be unique, the following conditions must be met (Cesaro 1906; Fung 1965; Amrouche et al. 2006):

$$\varepsilon_{il,\,jk} + \varepsilon_{jk,\,il} - \varepsilon_{ik,\,jl} - \varepsilon_{jl,\,ik} = 0 \quad \forall i,j,k,l \tag{11.50}$$

The above equations are called *the Saint-Venant strain compatibility equations* (after Barré de Saint-Venant, who provided the first derivation of them in 1860, although they were proven conclusively by Eugenio Beltrami (1835–99) in 1886). Though somewhat daunting on first glimpse, there are actually only six unique equations represented by (11.50).

These may be substituted into the Lamé equations to produce *the Beltrami-Michel stress compatibility equations*:

$$\sigma_{ij,\,kk} + \frac{2(\lambda+\mu)}{(3\lambda+2\mu)}\sigma_{kk,\,ij} + \rho f_{i,\,j} + \rho f_{j,\,i} + \frac{\lambda}{(\lambda+2\mu)}\rho f_{k,\,k}\delta_{ij} = 0 \tag{11.51}$$

The above represent three independent equations. These equations, together with Eq. (11.5) (quasi-static case), represent six equations in terms of the six unknown components of the stress tensor. Thus, these six equations, together with Cauchy's formula Eq. (10.48), may be used to solve for the six primary unknown stress components within the object. Thereafter, the strains may be determined by using Eq. (11.36), and the displacements may be determined by integrating Eq. (11.4).

Both the Dirichlet and the Neumann problem represent significant simplifications of the three-dimensional elasticity boundary value problem. Nonetheless, mathematicians encountered great difficulty in solving practical elasticity problems in the nineteenth century. One strategy was to *both* reduce the dimensionality *and* solve either the Dirichlet or the Neumann problem, in which case the number of equations was reduced still further. Despite this approach, solutions to many problems continued to elude mathematicians and engineers.

11.2.4.3 Analytic Solution Methods

By the middle of the nineteenth century, mathematicians had succeeded in obtaining solutions to only a small number of elasticity problems. The difficulty centered around one of two issues: either the boundary of the object in question was irregular or the loading was applied irregularly in the coordinate dimensions. In either case, analytic solutions were slow in coming. However in 1862 George Biddell Airy (1801–92) proposed a technique for solving the Neumann problem that is today called *the Airy stress function*. This technique essentially transforms the elasticity problem into a so-called dual space in which the physics of the problem are lost, but the mathematics are somewhat simpler. Once the solution is obtained in this dual space, it can be transformed back into the physical space, thereby determining the unknown stresses, strains, and displacements.

This technique quickly gained momentum, and it became one of the most utilized methods for obtaining analytic solutions to elasticity problems, spawning a number of other dual space methods along the way.

Another method for obtaining approximate, but nonetheless analytic solutions emanated from the variational calculus invented by the Bernoulli brothers (Johann 1667–1748 and Jacob 1654–1705) and Leonhard Euler (1707–83) in the early eighteenth century, expanded by Joseph-Louis Lagrange (1736–1813), and deployed in the field of elasticity by Lord Rayleigh (John William Strutt 1842–1919), among others, in the latter part of the nineteenth century.

Still another method utilized was based on making certain kinematically based simplifying assumptions that are based on experimental observations. This approach seems to have originated once again with the Bernoulli brothers Jacob and Johann. Focusing on the cantilever beam problem posed earlier by Galilei (1638), these two mathematicians proposed the first theory of beams. Although their approach was incomplete, Johann's son Daniel Bernoulli (1700–1782) later took up the problem with Leonhard Euler, the pair producing the Euler–Bernoulli theory of beams (Euler 1744). This method, which is still in use by practicing engineers today, takes an ingenious approach to an otherwise intractable problem in mathematics by proposing an *approximation* of the exact linear elasticity problem for which an *exact* solution may be obtained. This approach has been used to great extent right up to the present time, resulting in the field of mechanics termed *structural mechanics*.

However, all of these methods would eventually fall by the wayside with the invention of computational methods in the latter half of the twentieth century.

11.2.4.4 Computational Methods

In the middle part of the twentieth century, a confluence of three disparate activities began to come together in what would produce a dramatic improvement in the ability to model the mechanics of deformable bodies. First, and perhaps foremost, the world was at war for much of the first half of the twentieth century. Technological advances in infrastructure, ground-based transport, aviation, rocketry, and even space flight drove the scientific community to develop more advanced means of modeling the linear elastic boundary value problem. Simultaneously, mathematicians such as Richard Courant (1888–1972) and engineers such as Alexander Hrennikoff (1896–1984) were developing approximate methods for solving sets of differential equations. Serendipitously, computer scientists kept up a frenetic pace of computer developments throughout the twentieth century. These three efforts converged in the late 1950s, resulting in the development of the now-familiar *finite element method* (coined by Ray William Clough (b. 1920) in 1960). This approach to solving sets of differential equations was actually first applied to problems in solid mechanics, including the linear elasticity problem.

The finite element method in a sense takes the mathematically opposite approach to what had previously been taken with the structural mechanics approach. Whereas *structural mechanists formulated an approximation of the linear elastic BVP wherein exact solutions were achievable*, *the finite element method formulates a method for obtaining approximate solutions to the exact linear elasticity BVP*. This new method, accompanied by the rise of the high-speed computer has now supplanted all other methods for solving boundary value problems. And while so-called exact analytic solutions are still sought in certain special circumstances, the need for such solutions is rare. Thus, the finite element method will be covered in some detail later in this text (see Chap. 14).

11.2.5 Micromechanics

One of the seminal technological advancements of the twentieth century was the development of components made from two chemically and physically different materials that are mechanically (as opposed to chemically) combined so as to improve the structural performance of the *composite material*. Although this concept actually dates to ancient times (Allen 2014), it has been explored scientifically to full advantage only in recent times.

An example is Portland cement concrete, in which aggregate is introduced into the mastic for the purpose of increasing the overall stiffness of the composite concrete. A similar approach is used with asphalt concrete (see Chap. 7). In fact,

this approach is now commonplace in our world, having been adapted for use not only in civil engineering infrastructures, but also in aerospace, manufacturing, automotive, and even sports applications, to name a few (Allen 2014).

One barrier to deployment of such composite materials (aside from processing cost) is the determination of exactly what mixture of materials will optimize the structural performance of the resulting composite (see Chap. 7). One (although certainly not the only) measure of performance is the composite modulus tensor that results from the combination of the disparate materials involved in creating the composite. Conceptually speaking, the determination of the composite modulus tensor resulting from the mechanical combination of two or more linear elastic materials involves the solution and subsequent homogenization of a heterogeneous linear elastic boundary value problem. However, solving such problems has until recent times proven to be exceedingly complicated due to the stress and strain concentrations introduced by the irregularity of the phase geometry therein.

Accordingly, engineers have adopted a simplified technique in which the material is assumed to be macroscopically homogeneous. Consider, for example, the two-phase composite shown in Fig. 11.2. Certainly, one way of determining the modulus tensor of the resulting composite is to fabricate the material and then perform constitutive experiments on statistically homogeneous specimens such as that described earlier in this chapter. Unfortunately, the process of determining the composite modulus as a function of the volume fraction of the particles will necessarily require that a number of specimens be tested, each with a different volume fraction of embedded particles (or fibers). This process could in fact become quite expensive, since both the processing and the experiments could be cumbersome.

With the increased development of composites after World War II, numerous mathematical techniques were developed for predicting the average modulus of the composite resulting from the combination of two or more materials into a single composite. All of these methods are essentially applications of linear elasticity theory. For example, suppose that the response of the composite shown in Fig. 11.2 is to be predicted for the uniaxial test shown in Fig. 10.2. Whereas the tractions on the boundary of the specimen may be applied spatially homogeneously, it does not follow that the displacements, or the stresses for that matter, will be spatially homogeneous within the specimen. Thus, a rather complicated boundary value problem results, in which the particles cause *stress concentrations*. In cases, wherein mathematicians and engineers have been able to obtain analytic solutions for the stresses, strains, and displacements, these results are explicit functions of the volume fraction of the particles, and the resulting stress and strain may be boundary averaged, called *homogenization* because the resulting properties are equivalent to the macroscopic properties of a homogeneous material, to obtain the average moduli of the composite (such as asphalt with additives (see Chap. 6) and/or asphalt mixtures (see Chap. 7)). When possible, this approach obviates the necessity to perform laboratory tests for the purpose of determining the average composite modulus as a function of the volume fraction and shape of the aggregate.

Although the terminology is not very accurate (or satisfying), this field of applied elasticity is often called *micromechanics*. While much effort has been focused on obtaining analytic solutions to such heterogeneous elasticity problems, this has proven to be challenging when the geometry of the imbedded particles and/or fibers is complicated. Accordingly, micromechanics techniques often utilize analytic solutions for approximate geometric configurations.

As will be discussed in some detail in Chap. 15, the finite element method may also be used for this purpose. This has the advantage that the geometry of the inclusions can be described quite precisely, but it is hindered by the disadvantage that explicit equations relating the modulus to the inclusion geometry are not possible when the finite element method is employed. Still, once the composite moduli have been predicted by this method, it is possible to use the finite element method yet again to predict the response of a structural component made of this composite by assuming that it is macroscopically homogeneous. This is perhaps the origin of the term *micromechanics*, although the procedure indeed has little or nothing to do with the scale of the particles utilized to make the composite.

To see how this process may be formalized for the purpose of performing stress analysis of a heterogeneous structural component, consider the heterogeneous object shown in Fig. 11.6.

For convenience, the coordinate directions for the structural component have been superscripted with the letter G, implying that the coordinate system is *global*. Note that a small parallelepiped of the structural component has been expanded and

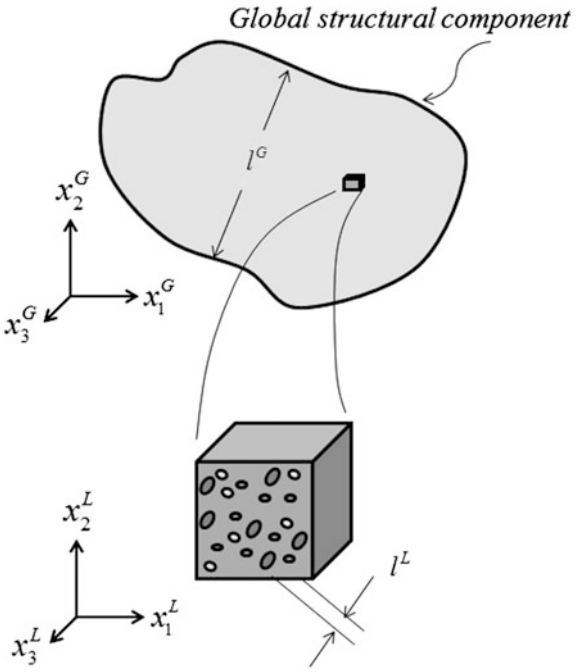


Fig. 11.6 Depiction of a heterogeneous structural component

shown below the global depiction, and the coordinate directions associated with this parallelepiped are superscripted with the letter L, implying that the coordinate system assigned to the parallelepiped is *local*.

It is apparent from the figure that if the length scale of the inclusions contained within the parallelepiped, l^L, is sufficiently small compared to the length scale of the structural component, l^G, then the inclusions will not be visible to the naked eye. This realization was never more apparent than when the founders of continuum mechanics were attempting in the first two decades of the nineteenth century to decide how to the treat the (at that time presumably) molecular nature of all materials (Maugin 2014). Indeed, several papers appeared around 1820, including the famous paper on elasticity by Navier (1821), wherein he attempted to construct a theory of elasticity by starting from a molecular scale and averaging upward to a length scale that could be approximated as continuous. In the following year, Cauchy (1822) did away with this necessity entirely by assuming that the structural component could be assumed to be *continuous* for the purpose of analysis, and thus was born *continuum mechanics* (Maugin 2014).

However, this issue relating to length scales was to be revisited in the latter part of the nineteenth century, initially for gases, by such icons as Maxwell (1860) and Boltzmann (1874), with the development of statistical mechanics and thermodynamics. Still, these works were focused on atomistic and molecular scale phenomena, whereas the practicing civil engineer deals normally with length scales that are many orders of magnitude larger than typical molecules.

In fact, the length scales of inclusions embedded within mastics deployed within civil infrastructures are typically at the smallest on the local length scale, l^L, of microns (such as binder additives) and at the largest local length scale centimeters (such as aggregates), as compared to the global length scale, l^G, of typical infrastructures (usually meters to tens of meters). Fortunately, these length scales all fall well within the size range for which continuum mechanics has been shown to be quite accurate for the purpose of structural analysis and design, so that molecular scale modeling is usually not employed.

Furthermore, it has been shown that when these length scales are widely separated, that is, when:

$$l^L \ll l^G \tag{11.52}$$

the heterogeneous boundary value problem may be replaced by one in which it is assumed that the material within the structural component is assumed to be spatially homogeneous. This may be accomplished by supposing that the exact solution to the heterogeneous boundary value problem is known, and the stress state for this problem is given by the following:

$$\sigma_{ij}(x_k^G, t) \equiv \sigma_{ij}^G(x_k^G, t) \quad \forall x_k^G \in V + S \tag{11.53}$$

Now, suppose that, given the above stress state, the volume average is computed as follows (Hill 1963; Mandel 1964):

$$\bar{\sigma}_{ij}^{G} \equiv \frac{1}{V^{L}} \int_{V^{L}} \sigma_{ij}^{G} dV \tag{11.54}$$

where V^{L} is the volume of the locally defined parallelepiped. Note that the above volume-averaged stress is actually not measured during a constitutive test. During such a test, the quantity that is measured is the boundary averaged stress, Σ_{ij}, given by (Boyd et al. 1993):

$$\Sigma_{ij} \equiv \frac{1}{V^{L}} \int_{S^{L}} \sigma_{ik}^{G} n_{k} x_{j} dS \tag{11.55}$$

However, by deploying the divergence theorem Eq. (9.31), it can be shown that:

$$\Sigma_{ij} = \bar{\sigma}_{ij}^{G} \tag{11.56}$$

Therefore, for purposes of calculating the average stress, Eq. (11.54) is more convenient than Eq. (11.55).

Similarly, during a constitutive test, the kinematic quantity that is measured is the boundary averaged strain, E_{ij}, given by (Boyd et al. 1993):

$$E_{ij} \equiv \frac{1}{V^{L}} \int_{S^{L}} \frac{1}{2}(u_{i} n_{j} + u_{j} n_{i}) dS \tag{11.57}$$

Once again, using the divergence theorem Eq. (9.31), it can be shown that (assuming that there are no internal cracks or voids):

$$E_{ij} = \bar{\varepsilon}_{ij}^{G} \tag{11.58}$$

where

$$\bar{\varepsilon}_{ij}^{G} \equiv \frac{1}{V^{L}} \int_{V^{L}} \varepsilon_{ij}^{G} dV \tag{11.59}$$

The exact global stress state can be expanded in a Taylor series as follows:

$$\sigma_{ij}^{G} = \bar{\sigma}_{ij}^{L} + \sum_{n=1}^{\infty} \frac{1}{V^{L}} \int_{V^{L}} \frac{\partial^{n} \sigma_{ij}^{L}}{\partial x_{k}^{n}} \left(x_{k}^{L} - \bar{x}_{k}^{L}\right)^{n} dV \tag{11.60}$$

where $\bar{x}_k$ are the coordinates of the centroid. In the case wherein all of the terms in the summation are small compared to the leading term in Eq. (11.55), the material will be defined to be *statistically homogeneous*, implying that

$$\sigma_{ij}^{G}(x_k^G, t) \cong \bar{\sigma}_{ij}^{L}(x_k^G, t) \tag{11.61}$$

The volume-averaged stress given by Eq. (11.54), as well as the volume-averaged strain given by Eq. (11.59) may now replace the actual global variables within all of the field equations (c.f. Table 11.1), thereby resulting in a dramatically simplified boundary value problem in which the structural component is viewed as macroscopically homogeneous.

It should be pointed out that with increasing simplicity there is invariably accompanying loss of accuracy, and such is the case with the simplification resulting from the homogenization of micromechanics solutions. Thus, this approach should be utilized with caution. Fortunately, as will be demonstrated in Chap. 14, it is possible to predict the error introduced when micromechanics is invoked by utilizing the technique termed *multi-scaling* (Allen and Searcy 2006).

11.3 Multidimensional Linear Thermoelasticity

When the temperature in the object of interest varies spatially and/or with time, the above elastic model must necessarily be generalized. Such a circumstance is all too common in flexible pavements (see Chap. 8), and in this case, a thermoelastic material is defined to be one in which the stress, σ_{ij}, is a single-valued function of the strain, ε_{kl}, and the temperature, T:

$$\sigma_{ij}(\vec{x}, t) = \sigma_{ij}(\varepsilon_{kl}(\vec{x}, t), T(\vec{x}, t)) \tag{11.62}$$

The above is adjoined with the following equations (from Table 9.7):

(1) strain-displacement equations:

$$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{11.63}$$

(2) equations of motion:

$$\sigma_{ji,j} + \rho f_i = \rho \frac{d^2 u_i}{dt^2} \tag{11.64}$$

(3) conservation of energy:

$$\rho \frac{du^I}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} - q_{i,i} + \rho r \tag{11.65}$$

(4) entropy production inequality:

$$\rho \frac{ds}{dt} + \frac{\partial}{\partial x_i}\left(\frac{q_i}{T}\right) - \frac{\rho r}{T} \geq 0 \tag{11.66}$$

For reasons that will become apparent below, a new kinematic variable is now introduced, similar to Eq. (9.40):

$$\gamma_i \equiv \frac{\partial T}{\partial x_i} \tag{11.67}$$

called the temperature gradient, $\gamma_i = \gamma_i(\vec{x}, t)$, which has been included because it is observed experimentally that the flux of heat is caused primarily by temperature gradient (just as stress is caused by *displacement gradient*).

The reader will recall that Table 9.6 listed 25 unknown variables to be predicted as functions of position, $\vec{x}$, and time, t, in the continuous body $V + S$. Using Eq. (11.67), this list is now expanded to 28 unknowns, and these are as follows:

Displacement vector	u_i
Strain tensor	ε_{ij}
Stress tensor	σ_{ij}
Body force per unit mass	f_i
Internal energy per unit mass	u^I
Heat flux vector	q_i
Heat source per unit mass	r
Entropy per unit mass	s
Temperature	T
Temperature gradient	γ_i

The prediction of the above variables is afforded with the use of the 13 equations and one inequality (11.63)–(11.67) listed above, plus the linear elastic constitutive Eq. (11.2).

The thermoelasticity problem is considerably more complicated than the elasticity problem, and this is due at least in part to the coupling between mechanics and thermodynamics. Thus, a more formal procedure is required to construct this model.

11.4 Thermodynamic Constraints on Thermoelastic Material Behavior

To see how this works, a method introduced by Coleman and Noll (1963) is next employed. The challenge is to solve the above set in terms of the independent variables $\vec{x}$ and t, implying that a solution is sought for the 28 unknowns $u_i, \varepsilon_{ij}, \sigma_{ij}, f_i, u^I, T, \gamma_i, q_i, r$ and s as functions of $\vec{x}$ and t. The solution procedure is described in Table 11.2.

Table 11.2 Procedure for obtaining thermodynamic constraints on linear thermoelastic media

Step 1: Note that normally the body force per unit mass, $f_i = f_i(\vec{x}, t)$, can be specified (by using Newton's gravitational law) as well as the heat source, $r = r(\vec{x}, t)$. However, in this case specify a priori instead $u_i = u_i(\vec{x}, t)$, and $T = T(\vec{x}, t) \quad \forall \vec{x}, t \in V + S \Rightarrow$ **24 remaining unknowns**	
Step 2: Use Eq. (11.63) to solve for $\varepsilon_{ij} = \varepsilon_{ij}(\vec{x}, t) \Rightarrow$ **18 remaining unknowns**	
Step 3: Use Eq. (11.67) to solve for $\gamma_i = \gamma_i(\vec{x}, t) \Rightarrow$ **15 remaining unknowns**	
Step 4: Assume that there exist 11 constitutive equations for a thermoelastic material as follows:	
$\sigma_{ij}(\vec{x}, t) = \sigma_{ij}(\varepsilon_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t))$	(11.68)
$u^I(\vec{x}, t) = u^I(\varepsilon_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t))$	(11.69)
$q_i(\vec{x}, t) = q_i(\varepsilon_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t))$	(11.70)
$s(\vec{x}, t) = s(\varepsilon_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t))$	(11.71)
where according to *the principle of equipresence* (Truesdell et al. 2004), since the temperature gradient is included in one constitutive equation, it must be included in all of them. Note that the above relate the kinetic variables to the kinematic variables. Use (11.68)–(11.71) to solve for $\sigma_{ij} = \sigma_{ij}(\vec{x}, t), u^I = u^I(\vec{x}, t), q_i = q_i(\vec{x}, t), s = s(\vec{x}, t) \Rightarrow$ **4 remaining unknowns**	
Step 5: Use Eq. (11.64) to solve for $f_i = f_i(\vec{x}, t) \Rightarrow$ **1 remaining unknown**	
Step 6: Use Eq. (11.65) to solve for $r = r(\vec{x}, t) \Rightarrow$ **0 remaining unknowns**	
Step 7: Introduce *the Helmholtz free energy* (after Hermann von Helmholtz 1821–94), defined as follows:	
$h \equiv u^I - Ts$	(11.72)
Step 8: It is necessary to impose the constraint imposed by the entropy production inequality. To do this, first recall that due to Eq. (11.72) the time derivative of the Helmholtz free energy per unit mass, $h = h(\vec{x}, t)$, may be written as follows:	
$\frac{dh}{dt} = \frac{du^I}{dt} - s\frac{dT}{dt} - T\frac{ds}{dt}$	(11.73)
Step 9: Substitute Eqs. (11.68)–(11.71) and (11.73) into inequality (11.66) and employ the chain rule of differentiation to obtain:	
$-\left[\rho\frac{\partial h}{\partial \varepsilon_{ij}}(\varepsilon_{kl}, T, \gamma_k) - \sigma_{ij}(\varepsilon_{kl}, T, \gamma_k)\right]\frac{d\varepsilon_{ij}}{dt}$ $-\left[\rho\frac{\partial h}{\partial T}(\varepsilon_{kl}, T, \gamma_k) + \rho s(\varepsilon_{kl}, T, \gamma_k)\right]\frac{dT}{dt} - \left[\rho\frac{\partial h}{\partial \gamma_i}(\varepsilon_{kl}, T, \gamma_k)\right]\frac{d\gamma_i}{dT} - \frac{q_i\gamma_i}{T} \geq 0$	(11.74)
Step 10: Now note that the quantities inside the brackets in inequality (11.74) are independent of the quantities outside the brackets, and the quantities outside the brackets are independent of one another, so that the above produces the followings constraints on the allowable material behavior for a thermoelastic material:	
$\sigma_{ij} = \rho\frac{\partial h}{\partial \varepsilon_{ij}}$	(11.75)
$s = -\frac{\partial h}{\partial T}$	(11.76)
$\frac{\partial h}{\partial \gamma_i} = 0$	(11.77)
$-\frac{q_i\gamma_i}{T} \geq 0$	(11.78)
Step 11: It follows from Eq. (11.77) together with (11.69) and (11.71) that:	
$h = h(\varepsilon_{kl}, T)$	(11.79)

(continued)

Table 11.2 (continued)

Expand (11.79) in a Taylor series as follows:	
$h = \frac{1}{\rho}\left(h^{00} + h_{ij}^{10}\varepsilon_{ij} + h^{01}\theta + h_{ij}^{11}\varepsilon_{ij}\theta + \frac{1}{2}h_{ijkl}^{20}\varepsilon_{ij}\varepsilon_{kl} + \frac{1}{2}h^{02}\theta^2 + \cdots\right)$	(11.80)
where $h^{00}, h_{ij}^{10}, h^{01}, h_{ij}^{11}, h_{ijkl}^{20}$ and h^{02} are material constants, and	
$\theta \equiv T - T_R$	(11.81)
Note that T_R is defined to be the reference temperature at which no strain is observed when the object is unstressed	
Step 12: Substitute Eq. (11.80) into Eq. (11.75) to obtain	
$\sigma_{ij} = h_{ij}^{10} + h_{ijkl}^{20}\varepsilon_{kl} + h_{ij}^{11}\theta$	(11.82)
The above may be rearranged to the following form (assuming there is no residual stress)	
$\sigma_{ij} = E_{ijkl}(\varepsilon_{kl} - \alpha_{kl}\theta)$	(11.83)
assuming that $a_{10} = 0$ (meaning that there is no residual stress), and where E_{ijkl} is the modulus tensor defined above, and α_{ij} is termed *the thermal expansion tensor*, which can be obtained from simple experiments. In the isotropic case, Eq. (11.83) reduces to the following (see Sect. 11.3):	
$\sigma_{ij} = \frac{vE}{(1+v)(1-2v)}\varepsilon_{kk}\delta_{ij} + \frac{E}{(1+v)}\varepsilon_{ij} - \frac{E\alpha}{(1-2v)}\Delta T\delta_{ij}$	(11.84)
Step 13: Substitute Eq. (11.80) into Eq. (11.76) to obtain:	
$s = -\frac{1}{\rho}(h^{01} + h_{kl}^{11}\varepsilon_{kl} + h^{02}\Delta T)$	(11.85)
Step 14: Expand the heat flux vector in terms of the temperature gradient to obtain:	
$q_i(\varepsilon_{kl}, T, \gamma_k) = -k_{ij}(\varepsilon_{kl}, T)\gamma_j$	(11.86)
where k_{ij} is a material property called *the thermal conductivity tensor*. The above is called *Fourier's law of heat conduction* (Fourier 1822) (after Jean-Baptiste Joseph Fourier 1768–1830) when the thermal conductivity tensor is constant. Although this is not always the case, in most materials it is indeed a rather weak function of strain and temperature	
Step 15: Substituting Eq. (11.86) into inequality (11.78) will result in the following:	
$\frac{k_{ij}(\varepsilon_{kl}, T)\gamma_i\gamma_j}{T} \geq 0$	(11.87)
Note that due to the final form of the entropy production inequality (11.87), the determinant of the thermal conductivity tensor, k_{ij}, must necessarily be nonnegative, that is,	
$\left\|k_{ij}\right\| \geq 0$	(11.88)
This completes the construction of thermodynamic constraints on linear thermoelastic media, thus demonstrating that it is possible to solve for all of the unknowns using the thermoelastic constitutive model, while at the same time obtaining powerful constraints on the constitutive behavior of linear thermoelastic media	

11.5 The Linear Thermoelastic Initial Boundary Value Problem

The results of the procedure described in Table 11.2 may now be utilized to simplify the thermoelastic initial boundary value problem. To accomplish this, first note that the temperature, T, appears in the thermoelastic constitutive Eq. (11.83).

As such, it will be necessary to treat it as a primary variable that must be adjoined to the fifteen primary variables $u_i, \sigma_{ij}, \varepsilon_{ij}$ required to solve the elastic boundary value problem. The means whereby this new primary unknown is predicted is the conservation of energy Eq. (11.65). This equation may be manipulated so as to remove the secondary unknowns u^I, s, h, γ_i by first employing Eq. (11.68) to solve for the internal energy. The time derivative of the resulting equation may then be taken, and this result substituted into Eq. (11.65) to obtain the following:

$$\rho\left(\frac{\partial h}{\partial t}+T\frac{\partial s}{\partial t}+s\frac{\partial T}{dt}\right)=\sigma_{ij}\frac{d\varepsilon_{ij}}{dt}-q_{i,i}+\rho r \tag{11.89}$$

Now, Eq. (11.79) may be utilized to take the time derivative of the Helmholtz free energy, thus altering Eq. (11.89) to the following form:

$$\left(\rho\frac{\partial h}{\partial \varepsilon_{ij}}-\sigma_{ij}\right)\frac{d\varepsilon_{ij}}{dt}+\rho\left(\frac{\partial h}{\partial T}+s\right)\frac{dT}{dt}+\rho T\frac{ds}{dt}=-q_{i,i}+\rho r \tag{11.90}$$

Using Eqs. (11.75) and (11.76), it can be seen that the two parenthetical terms in Eq. (11.90) are identically zero.

Now, note from Eqs. (11.71) and (11.76) that the time derivative of the entropy may be written as follows:

$$\frac{ds}{dt}=\frac{\partial s}{\partial \varepsilon_{ij}}\frac{d\varepsilon_{ij}}{dt}+\frac{\partial s}{\partial T}\frac{dT}{dt}=-\frac{\partial^2 h}{\partial \varepsilon_{ij}\partial T}\frac{d\varepsilon_{ij}}{dt}-\frac{\partial^2 h}{\partial T^2}\frac{dT}{dt} \tag{11.91}$$

Substituting Eq. (11.75) into Eq. (11.91) and this result into Eq. (11.90) results in the following:

$$-\rho T\left(\frac{1}{\rho}\frac{\partial \sigma_{ij}}{\partial T}\frac{d\varepsilon_{ij}}{dt}+\frac{\partial^2 h}{\partial T^2}\frac{dT}{dt}\right)=-q_{i,i}+\rho r \tag{11.92}$$

From isotropic constitutive Eq. (11.84), it can be seen that:

$$\frac{\partial \sigma_{ij}}{\partial T}=-\beta\delta_{ij}=-\alpha(3\lambda+2\mu)\delta_{ij} \tag{11.93}$$

where α is called *the coefficient of thermal expansion*, discussed in Chap. 11 (Boley and Weiner 1960). Finally, let *the specific heat at constant volume* (or equivalently —dilatation), c_v, be defined as follows:

$$c_v\equiv-\frac{\partial^2 h}{\partial T^2}T=-h^{02}T \tag{11.94}$$

Thus, substituting Eqs. (11.86), (11.93) and (11.94) into Eq. (11.92) results in the following (Boley and Weiner 1960):

$$\rho c_v \frac{dT}{dt} = (kT_{,i})_{,i} + \rho r - \alpha(3\lambda + 2\mu)T\frac{d\varepsilon_{kk}}{dt} \tag{11.95}$$

The above equation now has the secondary variables removed, so that it may be used to obtain the temperature field, $T(\vec{x}, t)$.

11.5.1 Two-Way Coupled Thermoelasticity

Unfortunately, the above equation also contains the dilatational component of the strain tensor, ε_{kk}. This implies that the mechanics and the thermodynamics are *coupled*, meaning that neither can be solved separately from one another in a thermoelastic object. This coupling seems to have been first studied by Jean-Marie-Constant Duhamel (1787–1872) (Duhamel 1837). This then results in the so-called *coupled thermoelastic initial boundary value problem*, thus implying that when temperature change occurs in a thermoelastic object, the deformation changes, *and* vice versa, that is, mechanical loading induces temperature change. For the linear isotropic case, this problem is described in Table 11.3.

11.5.2 One-Way Coupled Thermoelasticity

The initial boundary value problem described in Table 11.3 can be quite challenging to solve due to the thermoelastic coupling. Fortunately, this coupling is in most cases quite small (Boley and Weiner 1960). In fact, for most practical circumstances "the coupling term appearing in the heat equation can be disregarded for all problems except those in which the thermoelastic dissipation is of primary interest" (Boley and Weiner 1960). This assumption appears to be accurate for the purpose of modeling flexible pavement structures. Therefore, this will be assumed to be the case throughout the remainder of this text, thus resulting in the following simplification of Eq. (11.95):

$$\rho c_v \frac{dT}{dt} = (kT_{,i})_{,i} + \rho r \tag{11.96}$$

It can be seen that the above is an uncoupled form of the conservation of energy, in which the temperature is the single unknown field variable, thus earning this simplification the name *uncoupled heat transfer*. It is in fact the theory first

Table 11.3 Two-way coupled linear thermoelastic initial boundary value problem

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads:	$\hat{t}_i(\text{on } S), f_i(\text{in } V+S), r(\text{in } V+S)$
Geometry:	$\vec{x}(\text{on } S),\ n_i(\text{on } S)$
Material Properties:	$\rho, \lambda, \mu, \alpha, k, c_v$

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}, T$ **= 16 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = \lambda \varepsilon_{kk}\delta_{ij} + 2\mu\varepsilon_{ij} - \alpha(3\lambda + 2\mu)\theta\delta_{ij}$	6
$\rho c_v \frac{dT}{dt} = (kT_{,i})_{,i} + \rho r - \alpha(3\lambda + 2\mu)T\frac{d\varepsilon_{kk}}{dt}$	1
Total	**16**

Initial Conditions:

Displacements:	$u_i(\vec{x}, 0) = \hat{u}_i$	$\forall \vec{x} \in V + S$
Stresses:	$\sigma_{ij}(\vec{x}, 0) = \hat{\sigma}_{ij}$	$\forall \vec{x} \in V + S$
Strains:	$\varepsilon_{ij}(\vec{x}, 0) = \hat{\varepsilon}_{ij}$	$\forall \vec{x} \in V + S$
Temperature:	$T(\vec{x}, 0) = \hat{T}$	$\forall \vec{x} \in V + S$

Boundary Conditions:

Tractions:	$t_i = \hat{t}_i \quad on \quad S_1$	
Displacements:	$u_i = \hat{u}_i \quad on \quad S_2$	$S = S_1 + S_2$
Temperature:	$T = \hat{T} \quad on \quad S_3$	
Heat Flux:	$q_i = \hat{q}_i \quad on \quad S_4$	$S = S_3 + S_4$

developed by Fourier (1822). In this case, the procedure described in Table 11.3 simplifies to two separate problems. First, the temperature field, $T(\vec{x},t)$ is predicted, subject to the thermodynamic initial and boundary conditions described in Table 11.3. This part of the problem may be predicted using the finite element method (see Chap. 14). Thereafter, the temperature may be deployed as a known quantity to predict the remaining fifteen unknowns, $u_i(\vec{x},t), \sigma_{ij}(\vec{x},t), \varepsilon_{ij}(\vec{x},t)$. The special case of a linear isotropic object subjected to quasi-static loading is described in Tables 11.4 and 11.5.

Table 11.4 Part 1: Procedure for solving for the temperature in uncoupled thermoelasticity

Part 1: Solve for $T = T(\vec{x},t)$

Independent Variables: $\vec{x},t$

Known Inputs:

Loads:	r
Geometry:	$\vec{x}\,(\text{on } S),\ n_i\,(\text{on } S)$
Material Properties:	ρ, k, c_v

Unknowns: T = 1 unknown

Field Equation:

	No. of Equations
$\rho c_v \dfrac{dT}{dt} = (kT_{,i})_{,i} + \rho r$	1
Total	**1**

Initial Conditions:

Temperature: $T(\vec{x},0) = \hat{T} \quad \forall \vec{x} \in V + S$

Boundary Conditions:

Temperature: $T = \hat{T} \quad on \quad S_3$

Heat Flux: $q_i = \hat{q}_i \quad on \quad S_4 \quad S = S_3 + S_4$

Table 11.5 Part 2: Procedure for solving for stresses, strains, and displacements in uncoupled thermoelasticity

PART 2: Solve for $u_i = u_i(\vec{x},t), \sigma_{ij} = \sigma_{ij}(\vec{x},t), \varepsilon_{ij} = \varepsilon_{ij}(\vec{x},t)$

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S), f_i(\text{in } V+S)$

Geometry: $\vec{x}(\text{on } S),\ n_i(\text{on } S)$

Material Properties: $\rho, \lambda, \mu, \alpha$

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}$ = **15 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = \lambda \varepsilon_{kk}\delta_{ij} + 2\mu\varepsilon_{ij} - \alpha(3\lambda + 2\mu)\theta\delta_{ij}$	6
Total	**15**

Initial Conditions:

Displacements: $u_i(\vec{x},0) = \hat{u}_i \quad \forall \vec{x} \in V+S$

Stresses: $\sigma_{ij}(\vec{x},0) = \hat{\sigma}_{ij} \quad \forall \vec{x} \in V+S$

Strains: $\varepsilon_{ij}(\vec{x},0) = \hat{\varepsilon}_{ij} \quad \forall \vec{x} \in V+S$

Boundary Conditions:

Tractions: $t_i = \hat{t}_i \quad on \quad S_1$

Displacements: $u_i = \hat{u}_i \quad on \quad S_2 \quad S = S_1 + S_2$

11.6 Modeling the Effects of Moisture on Roadway Performance

Moisture effects road way performance via both chemistry and mechanics (see Chap. 8). While the physics of how this occurs is quite different from the effects of temperature on road way performance, the procedure for modeling the effects of moisture is quite similar to that used to model the effects of temperature. Consider the case of a road way subjected to both moisture and heating (or cooling) on the road way surface, as shown in Fig. 11.7.

A close-up view of Fig. 11.7 will reveal that both temperature, T, and moisture per unit volume, M, diffuse into the road way as a function of both time and the location within the road way, as shown in Fig. 11.8.

The methodology for predicting the temperature distribution within the road way was discussed in the previous section. In this section, the procedure for predicting the moisture distribution is discussed briefly. This is accomplished by deploying conservation of mass, which states that the rate that mass is added to an object is balanced by the rate of change of mass within the object (in the absence of chemical changes within the object). Now, consider the application of conservation of mass to a tiny element taken from the road way, as shown in Fig. 11.9, where the vector

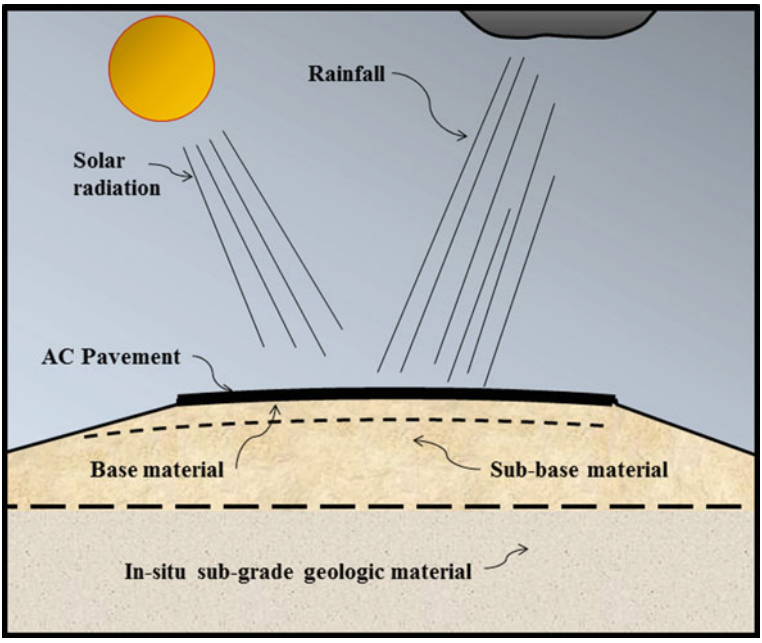


Fig. 11.7 Two-dimensional depiction of roadway heating and moisture

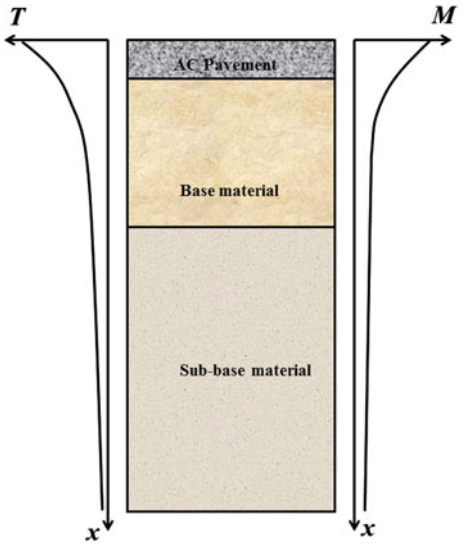


Fig. 11.8 Close-up of roadway depicting distribution of temperature and moisture within the roadway in a given time

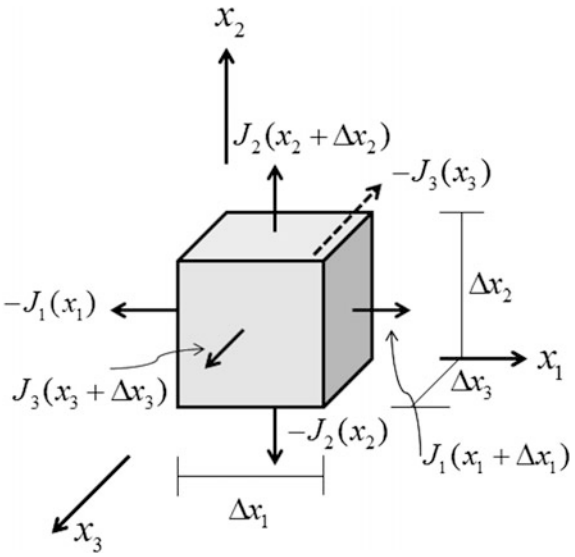


Fig. 11.9 Depiction of flux of moisture within a tiny element

$$\vec{J} = J_1\vec{e}_1 + J_2\vec{e}_2 + J_3\vec{e}_3 \tag{11.97}$$

is the flux per unit area of moisture.

Employing conservation of mass within the element will thus result in the following equation:

$$-[J_1(x_1+\Delta x_1)-J_1(x_1)]\Delta x_2\Delta x_3-[J_2(x_2+\Delta x_2)-J_2(x_2)]\Delta x_1\Delta x_3 - [J_3(x_3+\Delta x_3)-J_3(x_3)]\Delta x_1\Delta x_2=\left[\frac{M(t+\Delta t)-M(t)}{\Delta t}\right]\Delta x_1\Delta x_2\Delta x_3 \tag{11.98}$$

Taking the limits as the dimensions and time increments go to zero in the above will result in the following:

$$-J_{i,i}=\frac{\partial M}{\partial t} \tag{11.99}$$

Note that the above is similar to that obtained for conservation of energy of thermoelastic media. In further similarity to thermal effects, it is now assumed that (for isotropic media):

$$J_i=-DM_{,i} \tag{11.100}$$

The above is called Fick's first law of diffusion (after Adolf Eugen Fick 1829–1901). Substituting Eq. (11.100) into Eq. (11.99) results in the following:

$$\left(DM_{,i}\right)_{,i}=\frac{\partial M}{\partial t} \tag{11.101}$$

which is called Fick's second law.

Note that the above is mathematically equivalent to the previously derived heat conduction model (Eq. (11.86)) when there is no internal heat source. Thus, it may be deduced that the same mathematical approach utilized to predict the temperature field may be used to predict the moisture field within the road way. This part of the problem may be predicted using the finite element method (see Chap. 14). The spatial and temporal distribution of moisture within the road way may then be used as a known quantity within the road way for the purposes of predicting the road way response.

It remains to determine exactly what effect the moisture distribution has on the road way, and to that end there are two distinct components: a chemical one and a mechanical one. Moisture can induce chemical changes within the road way, and these changes will necessarily result in changes in the material properties of the road way. Predicting this part of the effect of moisture falls under the domain of materials science and is therefore discussed in Part I.

The mechanical effect on the road way due to moisture, as it turns out, is quite similar to that caused by temperature, as described for the uniaxial case in Chap. 10. Thus, it can be seen that for isotropic linear hygrothermoelastic materials, the constitutive equations are as follows:

$$\sigma_{ij}=\frac{vE}{(1+v)(1-2v)}\varepsilon_{kk}\delta_{ij}+\frac{E}{(1+v)}\varepsilon_{ij}-\frac{E\alpha}{(1-2v)}\Delta T\delta_{ij}-\frac{E\beta}{(1-2v)}\Delta M\delta_{ij} \tag{11.102}$$

Furthermore, note that, due to Eq. (11.101), the moisture is now known both spatially and temporally within the road way.

11.7 Summary

This chapter has presented an overview of the historical development of the theories of elasticity and thermoelasticity. While it is not an exhaustive coverage of this subject, the material presented herein has been introduced as a means of preparing the reader for the subjects to come that deal with the inelastic deformations that occur in flexible road ways. As such, this material is an essential part of the knowledge necessary to design modern flexible road ways.

11.8 Problems

Problem 11.1
Given: A particular linear elastic object may be treated as two-dimensional because the stresses, strains, and displacements do not vary in the x_3 coordinate direction within the object
Required: Reconstruct Table 11.1, enumerating all of the component equations and all of the unknowns without the use of indicial notation

Problem 11.2
Given: The following objects are brought into contact with one another

(a) A typical seat cushion and a wooden board made of oak
(b) A steel reinforcement bar and a piece of asphalt binder
(c) A piece of asphalt binder and a rubber ball

Required: Determine in each case the type of boundary conditions applied to each object by its counterpart

Problem 11.3
Given: A linear elastic material is transversely isotropic about the $x_2 - x_3$-plane
Required: Use Eq. (11.29) and prove that it reduces to Eq. (11.30)

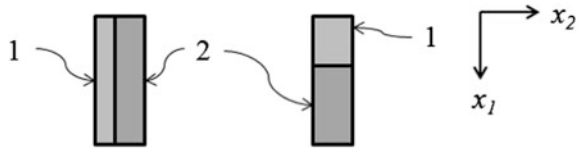
Problem 11.4
Given: Eq. (11.31).
Required: Prove Eq. (11.32) is correct.

Problem 11.5
Given: Eq. (11.38).
Required: Invert to obtain Eq. (11.39).

Problem 11.6
Given: A given road way may be assumed to undergo a state of plane strain everywhere due to an applied loading
Required: Describe the resulting linear elastic boundary value problem, showing a diagram depicting the cross section of the road way, all governing equations, and all boundary conditions

Problem 11.7
Given: The linear elastic boundary value problem described in Table 11.1
Required: Derive the Navier equations (11.49) for an isotropic solid

Problem 11.8
Given: Two mechanical contraptions are made from two blocks of material (denoted material 1 and material 2), as shown below



Required: Use elasticity theory to derive the macroscopically averaged modulus of elasticity of each contraption in the x_1 coordinate direction as a function of the volume fraction of each material

Problem 11.9
Given: A particular object may be idealized as a one-dimensional linear thermoelastic continuum
Required: Utilize the method of Coleman and Noll to construct a complete thermomechanical model of the object

Problem 11.10
Given: The Gibbs free enthalpy is defined as follows:

$$g \equiv u - Ts - \frac{1}{\rho}\sigma_{ij}\varepsilon_{ij}$$

Required: Use the Gibbs free enthalpy to derive thermodynamic constraints on the following class of materials:

$$\begin{aligned}
\varepsilon_{ij} &= \varepsilon_{ij}(\sigma_{kl}, T, \gamma_k) \\
u^I &= u^I(\sigma_{kl}, T, \gamma_k) \\
q_i &= q_i(\sigma_{kl}, T, \gamma_k) \\
s &= s(\sigma_{kl}, T, \gamma_k)
\end{aligned}$$

References

Allen, D., & Searcy, C. (2006). A model for predicting the evolution of multiple cracks on multiple length scales in viscoelastic composites. *Journal Materials Science, 41,* 6510.

Allen. D. (2013). *Introduction to the mechanics of deformable solids: Bars and beams*. Springer.

Allen, D. (2014). *How mechanics shaped the modern world*. Springer.

Amrouche, C., Ciarlet, P., Gratie, L., & Kesavan, S. (2006). On Saint Venant's compatibility conditions and Poincaré's lemma. *Comptes Rendus de l'Académie des Sciences - Series I, 342,* 887–891.

Boley, B., & Weiner, J. (1960). *Theory of thermal stresses*. Wiley.

Boltzmann, L. (1874). Zur theorie der elastichen Nachwirkung. *Sitz K u K Akad Wien, 70,* 275.

Boyd, J., Costanzo, F., & Allen, D. (1993). A micromechanics approach for constructing locally averaged damage dependent constitutive equations in inelastic composites. *International Journal of Damage Mechanics*, *2*, 209.

Cesaro, E. (1906). Sulle formole del Volterra, fondamentali nella teoria delle dostorsioni elastiche. Rendiconto dell' Academia della Scienze Fisiche e Matematiche (Società Reale di Napoli).

Coleman, B., & Noll, W. (1963). The thermodynamics of elastic materials with heat conduction and viscosity. *Archive Rational Mechanics and Analysis, 13,* 167.

Duhamel, J. (1837), Second mémoire sur les phénomènes thermo-mécaniques. *Journal de l'École Polytechnique*.

Euler, L. (1744). *Methodus inveniendi lineas curvas*. St. Petersburg.

Fourier, J. (1822). *Theorie analytique de la chaleur*. Firmin Didot.

Fung, Y. (1965). *Foundations of solid mechanics*. Prentice Hall.

Galilei, G. (1638). *Dialogues concerning two new sciences*. University of Toronto Library.

Hill, R. (1963). Elastic properties of reinforced solids: Some theoretical principles. *Journal of the Mechanics and Physics of Solids, 13,* 89.

Love, A. (1906). *A treatise on the mathematical theory of elasticity*. University of Michigan library.

Malvern, L. (1969). *Introduction to the mechanics of a continuous medium*. Prentice-Hall.

Mandel, J. (1964). *Proceedings of the 11th International Congress of Applied Mechanics* (Vol. 502).

Maugin, G. (2014). *Continuum mechanics through the eighteenth and nineteenth centuries*. Springer.

Maxwell, J. (1860). Illustrations of the dynamical theory of gases, Parts I and II. *Philosophical Magazine, 19,* 9.

Newton, I. (1687). *Principia—Vol. 1—The motion of bodies*. University of California Press.

Saint-Venant, A. (1856). Mémoire sur la torsion des prismes, avec des considérations sur leur flexion ainsi que sur l'équilibre intérieur des solides élastiques en général, et des formules pratiques pour le calcul de leur résistance à divers efforts s'exercant simultanément. *Mémoires Présentés par divers savants à l'Académie des sciences de l'Institut de France, 14,* 233–560.

Timoshenko, S. (1970). *Theory of elasticity*. McGraw-Hill.

Todhunter, I., & Pearson, K. (1893). *A history of the theory of elasticity and the strength of materials from Galilei to the present time*. Cambridge University Press.

Truesdell, C., Noll, W., & Antman, S. (2004). *The non-linear field theories of mechanics*. Springer.

Chapter 12
Viscoelasticity and Thermoviscoelasticity

12.1 Introduction

The previous chapter focused on models for predicting the mechanical response of elastic media. It was shown that elastic materials do not dissipate energy. Although in many applications the amount of energy dissipated may be negligibly small, the fact is that some energy is dissipated in all materials. For the purpose of predicting failure, especially that of flexible pavements, it is essential to incorporate energy dissipation mechanisms into material models in order to predict pavement life. While elastic material models may be sufficiently accurate for predicting the response of some base layer(s), the dissipation of energy is especially pronounced in the layers near the surface of the road way (see Chap. 8). One means of predicting energy dissipation is to utilize a viscoelastic material model. A one-dimensional viscoelastic material model was introduced in Chap. 10. In this chapter, the discussion focuses on three-dimensional viscoelastic material behavior.

Charles-August de Coulomb (1736–1806) appears to be the first person to report the observation of viscoelastic material behavior (Coulomb 1784). Toward the middle of the nineteenth century, scientists began studying the behavior of various solids that do not behave elastically. The first model for a viscoelastic material was proposed by Wilhelm Eduard Weber (1804–91) in 1835 (Weber 1835, 1841). Weber's colleague, Rudolf Hermann Arndt Kohlrausch (1809–58), produced the first analysis of a polymer (rubber) that used viscoelasticity theory (as reported by his son, Kohlrausch 1863). The elder Kohlrausch also introduced the concept of relaxation in 1854, although it was first applied to electromagnetism. James Clerk Maxwell (1831–79) studied relaxation phenomena in gases (Maxwell 1867, 1875). Ludwig Boltzmann (1844–1906) introduced the first three-dimensional isotropic viscoelasticity model in 1874 (Boltzmann 1874), and this was extended to anisotropic media by Vito Volterra (1860–1940) in 1909 (Volterra 1928, 1959). The concept of an intrinsic timescale was introduced by Peter Debye (1884–1966) in 1913 (Debye 1913).

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_12

Marcus Reiner introduced the Deborah constant (from the prophetess Deborah in *The Bible*: Judges V, 5: "The mountains flow before the Lord."), N_D, defined as follows:

$$N_D \equiv \frac{\tau_{mat}}{\tau_{\exp}} \tag{12.1}$$

where τ_{mat} is defined to be the time required for material rearrangements, and $\tau_{\exp}$ is defined to be the experimental timescale. While all materials are to one degree or another viscoelastic (Ferry 1980), it can be said that

$$\begin{aligned} N_D &\to \infty \Rightarrow \textit{elastic} \\ N_D &\to 0 \Rightarrow \textit{viscous} \end{aligned} \tag{12.2}$$

whereas for bounded values of N_D the material behavior is some combination of viscous and elastic. In many cases, a viscoelastic material model such as that to be developed in this chapter will serve this purpose.

Because the independent variable time, t, occurs explicitly in problems involving viscoelastic media even in the quasi-static case, they are termed *initial boundary value problems* (IBVPs), implying that the unknowns to be determined depend on both spatial coordinates, x_i, and time, t. This fact, together with the increased mathematical complexity required to model the constitutive behavior of viscoelastic media, necessarily results in an even more complicated IBVP than that encountered when dealing with elastic media. Thus, just as in the case of elasticity theory, solutions for viscoelastic IBVPs remained exceedingly difficult to obtain until the finite element method and large-scale computers were developed concurrently during the latter third of the twentieth century.

While it can be said that the theory of elasticity is now relatively mature, the field of viscoelasticity is still evolving at this writing. This may be traced at least in part to the development of plastics, as well as the application of viscoelasticity theory to the field of biomechanics. Furthermore, viscoelasticity is now finding its way into predictive methodologies for road ways, the subject of this text. Thus, it can be said that viscoelasticity is a subset of deformable body mechanics that is now in its heyday. And within the most recent decade, this rapidly growing field of mechanics has been utilized frequently to model both asphaltic binder and the resulting pavement.

12.2 Multi-dimensional Linear Viscoelasticity

The subject of viscoelasticity begins with the definition of a three-dimensional viscoelastic material, that is, a three-dimensional generalization of one-dimensional Eq. (10.11). Thus, *a viscoelastic material is defined (in this text) to be one in which*

the stress tensor is a single-valued functional of the strain tensor. Mathematically, this is represented as follows:

$$\sigma_{ij}(\vec{x},t) \equiv E_{ij\,\tau=0}^{\tau=t}\{\varepsilon_{kl}(\vec{x},t)\} \tag{12.3}$$

where the modulus functional, $E_{ij}\{\}$, maps the entire history of strain at a given material point into the current value of the stress at that point. Recall from Chap. 10 that the above is termed a *strain (or modulus) formulation* because the strain is the input variable and the stress is the output variable. *A stress (or compliance) formulation* is given by the following form:

$$\varepsilon_{ij}(\vec{x},t) = D_{ij\,\tau=0}^{\tau=t}\{\sigma_{kl}(\vec{x},t)\} \tag{12.4}$$

where $D_{ij}\{\}$ is termed the compliance functional.

The decision as to which of the above two formulations the modeler chooses to use normally depends on the particular problem to be solved. Since the mathematics necessary to predict the mechanical response of viscoelastic materials is notoriously complicated, this complexity will often dominate the choice of formulation. Fortunately, in most practical circumstances there exists a mathematical mapping between the material property functionals $E_{ij}\{\}$ and $D_{ij}\{\}$, so that only one need be formulated and characterized experimentally.

The mathematics describing mappings (12.3) and (12.4) may be exceedingly complicated for some materials, depending on such factors as loading conditions, temperature, and relative humidity. However, when the material is linear (as described in Chaps. 9 and 10), both Eqs. (12.3) and (12.4) may be written in the form of convolution integral equations such as the following:

$$\sigma_{ij}(\vec{x},t) = \int_0^t E_{ijkl}(t-\tau)\frac{\partial \varepsilon_{kl}(\vec{x},\tau)}{\partial \tau}d\tau \tag{12.5}$$

or

$$\varepsilon_{ij}(\vec{x},t) = \int_0^t D_{ijkl}(t-\tau)\frac{\partial \sigma_{kl}(\vec{x},\tau)}{\partial \tau}d\tau \tag{12.6}$$

where $E_{ijkl}(t)$ is called the relaxation modulus tensor, and $D_{ijkl}(t)$ is called the creep compliance tensor. Note that while these are fixed quantities for a given material, similar to that in elastic media, they differ from their elastic counterparts in that they are time dependent rather than constant in time, t. Thus, in viscoelastic media they are termed *material properties* rather than material constants.

As described in Chap. 10, the various terms in the moduli or compliances may be obtained from relaxation or creep tests, respectively. Note that Eqs. (12.5) and (12.6) do not include the effects of moisture and temperature, as these will be the subjects of later developments.

12.2.1 The Linear Viscoelastic Initial Boundary Value Problem

The model that is produced by adjoining the linear viscoelastic constitutive Eq. (12.5) with the other parts of the model is called *the linear viscoelastic initial boundary value problem.* The construction of this model begins with Tables 9.6 and 9.7, wherein the list of unknowns and available equations is described in detail. The reader will recall that Table 9.6 lists 25 unknown variables to be predicted as functions of position, $\vec{x}$, and time, t, in the continuous body $V + S$. These are as follows:

Displacement vector	u_i
Strain tensor	ε_{ij}
Stress tensor	σ_{ij}
Body force per unit mass	f_i
Internal energy per unit mass	u
Heat flux vector	q_i
Heat source per unit mass	r
Entropy per unit mass	s
Temperature	T

The prediction of these variables is afforded with the use of the 10 equations and one inequality shown in Table 9.7, plus the linear viscoelastic constitutive Eq. (12.5) (or equivalently (12.6)).

Recall from Chap. 10 that when dealing with a viscoelastic object the body is by definition *at constant temperature* in both time and space, thus obviating the necessity to include temperature, T, in the model. Since the temperature is constant, the heat flux vector, $\vec{q}$, and the heat source term, r, must necessarily be identically zero, and it follows that for the linear viscoelastic case the internal energy, u, can be treated as an ancillary variable due to the fact that the conservation of energy simplifies to a means of calculating the internal energy once the stress field, $\sigma_{ij}(\vec{x}, t)$, and strain field, $\varepsilon_{ij}(\vec{x}, t)$, are predicted with the model.

In addition, the entropy, s, although not constant, will be shown below to be an ancillary variable. This reduces the above total of 25 variables to the following 18 variables: $u_i, \varepsilon_{ij}, \sigma_{ij}$ and f_i. The remaining field equations for a linear viscoelastic body listed in Table 9.7 (along with Eq. (12.5)) are therefore (renumbered) as follows:

Linear viscoelastic constitutive equations:

$$\sigma_{ij}(t) = \int_0^t E_{ijkl}(t-\tau)\frac{\partial \varepsilon_{kl}(\tau)}{\partial \tau}d\tau \tag{12.7}$$

Strain-displacement equations:

$$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{12.8}$$

Equations of motion:

$$\sigma_{ji,j} + \rho f_i = \rho \frac{d^2 u_i}{dt^2} \tag{12.9}$$

where the rate of change of the displacement has been substituted for the velocity vector, $\vec{v}$.

The above set totals to 15 equations. This would appear at first glance to be three to few equations (15 equations in 18 unknowns). However, a careful examination of the unknown field variables will reveal that the body force vector, f_i, can be predetermined using Newton's gravitational law when this is the source of the body force, which is normally the case when dealing with road ways. Thus, the above 15 equations are sufficient to predict the 15 unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}$, as functions of space and time at all points within a linear viscoelastic body.

In many circumstances, the momentum terms in Eq. (12.9) may be neglected, called *quasi-static*, and in this case the governing field equations simplify somewhat, although the problem remains time dependent (even for loading that is constant in time). Also, the following *initial conditions* are required:

$$\text{Displacements:} \quad u_i(\vec{x}, t=0) = \widehat{u}_i \quad \forall \vec{x} \in V+S \tag{12.10a}$$

$$\text{Strains:} \quad \varepsilon_{ij}(\vec{x}, t=0) = \widehat{\varepsilon}_{ij} \quad \forall \vec{x} \in V+S \tag{12.10b}$$

$$\text{Stresses:} \quad \sigma_{ij}(\vec{x}, t=0) = \widehat{\sigma}_{ij} \quad \forall \vec{x} \in V+S \tag{12.10c}$$

where the symbol $\widehat{\cdot}$ implies that the quantity is known a priori. In addition, the *boundary conditions* must be known:

$$\text{Tractions:} \quad t_i = \hat{t}_i \quad on \quad S_1 \tag{12.10d}$$

$$\text{Displacements:} \quad u_i = \hat{u}_i \quad on \quad S_2 \tag{12.10e}$$

and

$$S = S_1 + S_2 \quad S_1 \cap S_2 = \emptyset \tag{12.11}$$

The above set of field Eqs. (12.7)–(12.9), together with initial and boundary conditions (12.10a, 12.10b, 12.10c, 12.10d, and 12.10e), constitutes a so-called *well-posed initial boundary value problem* (IBVP), and this is due to the fact that there exists a solution for the unknown field variables in a linear viscoelastic body for any given set of input *loads*, *geometry*, *and material properties*. Furthermore, because all of the governing equations and the boundary conditions are linear, it can be shown that any solution that satisfies all of the field equations and the boundary conditions is unique (Christensen 1982). This IBVP is therefore a very powerful tool for predicting the mechanical response of a linear viscoelastic body, and for this reason, the *linear viscoelastic boundary value problem* is restated in Table 12.1.

Methods for solving the linear viscoelastic boundary value problem will be discussed in later sections.

12.2.2 Thermodynamic Constraints on Linear Viscoelastic Material Behavior

In Chap. 11, it was shown that elastic materials do not dissipate energy. By contrast, viscoelastic materials do normally dissipate energy, and this dissipation can be significant. In order to determine the effect of this dissipation, consider a body that is at spatially and temporally constant temperature. In this case, the conservation of energy, Eq. (9.89), simplifies to the following:

$$\rho \frac{du^I}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} \tag{12.12}$$

Similarly, the entropy production law, inequality (9.92), simplifies to

$$\rho \frac{ds}{dt} \geq 0 \tag{12.13}$$

Now, similar to the approach taken in Chap. 11, introduce the Helmholtz free energy, h, defined as follows:

$$h \equiv u^I - Ts \tag{12.14}$$

Table 12.1 Quasi-static linear viscoelastic initial boundary value problem

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S), f_i(\text{in } V+S)$

Geometry: $\vec{x}(\text{on } S),\ n_i(\text{on } S)$

Material Properties: ρ, E_{ijkl}

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}$ **= 15 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij}(t) = \int_0^t E_{ijkl}(t-\tau)\frac{\partial \varepsilon_{kl}(\tau)}{\partial \tau} d\tau$	6
Total	**15**

Initial Conditions:

Displacements:	$u_i(\vec{x}, t=0) = \bar{u}_i$	$\forall \vec{x} \in V+S$
Strains:	$\varepsilon_{ij}(\vec{x}, t=0) = \bar{\varepsilon}_{ij}$	$\forall \vec{x} \in V+S$
Stresses:	$\sigma_{ij}(\vec{x}, t=0) = \bar{\sigma}_{ij}$	$\forall \vec{x} \in V+S$

Boundary Conditions:

Tractions:	$t_i = \hat{t}_i$	$on \quad S_1$
Displacements:	$u_i = \hat{u}_i$	$on \quad S_2$

It follows that for isothermal conditions

$$\frac{dh}{dt} = \frac{du^I}{dt} - T\frac{ds}{dt} \Rightarrow \frac{du^I}{dt} = \frac{dh}{dt} + T\frac{ds}{dt} \tag{12.15}$$

Substituting (12.15) into (12.12) results in the following:

$$\rho\left(\frac{dh}{dt}+T\frac{ds}{dt}\right)=\sigma_{ij}\frac{d\varepsilon_{ij}}{dt} \tag{12.16}$$

The above may be substituted into inequality (12.13) with the following result (since T may be assigned to be nonnegative):

$$\sigma_{ij}\frac{d\varepsilon_{ij}}{dt}-\rho\frac{dh}{dt}\geq 0 \tag{12.17}$$

Consistent with constitutive Eq. (12.3), let $\rho h=\rho h\{\varepsilon_{ij}\}$ and expand h in convolution integrals in ε_{ij} as follows (Christensen 1982):

$$\begin{aligned}\rho h=\rho h_0+&\int_{-\infty}^{t}D_{ij}(t-\tau)\frac{\partial\varepsilon_{ij}}{\partial\tau}d\tau\\&+\frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t}E_{ijkl}(t-\tau,t-n)\frac{\partial\varepsilon_{ij}}{\partial\tau}\frac{\partial\varepsilon_{kl}}{\partial\eta}d\tau\,d\eta+H.O.T.\end{aligned} \tag{12.18}$$

In order to place the above expansion within the second law of thermodynamics, inequality (12.17), it is necessary to take the time derivative of Eq. (12.18). Since the limits of the integrals depend on time, t, it is necessary to utilize *Leibniz' rule* (Greenburg 1978), after Gottfried Wilhelm Leibniz. This rather complicated differentiation is accomplished by first defining a generic integral, $I(t)$, as follows:

$$I(t)=\int_{-\infty}^{t}F(\tau,t)d\tau \tag{12.19}$$

The time derivative of $I(t)$ is then given by

$$\frac{dI}{dt}=\int_{-\infty}^{t}\frac{\partial f(\tau,t)}{\partial t}d\tau+f(t,t) \tag{12.20}$$

Utilizing Leibniz' rule to differentiate Eq. (12.18) and substituting this result into inequality (12.17) thus results in the following:

$$\left[\sigma_{ij} - D_{ij}(0) - \int_{-\infty}^{t} E_{ijkl}(t-\tau,0)\frac{\partial \varepsilon_{kl}}{\partial \tau}d\tau\right]\frac{d\varepsilon_{ij}(t)}{dt} - \int_{-\infty}^{t}\frac{\partial D_{ij}(t-\tau)}{\partial t}\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau - \frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t}\frac{\partial E_{ijkl}(t-\tau,t-\eta)}{\partial t}\frac{\partial \varepsilon_{ij}\partial \varepsilon_{kl}}{\partial \tau\,\partial \eta}d\tau d\eta \geq 0 \tag{12.21}$$

In order for the above inequality to be true for any and all conditions, it follows that

$$\sigma_{ij} = D_{ij}(0) + \int_{-\infty}^{t} E_{ijkl}(t-\tau)\frac{\partial \varepsilon_{kl}}{\partial \tau}d\tau \tag{12.22}$$

and

$$\frac{\partial D_{ij}(t)}{\partial t} = 0 \tag{12.23}$$

It can be seen from the above that the term D_{ij} is the residual stress, and due to (12.23), this stress must necessarily be independent of time in order to satisfy the entropy production inequality. As a result, Eq. (12.22) is a generalization of Eq. (12.5) introduced above.

The entropy production inequality therefore simplifies to the following:

$$-\frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t}\frac{\partial E_{ijkl}(t-\tau,t-\eta)}{\partial t}\frac{\partial \varepsilon_{ij}}{\partial \tau}\frac{\partial \varepsilon_{kl}}{\partial \eta}d\tau d\eta \geq 0 \tag{12.24}$$

It can be seen from inequality (12.17) that the above represents the rate of entropy production per unit volume within a linear viscoelastic body, which is also the temperature-multiplied rate of energy dissipation. As such, this rate of energy dissipation can never be negative. This implies that since the terms containing the strain tensor are quadratic the rate of change of the determinant of the modulus tensor must be positive semidefinite. This is a rather complicated constraint on the viscoelastic material properties, but in the case wherein the material is isotropic, this result will be much simpler, as will be shown later in this chapter.

The above results comprise the constraints imposed on linear viscoelastic media by thermodynamics.

12.2.3 Material Symmetry

As discussed in Chap. 11, the material may possess certain degrees of symmetry, and for a viscoelastic material, the moduli or compliance functions will normally simplify according to the same procedure as that employed for elastic media.

Although road way materials can exhibit anisotropy, the constituents utilized in road ways are for the most part isotropic, so that any anisotropy observed in road way materials is due to aligning of the principal axes of the aggregate during the construction of the road way. Assuming that this alignment is minimal, road way materials can be assumed to be (at least initially) isotropic. Thus, for the case of isotropic linear viscoelastic media, Eq. (12.5) may be simplified by substituting Eq. (11.32) into (12.5), thereby resulting in the following *strain formulation*:

$$\begin{aligned}\sigma_{ij}(t) &= \frac{1}{3}\int_0^t [G_2(t-\tau) - G_1(t-\tau)]\delta_{ij}\frac{\partial \varepsilon_{kk}}{\partial \tau}d\tau + \int_0^t G_1(t-\tau)\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau \\ &= \int_0^t \lambda(t-\tau)\delta_{ij}\frac{\partial \varepsilon_{kk}}{\partial \tau}d\tau + \int_0^t 2\mu(t-\tau)\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau\end{aligned} \tag{12.25}$$

where

$$\begin{aligned}\lambda(t) &\equiv \frac{1}{3}[G_2(t) - G_1(t)] \\ \mu(t) &\equiv \frac{G_1(t)}{2}\end{aligned} \tag{12.26}$$

The above may also be written equivalently as follows:

$$\sigma_{kk} = 3\int_0^t K(t-\tau)\frac{\partial \varepsilon_{kk}}{\partial \tau}d\tau \tag{12.27}$$

where K is *the bulk relaxation modulus*, given by

$$K(t) \equiv \frac{G_2(t)}{3} = \frac{3\lambda(t) + 2\mu(t)}{3} \tag{12.28}$$

and

$$\sigma'_{ij} = \int_0^t 2\mu(t-\tau)\frac{\partial \varepsilon'_{ij}}{\partial \tau}d\tau \tag{12.29}$$

where $\mu(t)$ is *the shear relaxation modulus*, and the deviatoric stress tensor, σ'_{ij}, is defined by Eq. (9.62). Similarly, the deviatoric strain tensor is defined as follows:

$$\varepsilon'_{ij} \equiv \varepsilon_{ij} - \frac{\varepsilon_{kk}}{3}\delta_{ij} \tag{12.30}$$

Conversely, the isotropic *stress formulation* is given by

$$\begin{aligned}\varepsilon_{ij} &= \frac{1}{3}\int_0^t [J_2(t-\tau) - J_1(t-\tau)]\delta_{ij}\frac{\partial \sigma_{kk}}{\partial \tau}d\tau + \int_0^t J_1(t-\tau)\frac{\partial \sigma_{ij}}{\partial \tau}d\tau \\ &= \frac{1}{3}\int_0^t [B(t-\tau) - J(t-\tau)]\delta_{ij}\frac{\partial \sigma_{kk}}{\partial \tau}d\tau + \int_0^t J(t-\tau)\frac{\partial \sigma_{ij}}{\partial \tau}d\tau\end{aligned} \tag{12.31}$$

where $J(t) \equiv J_1(t)$ is called *the shear creep compliance*, and $B(t) \equiv J_2(t)$ is called *the bulk creep compliance*.

The above may be written equivalently in deviatoric form as follows:

$$\varepsilon_{kk} = \int_0^t B(t-\tau)\frac{\partial \sigma_{kk}}{\partial \tau}d\tau \tag{12.32}$$

$$\varepsilon'_{ij} = \int_0^t J(t-\tau)\frac{\partial \sigma'_{ij}}{\partial \tau}d\tau \tag{12.33}$$

Note also that by utilizing line 14 of Table 9.2, Laplace transforming Eqs. (12.27), (12.29), (12.32), and (12.33) will result in the following identities:

$$3\tilde{K} = \frac{1}{\tilde{B}} \quad 2\tilde{\mu} = \frac{1}{\tilde{J}} \tag{12.34}$$

where it should be noted that the peculiar numeric factors (2 and 3) occurring in Eq. (12.34) result from nothing more than incongruities in the definitions of the bulk and shear moduli and compliances.

Note that only one of the above four formulations is required in order to solve an IBVP. Each form is reported herein only for the sake of completeness, as the choice of which formulation is deployed is simply a matter of convenience that depends on the particular IBVP to be solved.

The determination of material properties will be discussed in further detail below, but note that it is often simplest to obtain properties from uniaxial tests in stress control (although this is unfortunately not always the case for asphalt binder

or asphalt concrete!). In this case, $v(t)$ and $D(t)$, as defined in Chap. 11, result from the test. From these results, it follows that

$$\tilde{E} = 1/\tilde{D} \tag{12.35}$$

Thus, if the quantity on the right can be inverse transformed (see below), one can obtain the relaxation modulus. Similarly,

$$\tilde{K} = \frac{\tilde{E}}{3(1-2\tilde{v})} \tag{12.36}$$

$$\tilde{\mu} = \frac{\tilde{E}}{2(1+\tilde{v})} \tag{12.37}$$

Note that the Laplace transform inversion of the above equations may or may not be straightforward (to be discussed further below). However, in some materials, it will be observed that v(t) will be nearly constant in time, and for this circumstance, the inversion is often quite simple. In this case, the resulting isotropic linear viscoelastic constitutive equations are given by

$$\sigma_{ij}(t) = \frac{v}{(1+v)(1-2v)} \int_0^t E(t-\tau)\delta_{ij}\frac{\partial \varepsilon_{kk}}{\partial \tau}d\tau + \frac{1}{(1+v)} \int_0^t E(t-\tau)\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau \tag{12.38}$$

Furthermore, for the stress formulation, when ν is constant, it follows that

$$J(t) = (1+v)D(t),\ B(t) = (1-2v)D(t) \tag{12.39}$$

These then are the properties that must be determined experimentally in order to model an isotropic linear viscoelastic material.

In order to determine the thermodynamic constraints on isotropic linear viscoelastic materials, Eq. (12.25) may now be substituted into inequality (12.24), thus resulting in the following:

$$\begin{aligned} &-\frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t}\frac{\partial \lambda(t-\tau, t-\eta)}{\partial t}\frac{\partial \varepsilon_{ii}}{\partial \tau}\frac{\partial \varepsilon_{kk}}{\partial \eta}d\tau d\eta \\ &-\int_{-\infty}^{t}\int_{-\infty}^{t}\frac{\partial \mu(t-\tau, t-\eta)}{\partial t}\frac{\partial \varepsilon_{ij}}{\partial \tau}\frac{\partial \varepsilon_{kl}}{\partial \eta}d\tau d\eta \geq 0 \end{aligned} \tag{12.40}$$

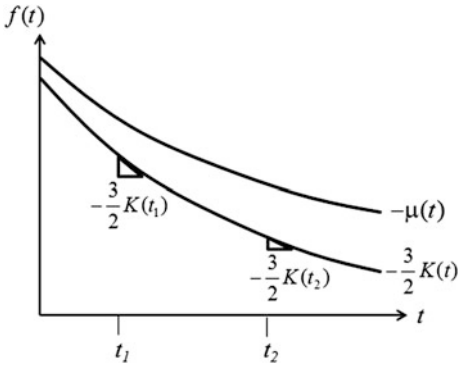


Fig. 12.1 Results of thermodynamic constraints on isotropic linear viscoelastic properties

Since both terms in the above are quadratic in strain, it follows that

$$\begin{aligned} -\frac{\partial \lambda}{\partial t} &\geq 0 \quad \forall t \\ -\frac{\partial \mu}{\partial t} &\geq 0 \quad \forall t \end{aligned} \tag{12.41}$$

Substituting (12.28) into the above thus results in

$$\begin{aligned} \frac{\partial \mu}{\partial t} &\leq 0 \qquad \forall t \\ \frac{3}{2}\frac{\partial K}{\partial t} &\leq \frac{\partial \mu}{\partial t} \quad \forall t \end{aligned} \tag{12.42}$$

The above thermodynamic constraints are depicted graphically in Fig. 12.1.

Figure 12.1 also depicts the phenomenon termed *fading memory*, defined as follows (Christensen 1982):

$$|f(t_2)| \leq |f(t_1)| \quad \forall t_2 > t_1 \tag{12.43}$$

12.3 Methods for Solving Viscoelastic IBVPs

In Sect. 12.2 of this chapter, the linear viscoelastic boundary value problem was described. It remains to develop methods for solving specific linear viscoelastic initial boundary value problems. These methods generally fall into one of four distinct categories:

(1) Direct analytic methods,
(2) Separable correspondence principles,
(3) Laplace Transform correspondence principles, and
(4) Computational methods.

In this section, the first three of these methods will be briefly discussed. Because the last of these methods is quite complicated (as well as powerful), it will be discussed in a later chapter.

12.3.1 Direct Analytic Method

The linear viscoelastic IBVP is described in Table 12.1. The objective of this problem is to predict the stresses, strains, and displacement components at all points and for all times in a linear viscoelastic object of known shape and material makeup subjected to specified loads. As described in Chap. 11, there are a number of simplifying steps that may be undertaken to simplify this problem for linear elastic media, such as reducing dimensionality, applying symmetry conditions, and solving either the Dirichlet or the Neumann problem. Those techniques apply equally well to linear viscoelastic IBVPs. Furthermore, dual methods such as those described in Chap. 11 may also be utilized in viscoelasticity problems.

In a very few cases, it is possible to "guess" a solution for the unknown variables as functions of space and time and show that they satisfy all of the initial conditions, boundary conditions, and field equations. This technique, termed *the direct analytic method*, may be deployed for problems that are quite simple geometrically, such as a rectangular parallelepiped subjected to evenly distributed boundary tractions.

As will be described below, the so-called *inverse solutions* may be obtained for the purpose of extracting material properties from laboratory tests for such problems. However, for the vast majority of linear viscoelastic IBVPs direct analytic solutions have thus far not been obtained in the literature. An example problem is solved below as a means of demonstrating the direct analytic method.

Example Problem 12.1: Tapered Bar Subjected to a Uniaxial Loading
Given: Consider the two-dimensional depiction of a linear viscoelastic tapered bar subject to loading in the long direction, as shown in Fig. 12.2.

Required: Predict the stresses, strains, and displacements in the bar as functions of the load, geometry of the bar, and the material properties.

Solution: Because the bar is long compared to the lateral dimension, and because it is loaded only in the long direction, it is called a uniaxial bar. The solution to this type of problem can be notoriously complicated because the boundary conditions are mixed. This problem is also complicated by the fact that the bar is not prismatic in x_1, not to mention the fact that the bar is viscoelastic. Nevertheless, it is possible to obtain an analytic solution that is quite accurate by simply *assuming* the solution and checking to ensure that it satisfies all of the field equations, as well as the initial and boundary conditions, as described in Table 12.1.

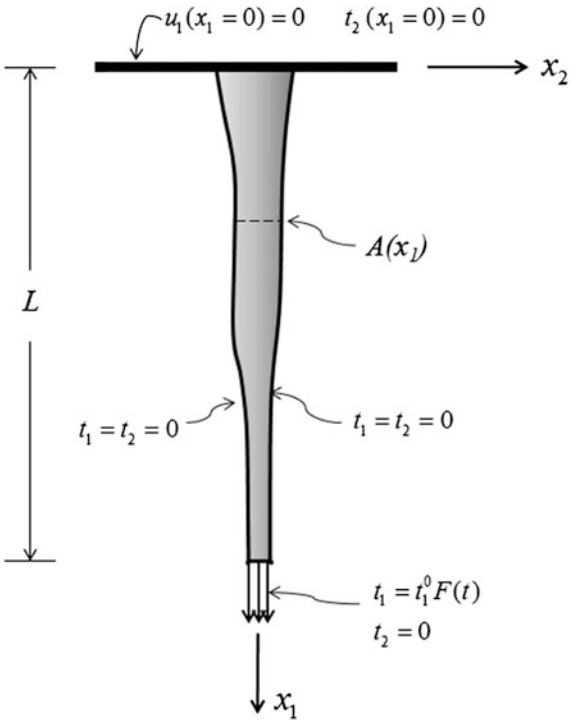


Fig. 12.2 Tapered linear viscoelastic bar subjected to uniaxial loading

It is assumed as a first approximation that none of the field variables are functions of x_3, except for the displacement component, u_3, which is everywhere linear in x_3. As a guess, assume that three of the unknown field variables are as follows:

$$\sigma_{11}(\vec{x}, t) = \frac{t_1^0 F(t) A(x_1 = L)}{A(x_1)} \tag{12.44}$$

$$\varepsilon_{11}(t) = \int_0^t D(t - \tau) \frac{\partial \sigma_{11}(\vec{x}, \tau)}{\partial \tau} d\tau \tag{12.45}$$

$$u_1(\vec{x}, t) = \int_0^{x_1} \varepsilon_{11}(\vec{x}, t) dx_1 \tag{12.46}$$

Furthermore, assume that all remaining components of the stress tensor are zero and that the applied load at the lower end of the bar is large compared to the weight of the bar itself, so that gravitational loads may be neglected. Finally, assume that the load is applied slowly enough that inertial effects may be neglected.

The above proposed solution will be correct if it satisfies all of the equations shown in Table 12.1. Under the above conditions, the first equilibrium equation will reduce to the following:

$$\frac{\partial \sigma_{11}}{\partial x_1} = 0 \tag{12.47}$$

Assumed solution (12.44) does not satisfy the above equation. However, if the above equation is integrated over the cross-sectional area, A, the above may be written (approximately, assuming that the cross-sectional area varies slowly in x_1) as follows (Allen and Haisler 1985):

$$\frac{\partial P}{\partial x_1} = 0 \tag{12.48}$$

where

$$P(x_1, t) \equiv \sigma_{11}(x_1, t) A(x_1) \tag{12.49}$$

is the internal axial force carried within the bar. Substituting Eq. (12.44) into Eq. (12.49) and this result into approximate equilibrium Eq. (12.48) now result in the following:

$$\frac{\partial}{\partial x_1}\left[\frac{t_1^0 F(t) A(x_1 = L)}{A(x_1)} A(x_1)\right] = \frac{\partial}{\partial x_1}\left[t_1^0 F(t) A(x_1 = L)\right] = 0 \tag{12.50}$$

Equilibrium is therefore satisfied approximately in the x_1 coordinate direction. Thus, since the remaining components of the equilibrium equations are identically zero, the assumed solution satisfies equilibrium in the remaining two coordinate directions identically, so that equilibrium is satisfied (approximately) by assumed Eq. (12.44) within the bar. Furthermore, the assumed stress field satisfies the initial and boundary conditions (approximately) by direct observation.

Substituting the stress into the uniaxial form of constitutive Eq. (12.6) now immediately verifies Eq. (12.45), and integration of the first component of the strain-displacement equations, subject to the displacement boundary condition at $x_1 = 0$, verifies Eq. (12.46). The remaining components of displacement, $u_2 = u_3$, as well as $\varepsilon_{22} = \varepsilon_{33}$, may be determined by a similar process.

The above very simple example demonstrates how direct analytic methods may on rare occasions be utilized to obtain solutions to viscoelastic IBVPs. Note also that this type of experiment contains a stress gradient in the x_1 coordinate direction, which may be useful for the purpose of predicting where microcracking will occur as a function of stress within asphalt road ways. More importantly, this problem can be utilized to obtain viscoelastic material properties, as will be demonstrated later in this chapter.

12.3.2 Separable Correspondence Principle

When the IBVP is either of Neumann or Dirichlet type (see Chap. 11), it is often possible to obtain solutions that are separable in time and space if all of the boundary conditions are proportional to one another in time (Alfrey 1944). In order to demonstrate how this simplification may be deployed, consider the case of a Neumann problem. The boundary conditions and body forces are assumed to be proportional to one another in time as follows:

$$t_i(\vec{x},t) = t_i^E(\vec{x})F(t) = known \quad \forall t \geq 0 \quad \forall \vec{x} \in S \tag{12.51a}$$

$$f_i(\vec{x},t) = f_i^E(\vec{x})F(t) = known \quad \forall t \geq 0 \quad \forall \vec{x} \quad \in V+S \tag{12.51b}$$

where $F(t)$ is an arbitrary function of time. In this case, the solution for the stress is assumed to be of the following separable form:

$$\sigma_{ij}(\vec{x},t) = \sigma_{ij}^E(\vec{x})F(t) \tag{12.52}$$

Note that the stress tensor is assumed to exhibit the same time dependence as the boundary tractions. Substitution of Eqs. (12.51b) and (12.52) into the equilibrium equations in Table 12.1 will result in the following:

$$\sigma_{ji,j}^E + \rho f_i^E = 0 \tag{12.53}$$

Similarly, substituting Eqs. (12.51b) and (12.52) into stress compatibility equations will result in the following:

$$\sigma_{ij,kk}^E + \frac{2(\lambda+\mu)}{(3\lambda+2\mu)}\sigma_{kk,ij}^E + \rho f_{i,j}^E + \rho f_{j,i}^E + \frac{\lambda}{(\lambda+2\mu)}\rho f_{k,k}^E \delta_{ij} = 0 \tag{12.54}$$

Since Eqs. (12.54) represent three independent equations (see Chap. 11), it can be shown that Eqs. (12.53) and (12.54), together with the time-independent part of boundary and body force condition Eqs. (12.51a, 12.51b), represents a time-independent boundary value problem that corresponds to the Neumann elasticity problem described in Chap. 11. Therefore, the solution to the above problem results in the determination of the unknown spatial part of the stress, $\sigma_{ij}^E(\vec{x})$, in Eq. (12.52). Since this amounts to a mathematics problem in applied elasticity, the result can oftentimes be found in an elasticity text. Assuming that this can be done, the elastic stress field, $\sigma_{ij}^E(\vec{x})$, may now be substituted into constitutive Eqs. (12.32) and (12.33) to obtain the following result:

$$\varepsilon_{kk}(\vec{x},t) = \int_0^t B(t-\tau)\frac{\partial\left[\sigma_{kk}^E(\vec{x})F(\tau)\right]}{\partial\tau}d\tau \tag{12.55}$$

$$\varepsilon'_{ij}(\vec{x},t) = \int_0^t J(t-\tau)\frac{\partial\left[\sigma_{ij}^{E\prime}(\vec{x})F(\tau)\right]}{\partial\tau}d\tau \tag{12.56}$$

Rearranging Eqs. (12.55) and (12.56) will thus result in

$$\varepsilon_{kk}(\vec{x},t) = \varepsilon_{kk}^E(\vec{x})M_1(t) \tag{12.57}$$

$$\varepsilon'_{ij}(\vec{x},t) = \varepsilon_{ij}^{E\prime}(\vec{x})M_2(t) \tag{12.58}$$

where

$$\varepsilon_{kk}^E(\vec{x}) = B^E\sigma_{kk}^E(\vec{x}) \tag{12.59}$$

$$\varepsilon_{ij}^{\prime E}(\vec{x}) = J^E\sigma_{ij}^{\prime E} \tag{12.60}$$

where B^E and J^E are the bulk and shear compliance, respectively, of the corresponding elastic BVP, and

$$M_1(t) \equiv \frac{1}{B^E}\int_0^t B(t-\tau)\frac{\partial[F(\tau)]}{\partial\tau}d\tau \tag{12.61}$$

and

$$M_2(t) \equiv \frac{1}{J^E}\int_0^t J(t-\tau)\frac{\partial[F(\tau)]}{\partial\tau}d\tau \tag{12.62}$$

Thus, the strain field may be determined by multiplying the solution to the elastic BVP by two rather straightforward functions of time, $M_1(t)$ and $M_2(t)$. Finally, the displacement may be obtained by integrating the strain-displacement equations in spatial coordinates.

Thus, it can be seen that when the boundary conditions and body forces are proportional to one another in time, the Neumann problem reduces to an elasticity problem multiplied by straightforward functions of time. It can be shown that a similar scenario results for the Dirichlet problem. The following example problem demonstrates the application of this technique.

Example Problem 12.2: Pressurized Linear Viscoelastic Cylinder
Given: Consider the right circular cylinder shown in Fig. 12.3.

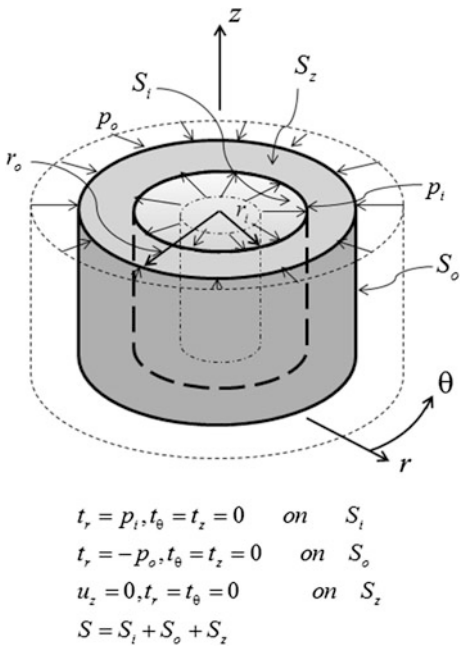


Fig. 12.3 Axisymmetric viscoelastic cylinder subjected to lateral pressures

Required: Predict the stresses, strains, and displacements as functions of the load, geometry of the cylinder, and the material properties.

Solution: By utilizing cylindrical coordinates, as described in Fig. 12.3, the field problem can be reduced mathematically to a single spatial dimension, r, so that the viscoelastic solution is a function of r and t. As shown in the figure, the cylinder is subjected to both internal pressure, p_i, and external pressure, p_o, normal to the inner and outer surfaces of the cylinder, respectively. In addition, it is assumed that the cylinder is restrained against motion in the z coordinate direction. This problem is therefore of mixed type. However, as pointed by Gabriel Lamé (1795–1870), who solved the elasticity problem in 1852 (Lamé 1852), the problem is a Neumann problem within the $r - \theta$ plane. As such, the solution to the viscoelastic IBVP may be obtained by correspondence to the elastic BVP.

Suppose that the internal and external pressures are functions of time, given by the following:

$$\begin{aligned} p_o &= -t_r(r = r_o) = p_o^E F(t) \\ p_i &= t_r(r = r_i) = p_i^E F(t) \end{aligned} \tag{12.63}$$

It is assumed that the solution is separable in space and time, and the spatial dependence can be modeled with the solution to the corresponding elastic

Table 12.2 Corresponding elastic field equations in cylindrical coordinates

Equilibrium
$\frac{\partial \sigma_{rr}^E}{\partial r} + \frac{\sigma_{rr}^E - \sigma_{\theta\theta}^E}{r} + \frac{1}{r}\frac{\partial \sigma_{r\theta}^E}{\partial \theta} + \frac{\partial \sigma_{rz}^E}{\partial z} = 0$ $\frac{\partial \sigma_{r\theta}^E}{\partial r} + \frac{1}{r}\frac{\partial \sigma_{\theta\theta}^E}{\partial \theta} + \frac{\partial \sigma_{\theta z}^E}{\partial z} + \frac{2\sigma_{r\theta}^E}{r} = 0$ $\frac{\partial \sigma_{zr}^E}{\partial r} + \frac{1}{r}\frac{\partial \sigma_{\theta z}^E}{\partial \theta} + \frac{\partial \sigma_{zz}^E}{\partial z} + \frac{\sigma_{zr}^E}{r} = 0$
Strain-displacement
$\varepsilon_{rr} = \frac{\partial u_r^E}{\partial r} \quad \varepsilon_{\theta\theta} = \frac{1}{r}\frac{\partial u_\theta^E}{\partial \theta} + \frac{u_r^E}{r} \quad \varepsilon_{zz} = \frac{\partial u_z^E}{\partial z}$ $\varepsilon_{r\theta} = \frac{1}{2}\left[\frac{\partial u_\theta^E}{\partial r} + \frac{1}{r}\frac{\partial u_r^E}{\partial \theta} - \frac{u_\theta^E}{r}\right]$ $\varepsilon_{rz} = \frac{1}{2}\left[\frac{\partial u_z^E}{\partial r} + \frac{\partial u_r^E}{\partial z}\right] \quad \varepsilon_{\theta z} = \frac{1}{2}\left[\frac{1}{r}\frac{\partial u_z^E}{\partial \theta} + \frac{\partial u_\theta^E}{\partial z}\right]$
Elastic constitutive equations
$\sigma_{rr}^E = \lambda\varepsilon_{kk}^E + 2G\varepsilon_{rr}^E$ $\sigma_{\theta\theta}^E = \lambda\varepsilon_{kk}^E + 2G\varepsilon_{\theta\theta}^E$ $\sigma_{zz}^E = \lambda\varepsilon_{kk}^E + 2G\varepsilon_{zz}^E$ $\sigma_{r\theta}^E = 2G\varepsilon_{r\theta}^E$ $\sigma_{rz}^E = 2G\varepsilon_{rz}^E$ $\sigma_{\theta z}^E = 2G\varepsilon_{\theta z}^E$ $\Leftrightarrow$ $\varepsilon_{rr}^E = (B-J)\sigma_{kk}^E + J\sigma_{rr}^E$ $\varepsilon_{\theta\theta}^E = (B-J)\sigma_{kk}^E + J\sigma_{\theta\theta}^E$ $\varepsilon_{zz}^E = (B-J)\sigma_{kk}^E + J\sigma_{zz}^E$ $\varepsilon_{r\theta}^E = J\sigma_{r\theta}^E$ $\varepsilon_{rz}^E = J\sigma_{rz}^E$ $\varepsilon_{\theta z}^E = J\sigma_{\theta z}^E$

BVP. Accordingly, the governing equations for the corresponding elasticity problem (see Table 11.1) are provided in cylindrical coordinates in Table 12.2 (assuming that body forces and inertial effects are negligible).

First assume that the in-plane stresses are separable in time and space as follows:

$$\begin{aligned} \sigma_{rr}(r,t) &= \sigma_{rr}^E(r)F(t) \\ \sigma_{\theta\theta}(r,t) &= \sigma_{\theta\theta}^E(r)F(t) \end{aligned} \tag{12.64}$$

All shear components of the stress tensor are zero due to symmetry. Under these assumptions, the equilibrium equations simplify to the following:

$$\frac{\partial \sigma_{rr}^E}{\partial r} + \frac{\sigma_{rr}^E - \sigma_{\theta\theta}^E}{r} = 0, \quad \frac{\partial \sigma_{\theta\theta}^E}{\partial \theta} = 0, \quad \frac{\partial \sigma_{zz}^E}{\partial z} = 0 \tag{12.65}$$

The third equilibrium equation, together with the boundary conditions on S_z(see Fig. 12.3), leads to the conclusion that the corresponding elastic stress component σ_{zz}^E is constant everywhere in the cylinder. The first two equations lead to the conclusion that the two remaining elastic stress components, σ_{rr}^E and $\sigma_{\theta\theta}^E$, depend only on r.

Substituting the first two constitutive equations into the first equilibrium equation and the first two strain-displacement equations into this equation results in the

following *Euler–Cauchy equation* (Kreyszig 2006) (after Leonhard Euler and Augustin-Louis Cauchy):

$$\frac{\partial^2 u_r^E}{\partial r^2} + \frac{1}{r}\frac{\partial u_r^E}{\partial r} - \frac{u_r^E}{r^2} = 0 \tag{12.66}$$

where u_r^E is the only nonzero component of the corresponding elastic displacement field.

The solution to Eq. (12.66) may be shown to be of the following form (Kreyszig 2006):

$$u_r^E(r) = c_1 r + \frac{c_2}{r} \tag{12.67}$$

The above equations may be substituted into the elastic strain-displacement equations to obtain the following:

$$\begin{aligned} \varepsilon_{rr}^E &= \frac{du_r^E}{dr} = c_1 - \frac{c_2}{r^2} \\ \varepsilon_{\theta\theta}^E &= \frac{u_r^E}{r} = c_1 + \frac{c_2}{r^2} \end{aligned} \tag{12.68}$$

Substituting the above into the constitutive equations and using this result to satisfy the traction boundary conditions on the cylindrical boundaries S_i and S_o (together with Cauchy's formula, Eq. (9.48)) will provide the values of c_1 and c_2 as follows:

$$c_1 = \frac{1}{2(\lambda^E + G^E)}\left[\frac{(p_i^E r_i^2 - p_o^E r_o^2)}{(r_o^2 - r_i^2)}\right], \quad c_2 = \frac{1}{2G^E}\left[\frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{(r_o^2 - r_i^2)}\right] \tag{12.69}$$

Substituting Eqs. (12.69) into Eqs. (12.68) thus results in the following elastic strain components:

$$\begin{aligned} \varepsilon_{rr}^E &= \frac{1}{2(\lambda^E + G^E)}\left[\frac{(p_i^E r_i^2 - p_o^E r_o^2)}{(r_o^2 - r_i^2)}\right] - \frac{1}{2G^E}\left[\frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{(r_o^2 - r_i^2)}\right]\frac{1}{r^2} \\ \varepsilon_{\theta\theta}^E &= \frac{1}{2(\lambda^E + G^E)}\left[\frac{(p_i^E r_i^2 - p_o^E r_o^2)}{(r_o^2 - r_i^2)}\right] + \frac{1}{2G^E}\left[\frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{(r_o^2 - r_i^2)}\right]\frac{1}{r^2} \end{aligned} \tag{12.70}$$

Equations (12.69) may also be substituted into Eq. (12.67), thus resulting in the elastic displacement field:

$$u_r^E(r) = \frac{1}{2(\lambda^E + G^E)}\left[\frac{(p_i^E r_i^2 - p_o^E r_o^2)}{(r_o^2 - r_i^2)}\right] r + \frac{1}{2G^E}\left[\frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{(r_o^2 - r_i^2)}\right]\frac{1}{r}$$
$$= \frac{(B^E + 2J^E)}{3B^E J^E}\left[\frac{(p_i^E r_i^2 - p_o^E r_o^2)}{(r_o^2 - r_i^2)}\right] r + J^E\left[\frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{(r_o^2 - r_i^2)}\right]\frac{1}{r} \tag{12.71}$$

where λ^E and G^E are arbitrary constants representing the elastic constants in the corresponding elastic BVP.

Equation (12.70) may be substituted into the elastic constitutive equations, with the result that

$$\sigma_{rr}^E(r) = \frac{p_i^E r_i^2 - p_o^E r_o^2}{(r_o^2 - r_i^2)} - \frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{r^2 (r_o^2 - r_i^2)}$$
$$\sigma_{\theta\theta}^E(r) = \frac{p_i^E r_i^2 - p_o^E r_o^2}{(r_o^2 - r_i^2)} + \frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{r^2 (r_o^2 - r_i^2)} \tag{12.72}$$
$$\sigma_{zz}^E = \lambda^E(\varepsilon_{rr}^E + \varepsilon_{\theta\theta}^E) = C^E\left[\frac{p_i^E r_i^2 - p_o^E r_o^2}{r_o^2 - r_i^2}\right]$$

where C^E is an elastic material constant that may be written in several different forms as follows:

$$C^E = \nu^E = \frac{\lambda^E}{\lambda^E + G^E} = \left[\frac{2(J^E - B^E)}{2J^E + B^E}\right] \tag{12.73}$$

These different forms will become important for the purpose of evaluating material properties later in this chapter. Note also that the two in-plane components of stress do not depend on the material properties.

This completes the determination of the field variables for the corresponding elastic BVP. In order to determine the variables in the viscoelastic IBVP, first substitute Eqs. (12.72) into (12.64). This process provides the time-dependent stress field:

$$\sigma_{rr}(r,t) = \left[\frac{p_i^E r_i^2 - p_o^E r_o^2}{(r_o^2 - r_i^2)} - \frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{r^2 (r_o^2 - r_i^2)}\right] F(t)$$
$$\sigma_{\theta\theta}(r,t) = \left[\frac{p_i^E r_i^2 - p_o^E r_o^2}{(r_o^2 - r_i^2)} + \frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{r^2 (r_o^2 - r_i^2)}\right] F(t) \tag{12.74}$$

In order to obtain the time-dependent displacement field, the elastic solution in Eq. (12.71) must be convoluted, thus resulting in the following:

$$u_r(r,t) = \left[\frac{(p_i^E r_i^2 - p_o^E r_o^2)}{(r_o^2 - r_i^2)}\right] r \int_0^t \left[\frac{B(t-\tau)+2J(t-\tau)}{3B(t-\tau)J(t-\tau)}\right] \frac{\partial F(\tau)}{\partial \tau} d\tau + \left[\frac{r_i^2 r_o^2 (p_i^E - p_o^E)}{(r_o^2 - r_i^2)}\right] \frac{1}{r} \int_0^t J(t-\tau) \frac{\partial F(\tau)}{\partial \tau} d\tau \tag{12.75}$$

The above may be differentiated using the strain-displacement equations to obtain the strain field. The difference in the resulting strains will thus be as follows:

$$\varepsilon_{\theta\theta}(r,t) - \varepsilon_{rr}(r,t) = \frac{1}{r^2}\left[\frac{r_i^2 r_o^2 (p_i^E - p_o^E}{(r_o^2 - r_i^2)}\right] \int_0^t J(t-\tau) \frac{\partial F(\tau)}{\partial \tau} d\tau \tag{12.76}$$

This completes the solution of the viscoelastic IBVP for the viscoelastic cylinder problem.

The above solution can be utilized to obtain viscoelastic material properties, to be described below.

12.3.3 Laplace Transform Correspondence Principles

Under quasi-static conditions, the boundary conditions and governing field equations described in Table 12.1 may be Laplace transformed, thus resulting in the following boundary value problem in Laplace-transformed space:

$$\begin{aligned} &\bar{\sigma}_{ji,j} + \bar{X}_i = 0 \\ &\bar{\varepsilon}_{ij} = \frac{1}{2}\left(\bar{u}_{i,j} + \bar{u}_{j,i}\right) \\ &\bar{\sigma}_{ij} = \tilde{\lambda}\,\bar{\varepsilon}_{kk}\delta_{ij} + 2\tilde{\mu}\bar{\varepsilon}_{ij} \\ &\bar{T}_i = \bar{\hat{T}}_i \quad on\ \partial\Omega_1 \\ &\bar{u}_i = \bar{\hat{u}}_i \quad on\ \partial\Omega_2 \end{aligned} \tag{12.77}$$

The above is essentially equivalent to the quasi-static linear elastic boundary value problem described in Table 11.1. Thus, if the solution to the corresponding elasticity problem is known, one need only place bars over the unknown variables, as well as replace the elastic material properties with the Carson-transformed viscoelastic material properties, and this solution is then the solution to the viscoelastic IBVP *in Laplace-transformed space*.

This method may look simple, but it is not as easy as the separable correspondence principle described in the previous section because one still has to perform the Laplace transform inversion in order to obtain the time-dependent

viscoelastic solution. In some cases, the Laplace transform inversion can be performed in closed form using the inverse properties of Laplace transform tables. However, in most problems this will not be the case. When no closed-form inversion is possible, there are two approximate options available: (1) Schapery's direct method and (2) collocation. These will be briefly described below.

12.3.3.1 The Direct Method for Laplace Transform Inversion

Schapery (1962a, b) developed an approximate method of Laplace transform inversion. This method, called the direct method, will not be covered in detail here. In order for it to be accurate, the log of the variable being transformed must exhibit small curvature when plotted against log t. If this is true, then

$$f(t) \underset{\sim}{=} s\bar{f}(s)\Big|_{s=\frac{0.56}{t} \underset{\sim}{=} \frac{1}{2t}} \tag{12.78}$$

This technique is very powerful due to its simplicity. However, it should be used with care, as accuracy is lost when the boundary conditions are nonmonotonic in time, thus leading to large curvatures of the field variables in log-log space.

12.3.3.2 The Collocation Method for Laplace Transform Inversion

Although called collocation (Christensen 1982), this method is actually a least squares technique. It is based on the assumption that the function $f(t)$ may be approximated by a Prony series:

$$f_{\mathrm{A}}(t) = \sum_{i=1}^{N} \mathrm{A}_i e^{-t/t_i} \tag{12.79}$$

where A_i are undetermined coefficients and t_i are predetermined constants. The total squared error is thus

$$E^2 = \int_0^{\infty} [f(t) - f_{\mathrm{A}}(t)]^2 dt \tag{12.80}$$

Minimizing the error results in

$$\frac{\partial E^2}{\partial \mathrm{A}_j} = \int_0^{\infty} \frac{\partial}{\partial \mathrm{A}_j} \left[f(t) - \sum_{i=1}^{N} \mathrm{A}_i e^{-t/t_i} \right]^2 dt = 0, j = 1, \ldots, N \tag{12.81}$$

or

$$\frac{\partial E^2}{\partial \mathrm{A}_j} = \int_0^\infty [f(t) - f_\mathrm{A}](-1)e^{-t/t_j}dt = 0, j = 1, \ldots, N \tag{12.82}$$

It can be seen that the above is the Laplace transform of $[f(t) - f_\mathrm{A}(t)]$ with $1/t_j \rightarrow s$. Thus, Eq. (12.82) may be written as follows:

$$[\bar{f}(s) - f_\mathrm{A}(s)]s \rightarrow 1/t_j = 0\, j = 1, \ldots, N \tag{12.83}$$

Now recall that from Table 9.2

$$\bar{f}_\mathrm{A}(s) = \sum_{i=1}^{N} \frac{\mathrm{A}_i}{s + (1/t_i)} \tag{12.84}$$

Therefore, Eq. (12.83) may be written as follows:

$$[\bar{f}(s)]s \rightarrow 1/t_j = \left[\sum_{i=1}^{N} \frac{\mathrm{A}_i}{(s + 1/t_j)}\right]\Bigg|_{s \rightarrow 1/t_j} \qquad j = 1, \ldots, N \tag{12.85}$$

The above is a set of linear equations that can be used to solve for the coefficients, $\mathrm{A_i}$, thereby resulting in the approximate Laplace transform inverse given in Eq. (12.79). This technique should also be utilized with care, as it sometimes results in large oscillations of the approximated function in log-log space.

12.4 Material Property Characterization of Viscoelastic Media

Recall that the purpose of a *material property characterization experiment* is to construct a test within the laboratory so that the IBVP described in Table 12.1 can be solved for one or more material properties rather than the stresses, strain, and displacements. Such a problem is called *an inverse problem* because displacements measured on the boundary of the specimen can be utilized to determine material properties within the laboratory specimen.

As described in Chap. 11, characterization of material properties for linear elastic media is usually straightforward. This is due at least in part to the fact that for elastic media the properties are constant in time, thus resulting in the terminology *material constants*. Unfortunately, the same cannot be said for linear viscoelastic media. For this (as well as other) reasons, characterization of material

properties in viscoelastic media can be quite cumbersome. These tests usually come in one of five different types: creep tests; ramp tests; relaxation tests; frequency sweeps; and time–temperature superposition tests. Each of these tests will be covered in some detail below.

12.4.1 Creep Tests

12.4.1.1 Uniaxial Creep Test

In this test, a coupon is constructed and loads are applied in such a way as to produce spatially constant stresses throughout the specimen as follows:

$$\sigma_{11}(t) = \sigma^0 H(t) \quad \sigma^0 = const, \quad \sigma_{ij} = 0 \quad \forall \quad other\ i,j \tag{12.86}$$

where $H(t)$ is the Heaviside step function. It can be seen that this experiment is a special case of the IBVP described in Example Problem 12.1, where in this case the specimen is prismatic (meaning the cross section of the bar does not change in the x_1 coordinate direction). Note that whereas the cross-sectional shape of the specimen is arbitrary in Example Problem 12.1, it is advisable that the cross section be either rectangular or circular so that the transverse strains are easily determined from the measured displacements during the experiment.

Recall that the displacements may be measured on the boundary of the specimen, thus resulting in the determination of $\varepsilon_{11}(t)$, $\varepsilon_{22}(t)$, and $\varepsilon_{33}(t)$ via strain-displacement Eq. (12.8) (due to the fact that the stress in the specimen is uniaxial and spatially homogeneous). The longitudinal stress, $\sigma_{11}(t)$, may be calculated by utilizing Hooke's Law in conjunction with a load cell (see Fig. 10.2). Note that the transverse strains are equivalent and the shear strains are negligible in this test if the material is isotropic.

Substituting Eq. (12.86) into Eq. (12.31) results in the following:

$$\varepsilon_{11}(t) = [\frac{1}{3}B(t) + \frac{2}{3}J(t)]\sigma^0 \tag{12.87}$$

and

$$\varepsilon_{22}(t) = \varepsilon_{33}(t) = \frac{1}{3}[B(t) - J(t)]\sigma^0 \tag{12.88}$$

Solving Eqs. (12.87) and (12.88) for $J(t)$ and $B(t)$ results in

$$J(t) = \frac{\varepsilon_{11}(t) - \varepsilon_{22}(t)}{\sigma^0} \tag{12.89}$$

and

$$B(t) = \frac{\varepsilon_{11}(t) + \varepsilon_{22}(t) + \varepsilon_{33}(t)}{\sigma^0} \tag{12.90}$$

Thus, it can be seen that the experiment described by Eq. (12.44), called a uniaxial creep test, leads quite naturally to the determination of both the shear and bulk compliance. Using curve-fitting techniques (to be described below), these functions of time can be constructed in such a way that the shear and bulk modulus can be calculated using Eqs. (12.89) and (12.90). Note also that the axial loading can be applied in either tension or compression so long as the specimen does not buckle or crack, thereby introducing spatial gradients in the stress.

The reader will recall from Chap. 10 that a uniaxial creep test can be used to obtain the uniaxial creep compliance, $D(t)$, as follows:

$$D(t) \equiv \frac{\varepsilon_{11}(t)}{\sigma_0} \tag{12.91}$$

as well as Poisson's ratio:

$$\nu^C(t) \equiv \frac{-\varepsilon_{22}(t)}{\varepsilon_{11}(t)} \tag{12.92}$$

where the superscript, C, has been added to denote that the value is obtained from a creep test. Poisson's ratio, $\nu^R(t)$, may also be obtained from a uniaxial relaxation test, and these two parameters are not necessarily equivalent. The above two properties may be related to other viscoelastic properties, as described previously in this chapter.

The uniaxial creep test described by Eq. (12.86) is the simplest test to perform mechanically, and as shown above, it is also the simplest means of quantifying the viscoelastic material properties for an isotropic linear viscoelastic material. Unfortunately, it is often not pragmatic to perform a uniaxial creep test on geologic media because the material in question may not be capable of supporting uniaxial loading in either tension or compression. Thus, other more complicated tests may be useful for characterizing the material properties.

12.4.1.2 Multi-axial Creep Test

For geologic media, it is often useful to apply a lateral confining pressure in addition to the axial load in order to preserve the specimen against crumbling during the test. There are two slightly different configurations that are utilized, either a rectangular prismatic specimen, or a cylindrical prismatic specimen.

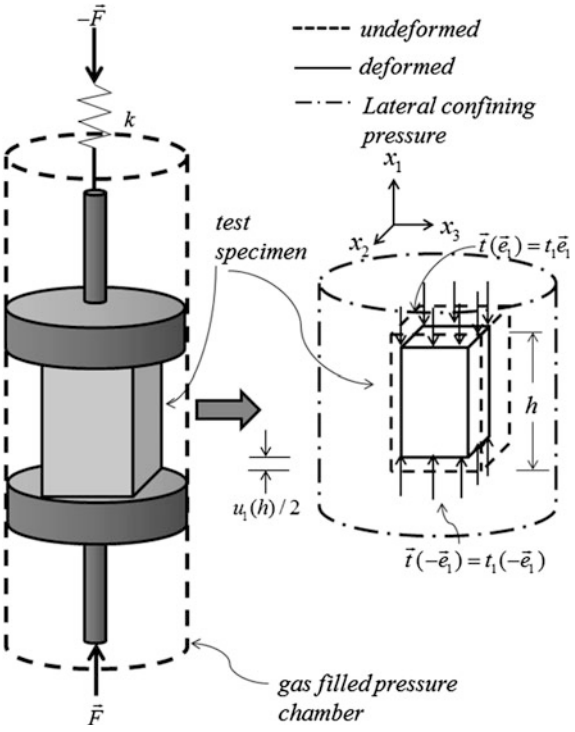


Fig. 12.4 Depiction of a multi-axial creep test using a rectangular prismatic specimen

Case 1: Rectangular Prismatic Specimen

In this test, the cross section of the specimen is rectangular, and the specimen is prismatic in the primary loading direction, as shown in Fig. 12.4. Loads are applied so as to produce the following stress state within the specimen:

$$\begin{aligned} &\sigma_{11}(t) = \sigma_0 H(t) && \sigma_0 = const \\ &\sigma_{22}(t) = \sigma_{33}(t) = -p_L H(t) && p_L = const \quad \sigma_{ij} = 0 \quad i \neq j \end{aligned} \tag{12.93}$$

where p_L is called the lateral pressure, which is assumed to be positive in compression (whereas the longitudinal stress, σ_0, is assumed to be positive in tension). Assuming that the specimen is perfectly lubricated at the platen interfaces, the stresses will be spatially homogeneous in the test specimen, as described by Eq. (12.93).

In this case, substituting Eq. (12.93) into Eq. (12.31) results in the following:

$$\varepsilon_{11}(t) = J(t)\sigma_0 + \frac{1}{3}[B(t) - J(t)](\sigma_0 - 2p_L) \tag{12.94}$$

$$\varepsilon_{22}(t) = \varepsilon_{33}(t) = -J(t)p_L + \frac{1}{3}[B(t) - J(t)](\sigma_0 - 2p_L) \tag{12.95}$$

Solving the above two equations for $J(t)$ and $B(t)$ results in the following:

$$J(t) = \frac{\varepsilon_{11}(t) - \varepsilon_{22}(t)}{(\sigma_0 + p_L)} \tag{12.96}$$

and

$$B(t) = \frac{\varepsilon_{11}(t) + \varepsilon_{22}(t) + \varepsilon_{33}(t)}{(\sigma_0 - 2p_L)} \tag{12.97}$$

Case 2: Cylindrical Prismatic Specimen

Geologic specimens are commonly cored using a saw with a cylindrical blade, thereby producing cylindrical specimens, as shown in Fig. 12.5. In this test, the specimen is subjected to the loading shown in the figure.

Tests are commonly performed on specimens of this shape because they are easy to produce. As it turns out, the stress state in Cartesian coordinates is identical to that found in the rectangular cylindrical specimen described above. This fact may be illustrated by superposing the results obtained in Example Problems 12.1 and 12.2. The resulting stresses, for the case wherein $p_i = r_i = 0$ in Example Problem 12.2 are found to be as follows:

$$\begin{aligned} \sigma_{zz}(t) &= p_z H(t) && p_z = const \\ \sigma_{rr}(t) &= \sigma_{\theta\theta}(t) = -p_r H(t) && p_r = const \end{aligned} \tag{12.98}$$

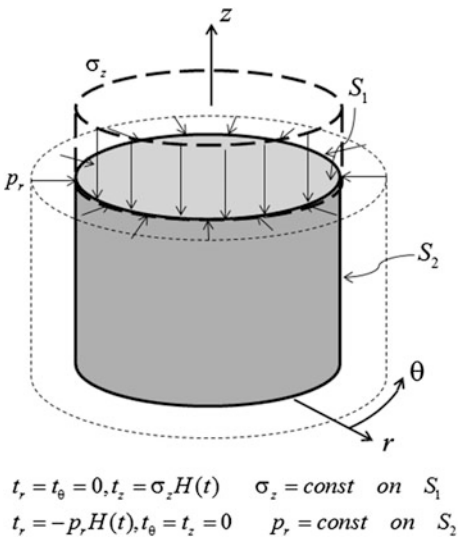


Fig. 12.5 Depiction of a multi-axial creep test using a cylindrical prismatic specimen

It follows that

$$\varepsilon_{zz}(t) = J(t)\sigma_z + \frac{1}{3}[B(t) - J(t)](\sigma_{zz} - 2p_r) \tag{12.99}$$

$$\varepsilon_{rr}(t) = \varepsilon_{\theta\theta}(t) = -J(t)p_r + \frac{1}{3}[B(t) - J(t)](\sigma_{zz} - 2p_r) \tag{12.100}$$

The resulting material properties are thus given by the following:

$$J(t) = \frac{\varepsilon_{zz}(t) - \varepsilon_{\theta\theta}(t)}{(\sigma_{zz} + p_r)} \tag{12.101}$$

$$B(t) = \frac{\varepsilon_{zz}(t) + \varepsilon_{rr}(t) + \varepsilon_{\theta\theta}(t)}{(\sigma_{zz} - 2p_r)} \tag{12.102}$$

Thus, it can be seen that while multi-axial creep tests are significantly more complicated experimentally, they lead to relatively straightforward formulas for determining the compliances. Furthermore, experiments of this type are relatively simple to perform on asphaltic pavement.

12.4.2 Ramp Tests

12.4.2.1 Uniaxial Ramp Test

In this experiment, a longitudinally applied traction is applied to the specimen, whereas the specimen is unconfined laterally in such a way that the stress state in the specimen is spatially homogeneous and uniaxial, as in the uniaxial creep test. However, contrary to the uniaxial creep test, the longitudinal traction is increased at constant rate, thereby producing the following stress state within the test specimen:

$$\sigma_{11}(t) = \dot{\sigma}^0 t H(t) \quad \dot{\sigma}^0 = const \quad \sigma_{ij} = 0 \quad \forall \quad \text{other i, j} \tag{12.103}$$

For this case, substituting Eq. (12.103) into Eq. (12.31) results in the following:

$$\varepsilon_{11}(t) = \int_0^t J(t-\tau)\dot{\sigma}^0 d\tau + \frac{1}{3}\int_0^t [B(t-\tau) - J(t-\tau)]\dot{\sigma}^0 d\tau \tag{12.104}$$

and

$$\varepsilon_{22}(t) = \varepsilon_{33}(t) = \frac{1}{3}\int_0^t [B(t-\tau) - J(t-\tau)]\dot{\sigma}^0 d\tau \tag{12.105}$$

It can be seen in this case that the integration cannot be carried out unless the precise form of the moduli is known a priori. Practice has shown that these may be modeled with a Prony series (to be discussed in further detail below). Thus, they are given by the following general equations:

$$B(t) = B_0 + \sum_{i=1}^{n} B_i(1 - e^{t/\eta_i^B}) \tag{12.106}$$

where B_0, B_i, η_i^B, i = 1, …, n are material properties, and

$$J(t) = J_0 + \sum_{i=1}^{n} J_i(1 - e^{t/\eta_i^J}) \tag{12.107}$$

where J_0 and J_i, η_i^J, i = 1, …, n, are material properties. The procedure for obtaining these properties from experimental data will be described below. For now, substituting Eqs. (12.106) and (12.107) into Eqs. (12.104) and (12.105) results in the following equations:

$$\varepsilon_{11}(t) = \left[\frac{2}{3}J'(t) + \frac{1}{3}B'(t)\right]\dot{\sigma}^0 \tag{12.108}$$

and

$$\varepsilon_{22}(t) = \varepsilon_{33}(t) = \frac{1}{3}[B'(t) - J'(t)]\dot{\sigma}^0 \tag{12.109}$$

where

$$B'(t) \equiv B_0 + \sum_{i=1}^{n}\left[\left(B_0 + \sum_{i=1}^{n} B_i\right)t - \sum_{i=1}^{n}\eta_i^B B_i(1 - e^{-t/\eta_i^B}\right] \tag{12.110}$$

and similarly

$$J'(t) \equiv J_0 + \sum_{i=1}^{n}\left[\left(J_0 + \sum_{i=1}^{n} J_i\right)t - \sum_{i=1}^{n}\eta_i^J J_i(1 - e^{-t/\eta_i^J}\right] \tag{12.111}$$

It can be seen by examination of Eqs. (12.106) and (12.107), together with Eqs. (12.110) and (12.111) that if the coefficients in Eqs. (12.110) and (12.111) are known, then Eqs. (12.106) and (12.107) are also characterized. Thus, the procedure

is to utilize the results of the uniaxial ramp test to characterize Eqs. (12.110) and (12.111). This can be accomplished by solving Eqs. (12.108) and (12.109) to obtain the following two equations:

$$B'(t) = \left[\frac{\varepsilon_{11}(t) + \varepsilon_{22}(t) + \varepsilon_{33}(t)}{\dot{\sigma}^0}\right] \tag{12.112}$$

and

$$J'(t) = \left[\frac{\varepsilon_{11}(t) - \varepsilon_{22}(t)}{\dot{\sigma}^0}\right] \tag{12.113}$$

Thus, it can be seen that it is possible to obtain the material properties from a uniaxial ramp test.

12.4.2.2 Multi-axial Ramp Test

In this experiment, the longitudinally applied stress is increased at a constant rate, similar to the uniaxial ramp test, except that in this test a lateral pressure is applied and held constant in time. Thus, the experimental profile is as follows:

$$\sigma_{11}(t) = \dot{\sigma}^0 t H(t) \quad \dot{\sigma}^0 = const$$
$$\sigma_{22}(t) = \sigma_{33}(t) = -p^L H(t) \quad p^L = const \quad \sigma_{ij} = 0 \quad i \neq j \tag{12.114}$$

Substituting Eq. (12.114) into Eq. (12.31) results in the following:

$$\varepsilon_{11}(t) = \int_0^t J(t-\tau)\dot{\sigma}^0 d\tau + \frac{1}{3}\int_0^t [B(t-\tau) - J(t-\tau)]\dot{\sigma}^0 d\tau - \frac{2}{3}[B(t) - J(t)]p^L \tag{12.115}$$

and

$$\varepsilon_{22}(t) = \varepsilon_{33}(t) = -J(t)p^L + \frac{1}{3}\int_0^t [B(t-\tau) - J(t-\tau)]\dot{\sigma}^0 d\tau - \frac{2}{3}[B(t) - J(t)]p^L \tag{12.116}$$

Simplifying the above results in the following:

$$\varepsilon_{11}(t) = \left[\frac{2}{3}J'(t) + \frac{1}{3}B'(t)\right]\dot{\sigma}^0 - \frac{2}{3}[B(t) - J(t)]p^L \tag{12.117}$$

and

$$\varepsilon_{22}(t) = \varepsilon_{33}(t) = \frac{1}{3}[B'(t) - J'(t)]\dot{\sigma}^0 - \left[\frac{2}{3}B(t) + \frac{1}{3}J(t)\right]p^L \tag{12.118}$$

The above two equations may be rewritten as follows:

$$\frac{\varepsilon_{11}(t)}{\dot{\sigma}^0} = \frac{2}{3}\left[J'(t) + \frac{p^L}{\dot{\sigma}^0}J(t)\right] + \frac{1}{3}\left[B'(t) - 2\frac{p^L}{\dot{\sigma}^0}B(t)\right] \tag{12.119}$$

and

$$\frac{\varepsilon_{22}(t)}{\dot{\sigma}^0} = -\frac{1}{3}\left[J'(t) + \frac{p^L}{\dot{\sigma}^0}J(t)\right] + \frac{1}{3}\left[B'(t) - 2\frac{p^L}{\dot{\sigma}^0}B(t)\right] \tag{12.120}$$

Now define two new quantities as follows:

$$J''(t) \equiv J'(t) + \frac{p^L}{\dot{\sigma}^0}J(t) \tag{12.121}$$

and

$$B''(t) \equiv B'(t) - 2\frac{p^L}{\dot{\sigma}^0}B(t) \tag{12.122}$$

Substituting Eqs. (12.121) and (12.122) into Eqs. (12.119) and (12.120) thus results in the following:

$$\frac{\varepsilon_{11}(t)}{\dot{\sigma}^0} = \frac{2}{3}J''(t) + \frac{1}{3}B''(t) \tag{12.123}$$

and

$$\frac{\varepsilon_{22}(t)}{\dot{\sigma}^0} = -\frac{1}{3}J''(t) + \frac{1}{3}B''(t) \tag{12.124}$$

Solving the above two equations results in the following:

$$B''(t) = \frac{9}{4}\left[\frac{\varepsilon_{11}(t) + \varepsilon_{22}(t)}{\dot{\sigma}^0}\right] \tag{12.125}$$

and

$$J''(t) = \frac{1}{8}\left[\frac{3\varepsilon_{11}(t) - 9\varepsilon_{22}(t)}{\dot{\sigma}^0}\right] \tag{12.126}$$

The above two equations can be used to obtain $B''(t)$ and $J''(t)$. It can be seen that characterizing these two equations also results in the properties required in $B(t)$ and $J(t)$. Thus, once again it is possible to characterize the material properties, this time utilizing a multi-axial ramp test. However, it should be noted that ramp tests, while they may be attractive to perform in the laboratory, result in complicated mathematical formulas that must be solved in order to characterize the linear viscoelastic properties. For this reason, creep tests, whether uniaxial or multi-axial, are preferable for the purpose of characterizing material properties in viscoelastic media.

12.4.3 Relaxation Tests

Relaxation tests are normally performed in displacement control. This type of test may be difficult to construct in the laboratory, but the rewards may justify the effort.

12.4.3.1 Uniaxial Relaxation Test

It is possible with modern testing machinery to perform a so-called uniaxial relaxation test, described by

$$\varepsilon_{11}(t) = \varepsilon^0 H(t) \quad \varepsilon^0 = const \quad \sigma_{22} = \sigma_{33} = 0 \quad \varepsilon_{ij} = 0 \quad i \neq j \tag{12.127}$$

While this test may appear at first glance to be somewhat straightforward, it is no simple matter to apply an instantaneous displacement and subsequently hold it constant in time in a laboratory test. Furthermore, if it is observed in this test that the lateral strain changes in time, then it will not be possible to extract material properties directly from the experiment. In the case where it is observed that the lateral strain is constant in time, then it will follow that

$$\varepsilon_{22} = \varepsilon_{33} = -\nu^R \varepsilon^0 H(t) \tag{12.128}$$

where v^R is the relaxation Poisson's ratio, which in this case is found to be constant in time.

In this experiment, Eqs. (12.127) and (12.128) are substituted in strain formulation (12.27) with the result that

$$K(t) = \frac{\sigma_{11}(t)}{3(1-2\nu^R)\varepsilon^0} \tag{12.129}$$

Utilizing Eqs. (12.25) and (12.28) will result in

$$\mu(t) = \frac{3(1-2\nu^R)}{2(1+\nu^R)} K(t) \tag{12.130}$$

The above test thus results in a relatively simple set of formulas for evaluating the relaxation moduli. However, it should be pointed out that this test only works when the relaxation Poisson's ratio, ν^R, is constant in time.

12.4.3.2 Pure Shear Relaxation Test

Rather than perform a uniaxial relaxation test, it is much simpler to perform relaxation tests in pure shear, and for this reason, pure shear tests are much more common for the purpose of obtaining viscoelastic moduli. This test is depicted in Fig. 11.3a, and in this case, the specimen is subjected to an instantaneous rotation such that the strain state in the specimen is spatially homogeneous at the surface of the specimen and given by the following:

$$\varepsilon_{12} = \varepsilon^0 t \quad \varepsilon^0 = const \quad \varepsilon_{ij} = 0 \quad \forall \quad other\, i,j \tag{12.131}$$

The output stress as a function of time can be calculated using Eq. (11.41), where it should be apparent that the torque, T, required to hold the rotational displacement constant in time, will gradually diminish in time. In this case, substitution of Eq. (12.131) into Eq. (12.25) will result in the following:

$$\mu(t) = \frac{\sigma_{12}(t)}{2\varepsilon^0} \tag{12.132}$$

This is one of the simplest tests that can be performed in the laboratory if the aim is to extract moduli directly rather than compliances.

12.4.4 Accelerated Characterization Tests

The material property tests discussed up to this point have necessarily suffered from the limitation that the material properties are determined only up to the time when the test is halted. This is a significant problem when attempting to model long-term performance of flexible pavements because these pavements can last ten, twenty, or even up to fifty years. Thus, in order to characterize the viscoelastic material properties using creep, ramp, and relaxation tests, it is necessary to perform the

material characterization experiments for up to fifty years! Performing such long-term tests is of course not practical. Fortunately, for linear viscoelastic media there are some experimental methods that may be employed that result in what is termed *accelerated characterization.*

12.4.4.1 Frequency Sweeps

The first accelerated technique involves subjecting an experimental specimen to steady vibrations under displacement control, thereby resulting in strains that are separable in space and time, as given by the following (Christensen 1982):

$$\varepsilon_{ij}(\vec{x},t) = \varepsilon_{ij}^{0}(\vec{x})e^{i\omega t} \tag{12.133}$$

where ω is the frequency of vibration.

Of course, the solution to the resulting IBVP must be known so that the spatial dependence in Eq. (12.133) must necessarily be predetermined. For linear viscoelastic media, there exists a closed-form mathematical transformation from the frequency domain to the time domain, so that short-term tests can be performed in the frequency domain, and the results of these tests can be used to construct long-term properties in the time domain.

To see how this works, it is first necessary to introduce a bit of mathematics involving complex variables. First note that if Eq. (12.133) is substituted into (12.29), this result may be substituted into Eq. (12.30), thus resulting in the following equation (after some manipulation):

$$\sigma_{ij}'(t) = 2\mu_{\infty}\varepsilon_{ij}^{0\prime}e^{i\omega t} + 2i\omega\varepsilon_{ij}^{0\prime}\int_{-\infty}^{t}\mu_1(t-\tau)e^{i\omega\tau}d\tau \tag{12.134}$$

where the following notation has been introduced:

$$\mu(t) = \mu_{\infty} + \mu_1(t) \tag{12.135}$$

where, in addition, *the long-term shear modu*lus is given by

$$\mu_{\infty} \equiv \lim_{t\to\infty}\mu(t) \tag{12.136}$$

Next introduce the following change of variables:

$$\eta \equiv t - \tau \Rightarrow d\eta = -d\tau \tag{12.137}$$

Substituting (12.137) into (12.134) and performing further manipulation results in the following:

$$\sigma'_{ij}(t) = 2\mu_\infty \varepsilon^{0\prime}_{ij} e^{i\omega t} + 2i\omega\varepsilon^{0\prime}_{ij} e^{i\omega t} \int_0^\infty \mu_1(\eta) e^{-i\omega\eta} d\eta \tag{12.138}$$

Now recall from complex variables that

$$e^{-i\omega\eta} = \cos\omega\eta - i\,\sin\omega\eta \tag{12.139}$$

Thus, substituting (12.138) into (12.139) results in the following:

$$\sigma'_{ij}(t) = 2\mu^*(i\omega)\varepsilon'_{ij} \tag{12.140}$$

where *the complex modulus*, μ^*, is defined as follows:

$$\mu^*(i\omega) \equiv \mu'(\omega) + i\mu''(\omega) \tag{12.141}$$

where *the storage modulus*, μ', is given by

$$\mu'(\omega) \equiv \mu_\infty + \omega \int_0^\infty \mu_1(\eta)\sin\omega\eta d\eta \tag{12.142}$$

and *the loss modulus*, μ'', is given by

$$\mu''(\omega) = \omega \int_0^\infty \mu_1(\eta)\cos\omega\eta d\eta \tag{12.143}$$

Now recall that Eq. (12.141) may be written alternatively as follows:

$$\mu^* = \mu' + i\mu'' = |\mu^*| e^{i\varphi} \tag{12.144}$$

where the angle, φ, called *the loss tangent*, is given by

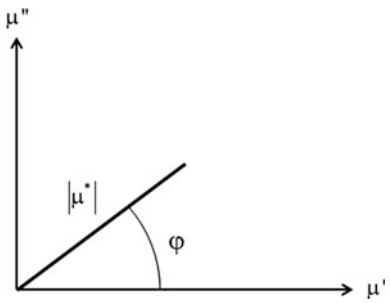


Fig. 12.6 Argand diagram for the complex modulus

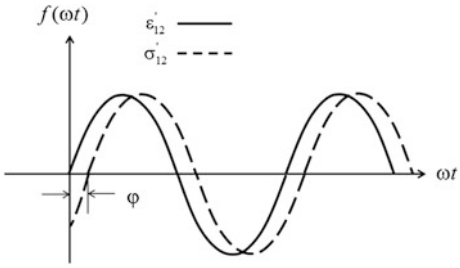


Fig. 12.7 Illustration of the loss tangent in a linear viscoelastic material

$$\varphi \equiv \tan^{-1} \frac{\mu''}{\mu'} \tag{12.145}$$

and

$$|\mu^*| = \left[(\mu')^2 + (\mu'')^2\right]^{1/2} \tag{12.146}$$

The above representation for the complex modulus can perhaps be better understood in terms of an *Argand diagram* (after Jean-Robert Argand 1768–1822) for complex variables (Greenberg 1978), as shown in Fig. 12.6.

Substituting Eq. (12.144) into Eq. (12.140) now results in the following:

$$\sigma'_{ij}(t) = 2|\mu^*|\varepsilon^{0\prime}_{ij} e^{i(\omega t + \varphi)} \tag{12.147}$$

The above equation illustrates the fact that in linear viscoelastic media subjected to steady state displacement imposed vibrations the stress lags the strain by the loss tangent, φ, as shown in Fig. 12.7.

It is possible to perform laboratory tests in such a way as to control the frequency of the input displacements in time, so that a sweep over a very broad range of frequencies can be performed. This test, called *a frequency sweep test*, will result in output storage and loss modulus curves of the general shape shown in Fig. 12.8, where it can be shown that

$$\mu_\infty = \lim_{\omega \to 0} \mu'(\omega) \tag{12.148}$$

as shown in the figure.

Either of the curves (they are uniquely related) shown in Fig. 12.8 can be curve fit within the frequency domain, and the procedure for doing this will be described later. However, once this process has been completed, it is possible to convert the solution to the time domain using the following transformations (Christensen 1982):

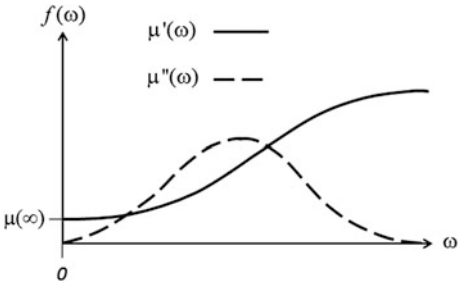


Fig. 12.8 Depiction of output storage and loss moduli for a typical frequency sweep performed on a linear viscoelastic material

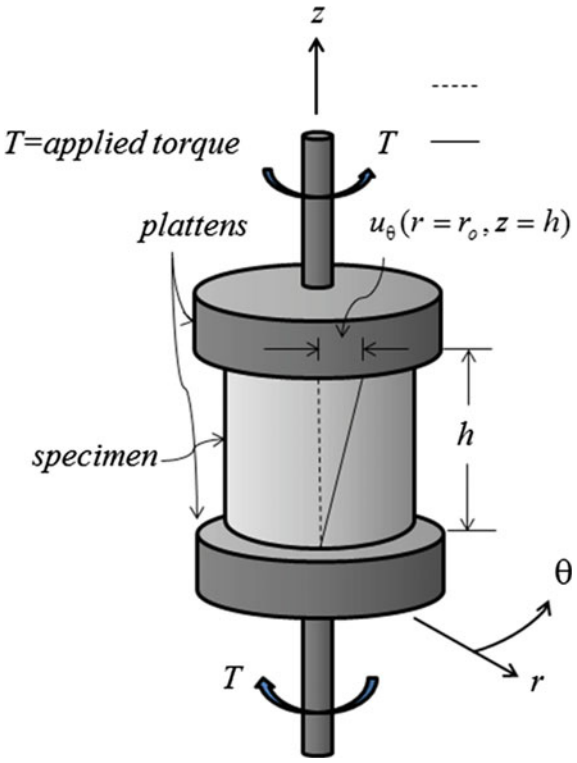


Fig. 12.9 Test configuration for a frequency sweep in pure shear

$$\mu_1(t) = \frac{1}{\pi}\int_0^\infty \left[\frac{\mu'(\omega) - \mu_\infty}{\omega}\right] \sin \omega t d\omega \tag{12.149a}$$

and

$$\mu_1(t) = \frac{2}{\pi}\int_0^\infty \frac{\mu''(\omega)}{\omega}\cos\omega t d\omega \tag{12.149b}$$

Thus, it can be seen that by performing a short-term frequency sweep in the laboratory, long-term properties can be obtained for linear viscoelastic media. This then is a powerful tool for obtaining viscoelastic properties. The reader will note that Eqs. (12.149a and 12.149b) may be circumvented by utilizing the procedure described previously in this chapter.

The most commonly used test configuration for a frequency sweep is a displacement-controlled torsional oscillation test performed on a right circular cylinder, as shown in Fig. 12.9.

The circumferential displacement component, u_θ, is assumed to be of the form:

$$u_\theta(r,z,t) = rf(z)e^{i\omega t} \tag{12.150}$$

The governing differential equation for the circumferential displacement component, u_θ, is given by Christensen (1982):

$$\frac{\partial^2 u_\theta}{\partial r^2} + \frac{1}{r}\frac{\partial u_\theta}{\partial r} + \frac{\partial^2 u_\theta}{\partial z^2} - \frac{u_\theta}{r^2} = \frac{\rho}{\mu*}\frac{\partial^2 u_\theta}{\partial t^2} \tag{12.151}$$

Substituting Eq. (12.150) into Eq. (12.151) thus results in the following equation:

$$\frac{\partial^2 f}{\partial z^2} + \frac{\rho\omega^2}{\mu*} = 0 \tag{12.152}$$

The solution to the above equation is assumed to be of the form:

$$f(z) = A\,\sin\frac{az}{h} + B\,\cos\frac{bz}{h} \tag{12.153}$$

Substituting Eq. (12.153) into Eq. (12.150) results in

$$f(z) = A\,\sin\frac{\Omega z}{h} + B\,\cos\frac{\Omega z}{h} \tag{12.154}$$

where

$$\Omega \equiv \sqrt{\frac{\rho\omega^2 h^2}{\mu*}} \tag{12.155}$$

Substituting Eq. (12.154) into Eq. (12.150) now results in the following:

$$u_\theta(r,z,t) = r\left(A\,\sin\frac{\Omega z}{h} + B\,\cos\frac{\Omega z}{h}\right)e^{i\omega t} \tag{12.156}$$

Applying the boundary condition $u_\theta(z=0)=0$ simplifies the above to the following:

$$u_\theta(r,z,t) = r\left(A\,\sin\frac{\Omega z}{h}\right)e^{i\omega t} \tag{12.157}$$

Now note that the only nonzero component of stress is $\sigma_{\theta z}$, which can be obtained from the constitutive equation:

$$\sigma_{\theta z} = 2\mu^*\varepsilon_{\theta z} = \mu^*\frac{\partial u_\theta}{\partial z} \tag{12.158}$$

The applied torque, T, can thus be obtained by integrating the stress over the cross section as follows:

$$T(z,t) = \int_0^{2\pi}\int_{r_i}^{r_o}\mu^*\frac{\partial u_\theta}{\partial z}r^2 dr d\theta \tag{12.159}$$

Substituting Eq. (12.157) into Eq. (12.159) thus results in the following:

$$T(z,t) = \mu^* J\left(\frac{A\Omega}{h}\cos\frac{\Omega z}{h}\right) \tag{12.160}$$

where J is the polar moment of inertia, given by

$$J \equiv \int_0^{2\pi}\int_{r_i}^{r_o} r^3 dr d\theta = \frac{\pi}{2}\left(r_o^4 - r_i^4\right) \tag{12.161}$$

The angle of twist, Φ, of the upper platen is found to be as follows:

$$\Phi(z=h,t) = \frac{u_z}{r} = A(\sin\Omega)e^{i\omega t} \tag{12.162}$$

The complex modulus can therefore be obtained by dividing Eq. (12.160) by Eq. (12.162), thus resulting in the following:

$$\mu^* = \frac{T(h,t)}{\Phi(h,t)}\frac{h\tan\Omega}{J} \tag{12.163}$$

Since both T and Φ can be measured during the experiment, the above equation presents a means of evaluating the complex shear modulus. As described below, a convenient procedure is available for obtaining the shear relaxation modulus from the complex shear modulus.

In some cases, researchers have constructed experiments for determining viscoelastic material properties in which cyclic displacements are applied in the axial direction and constant pressure is applied in the lateral directions. It can be seen that this type of test is mixed, in the sense that neither the strain nor the stress formulation for the material properties can be utilized to extract material properties. Thus, this type of experiment is not pragmatic when attempting to extract viscoelastic properties experimentally.

12.4.5 Time–Temperature Superposition Tests

As mentioned above, long-term laboratory tests should be avoided for reasons of cost and practicality. Therefore, it is advisable to develop methodologies that allow short-term tests to be utilized to predict long-term response. The previous section introduced one method. In this section, an entirely different approach is taken. Instead, it is assumed that a series of relatively short-term tests may be performed at different but constant temperatures, material properties may be extracted from the tests performed at these temperatures, and these may be subsequently shifted along the time axis to produce a single long-term material property at a single temperature. Materials that demonstrate this type of behavior are called *thermorheologically simple* (Leaderman 1943; Ferry 1980).

In order to illustrate how this procedure works, it will be demonstrated for the case of uniaxial relaxation tests, as described by Eq. (12.127). Suppose that several specimens are subjected to relaxation tests at various temperatures, and the results

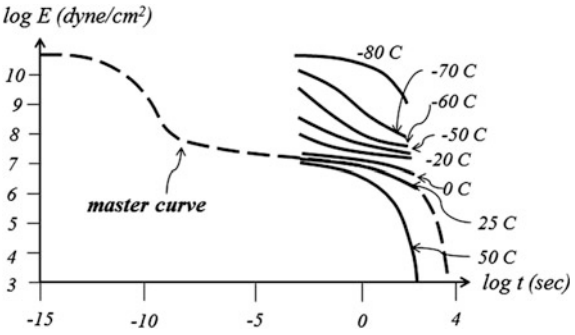


Fig. 12.10 Typical experimental creep data for a thermorheologically simple material (Christensen 1982)

of these tests are as shown in Fig. 12.10, where it should be noted that the horizontal scale is the logarithm of time.

Note from the figure that all of the tests are performed over the range of $10^{-3} < t < 10^{2}$ s. In the case shown, the data are shifted horizontally to a so-called *master curve* at a temperature of 25 °C. The data obtained for temperatures above 25 °C are shifted to the right, whereas data taken at temperatures below 25 °C are shifted to the left.

The above results may be represented mathematically at the temperature 25 °C as follows:

$$\sigma(t) = \int_0^{\xi} E(\xi - \xi') \frac{\partial \varepsilon}{\partial \xi'} d\xi' \tag{12.164}$$

where *the reduced timescale*, ξ, is given by

$$\log \xi \equiv \log t - \log a_T \Rightarrow d\xi = \frac{dt}{a_T} \tag{12.165}$$

and a_T is called *the shift factor*. There are physical reasons that some viscoelastic materials are thermorheologically simple, and for that reason, there are models that may be used for the purpose of describing the shift factor, a_T. Among these is the Arrhenius equation (after Svante August Arrhenius 1859–1927), given by:

$$\log a_T = \frac{\Delta F}{2.303R} \left(\frac{1}{T_R} - \frac{1}{T_M} \right) \tag{12.166}$$

where R is the universal gas constant, ΔF is the activation energy per gram mole, T_R is the reference temperature, and T_M is the melting temperature. The above equation often works well for polymers below the glass transition temperature, T_G.

Another model that is used for polymers above the glass transition temperature is *the WLF equation*, given by (Ferry 1980):

$$\log a_T = -C_1 \frac{(T_R - T_M)}{(C_2 + T_R - T_M)} \tag{12.167}$$

where C_1 and C_2 are curve-fitting coefficients.

Equations such as the two above may be used to fit to experimental data for the purpose of constructing long-term properties from short-term experiments.

As a final note regarding accelerated testing, when frequency sweeps are performed for the purpose of material characterization, as described in the previous sections, equipment limitations may preclude the ability to perform tests over a sufficiently broad range of frequencies to obtain robust viscoelastic material properties. In this case, it may often be possible to perform frequency sweeps at various temperatures over a relatively narrow range of frequencies and utilize the shifting

procedure introduced within this section to produce a master curve in the frequency domain. This procedure (to be considered in further detail below) can be utilized as a means of obtaining both the complex modulus and the relaxation modulus.

12.5 Mechanical Analogs for Creep Compliances and Relaxation Moduli

In Chap. 10, a simple mechanical analog called a Voigt model (see Fig. 10.10) was introduced for the purpose of constructing a concise mathematical form for the relaxation moduli or creep compliances. However, it was shown in Fig. 10.11 that this analog does not accurately capture the response of real viscoelastic materials because real materials exhibit much longer relaxation times than that represented within a single exponential decay function. As shown in Fig. 12.10, moduli and compliances of typical viscoelastic materials can be quite complicated. Without going into the physics and chemistry of polymers, suffice it to say that polymers exhibit a range of relaxation times, each of which can be associated with a viscous element (dashpot) within a mechanical analog representing the material response. Thus, for the purpose of constructing creep compliances or relaxation moduli, two more complicated analogs have been demonstrated to be accurate for most linear viscoelastic materials. These are discussed below.

12.5.1 The Kelvin Model for Creep Compliances

In order to improve upon the Voigt model, a number of Voigt elements may be strung together in series, to create a generalized Voigt model, termed *the Kelvin model* (after William Thompson, 1st Baron Kelvin 1824–1907), as shown in Fig. 12.11 (Flugge 1975; Tschoegl 1989).

The reader will recall that the special case of a single element (the Voigt model) in the above analog was developed in detail in Chap. 10. In that derivation, it was pointed out that a mechanical analog can be solved mathematically as a means of constructing either compliance or modulus. Lest it not be clear, this process is nothing more than a micromechanics problem (as described in Chap. 11), in which the mathematical solution of the IBVP for a heterogeneous medium is solved, and

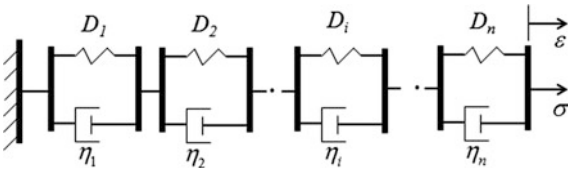


Fig. 12.11 The Kelvin model

this solution is then homogenized; that is, the object is viewed as if it were macroscopically homogeneous.

This is exactly the procedure taken with the Kelvin model shown in Fig. 12.11. The resulting compliance, which can be obtained from the Laplace transform of the governing differential equation for the above analog, is as follows (Flugge 1975):

$$D_K(t) = D_0 + \sum_{i=1}^{n} D_i\left(1 - e^{-\frac{t}{\eta_i D_i}}\right) \tag{12.168}$$

Despite the fact that polymers are significantly more complicated than the Kelvin analog shown in Fig. 12.11, researchers have demonstrated that the homogenized compliance given by Eq. (12.168) that is obtained from the Kelvin micromechanics model results in accurate predictions of the mechanical response of linear viscoelastic media subjected to complex loadings.

A sum of exponential terms is called *a Prony series* (after Gaspard Clair François Marie Riche de Prony 1755–1839). By carefully spacing the quotients on the exponential terms along the temporal axis, the above series can be utilized to create a smooth curve that overlaps experimental data for real linear viscoelastic media quite accurately, even when as many as 15 terms are needed in the Prony series. Note that the above mathematical form may be used for any component of the creep compliance, such $B(t)$ or $J(t)$.

12.5.2 The Wiechert Model for Relaxation Moduli

For the purpose of constructing relaxation moduli, the so-called Maxwell elements are constructed in parallel, to create a generalized Maxwell model, termed *the Wiechert model* (introduced by both Joseph John Thomson 1856–1940 and Johann Emil Wiechert 1861–1928), as shown in Fig. 12.12 (Flugge 1975; Tschoegl 1989).

The resulting Prony series for the above mechanical analog is as follows:

$$E(t) = E_\infty + \sum_{i=1}^{n} E_i e^{-\frac{E_i}{\eta_i}t} \tag{12.169}$$

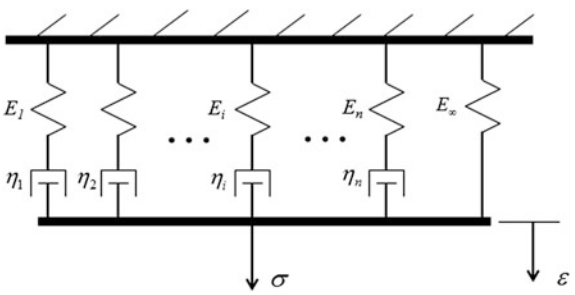


Fig. 12.12 The Wiechert model

The above two analogs can be shown to produce accurate representations of real material response when the material is observed to be linear viscoelastic.

Note that while these analogs have been introduced for the case of uniaxial properties, they may be deployed for any and all components of the viscoelastic property tensor.

12.5.3 Power Laws

Most materials exhibit some curvature when their moduli or relaxation curves are plotted in log-log space. As is well known, power laws plot linearly in log-log space. Thus, power law models may not be very accurate, especially for the purpose of predicting long-term response. However, power laws are often simpler to solve problems with than are Prony series. Thus, it is sometimes propitious to choose 5–7 decades of time and fit a power law to this restricted range. Furthermore, the Prony series representation, due to the inherent nature of the decaying exponential terms contained therein, tends to underpredict both short-term and long-term creep response, whereas a power law model tends to overpredict both short and long creep response, thereby producing contrasting results that may in fact bound the actual material response from above and below. Thus, power law, as given by the following, can be useful for predictive purposes.

$$D(t) = D_0 + D_1 \left(\frac{t}{\tau_0} \right)^n \tag{12.170}$$

$$E(t) = E_\infty + E_1 \left(\frac{t}{\tau_0} \right)^n \tag{12.171}$$

Both of the above models can be fit to experimental data by methods to be discussed in the next section.

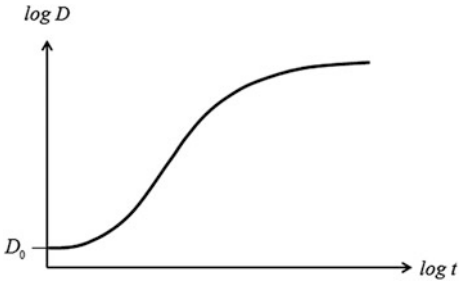


Fig. 12.13 Typical results of a uniaxial creep test

12.6 Procedures for Curve Fitting

Suppose that a typical creep test has produced the curve shown in Fig. 12.13. The next challenge is to fit one of the models described above to the data as accurately as possible. Methods for fitting a Prony series and a power law are described in the next two sections.

12.6.1 Prony Series Model

In this case, the objective is to fit the experimental data shown in Fig. 12.13 with the Kelvin model, given by Eq. (12.168). Equation (12.168) is first rewritten as follows:

$$D_K(t) - D_0 = \sum_{i=1}^{n} D_i\left(1 - e^{-\frac{t}{\tau_i}}\right) \tag{12.172}$$

where $\tau_i \equiv D_i \eta_i$. The above equation can be fit to the experimental data by a collocation technique (Schapery 1962a, b). This may be accomplished by first choosing the n values of τ_i. Experience indicates that the τ_i values should be separated by no more than one decade in time in order to obtain an accurate curve fit. (Otherwise, the resulting Prony series may predict physically unrealistic oscillatory behavior.) The data may then be used to evaluate Eq. (12.172) at n times t_i, $i = 1, \ldots, n$. Note that in order to maintain numerical stability, it is also advisable to require that $\tau_1 < t_1$. This process results in the following set of n equations in the unknown coefficients, $D_j, j = 1, \ldots, n$:

$$D(t_i) - D_0 = \sum_{j=1}^{n} D_j(1 - e^{-\frac{t_i}{\tau_j}}) \quad i = 1, \ldots, n \tag{12.173}$$

The form of these equations will be as follows:

$$A_{ij} D_j = C_i \quad i = 1, \ldots, n \quad j = 1, \ldots, n \tag{12.174}$$

where due to Eq. (12.173)

$$A_{ij} = 1 - e^{\frac{-t_i}{\tau_j}} \tag{12.175}$$

and

$$C_i = D(t_i) - D_0 \tag{12.176}$$

Since the above equations are linear, Gauss elimination or any available linear equation solver can be used to solve for the unknown coefficients $D_i, i = 1, \ldots, n$.

While the choice of both t_i and τ_j is somewhat arbitrary, it will develop that despite the nonuniqueness of the coefficients produced by this approach, when the procedure is performed properly, the resulting curve fit will indeed be both smooth and unique. For the purpose of ensuring that a smooth result is obtained, it is propitious to introduce the parameter a_S, defined by the following relationship (Tschoegl 1989):

$$\tau_k = a_S t_k \tag{12.177}$$

In order to ensure smoothness, it is recommended that a value of $a_S = 1/\ln 2$ be deployed, thereby resulting in values of ½ on the diagonal of the coefficient matrix A_{ij} (Tschoegl 1989).

In order to ensure additional accuracy, it is also recommended that an additional constraint equation be added to the above set, where, from Eq. (12.173) (Tschoegl 1989):

$$D_\infty \equiv D(t_{n+1} \to \infty) = D_0 + \sum_{i=1}^{n} D_i \tag{12.178}$$

The above results in values of unity for all of the coefficients in the last row of the coefficient matrix, that is,

$$A_{ij} = 1.0 \quad i = n+1, j = 1, \ldots, n+1 \tag{12.179}$$

and

$$C_{n+1} = D_\infty - D_0 \tag{12.180}$$

The above procedure may be applied for the purpose of determining components of both the relaxation modulus and creep compliance tensors. Furthermore, as will be demonstrated below, a similar collocation procedure may be utilized for the purpose of curve-fitting frequency sweeps.

This completes the procedure for obtaining the coefficients in the Prony series. However, once this is completed, it is advisable to construct a graph of the predicted curve versus the experimental data in order to ensure that the procedure has been properly deployed.

12.6.2 Power Law Model

As mentioned above, power laws are sometimes utilized, as they are often sufficiently accurate and much easier to use than a Prony series. This can be accomplished quite easily by using a least squares technique. To see how this may be accomplished, consider the power law creep compliance, given by Eq. (12.170). This equation may be rearranged to the following form:

$$\log[D(t) - D_o] = \log D_1 \; + \; n \; \log\left(\frac{t}{\tau_0}\right) \tag{12.181}$$

Note that the quantity τ_0 is a free parameter that may be arbitrarily assigned, typically to a time near the middle of the range of times under consideration. Now note that Eq. (12.181) is a straight line in log-log space. To fit a curve to the data, first define the error at the ith time, e_i, by

$$e_i \equiv \; \log[D(t_i) - D_o] - \log \; D_1 - n \; \log\left(\frac{t_i}{\tau_0}\right) \tag{12.182}$$

Suppose there are m data points. Then, the total squared error is

$$S \equiv \sum_{i=1}^{m} (e_i)^2 \tag{12.183}$$

In order to minimize this error, it follows that

$$\frac{\partial S}{\partial a_j} = 0 \quad j = 1, 2 \tag{12.184}$$

where

$$\begin{aligned} a_1 &= \log D_1 \\ a_2 &= n \end{aligned} \tag{12.185}$$

Carrying the above differentiation out results in the following two equations in two unknowns:

$$\begin{bmatrix} m & \sum_{j=1}^{m} \log(t_j/\tau_n) \\ \sum_{j=1}^{m} \log(t_j/\tau_n) & \sum_{j=1}^{m} \left(\log(t_j/\tau_n)\right)^2 \end{bmatrix} \begin{Bmatrix} a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum_{j=1}^{m} \log\left[D(t_j) - D_o\right] \\ \sum_{j=1}^{m} \log\left[D(t_j) - D_o\right] \log(t_j/\tau_n) \end{Bmatrix} \tag{12.186}$$

The above equations may be solved for the unknowns a_1 and a_2. Finally, from Eq. (12.185), it follows that

$$\begin{aligned} D_1 &= 10^{a_1} \\ n &= a_2 \end{aligned} \tag{12.187}$$

This completes the determination of the power law material parameters necessary to characterize Eq. (12.170). A similar procedure may be deployed for the purpose of characterizing Eq. (12.171).

12.6.3 Frequency Sweeps

An examination of Eq. (12.134) will result in the following form of the complex shear modulus:

$$\mu * (i\omega) = \mu_\infty + i\omega \int_0^\infty \mu_1(\omega) e^{-i\omega\eta} d\eta \tag{12.188}$$

Consider also the Carson transform of the shear modulus, given by the following:

$$\tilde{\mu}(s) = s\bar{\mu}(s) = \mu_\infty + s \int_0^\infty \mu_1(t) e^{-st} dt \tag{12.189}$$

The above two formulas present an interesting analogy, given by:

$$\mu^*(i\omega) = \tilde{\mu}(s)_{s \to i\omega} \tag{12.190}$$

Based on the above analogy, consider the Carson transform of a Prony series representation of the shear modulus (similar to Eq. (12.169):

$$\tilde{\mu}(s) = \mu_\infty + \sum_{j=1}^{n} \frac{s\mu_j}{(s + 1/\tau_j)} \tag{12.191}$$

where $\tau_i \equiv \mu_i/\eta_i$. Applying Eq. (12.190) to Eq. (12.191) thus results in the following:

$$\mu^*(i\omega) = \mu_\infty + \sum_{j=1}^{n} \frac{i\omega\mu_j}{(i\omega + 1/\tau_j)} \tag{12.192}$$

The above equation may be written as follows:

$$\mu^*(i\omega) = \mu' + i\mu'' \tag{12.193}$$

where

$$\mu' = \mu_\infty + \sum_{j=1}^{n} \frac{\mu_j' \omega^2}{\left(\omega^2 + 1/\tau_j^2\right)} \tag{12.194}$$

where it can be shown that $\mu_\infty \equiv \mu(t \to \infty) = \mu'(\omega \to 0)$ and

$$\mu'' = \sum_{j=1}^{n} \frac{\omega \mu_J' / \tau_j}{\left(\omega^2 + 1/\tau_j^2\right)} \tag{12.195}$$

It is therefore apparent that either Eq. (12.194) or (12.195) can be curve fit to dynamic test data such as that shown in Fig. 12.8 by utilizing the method described above. Thus, by choosing n values of the $\tau_j, j = 1, \ldots, n$, the coefficients $\mu_j', j = 1, \ldots, n$ in Eq. (12.194) can be fit to experimental data obtained from a frequency sweep. In order for the diagonal components of the coefficient matrix A_{ij} to be ½, it is recommended that for frequency sweep data $a_S = 1$ (Tschoegl 1989).

Example Problem 12.3
Given: Consider the frequency sweep data for a typical partially aged bitumen specimen shown in Table 12.3.

Required:
(a) Use time–temperature superposition to shift the data to a single master curve at $T = 35\,^\circ\text{C}$.

(b) Fit the real part of the complex modulus using Eq. (12.194) and compare to the experimental data graphically.

(c) Plot the resulting relaxation modulus in the time domain.

Solution: (a) As shown in Fig. 12.14, the experimental data for the real part of the complex modulus have been shifted to create a master curve using $a_T = 1.2$ at $T = 15\,^\circ\text{C}$ and $a_T = -1.1$ at $T = 55\,^\circ\text{C}$.

(b) The master curve was then fit with a five-term series using Eq. 12.194. The resulting algebraic equations are as follows:

Table 12.3 Frequency sweeps at three different temperatures showing real and imaginary parts of the complex modulus versus frequency (Courtesy Pavement Scientific International)

Frequency	Temperature	Real part	Imaginary part
rad/s	°C	MPa	MPa
4.991	15	27.19	12.49
6.283	15	28.32	12.51
7.91	15	29.47	14.63
9.958	15	30.67	15.86
12.54	15	32.76	17.16
15.78	15	35.39	18.52
19.87	15	36.08	19.78
25.01	15	35.75	21.15
31.49	15	35.38	21.96
39.65	15	35.88	22.12
49.91	15	36.51	20.98
4.991	35	27.19	12.49
6.283	35	28.32	12.51
7.91	35	29.47	14.63
9.958	35	30.67	15.86
12.54	35	32.76	17.16
15.78	35	35.39	18.52
19.87	35	36.08	19.78
25.01	35	35.75	21.15
31.49	35	35.38	21.96
39.65	35	35.88	22.12
49.91	35	36.51	20.98
4.991	55	27.19	12.49
6.283	55	28.32	12.51
7.91	55	29.47	14.63
9.958	55	30.67	15.86
12.54	55	32.76	17.16
15.78	55	35.39	18.52
19.87	55	36.08	19.78
25.01	55	35.75	21.15
31.49	55	35.38	21.96
39.65	55	35.88	22.12
49.91	55	36.51	20.98

$$\begin{bmatrix} 0.50 & 0.01 & 0. & 0. & 0. & 0. \\ 0.99 & 0.50 & 0.01 & 0. & 0. & 0. \\ 1.0 & 0.99 & 0.50 & 0.01 & 0. & 0. \\ 1.0 & 1.0 & 0.99 & 0.50 & 0.01 & 0. \\ 1.0 & 1.0 & 1.0 & 0.99 & 0.50 & 0.01 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix} \begin{Bmatrix} \mu_1' \\ \mu_2' \\ \mu_3' \\ \mu_4' \\ \mu_5' \\ \mu_\infty' \end{Bmatrix} = \begin{Bmatrix} 0.50 \\ 3.0 \\ 14.75 \\ 29.0 \\ 35.5 \\ 39.0 \end{Bmatrix} \tag{12.196}$$

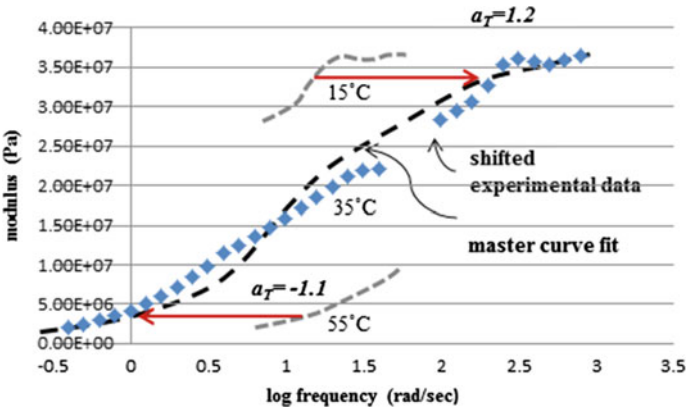


Fig. 12.14 Graphical depiction of experimental and curve fit master curves for real part of complex modulus for a typical asphalt binder

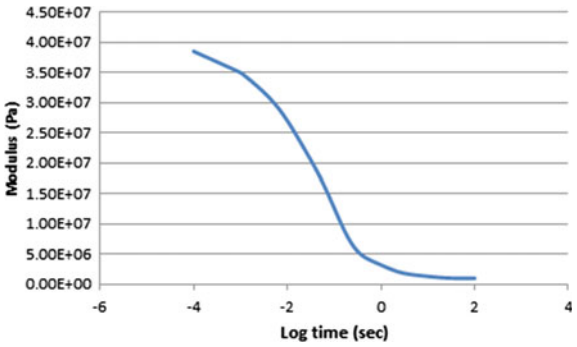


Fig. 12.15 Predicted relaxation modulus for asphalt binder described in Table 12.3

where by extrapolation of the experimental data it has been assumed that $\mu'(\omega \to 0) = \mu_\infty \simeq 1.0\,\text{MPa}$ and $\mu'(\omega \to \infty) \simeq 40.0\,\text{MPa}$, and Eqs. (12.179) and (12.180) have been utilized to produce the sixth equation. Solving the above set of linear algebraic equations results in the following values for the real moduli (in MPa) in Eq. (12.194):

$$\{\mu_i'\} = \begin{Bmatrix} 0.925 \\ 3.77 \\ 20.0 \\ 8.92 \\ 3.92 \end{Bmatrix} \tag{12.197}$$

corresponding to the frequencies $\omega_i = 10^{-1}, 10^0, 10^1, 10^2, 10^3$, thereby producing the master curve fit shown in Fig. 12.14.

(c) In accordance with Eq. (12.190), the above coefficients then become the coefficients for the Prony series, that is, $\mu_i' \to \mu_i$ and $\tau_i = 1/\omega_i$ (note that $\mu_\infty = \mu_0'$), thereby resulting in the following representation of the shear modulus in the time domain:

$$\mu(t) = \mu_\infty + \sum_{i=1}^{n} \mu_i e^{-\frac{t}{\tau_i}} \tag{12.198}$$

Equation (12.198) is plotted in Fig. 12.15.

The above procedure presents a simple means of transforming the dynamic modulus from the frequency domain to the time domain.

12.7 Multi-dimensional Linear Thermoviscoelasticity

When the temperature in the object of interest varies spatially and/or with time the above viscoelastic model must necessarily be generalized. In this case, a thermoviscoelastic material is defined to be one in which the stress, σ_{ij}, is a single-valued functional of the strain history, ε_{kl}, and the temperature, T, as follows:

$$\sigma_{ij}(\vec{x}, t) = E_{ij\,\tau=0}^{\tau=t}\{\varepsilon_{kl}(\vec{x}, \tau), T(\vec{x}, \tau)\} \tag{12.199}$$

The above is adjoined with the following (from Table 9.7):

(1) strain-displacement equations:

$$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{12.200}$$

(2) equations of motion:

$$\sigma_{ji,j} + \rho f_i = \rho \frac{d^2 u_i}{dt^2} \tag{12.201}$$

(3) conservation of energy:

$$\rho \frac{du^I}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} - q_{i,i} + \rho r \tag{12.202}$$

(4) entropy production inequality:

$$\rho \frac{ds}{dt} + \frac{\partial}{\partial x_i}\left(\frac{q_i}{T}\right) - \frac{\rho r}{T} \geq 0 \tag{12.203}$$

Similar to that introduced for thermoelastic media in Chap. 11, the temperature gradient, γ_i, as described in Eq. (11.67), is defined as follows:

$$\gamma_i \equiv \frac{\partial T}{\partial x_i} \tag{12.204}$$

The list of 28 unknowns for a thermoviscoelastic continuum is thus as follows:

Displacement vector	u_i
Strain tensor	ε_{ij}
Stress tensor	σ_{ij}
Body force per unit mass	f_i
Internal energy per unit mass	u^I
Heat flux vector	q_i
Heat source per unit mass	r
Entropy per unit mass	s
Temperature	T
Temperature gradient	γ_i

The prediction of the above variables is afforded with the use of the 13 equations and one inequality (12.200)–(12.204) listed above.

The thermoviscoelasticity problem is considerably more complicated than the viscoelasticity problem, and this is due at least in part to the coupling between mechanics and thermodynamics. Thus, a more formal procedure similar to that employed for thermoelastic media (see Chap. 11) is required to construct this model.

12.7.1 Thermodynamic Constraints on Thermoviscoelastic Material Behavior

The development of thermodynamic constraints on thermoviscoelastic materials will follow the method developed by Coleman and Noll in Chap. 11, but will be extended for the case of materials with memory (Coleman 1964; Christensen and Naghdi 1967; Day 1972). The challenge is to solve the above set of equations in terms of the independent variables $\vec{x}$ and t, implying that a solution is sought for the 28 unknowns $u_i, \varepsilon_{ij}, \sigma_{ij}, f_i, u^I, T, \gamma_i, q_i, r$ and s as functions of $\vec{x}$ and t. The solution procedure is described in Table 12.4.

Table 12.4 Procedure for obtaining thermodynamic constraints on thermoviscoelastic media

Step 1: Note that normally the body force per unit mass, $f_i = f_i(\vec{x}, t)$, can be specified (by using Newton's gravitational law) as well as the heat source, $r = r(\vec{x}, t)$. However, in this case specify a priori instead $u_i = u_i(\vec{x}, t)$, and $T = T(\vec{x}, t) \forall \vec{x}, t \in V + S \Rightarrow$ **24 remaining unknowns**	
Step 2: Use Eq. (12.202) to solve for $\varepsilon_{ij} = \varepsilon_{ij}(\vec{x}, t) \Rightarrow$ **18 remaining unknowns**	
Step 3: Use Eq. (12.204) to solve for $\gamma_i = \gamma_i(\vec{x}, t) \Rightarrow$ **15 remaining unknowns**	
Step 4: Assume that there exist 11 constitutive equations for a thermoviscoelastic material as follows:	
$\sigma_{ij}(\vec{x}, t) = E_{ij\,\tau=0}^{\tau=t}\{\varepsilon_{kl}(\vec{x}, \tau), T(\vec{x}, \tau), \gamma_k(\vec{x}, \tau)\}$	(12.205)
$u^I(\vec{x}, t) = u_{\tau=0}^{I\,\tau=t}\{\varepsilon_{kl}(\vec{x}, \tau), T(\vec{x}, \tau), \gamma_k(\vec{x}, \tau)\}$	(12.206)
$q_i(\vec{x}, t) = q_{i\,\tau=0}^{\tau=t}\{\varepsilon_{kl}(\vec{x}, \tau), T(\vec{x}, \tau), \gamma_k(\vec{x}, \tau)\}$	(12.207)
$s(\vec{x}, t) = s_{\tau=0}^{\tau=t}\{\varepsilon_{kl}(\vec{x}, \tau), T(\vec{x}, \tau), \gamma_k(\vec{x}, \tau)\}$	(12.208)
where, according to *The Principle of Equipresence* (Truesdell et al. 2004), since the temperature gradient is included in one constitutive equation, it must be included in all of them. Note that the above relate the kinetic variables to the kinematic variables (which have been determined above). Use (12.203)–(12.206) to solve for $\sigma_{ij} = \sigma_{ij}(\vec{x}, t), u^I = u^I(\vec{x}, t), q_i = q_i(\vec{x}, t), s = s(\vec{x}, t) \Rightarrow$ **4 remaining unknowns**	
Step 5: Use Eq. (12.201) to solve for $f_i = f_i(\vec{x}, t) \Rightarrow$ **1 remaining unknown**	
Step 6: Use Eq. (12.202) to solve for $r = r(\vec{x}, t) \Rightarrow$ **0 remaining unknowns**	
Step 7: Recall that the Helmholtz free energy is defined as follows:	
$h \equiv u^I - Ts$	(12.209)
Thus, as a result of constitutive Eqs. (12.206) and (12.208), it follows that	
$h(\vec{x}, t) = h_{\tau=0}^{\tau=t}\{\varepsilon_{kl}(\vec{x}, \tau), T(\vec{x}, \tau), \gamma_k(\vec{x}, \tau)\}$	(12.210)
Expanding the free energy in convolutions of the dependent variables thus results in the following:	
$\rho h = \rho h_0 + \int_{-\infty}^{t} D_{ij}(t-\tau)\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau - \int_{-\infty}^{t} \beta(t-\tau)\frac{\partial \theta}{\partial \tau}d\tau + \frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t} E_{ijkl}(t-\tau, t-n)\frac{\partial \varepsilon_{ij}}{\partial \tau}\frac{\partial \varepsilon_{kl}}{\partial \eta}d\tau\, d\eta \int_{-\infty}^{t}\int_{-\infty}^{t} \phi(t-\tau, t-\eta)\frac{\partial \varepsilon_{ij}}{\partial \tau}\frac{\partial \theta}{\partial \eta}d\tau d\eta - \frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t} m(t-\tau, t-\eta)\frac{\partial \theta}{\partial \tau}\frac{\partial \theta}{\partial \eta}d\tau d\eta + H.O.T.$	(12.211)
where it should be recalled that $\theta \equiv T - T_R$ where T_R is the reference temperature at which no strain is observed	
Step 8: It is necessary to invoke the constraint imposed by the entropy production inequality. To do this, first recall that due to Eq. (12.209) the time derivative of the Helmholtz free energy per unit mass, $h = h(\vec{x}, t)$, may be written as follows:	
$\frac{dh}{dt} = \frac{du^I}{dt} - s\frac{dT}{dt} - T\frac{ds}{dt}$	(12.212)
Step 9: Utilize Leibniz' rule (Eq. (12.20)) to obtain the time derivative of Eq. (12.211). Substitute the resulting equation into Eq. (12.212) and this result into inequality (12.203). Rearranging will result in the following:	

(continued)

Table 12.4 (continued)

$$\left[\sigma_{ij} - D_{ij}(0) - \int_{-\infty}^{t} E_{ijkl}(t-\tau,0)\frac{\partial \varepsilon_{kl}}{\partial \tau}d\tau + \int_{-\infty}^{t} \phi_{ij}(0,t-\tau)\frac{\partial \theta}{\partial \tau}d\tau\right]\frac{d\varepsilon_{ij}(t)}{dt}$$
$$+\left[\beta(0) + \int_{-\infty}^{t} m(t-\tau,0)\frac{\partial \theta}{\partial \tau}d\tau + \int_{-\infty}^{t} \phi_{ij}(t-\tau,0)\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau - \rho s\right]\frac{dT(t)}{dt}$$
$$-\int_{-\infty}^{t} \frac{\partial D_{ij}(t-\tau)}{\partial t}\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau + \int_{-\infty}^{t} \frac{\partial \beta(t-\tau)}{\partial \tau}\frac{\partial \theta}{\partial \tau}d\tau$$
$$-\frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t} \frac{\partial E_{ijkl}(t-\tau,t-\eta)}{\partial t}\frac{\partial \varepsilon_{ij}\partial \varepsilon_{kl}}{\partial \tau\,\partial \eta}d\tau d\eta$$
$$+\int_{-\infty}^{t}\int_{-\infty}^{t} \frac{\partial \phi_{ij}(t-\tau,t-\eta)\partial \varepsilon_{ij}}{\partial t \qquad \partial \tau}\frac{\partial \theta}{\partial \eta}d\tau d\eta$$
$$+\frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t} \frac{\partial m(t-\tau,t-\eta)\partial \theta}{\partial t \qquad \partial \tau}\frac{\partial \theta}{\partial \eta}d\tau d\eta - q_i\left(\frac{T,_i}{T}\right) \geq 0 \tag{12.213}$$

Step 10: Now note that the quantities inside the brackets in inequality (12.213) are independent of the quantities outside the brackets, and the quantities outside the brackets are independent of one another, so that the above produces the following constraints on the allowable material behavior for a thermoelastic material

$$\sigma_{ij} = D_{ij}(0) + \int_{-\infty}^{t} E_{ijkl}(t-\tau,0)\frac{\partial \varepsilon_{kl}}{\partial \tau}d\tau - \int_{-\infty}^{t} \phi_{ij}(0,t-\tau)\frac{\partial \theta}{\partial \tau}d\tau \tag{12.214}$$

and

$$\rho s = \beta(0) + \int_{-\infty}^{t} m(t-\tau,0)\frac{\partial \theta}{\partial \tau}d\tau + \int_{-\infty}^{t} \phi_{ij}(0,t-\tau)\frac{\partial \theta}{\partial \tau}d\tau \tag{12.215}$$

Since the first-order terms in inequality (12.213) must also be zero, it follows that

$$\frac{\partial D_{ij}}{\partial t} = 0 \tag{12.216}$$

and

$$\frac{\partial \beta}{\partial t} = 0 \tag{12.217}$$

thus implying that the above two material properties are time independent. The entropy production inequality therefore simplifies to the following form:

$$-\frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t} \frac{\partial E_{ijkl}(t-\tau,t-\eta)}{\partial t}\frac{\partial \varepsilon_{ij}\partial \varepsilon_{kl}}{\partial \tau\,\partial \eta}d\tau d\eta$$
$$+\int_{-\infty}^{t}\int_{-\infty}^{t} \frac{\partial \phi_{ij}(t-\tau,t-\eta)\partial \varepsilon_{ij}}{\partial t \qquad \partial \tau}\frac{\partial \theta}{\partial \eta}d\tau d\eta$$
$$+\frac{1}{2}\int_{-\infty}^{t}\int_{-\infty}^{t} \frac{\partial m(t-\tau,t-\eta)\partial \theta}{\partial t \qquad \partial \tau}\frac{\partial \theta}{\partial \eta}d\tau d\eta - q_i\left(\frac{T,_i}{T}\right) \geq 0 \tag{12.218}$$

(continued)

Table 12.4 (continued)

The above represents the rate of energy dissipation in linear thermoviscoelastic media. A strong form of the entropy production inequality, comprising only the last term in inequality (12.203), will lead to identical results for the heat flux constitutive equation as that found in Chap. 11 for a thermoelastic material (see Eqs. (11.76)–(11.78)).
This completes the construction of thermodynamic constraints on linear thermoviscoelastic media, thus demonstrating that it is possible to solve for all of the unknowns using the thermoviscoelastic constitutive model, while at the same time obtaining powerful constraints on the constitutive behavior of linear thermoviscoelastic media.

12.7.2 The Linear Thermoviscoelastic Initial Boundary Value Problem

The results of the procedure described in Table 12.4 may now be utilized to simplify the thermoviscoelastic initial boundary value problem. To accomplish this, first note that the temperature, T, appears in the thermoviscoelastic constitutive Eq. (12.214). As such, it will be necessary to treat it as a primary variable that must be adjoined to the fifteen primary variables $u_i, \sigma_{ij}, \varepsilon_{ij}$ required to solve the viscoelastic initial boundary value problem. The means whereby this new primary unknown is predicted is the conservation of energy Eq. (12.202). This equation may be manipulated so as to remove the secondary unknowns u^I, s, h, γ_i, thus resulting in the following coupled heat conduction equation:

$$\rho r - T_R \frac{\partial}{\partial t}\left[\int_{-\infty}^{t} \phi_{ij}(t-\tau,0)\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau + \int_{-\infty}^{t} m(t-\tau,0)\frac{\partial \theta}{\partial \tau}d\tau\right] + (k_{ij}T_{,j})_{,i} = 0 \tag{12.219}$$

The above equation now has the secondary variables removed, so that it may be used to obtain the temperature field, $T(\vec{x}, t)$.

12.7.3 Two-Way Coupled Linear Thermoviscoelasticity

Unfortunately, Eq. (12.219) contains the dilatational component of the strain tensor, ε_{kk}, in addition to the temperature, T. This implies that the mechanics and the thermodynamics are *coupled*, meaning that neither can be solved separately from one another, just as in the case of a thermoelastic object (see Chap. 11). This then results in the so-called *coupled thermoviscoelastic initial boundary value problem*, thus implying that when temperature change occurs in a thermoviscoelastic object, the deformation changes, *and* vice versa, that is, mechanical loading induces temperature change. For the linear isotropic case, this problem is described in Table 12.5.

Table 12.5 Coupled linear thermoviscoelastic initial boundary value problem

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S), f_i(\text{in } V+S), r(\text{in } V+S)$

Geometry: $\vec{x}(\text{on } S),\ n_i(\text{on } S)$

Material Properties: $\rho, \lambda, \mu, \alpha, k, c_v$

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}, T$ = **16 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = D_{ij}(0) + \int_{-\infty}^{t} E_{ijkl}(t-\tau,0)\frac{\partial \varepsilon_{kl}}{\partial \tau}d\tau - \int_{-\infty}^{t} \phi_{ij}(0,t-\tau)\frac{\partial \theta}{\partial \tau}d\tau$	6
$\rho r - T_R \frac{\partial}{\partial t}\left[\int_{-\infty}^{t} \phi_{ij}(t-\tau,0)\frac{\partial \varepsilon_{ij}}{\partial \tau}d\tau + \int_{-\infty}^{t} m(t-\tau,0)\frac{\partial \theta}{\partial \tau}d\tau\right] + (k_{ij}T_{,j})_{,i} = 0$	1
Total	**16**

Initial Conditions:

Displacements: $u_i(\vec{x},0) = \hat{u}_i \quad \forall \vec{x} \in V+S$

Stresses: $\sigma_{ij}(\vec{x},0) = \hat{\sigma}_{ij} \quad \forall \vec{x} \in V+S$

Strains: $\varepsilon_{ij}(\vec{x},0) = \hat{\varepsilon}_{ij} \quad \forall \vec{x} \in V+S$

Temperature: $T(\vec{x},0) = \hat{T} \quad \forall \vec{x} \in V+S$

Boundary Conditions:

Tractions: $t_i = \hat{t}_i \quad on \quad S_1$

Displacements: $u_i = \hat{u}_i \quad on \quad S_2 \quad S = S_1 + S_2$

Temperature: $T = \hat{T} \quad on \quad S_3$

Heat Flux: $q_i = \hat{q}_i \quad on \quad S_4 \quad S = S_3 + S_4$

Table 12.6 Part 1: Procedure for solving for the temperature in uncoupled thermoviscoelasticity

Part 1: Solve for $T = T(\vec{x}, t)$

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: r

Geometry: $\vec{x}\,(\text{on } S),\ n_i\,(\text{on } S)$

Material Properties: ρ, k, c_v

Unknowns: T = 1 unknown

Field Equation:

	No. of Equations
$\rho c_v \dfrac{dT}{dt} = (kT_{,i})_{,i} + \rho r$	1
Total	**1**

Initial Conditions:

Temperature: $T(\vec{x}, 0) = \hat{T} \quad \forall \vec{x} \in V + S$

Boundary Conditions:

Temperature: $T = \hat{T} \quad on \quad S_3$

Heat Flux: $q_i = \hat{q}_i \quad on \quad S_4 \quad S = S_3 + S_4$

12.7.4 One-Way Coupled Thermoviscoelasticity

The initial boundary value problem described in Table 12.5 can be quite challenging to solve due to the thermoviscoelastic coupling. Fortunately, this coupling is in many cases quite small. For many practical circumstances, the coupling term appearing in the heat equation can be disregarded. This assumption appears to be accurate for the purpose of modeling flexible pavement structures. Furthermore, it

will be assumed that the viscoelastic relaxation in the term involving the temperature is negligible. This will result in the following simplification of Eq. (12.219):

$$\rho c_v \frac{dT}{dt} = (kT_{,i})_{,i} + \rho r \tag{12.220}$$

It can be seen that the above is the same uncoupled form of the conservation of energy arrived at in Chap. 11, in which the temperature is the single unknown field variable. In this case, the procedure described in Table 12.4 simplifies to two separate problems. First, the temperature field, $T(\vec{x}, t)$ is predicted, subject to the thermodynamic initial and boundary conditions described in Table 12.5. Thereafter, the temperature may be deployed as a known quantity to predict the remaining fifteen unknowns, $u_i(\vec{x}, t), \sigma_{ij}(\vec{x}, t), \varepsilon_{ij}(\vec{x}, t)$. The special uncoupled case of a linear isotropic thermoviscoelastic object subjected to quasi-static loading is described in Tables 12.6 and 12.7.

12.8 Nonlinear Viscoelasticity

When modeling the constitutive behavior of the AC pavement material, it is often found that the material behavior is nonlinear. For example, if creep tests are performed at increasing load levels, it is found that when a threshold value of loading is exceeded the principle of homogeneity (see Chap. 9) does not hold. Instead, doubling the load, for example, results in more than a doubling of the measured strain. This type of observed behavior will result in some inaccuracy if a linear viscoelastic model is employed, and while at least some of this can be attributed to microcracking (see Chap. 8), it has been observed that in asphalt concrete this error can be significant (Kim and Little 1990).

In order to account for this error it is necessary to employ a nonlinear viscoelastic material model, one in which the higher order terms in stress must be included in the stress formulation, or conversely, higher order terms in strain must be included in the strain formulation. Such formulations will necessarily become mathematically complicated. Much of the complexity arises from the inclusion of finite strains (Eq. (9.39), thereby leading to formulations that can be quite cumbersome (Green and Rivlin 1957; Green et al. 1959; Day 1972; Christensen 1982).

In this text, in order to develop a theory that is both instructive and oftentimes sufficiently accurate for purposes of analysis and design, it will be assumed that the deformations may be accurately captured with the linearized strain tensor (Eq. (9.40)) as opposed to the finite deformation strain tensor (Eq. (9.39)). The nonlinearity is therefore intended to be captured via some pragmatic deployment of Eq. (12.3) or (12.4) with nonlinear dependence on the right hand side.

Table 12.7 Part 2: Procedure for solving for stresses, strains, and displacements in uncoupled thermoviscoelasticity

PART 2: Solve for $u_i = u_i(\vec{x},t), \sigma_{ij} = \sigma_{ij}(\vec{x},t), \varepsilon_{ij} = \varepsilon_{ij}(\vec{x},t)$

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S), f_i(\text{in } V+S), T(\text{in } V+S)$

Geometry: $\vec{x}(\text{on } S),\ n_i(\text{on } S)$

Material Properties: $\rho, \lambda, \mu, \alpha$

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}$ = **15 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = D_{ij}(0) + \int_{-\infty}^{t} E_{ijkl}(t-\tau,0)\frac{\partial \varepsilon_{kl}}{\partial \tau}d\tau - \int_{-\infty}^{t} \phi_{ij}(0,t-\tau)\frac{\partial \theta}{\partial \tau}d\tau$	6
Total	**15**

Initial Conditions:

Displacements: $u_i(\vec{x},0) = \hat{u}_i \quad \forall \vec{x} \in V+S$

Stresses: $\sigma_{ij}(\vec{x},0) = \hat{\sigma}_{ij} \quad \forall \vec{x} \in V+S$

Strains: $\varepsilon_{ij}(\vec{x},0) = \hat{\varepsilon}_{ij} \quad \forall \vec{x} \in V+S$

Boundary Conditions:

Tractions: $t_i = \hat{t}_i \quad on \quad S_1$

Displacements: $u_i = \hat{u}_i \quad on \quad S_2 \quad S = S_1 + S_2$

Toward this end, perhaps the simplest accurate single integral nonlinear viscoelastic model for isotropic media has been developed by Schapery (1962a, b, 1964, 1966, 1997). Proceeding from a thermodynamic formulation derived from an assemblage of springs and dashpots first proposed by Maurice Anthony Biot (1905–85) for the linear case (Biot 1954, 1965), Schapery defined the spring and dashpot coefficients to be nonlinear, thereby producing a quite general nonlinear *single integral* formulation. The uniaxial compliance form of this model is given by:

$$\varepsilon(t) = g_0 D_0 \sigma(t) + g_1 \int_0^t \Delta D(\psi - \psi') \frac{\partial (g_2 \sigma)}{\partial \tau} d\tau \tag{12.221}$$

where D_0 is the initial compliance, $\Delta D(t)$ is the time-dependent portion of the compliance, and ψ is the so-called reduced time, defined by

$$\psi \equiv \int_0^t \frac{dt'}{a_\sigma} \tag{12.222}$$

and

$$\psi' \equiv \psi(\tau) = \int_0^\tau \frac{dt'}{a_\sigma} \tag{12.223}$$

and g_0, g_1, g_2 and a_σ are functions of stress.

Note both the similarities and dissimilarities between Eqs. (12.25) and (12.221). It can be seen that in the linear case $g_0 = g_1 = g_2 = a_\sigma = 1$. Conversely, in the nonlinear case they are not equal to unity so that, despite the fact that the equations are still single integral convolutions, the nonlinear model is significantly more complex, implying that the experimental protocols for determining the material properties are far more intricate and therefore costly.

Nonetheless, the above formulation is constructed in such a way that a reasonable set of experiments can be utilized to construct all of the required material properties using creep and creep and recovery tests (Schapery 1969; Lou and Schapery 1971), or frequency sweeps (Golden et al. 1999). In the case where Poisson's ratio is constant, the nonlinear compliance (as well as the associated functions in Eq. (12.221)) can be obtained and then inverted computationally to produce the nonlinear relaxation modulus and corresponding associated functions, and these can be deployed in a corresponding nonlinear form of Eq. (12.38) for the purpose of modeling with the nonlinear strain formulation. Accordingly, the above model has in recent years been applied by several researchers to asphalt (Park et al. 1996; Lee et al. 2000; Saadeh et al. 2007; Masad et al. 2008; Huang et al. 2010; Ban et al. 2013).

It should be noted that this type of nonlinear constitutive model is extremely complicated to deploy computationally. Furthermore, because it is typically employed at the level of asphalt concrete, the material properties must be regenerated for each set of aggregates and asphalt mastic under consideration. Thus, for example, in order to determine the optimum volume fraction of aggregate for a particular road way design, it is necessary to experimentally determine the material properties for several different volume fractions of aggregates. The experiments are in themselves sufficiently expensive to perform as to render it untenable by many engineers for the purpose of road way design and analysis. Nonetheless, this type of constitutive model can be incorporated into a finite element model so that predictions of the stress, strain, and deformations as functions of time and space can be made in flexible pavements (see Chaps. 14 and 15).

12.9 Summary

This chapter has presented an overview of the development of the theories of viscoelasticity and thermoviscoelasticity. An understanding of these models is essential to the ability to predict the response of flexible pavements containing asphalt binder (see Chaps. 2 and 7), as well as rate-dependent base materials.

12.10 Problems

Problem 12.1
Given: The linear viscoelastic initial boundary value problem described in Table 12.1.

Required:

(a) Reduce to the IBVP for the case of an object subjected to plane stress conditions.

(b) Reduce to the IBVP for the case of an object subjected to plane strain conditions.

(c) Further reduce the results obtained in part (a) for the case of the Dirichlet problem.

(d) Further reduce the results obtained in part (b) for the case of the Neumann problem.

Problem 12.2
Given: Gibb's free energy is given by the following:

$$g \equiv u^I - Ts - \frac{1}{\rho}\sigma_{ij}\varepsilon_{ij}$$

Required: Postulate the existence of a stress formulation for viscoelastic media of the following form:

$$\varepsilon_{ij}(\vec{x},t) \equiv D_{ij\,\tau=0}^{\tau=t} D\{\sigma_{kl}(\vec{x},t)\}$$
$$\rho g = \rho g\{\sigma_{kl}(\vec{x},t)\}$$

Using the above, derive thermodynamic constraints on the stress formulation.

Problem 12.3
Given: The relaxation modulus Eq. (10.41) and the creep compliance Eq. (10.39) for a Maxwell model.

Required:

(a) Use Eq. (9.25) to prove Eq. (12.35).

(b) Show that Eq. (9.25) holds true for a Maxwell model.

(c) Use Eq. (9.25) to show that (10.41) is linear for a Maxwell model.

Problem 12.4
Given: The results of Problem 12.2

Required:

(a) Reduce thermodynamic constraints on the linear viscoelastic stress formulation for the case of an isotropic material.

(b) Graph these constraints similar to that shown in Fig. 12.1 for the isotropic stress formulation.

Problem 12.5
Given: Suppose that the bar shown in Fig. 12.2 is tapered linearly such that $A(x_1) = A_1 - A_2 x_1$ where A_1 and A_2 are geometric constants, and in addition, the bar is subjected to an evenly distributed load per unit length, p_0 along the length of the bar in the x_1 coordinate direction. Suppose furthermore that the material may be modeled with a Maxwell model.

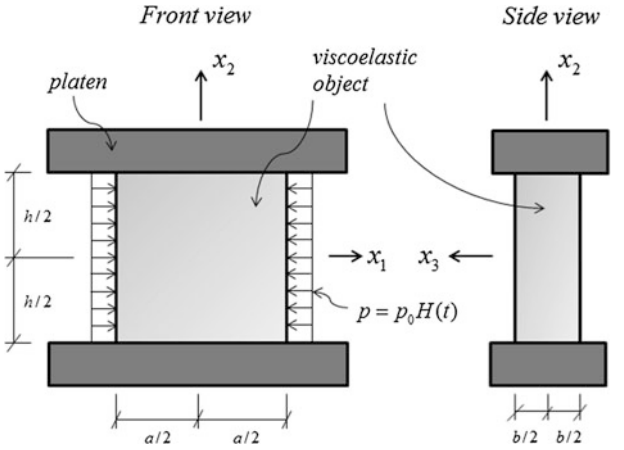
Required:

(a) Obtain the analytic solution to the problem assuming that it is one-dimensional in x_1.

(b) Plot the deflection $u_1(x_1 = L)$ as a function of time.

Problem 12.6

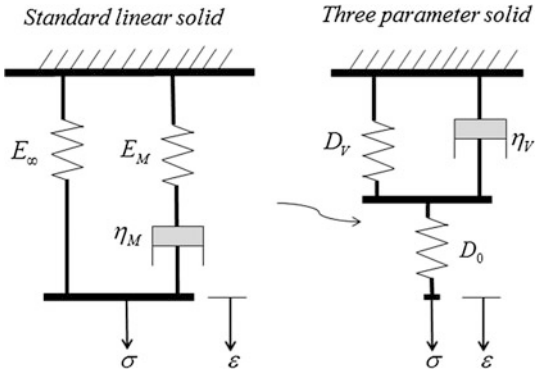
Given: A rectangular parallelepiped of isotropic linear viscoelastic material with $\nu = const.$ and modulus $E(t)$ is placed between rigid platens and subjected to lateral pressure as shown below:



Required: Determine all components of the stress, strain, and displacement as functions of the input loads, geometry, and material properties.

Problem 12.7

Given: The two viscoelastic analogs shown below.



Required:

(a) Derive the governing differential equation for the standard linear solid.

(b) Derive the relaxation modulus for the standard linear solid.

(c) Derive the governing differential equation for the three-parameter solid.

(d) Derive the creep compliance for the three-parameter solid.

(e) Show by using the results of (a) and (c) that the two analogs are mathematically equivalent.

Problem 12.8
Given: The creep data shown below are for a typical asphalt concrete.

Compliance (GPa^{-1})	Time (s)
1.00	1.0×10^{-5}
1.02	3.16×10^{-5}
1.05	1.0×10^{-4}
1.10	3.16×10^{-4}
1.18	1.0×10^{-3}
1.29	3.16×10^{-3}
1.41	1.0×10^{-2}
1.55	3.16×10^{-2}
1.70	1.0×10^{-1}
1.88	3.16×10^{-1}
2.10	1.0×10^{0}
2.35	3.16×10^{0}
2.60	1.0×10^{1}
2.93	3.16×10^{1}
3.25	1.0×10^{2}
3.58	3.16×10^{2}
3.88	1.0×10^{3}
4.15	3.16×10^{3}
4.25	1.0×10^{4}
4.30	3.16×10^{4}
4.32	1.0×10^{5}

Required:

(a) Use collocation to fit the data over the entire time span of the experiment.

(b) Plot the predicted creep compliance and the data on the same graph.

(c) Fit the data using a power law over the time span $10^{-2} \leq t \leq 10^{2}$.

Problem 12.9
Given: The creep data shown below are obtained at four different temperatures for an asphalt binder with fines but no aggregate

Time (s)	T = 0 °C	T = 10 °C	T = 20 °C	T = 30 °C
	Compliance (GPa^{-1})	Compliance (GPa^{-1})	Compliance (GPa^{-1})	Compliance (GPa^{-1})
1.0×10^{-1}	0.145	0.210	0.400	0.680
3.16×10^{-1}	0.150	0.250	0.470	0.750
1.0×10^{0}	0.160	0.300	0.540	0.780
3.16×10^{0}	0.180	0.360	0.600	0.795
1.0×10^{1}	0.210	0.420	0.650	0.800

Required:

(a) Plot the data on a single graph.

(b) Use time–temperature superposition to create a single master curve and plot on the graph.

(c) Plot the shift factor versus temperature and determine whether either the Arrhenius or the WLF equation represents the shift factor.

(d) Use the collocation method to fit the master curve.

(e) Plot the predicted master curve on the graph and compare to the experimental master curve.

Problem 12.10
Given: The frequency sweep data for a typical asphalt binder shown in Table 12.3.

Required:

(a) Using the procedure for fitting frequency sweep data described in this chapter fit the imaginary part of the complex modulus as a function of frequency.

(b) Plot the result on a graph and compare to the experimental data.

(c) Transform the result to the time domain and compare to the prediction obtained using the real part of the complex modulus shown in Fig. 12.15.

References

Allen, D., & Haisler, W. (1985). *Introduction to aerospace structural analysis*. Wiley.
Alfrey, T. (1944). Nonhomogeneous stresses in viscoelastic media. *Quarterly of Applied Mathematics, 2,* 113.

Ban, H., Im, S., & Kim, Y. (2013). Nonlinear viscoelastic approach to model damage-associated performance behavior of asphaltic mixture and pavement structure. *Canadian Journal of Civil Engineering, 40,* 313.

Biot, M. (1954). Theory of stress-strain relations in anisotropic viscoelasticity and relaxation phenomena. *Journal of Applied Physics, 25,* 1385.

Biot, M. (1965). *Mechanics of incremental deformations: Theory of elasticity and viscoelasticity of initially stressed solids and fluids, including thermodynamic foundations and applications to finite strains*. Wiley.

Boltzmann, L. (1874). Zur theorie der elastischen nachwirkungen. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften Mathematisch-Naturwissenschaftliche, 70,* 275–306.

Christensen, R., & Naghdi, P. (1967). Linear non-isothermal viscoelastic solids. *Acta Mechanica, 3,* 1.

Christensen, R. (1982). *Theory of viscoelasticity—An introduction* (2nd ed.). New York: Academic Press.

Coleman, B. (1964). Thermodynamics of materials with memory. *Journal of Chemical Physics, 47,* 597.

Coulombe, C. (1784). Recherches théoretiques et experimentales sur la force de torsion et sur l'élasticité des fils de metal. *Mémoires de l'Académie Royale des Sciences*.

Day, W. (1972). *The thermodynamics of simple materials with fading memory*. Springer.

Debye, P. (1913). *Ver. Deut. Phys. Gesell.*, *15*, 777, New York.

Ferry, J. (1980). *Viscoelastic properties of polymers* (3rd ed.). New York: Wiley.

Flugge, W. (1975). *Viscoelasticity* (Second Revised ed.). Berlin: Springer.

Golden, H., Strganac, T., & Schapery, R. (1999). An approach to characterize nonlinear viscoelastic material behavior using dynamic mechanical tests and analyses. *Journal of Applied Mechanics, 66,* 872.

Green, A., & Rivlin, R. (1957). The mechanics of non-linear materials with memory; Part I. *Archive for Rational Mechanics and Analysis, 1,* 1.

Green, A., Rivlin, R., & Spencer, A. (1959). The mechanics of non-linear materials with memory: Part II. *Archive for Rational Mechanics and Analysis*, *3*, 82.

Greenburg, M. (1978). *Foundations of applied mathematics*. Prentice Hall.

Huang, C., Abu Al-Rub, R., & Masad, E. (2010). *Journal of Materials in Civil Engineering*, *23*, 56.

Kim, R., & Little, D. (1990). One-dimensional constitutive modeling of asphalt concrete. *Journal of Engineering Mechanics, 116,* 751.

Kohlrausch, F. (1863). Pogg. *Annals of Physics, 4,* 337.

Kreyszig, E. (2006). *Advanced engineering mathematics*. Wiley.

Leaderman, H. (1943). *Elastic and creep properties of filamentous materials and other High polymers*. Washington, D.C.: Textile Foundation. 175.

Lee, H., Daniel, J., & Kim, Y. (2000). Continuum damage mechanics-based fatigue model of asphalt concrete. *Journal of Materials in Civil Engineering, 12,* 105.

Lou, Y., & Schapery, R. (1971). Viscoelastic characterization of a nonlinear fiber-reinforced plastic. *Journal of Composite Materials, 5,* 208.

Masad, E., Huang, C., Airey, A., & Muliana, A. (2008). Nonlinear viscoelastic analysis of unaged and aged asphalt binders. *Construction and Building, 22,* 2170.

Maxwell, J. (1867). On the dynamical theory of gases. *Philosophical Transactions of the Royal Society of London, 157,* 49–88.

Maxwell, J. (1875). On the dynamical evidence of the molecular constitution of bodies. *Journal of the Chemical Society London, 28,* 493–508.

Park, S., Kim, R., & Schapery, R. (1996). A viscoelastic continuum damage model and its application to uniaxial behavior of asphalt concrete. *Mechanics of Materials, 24,* 241.

Reiner, M. (1964). The Deborah number. *Physics Today, 17,* 62.

Saadeh, S., Masad, E., & Little, D. (2007). Characterization of asphalt mix response under repeated loading using anisotropic nonlinear viscoelastic-viscoplastic model. *Journal of Materials in Civil Engineering, 19,* 912.

Schapery, R. (1962). Approximate methods of transform inversion for viscoelastic stress analysis. In *Proceedings of 4th U.S. National Congress of Applied Mechanics* (Vol. 1075).

Schapery, R. (1962). Irreversible thermodynamics and variational principles with applications to viscoelasticity. Cal Tech Thesis.

Schapery, R. (1964). Application of thermodynamics to thermomechanical, fracture, and birefringent phenomena in viscoelastic media. *Journal of Applied Physics, 35,* 1451.

Schapery, R. (1966). An engineering theory of nonlinear viscoelasticity with applications. *International Journal of Solids and Structures, 2,* 407.

Schapery, R. (1969). On the characterization of nonlinear viscoelastic materials. *Polymer Engineering & Science, 9,* 295.

Schapery, R. (1997). Nonlinear viscoelastic and viscoplastic constitutive equations based on thermodynamics. *Mechanics of Time-Dependent Materials*, *1*, 209.

Truesdell, C., Noll, W., & Antman, S. (2004). *The non-linear field theories of mechanics*. Springer.

Tschoegl, N. (1989). *The phenomenological theory of linear viscoelastic behavior an introduction*. Berlin: Springer.

Volterra, V. (1928). Sur la théorie mathématique des phénomènes héréditaires. *Journal de Mathématiques Pures et Appliquées, 7,* 249–298.

Volterra, V. (1959). *Theory of functionals and of integral and integro-differential equations*. Dover.

Weber, W. (1835). Über die elastizität der seidenfäden. *Annalen der Physik und Chemie Poggendorf's, 4,* 247–257.

Weber, W. (1841). *Poggendorf Annalen der Physik*, *24*, 1–26.

Chapter 13
Plasticity, Viscoplasticity, and Fracture

13.1 Introduction

The previous chapter focused on models for predicting the behavior of viscoelastic media. Viscoelastic material models have been demonstrated to be effective for modeling the material behavior of a wide range of materials, including both asphalt binder (see Chap. 2) and asphalt concrete (see Chaps. 7 and 12). However, viscoelastic material models are typically not utilized for the purpose of modeling other geologic media, including road way base materials.

In many cases, road way failure is caused by excessive deformations and/or fracture within the surface layer of the road way (see Chap. 8). However, in some cases the road way fails due to excessive permanent deformations (usually termed rutting), and in these circumstances, it is often observed that a primary component of the rutting is the result of permanent deformations within the road way base and/or sub-base material (see Chap. 8). This chapter focuses on the development of constitutive models for predicting permanent deformations in these layers of the road way, as well as fracture within the road way (see Chap. 8).

There are several reasons that specialized material models are required for road way base materials, and they are related to their fundamental material makeup (see Chaps. 2 through 6). These reasons are described briefly below.

Firstly, as shown in Fig. 13.1, when road way materials are observed under low magnification they are typically observed to be composed of large grains embedded within an intergranular material (see Chap. 7). When subjected to mechanical loading, this granular structure is typically locked in place up to significant and somewhat predictable stress states. Up to the point at which the grains begin sliding with respect to one another and/or undergo fracture, the macroscopic response of the medium will result primarily from linear elastic deformations within the individual grains, and little if any permanent deformations (as well as energy dissipation) are observed in the material. For this reason, up to the point at which sliding

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_13

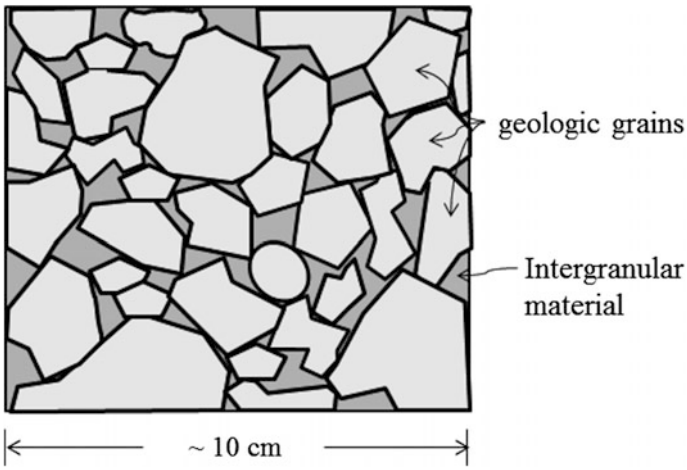


Fig. 13.1 Typical granular structure of geologic media

is initiated the material can be modeled quite accurately with the use of a linear elastic material model, as described in detail in Chap. 11.

Secondly, in many geologic media the grains are sufficiently tough that grain sliding is initiated well before the onset of fracture within individual grains. Beyond the loading level at which relative sliding between grains is initiated, the response includes an energy dissipative viscous effect caused by grain sliding, and this effect is typically observed to be highly nonlinear. Grain sliding can typically be modeled at least approximately at the macroscopic scale by incorporating a single nonlinear viscous term within the constitutive model. This type of material model is not designed to account for intragranular fracture because this phenomenon is not a commonplace in typical road way base materials.

When there is little or no moisture in the material, the grain sliding mechanism is typically observed to be rate independent, and in this case, the material behavior can be modeled with an elastoplastic constitutive material model. Alternatively, when the medium contains significant moisture, both the grain sliding and the bulk material exhibit rate dependence, so that the macroscopic material behavior will typically be observed to be rate dependent, and in this case, the material behavior must be modeled with a viscoplastic material model (see Chap. 8).

In Chap. 10, one-dimensional material models were introduced for elastoplastic and viscoplastic materials. In this chapter, the discussion focuses on three-dimensional elastoplastic and viscoplastic material behavior. These models are termed "phenomenological models," meaning that they are constructed at the macroscopic scale of the road way, assuming that the base and sub-base materials may be treated as macroscopically homogeneous. This assumption is of course an approximation of reality that necessarily results in some error in the models. For the purpose of modeling permanent deformations within road ways (see Chap. 8), this

error is normally not excessive. However, in cases where it is, it is advisable to utilize either micromechanics or multi-scaling approaches, as detailed in Chaps. 11 and 15.

13.2 Multi-dimensional Plasticity

As described in Fig. 10.13, uniaxial tests performed on geologic media often result in elastic material response up to the yield point, σ_Y, and the material undergoes permanent deformation if this level of stress is exceeded. This permanent deformation is typically nonlinear, and, as described above, it is often caused by microscopic slip between particles within the geologic medium. When this process is rate independent, the material is termed elastoplastic. In accordance with one-dimensional definition (10.43), for the three-dimensional case *an elastoplastic material* is defined as follows:

$$\begin{aligned} \sigma_{ij} &= E_{ijkl}\varepsilon_{kl} \quad \forall F(\sigma_{ij}) - F^Y < 0 \\ \sigma_{ij}(\vec{x},t) &= \overset{\tau=t}{\underset{\tau=0}{E}}\{\varepsilon_{kl}(\vec{x},\tau)\} = \overset{\tau=ct}{\underset{\tau=0}{E}}\{\varepsilon_{kl}(\vec{x},c\tau)\} \quad \tau \to c\tau \quad \forall F(\sigma_{ij}) - F^Y \geq 0 \end{aligned} \tag{13.1}$$

where $F(\sigma_{ij})$ is called *the yield function*, and F^Y is a history-dependent material parameter. These two quantities together make up what is termed *the yield criterion*.

Historically, the subject of elastoplasticity traces its roots to the work of Leonardo da Vinci (1452–1519), Allen (2014). Unfortunately, the results described within his journals were not published until much later (1699), so that the first scientific reference on the subject was reported by Guillaume Amontons (1663–1705), who recorded three laws of friction in 1699 (l'Académie Française 1699). Charles-Augustin de Coulomb (1736–1806), who developed a model for predicting the maximum load that retaining walls could withstand (Coulomb 1776), later expanded on Amontons' laws. A modern amalgamation of the above two laws, termed *Coulomb's law of friction*, can be stated as follows (see Fig. 13.2):

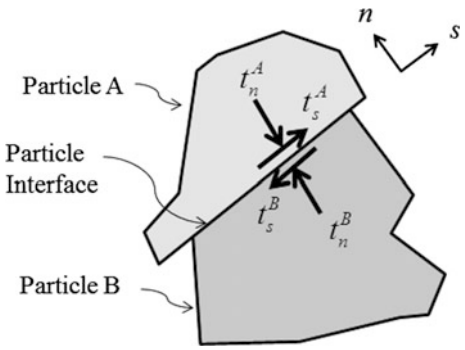


Fig. 13.2 Depiction of Coulomb's law of friction

$$\begin{aligned} t_s \leq \mu t_n &\Rightarrow u_s^A = u_s^B \\ t_s > \mu t_n &\Rightarrow u_s^A \neq u_s^B \end{aligned} \tag{13.2}$$

where the subscripted t denotes traction components, the subscripted u denotes displacement components, and the subscripts s imply components parallel to the surfaces in contact with one another, whereas the subscripts n imply components normal to the surfaces in contact with one another. In addition, the term μ is called *the coefficient of friction*, a constant empirical material property describing the threshold value at which sliding is initiated between the two contact surfaces. Thus, as shown in Eq. (13.2), when the shear traction at the interface exceeds the coefficient of friction, μ, multiplied by the normal traction component, t_n, at the interface, the shear component of displacement, u_s, along the interface is predicted by Coulomb's model to become discontinuous, thus implying interfacial slip between the two particles. As such, the above friction model represents the first scientifically reported model of failure to appear in the open literature.

The above model was later expanded into the so-called Mohr–Coulomb theory, to be discussed below. Further developments were due to Henri Édouard Tresca (1814–85), Tresca (1864), Christian Otto Mohr (1835–1918), Timoshenko (1953), Richard von Mises (1883–1953), Mises (1913), Rodney Hill (1921–2011) and Hill (1950), to name a few. During the latter half of the twentieth century, plasticity theory was extended to account for the effects of cyclic loading, as computers became sufficiently powerful to model the effects of complicated loading histories.

The theory of plasticity is a complex subject that typically requires significantly more study than is possible within the context of this text. Instead, a short overview will be developed within this chapter that is hopefully sufficient for the purpose of describing only that necessary for the purpose of designing road ways. The interested reader is referred to the literature on this complex subject for more details (Hill 1950; Prager 1959; Lubliner 2008; Negahban 2012).

In practice, a specific form of Eq. (13.1) is normally utilized, and this specific form is comprised of four distinct parts as follows:

(1) The stress–elastic strain relationship;
(2) The yield function;
(3) The flow rule; and
(4) The workhardening rule.

These four components of the model are described in the following sections.

13.2.1 The Stress–Elastic Strain Relationship

It is observed experimentally that elastoplastic materials exhibit a reproducible behavior that consists of two distinct ranges of behaviors, one elastic and the other inelastic, and these two ranges may be distinguished via the use of a yield function,

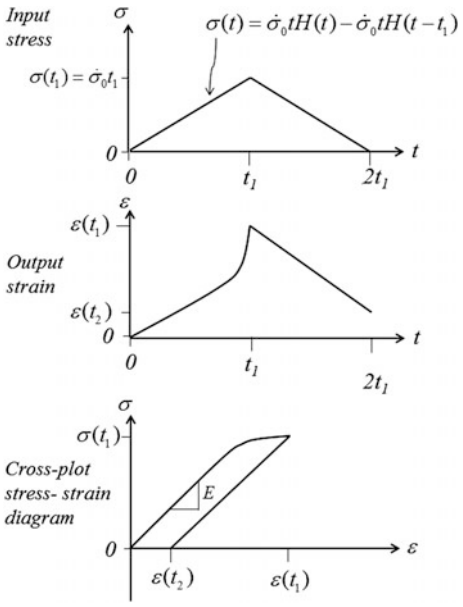


Fig. 13.3 Typical results of a uniaxial test performed on an elastoplastic material

to be described in detail in the next section. As an example of these two distinct ranges, consider the results of a uniaxial test, as described in Chap. 10. In this case, the specimen is both loaded and unloaded, and if the yield point is exceeded, the uniaxial stress–strain curve will be observed as shown in Fig. 13.3. Note that, according to Eq. (13.1), in order for the material to behave elastoplastically, the experimental results must be independent of the loading and unloading rates.

Now suppose that, by observation of the experimental result in Fig. 13.3, the *uniaxial plastic strain* is defined as follows:

$$\varepsilon^P(t) \equiv \varepsilon(t) - \frac{\sigma(t)}{E} \Rightarrow \sigma(t) = E\left[\varepsilon(t) - \varepsilon^P(t)\right] \tag{13.3}$$

By analogy, in the multi-axial case, it is assumed that the stress may be related to the strain at any point in time (for the isothermal case) via the following:

$$\sigma_{ij} = E_{ijkl}(\varepsilon_{kl} - \varepsilon^P_{kl}) \tag{13.4}$$

where ε^P_{ij} is called the *plastic strain tensor*. It should be apparent that this formulation is an assumption, but experimental evidence indicates that it is an accurate one for most elastoplastic materials, including many geologic materials used in road way bases and sub-bases. This may be explained physically by the observation that once the yield point is reached on loading, particles within the material begin sliding with respect to one another (thereby dissipating energy), but on load

reduction, the particles become locked into a fixed configuration with respect to one another (thereby dissipating little or no energy).

Where the material is observed to be initially isotropic, the formulation may be simplified still further to the following:

$$\sigma_{ij} = \lambda(\varepsilon_{kk} - \varepsilon^P_{kk})\delta_{ij} + 2\mu(\varepsilon_{ij} - \varepsilon^P_{ij}) \tag{13.5}$$

Assuming that the material has not been previously loaded, it can therefore be construed by comparison between Eqs. (13.5) and (11.33a) that *in the initial state all components of the plastic strain, ε^P_{ij}, are identically zero in an elastoplastic material*. In most circumstances, road way geologic materials can be modeled accurately with the initially isotropic formulation given by Eq. (13.5).

13.2.2 The Yield Criterion

As indicated in the uniaxial test shown in Fig. 13.3, the material behavior of elastoplastic media is initially linear elastic, so that the components of the plastic strain tensor are all initially zero, and they will not begin to evolve until initial yielding occurs. In the case wherein the material is loaded such that the stress state is uniaxial, it is a simple matter to ensure that the stress does not exceed the uniaxial yield point, σ_Y, obtained from experimental data, to determine when the plastic strain will begin to evolve. However, when the stress state is not uniaxial, it is apparent that the uniaxial yield stress is not sufficient to determine at what point yielding occurs. Toward this end, it is hypothesized that, consistent with Eq. (13.1), there exists a *yield function*, F, of the following generic form:

$$F = F(\sigma_{ij}, \varepsilon^P_{ij}) \tag{13.6}$$

such that

$$\begin{aligned} F < 0 &\Rightarrow \frac{d\varepsilon^P_{ij}}{dt} = 0 \\ F \geq 0 &\Rightarrow \frac{d\varepsilon^P_{ij}}{dt} \neq 0 \end{aligned} \tag{13.7}$$

Furthermore, *it is assumed that F is never greater than zero*, implying that the yield function evolves whenever the plastic strain evolves.

In the case of initial yielding, F is denoted by F^I, and since the initial plastic strain is necessarily zero, (13.6) simplifies to the following:

$$F^I = F^I(\sigma_{ij}) \tag{13.8}$$

The above can be viewed as a surface in six-dimensional stress hyperspace. It would stand to reason that the function would usually be simply connected, convex, and enclose the coordinate origin in the space of stress.

The precise form that Eq. (13.8) will take on will depend on the material to be modeled, and the equation describing yielding will necessarily have to be constructed based on experimental evidence. Since multi-axial experiments can be quite cumbersome and expensive to perform, it is advisable to deploy as much physical cognition as possible before resorting to experimentation for the purpose of describing Eq. (13.8). For example, in the case wherein the material is initially isotropic, Eq. (13.8) may be simplified by writing it in terms of either the principal stresses or the stress invariants as follows:

$$F^I = F^I(\sigma_1, \sigma_2, \sigma_3) = F^I(I_1, I_2, I_3) \tag{13.9}$$

It is apparent that Eq. (13.9) is a significant simplification of Eq. (13.8). Accordingly, when the material is statistically isotropic, it is possible to represent the initial yield function as a surface in a three-dimensional principal stress space, as shown in Fig. 13.4.

By manipulating Eqs. (9.57)–(9.59) and (9.65)–(9.67), it can be shown that

$$J_2 = \frac{I_1^2}{3} - I_2 \Rightarrow I_2 = \frac{I_1^2}{3} - J_2 \tag{13.10}$$

and

$$J_3 = I_3 - \frac{I_1 I_2}{3} + \frac{2}{27} I_1^3 \Rightarrow I_3 = J_3 + \frac{I_1 I_2}{3} - \frac{2}{27} I_1^3 \tag{13.11}$$

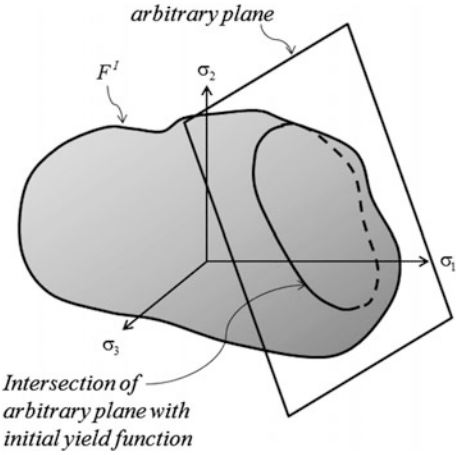


Fig. 13.4 Depiction of initial yield surface in principal stress space for an isotropic material

Substituting Eqs. (13.10) and (13.11) into Eq. (13.9) will therefore result in the following mathematically equivalent form for the initial yield function:

$$F^I = F^I(I_1, J_2, J_3) \tag{13.12}$$

While the above representation is mathematically equivalent, it is a physically intriguing representation for the initial yield function due to the fact that it is written explicitly in terms of the average hydrostatic stress, $\sigma_0 = I_1/3$, and the deviatoric stress invariants.

For some materials, further simplification of Eq. (13.8) (or equivalently, Eq. 13.12 for the case where the material isotropic) is possible. To see how this can be achieved, consider a particular plane in principal stress space termed *the π-plane*, defined by the following function:

$$f(\sigma_1, \sigma_2, \sigma_3) = \sigma_1 + \sigma_2 + \sigma_3 = 0 \tag{13.13}$$

A unit outer normal vector, $\vec{n}$, to this plane may be obtained by taking the gradient of the function, f, with respect to the three principal stress components, as follows:

$$\frac{\partial f}{\partial \sigma_1} = 1, \frac{\partial f}{\partial \sigma_2} = 1, \frac{\partial f}{\partial \sigma_3} = 1 \Rightarrow \vec{n} = \frac{1}{\sqrt{3}}\vec{e}_1 + \frac{1}{\sqrt{3}}\vec{e}_2 + \frac{1}{\sqrt{3}}\vec{e}_3 \tag{13.14}$$

Thus, the π-plane can be seen to be a flat (having zero curvature everywhere) plane that makes equal angles with each of the principal stress axes, as shown in Fig. 13.5. Note that due to Eq. (13.13), on this plane the hydrostatic pressure is zero,

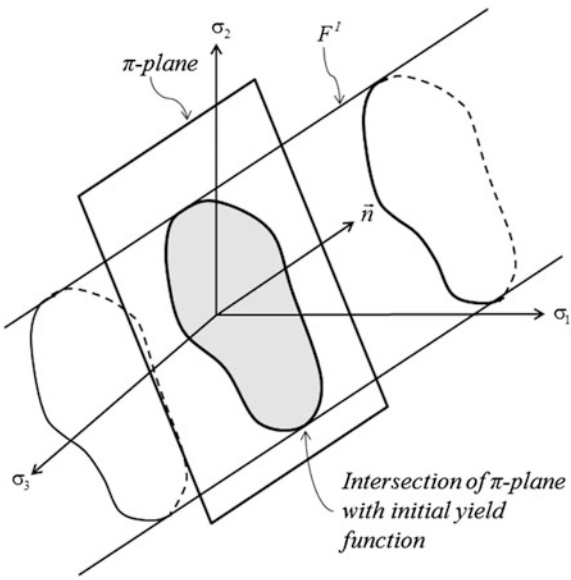


Fig. 13.5 Depiction of the π-plane together with hydrostatic pressure-independent initial yield function

implying that on the π-plane $I_1 = 0$. Thus, the initial yield function given by Eq. (13.12) depends only on the deviatoric stress invariants when the state of stress lies on the π-plane.

The above-assembled physical information can now be utilized to hypothesize several specific possibilities for the initial yield function for a variety of materials.

13.2.2.1 Von Mises Yield Criterion

Some materials, such as most metals, exhibit initial yielding that is observed to be independent of hydrostatic stress, so that, whatever shape the initial yield surface takes on in the π-plane, this shape is unchanged by translation normal to the π-plane, as shown in Fig. 13.5. Thus, the simplest approximation of Eq. (13.12) for a material that exhibits yielding that is independent of hydrostatic pressure is called the *von Mises yield criterion* (Mises 1913), given by the following mathematical expression:

$$F_{VM}^{I} \equiv \sqrt{J_2} - C_{VM}^{Y} = \left[\frac{1}{2}(\sigma_1^{'})^2 + (\sigma_2^{'})^2 + (\sigma_3^{'})^2\right]^{1/2} - C_{VM}^{Y} \tag{13.15}$$

where, from Eq. (13.1), C_{VM}^{Y} is a material constant to be determined experimentally.

To see how the above function appears in principal stress space, first consider the following function:

$$\left[\frac{1}{2}\left(\sigma_1^2 + \sigma_2^2 + \sigma_3^2\right)\right]^{1/2} - C_{VM}^{Y} = 0 \tag{13.16}$$

The square of the above equation is clearly the equation of a sphere in principal stress space with center at the coordinate origin. Since the hydrostatic stress is zero in the π-plane, Eqs. (13.15) and (13.16) coincide in the π-plane, thereby indicating that the square of Eq. (13.15) represents a circle in the π-plane. Accordingly, since the von Mises yield criterion is independent of hydrostatic pressure, it can be inferred that its square is represented by a right circular cylinder of constant radius, and with principal axis coincident with the unit vector, $\vec{n}$, as shown in Fig. 13.6.

The above yield criterion has been shown to be quite accurate for predicting initial yielding in a variety of initially isotropic polycrystalline materials, including many metals. This is due to the fact that these materials yield initially due to molecular-scale physical phenomena such as dislocation climb and cross-slip that do not behave according to friction law (13.2).

Note that, due to the fact that the cylinder is of constant radius and extends to infinity, it is characterized by a single constant that is equivalent to the cylinder radius. Thus, one may choose any experiment that is convenient for the purpose of evaluating this quantity, such as indeed a tensile uniaxial test, in which yielding is observed to occur at a uniaxial value of stress given by $\sigma_{11} = \sigma_Y^T$. Since all of the

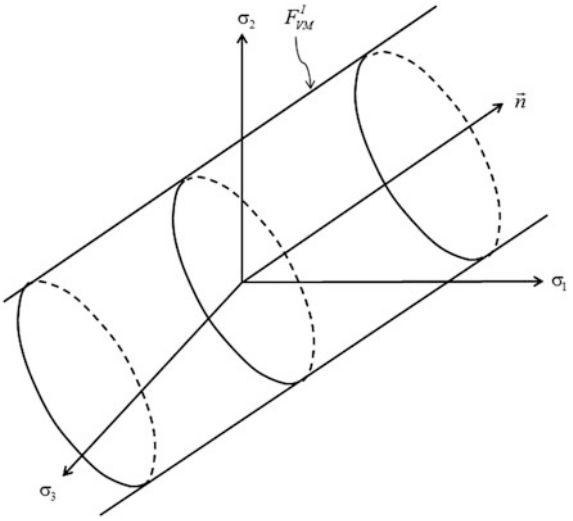


Fig. 13.6 Depiction of the von Mises yield criterion in principal stress space

remaining components of stress are zero in this experiment, it follows that, by direct substitution of the above state of stress into Eq. (13.15),

$$\left[\frac{1}{2}\left(\left(\sigma_Y-\frac{\sigma_Y}{3}\right)^2+\left(-\frac{\sigma_Y}{3}\right)^2+\left(-\frac{\sigma_Y}{3}\right)^2\right)\right]^{1/2}-C^Y_{VM}=0\Rightarrow C^Y_{VM}=\frac{\sigma^T_Y}{\sqrt{3}} \tag{13.17}$$

thus providing the value of the material constant, C^Y_{VM}, from a simple uniaxial test.

The above is of course a model for initial yielding, and as such, the accuracy obtained with it will depend on the material in question. In practice, only initially isotropic crystalline materials behave according to the von Mises yield criterion. Unfortunately, road way base materials do not behave according to this model, thus requiring development of more complex yield criteria, to be described below.

13.2.2.2 Tresca Yield Criterion

Another yield criterion that is independent of hydrostatic stress is actually the first criterion proposed specifically for yielding. This criterion, called the *Tresca yield criterion* (Tresca 1864), assumes that yielding will occur when the maximum shearing stress at a material point reaches a critical value. This assumption may be described mathematically as follows:

$$F^I_T\equiv\frac{1}{2}\max[|\sigma_1-\sigma_2|,|\sigma_2-\sigma_3|,|\sigma_3-\sigma_1|]-C^Y_T \tag{13.18}$$

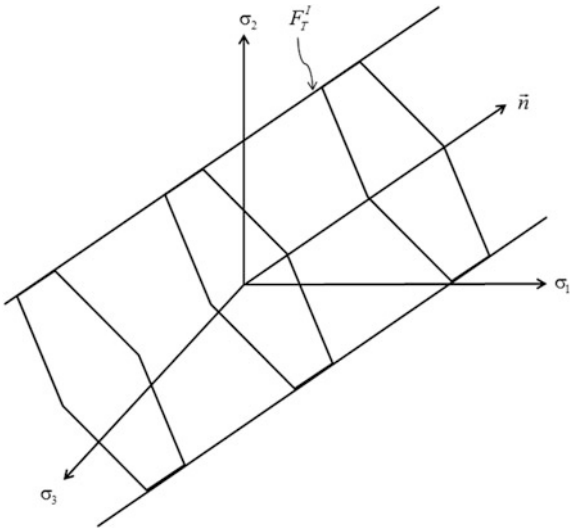


Fig. 13.7 Depiction of Tresca yield criterion in principal stress space

where C_T^Y is the material constant to be determined experimentally. It can be shown that the Tresca yield criterion is both hexagonal in the π-plane and independent of hydrostatic pressure, so that it produces the surface shown in Fig. 13.7.

For the case wherein the state of stress is plane stress ($\sigma_3 = 0$), the Tresca and von Mises yield criteria may be depicted graphically in a two-dimensional subspace of principal stress space, as shown in Fig. 13.8. Once again, any experiment that is expedient can be used to evaluate C_T^Y. For example, if a tensile uniaxial test is performed, substituting the uniaxial yield stress in tension, σ_Y^T, into Eq. (13.18) will result in the following:

$$\frac{1}{2}\max\left[\left|\sigma_Y^T - 0\right|, |0 - 0|, |0 - 0|\right] - C_T^Y = 0 \Rightarrow C_T^Y = \frac{\sigma_Y^T}{2} \tag{13.19}$$

thus providing a simple uniaxial experiment for obtaining the material constant necessary to complete the description of the Tresca yield criterion. Note that by comparison with the initial yield values obtained in Eqs. (13.17) and (13.19), the Tresca yield criterion provides a more conservative predictor of initial yielding than does the von Mises yield criterion when a uniaxial test is used to evaluate the constant C_T^Y.

13.2.2.3 Drucker–Prager Yield Criterion

Unlike that predicted by the von Mises and Tresca yield criteria, most geologic media exhibit initial yield behavior that is hydrostatic stress dependent, that is,

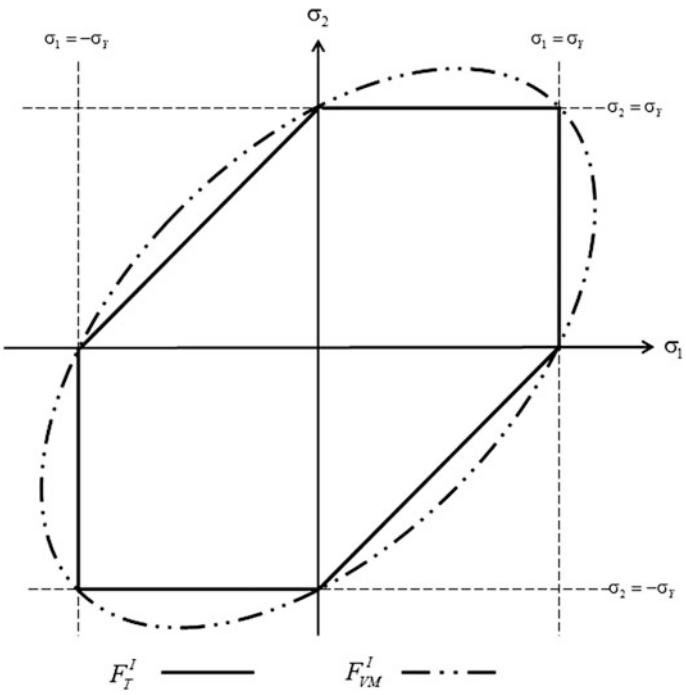


Fig. 13.8 Depiction of the von Mises and Tresca yield criteria in two-dimensional principal stress subspace

$$F^I = F^I(I_1) \tag{13.20}$$

This is due to the fact that initial yielding in these materials is at least partially due to sliding friction, described by friction law (13.2). Note that this law accounts for hydrostatic pressure implicitly via the normal traction component, t_n. The von Mises and Tresca yield criteria are therefore not sufficiently accurate for predicting initial yielding in geologic media. On the other hand, these media tend to be initially isotropic, and they do exhibit yielding that depends on the second deviatoric stress invariant, J_2.

Toward this end, perhaps the simplest initial yield criterion that accounts for both I_1 and J_2 dependence is called the *Drucker–Prager yield criterion* (Prager 1959), which is nothing more than a linear combination of these invariants:

$$F^I_{DP} \equiv \sqrt{J_2} + \frac{\mu}{\sqrt{6}} I_1 - C^Y_{DP} \tag{13.21}$$

where C^Y_{DP} and μ are material constants to be determined from experiments. It can be seen by comparison between Eqs. (13.15) and (13.21) that when μ is identically zero, the Drucker–Prager yield criterion degenerates to the von Mises yield criterion. Thus, since I_1 is identically zero in the π-plane, the Drucker–Prager yield

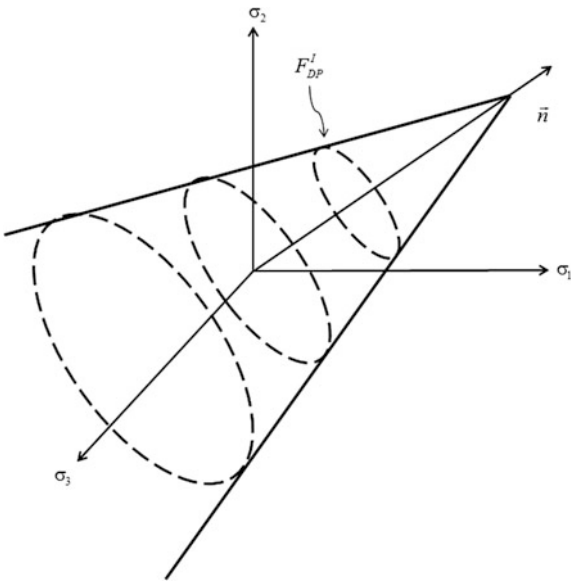


Fig. 13.9 Depiction of the Drucker–Prager yield criterion in principal stress space

criterion is identical to the von Mises yield criterion in this plane, implying that the Drucker–Prager yield criterion is circular in the π-plane. Outside the π-plane, the function remains circular for constant values of I_1, implying that the Drucker–Prager yield criterion is circular on any plane parallel to the π-plane. Finally, since the Drucker–Prager yield criterion is linear in I_1, it can now be established that the Drucker–Prager yield criterion is a right circular cone in principal stress space, with principal axis of the cone coincident with the unit normal vector, $\vec{n}$, to the π-plane, as shown in Fig. 13.9.

Various experiments may be used to determine the two material constants, C^Y_{DP} and μ, in Eq. (13.21). However, it should be evident that precisely two experiments are required, so long as they are linearly independent of one another. This is due to the fact that the orientation of the Drucker–Prager yield criterion in principal stress space is by necessity aligned normal to the π-plane, and once the orientation of a right circular cone is determined, only two additional values on the surface of the cone must be known in order to completely specify the cone geometrically.

These two experiments are arbitrary, except that they must be linearly independent (implying that they must be performed with different values of average hydrostatic stress), thus requiring more complicated experiments than those required for materials that do not exhibit hydrostatic stress dependence.

For example, first consider the case of a test performed in pure shear, given by the following:

$$\sigma_{12} \neq 0, \sigma_{ij} = 0 \quad \forall \quad other \quad i, j \Rightarrow \sigma_2 = -\sigma_1, \sigma_3 = 0 \tag{13.22}$$

Substituting the above into Eq. (13.21) for the case where yielding in pure shear occurs $(\sigma_{12} = \sigma_1 = -\sigma_2 = \sigma_S)$ will result in the following equation:

$$F_{DP}^{I} = 0 = \sqrt{\frac{1}{2}(\sigma_S)^2 + \frac{1}{2}(\sigma_S)^2} + 0 - C_{DP}^{Y} \Rightarrow C_{DP}^{Y} = \sigma_S \tag{13.23}$$

Next consider the case of a uniaxial compression test, given by the following:

$$\sigma_{11} = \sigma_1, \sigma_2 = \sigma_3 = 0, \sigma_{23} = \sigma_{13} = \sigma_{12} = 0 \tag{13.24}$$

Substituting the above into Eq. (13.21) for the case where uniaxial yielding occurs in compression $\left(\sigma_{11} = \sigma_Y^C\right)$ will result in the following equation:

$$0 = \frac{1}{\sqrt{3}}\sigma_Y^C + \frac{\mu}{\sqrt{6}}\sigma_Y^C - C_{DP}^{Y} \tag{13.25}$$

In order to solve for μ, substitute Eq. (13.23) into Eq. (13.25), thereby resulting in the following:

$$\mu = \sqrt{6}\frac{\sigma_S}{\sigma_Y^C} - \sqrt{2} \tag{13.26}$$

From Fig. 13.9, it can be seen that if pure hydrostatic pressure is applied to the material (moving in the negative direction along the vector $\vec{n}$) yielding will never be predicted by the Drucker–Prager yield criterion, no matter how large the hydrostatic pressure grows. This would seem to be unrealistic. Thus, the Drucker–Prager yield criterion should be used with caution when attempting to model materials undergoing pure hydrostatic pressure or large values of average hydrostatic pressure (meaning negative average hydrostatic stress). This circumstance arises oftentimes in road ways, and the Drucker–Prager yield criterion should be used when this circumstance arises. Fortunately, there is a modification to the Drucker–Prager model that may be utilized in this circumstance, and this will be described below.

Example Problem 13.1 Given: Suppose that the following two experiments are performed for the purpose of characterizing yielding. These are as follows:

Test #1 is a uniaxial compressive ramp test performed in load control.

Test #2 is a multi-axial compression test with axial ramp loading and lateral pressure $p = 100$ KPa.

Required: Determine the material constants in the Drucker–Prager yield criterion.

Solution: From test #1, it is found that the uniaxial yield stress is $\sigma_Y^C = -750$ KPa. This results in the following:

$$I_1 = \sigma_{kk} = (-750 + 0 + 0) = -750 \text{ KPa} \tag{13.27}$$

Accordingly,

$$\begin{aligned}\sigma_{11}^{'} &= \sigma_{11} - I_1/3 = -750 + 750/3 = -500 \text{ KPa} \\ \sigma_{22}^{'} &= \sigma_{33}^{'} = \sigma_{22} - I_1/3 = 0 + 750/3 = 250 \text{ KPa} \\ \sigma_{23}^{'} &= \sigma_{13}^{'} = \sigma_{12}^{'} = 0\end{aligned} \tag{13.28}$$

It follows that

$$J_2 = \frac{1}{2}\sigma_{ij}^{'}\sigma_{ij}^{'} = \frac{1}{2}\left[(-500)^2 + (250)^2 + (250)^2\right] = 187,500 \text{ KPa}^2 \tag{13.29}$$

Substituting the results of (13.27) and (13.29) into Eq. (13.21) therefore gives the following:

$$F_{DP}^{I} \equiv \sqrt{J_2} + \frac{\mu}{\sqrt{6}} I_1 - C_{DP}^{Y} \Rightarrow 0 = \sqrt{187500} + \frac{\mu}{\sqrt{6}}(-750) - C_{DP}^{Y} \tag{13.30}$$

From test #2, it is found that the axial stress at yield is $\sigma_{11}^{Y} = -800\ KPa$. This results in the following:

$$I_1 = \sigma_{kk} = (-800 - 100 - 100) = -1000 \text{ KPa} \tag{13.31}$$

Accordingly,

$$\begin{aligned}\sigma_{11}^{'} &= \sigma_{11} - I_1/3 = -800 + 1000/3 = -466.7 \text{ KPa} \\ \sigma_{22}^{'} &= \sigma_{33}^{'} = \sigma_{22} - I_1/3 = 0 + 1000/3 = 333.3 \text{ KPa} \\ \sigma_{23}^{'} &= \sigma_{13}^{'} = \sigma_{12}^{'} = 0\end{aligned} \tag{13.32}$$

It follows that

$$J_2 = \frac{1}{2}\sigma_{ij}^{'}\sigma_{ij}^{'} = \frac{1}{2}\left[(-466.7)^2 + (333.3)^2 + (333.3)^2\right] = 219,993 \text{ KPa}^2 \tag{13.33}$$

Substituting the results of (13.31) and (13.33) into Eq. (13.21) therefore gives the following:

$$F_{DP}^{I} \equiv \sqrt{J_2} + \frac{\mu}{\sqrt{6}} I_1 - C_{DP}^{Y} \Rightarrow 0 = \sqrt{219993} + \frac{\mu}{\sqrt{6}}(-1000) - C_{DP}^{Y} \tag{13.34}$$

Equations (13.30) and (13.34) may now be solved simultaneously for the unknown coefficients C_{DP}^{Y} and μ, resulting in the following material properties describing initial yielding:

$$C_{DP}^{Y} = 325 \text{ KPa}, \mu = 0.35 \tag{13.35}$$

One additional comment should be made concerning the Drucker–Prager yield criterion. Geotechnical engineers are sometimes known to plot yield criteria using the *negative values* of principal stresses rather than the positive values (contrary to Cauchy's definition of stress introduced in Chap. 9), as shown in Fig. 13.9, so that in this representation the cone described by the Drucker–Prager yield criterion would be expanding toward the reader, rather than shrinking, as shown in the figure.

13.2.2.4 Mohr–Coulomb Yield Criterion

Another yield criterion that exhibits hydrostatic pressure dependence is the so-called *Mohr–Coulomb yield criterion*, after Christian Otto Mohr (1835–1918) and Charles-Augustin de Coulomb (1736–1806), although neither actually proposed the criterion in its currently used form. Rather, it was shown later by several researchers (Lubliner 2008) that the essential features of models previously proposed by Mohr and Coulomb are reflected within this yield criterion, whereby it has acquired its name. The criterion is given mathematically by the following:

$$F_{MC}^{I} = \max[(|\sigma_1 - \sigma_2| + (\sigma_1 + \sigma_2)\sin\phi, |\sigma_2 - \sigma_3| + (\sigma_2 + \sigma_3)\sin\phi, |\sigma_3 - \sigma_1| + (\sigma_3 + \sigma_1)\sin\phi)] - 2C_{MC}^{Y}\cos\phi \tag{13.36}$$

where C_{MC}^{Y} and ϕ are material properties to be determined from experiments. It can be seen by comparison of the above equation to Eq. (13.18) that when the material parameter ϕ is zero, the above yield criterion reduces to the Tresca yield criterion. Furthermore, it can be shown that the above is linear in the average hydrostatic pressure, so that the Mohr–Coulomb yield criterion is a right hexagonal cone in principal stress space, as shown in Fig. 13.10. This hexagonal cone can be shown to fit neatly within the right circular Drucker–Prager yield surface.

As indicated for the Drucker–Prager model, only two experiments are required in order to completely specify the initial Mohr–Coulomb yield surface. This may be accomplished by substituting the two different states of stress that cause incipient yield directly into Eq. (13.36). The resulting two equations comprise a set of two equations that may be solved for the material constants C_{MC}^{Y} and ϕ.

For example, first consider the case of a test performed in pure shear, given by Eq. (13.22). Substituting this into Eq. (13.36) for the case where yielding in pure shear occurs $(\sigma_{12} = \sigma_1 = -\sigma_2 = \sigma_S)$ will result in the following equation:

$$0 = \max[(|\sigma_S + \sigma_S| + (\sigma_S - \sigma_S)\sin\phi, |\sigma_S - 0| + (\sigma_S + 0)\sin\phi, |0 - \sigma_S| + (0 + \sigma_S)\sin\phi] - 2C_{MC}^{Y}\cos\phi \Rightarrow 0 = 2\sigma_S - 2C_{MC}^{Y}\cos\phi \tag{13.37}$$

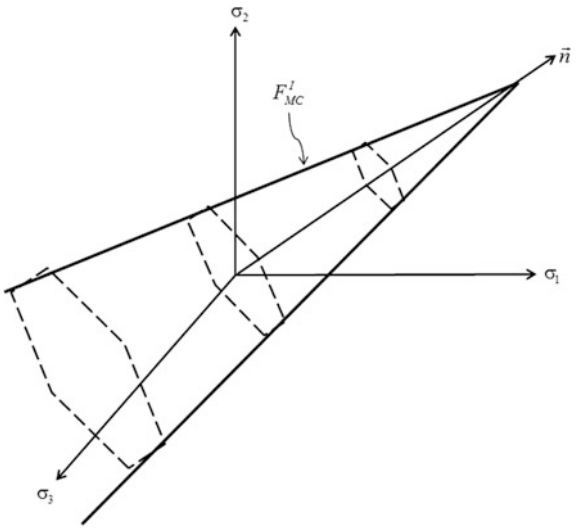


Fig. 13.10 Depiction of Mohr–Coulomb yield criterion in principal stress space

Next consider the case of uniaxial compression stress, given by Eq. (13.24). Substituting this into Eq. (13.36) for the case where uniaxial yielding occurs in compression $\left(\sigma_{11} = -\sigma_Y^C\right)$ will result in the following equation:

$$\begin{aligned} 0 &= \max\left[\left(\left|-\sigma_Y^C\right| - \sigma_Y^C \sin\phi\right), 0, \left|\sigma_Y^C\right| + \sigma_Y^C \sin\phi\right] - 2C_{MC}^Y \cos\phi \Rightarrow \\ 0 &= \sigma_Y^C(1 + \sin\phi) - 2C_{MC}^Y \cos\phi \end{aligned} \tag{13.38}$$

Solving Eqs. (13.37) and (13.38) simultaneously for the two unknowns C_{MC}^Y and ϕ will thus result in the following:

$$\phi = \sin^{-1}\left(\frac{2\sigma_s}{\sigma_Y^C} - 1\right) \tag{13.39}$$

and

$$C_{MC}^Y = \frac{\sigma_Y^C(1 + \sin\phi) + 2\sigma_S}{\cos\phi} \tag{13.40}$$

Note that the Mohr–Coulomb yield criterion possesses a similar shortcoming to the Drucker–Prager yield criterion in that this model fails to predict yielding for large values of average hydrostatic stress.

13.2.2.5 Cap Yield Criterion

The four yield criteria described in this section constitute the mainstream yield criteria that are utilized for the purpose of modeling initial yielding in isotropic media. For circumstances wherein they are found to be inaccurate, in most cases simple modifications are made to these models, such as in the case wherein yielding is to be predicted in geologic media undergoing large average hydrostatic pressure. In this case, it is customary to include a so-called cap on the compression side of the conic shown in Figs. 13.9 and 13.10, thereby creating a so-called *cap yield criterion*, as depicted graphically in Fig. 13.11 for the Drucker–Prager yield criterion. This cap is typically described as a flat plane parallel to the π-plane, and it should be apparent that a third experiment is required in order to characterize this part of the yield surface.

13.2.3 The Flow Rule

As described in Eq. (13.4), the plastic strain tensor, ε_{ij}^{P}, must be determined in order to complete the constitutive equation relating the stress and strain tensors in elastoplastic media. Prior to initial yielding, the plastic strain tensor is defined to be identically zero, thereby resulting in linear elastic behavior of the material in question. However, once initial yielding is predicted by one of the yield criteria described in the previous section, the plastic strain will necessarily evolve as a function of the loading history. The equation describing the evolution of the plastic strain tensor is termed *the flow rule*, and this equation is generally postulated to be in the form of a differential equation that is necessarily evolutionary in nature.

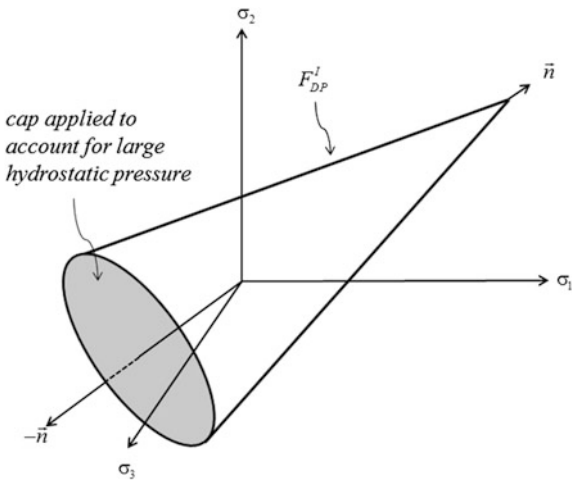


Fig. 13.11 Drucker–Prager yield criterion supplemented with hydrostatic pressure cap

The form chosen for the flow law has itself evolved significantly over time. However, for the sake of simplicity, only a brief overview of the various forms will be discussed herein.

13.2.3.1 Prandtl–Reuss Equations

As a preamble to the development of the actual equations, note that in geologic media a primary physical mechanism of inelastic deformation is due to relative slip between small granular particles and, although the length scale is considerably larger than in fluids, the physics are not unlike relative sliding of molecules in fluids. As described in Chap. 10, Newtonian fluids behave in shear according to a constitutive law of the following form:

$$\sigma'_{ij} = \eta \frac{d\varepsilon'_{ij}}{dt} \tag{13.41}$$

However, the physics of sliding in granular is nonlinear (in accordance with Coulomb's law of friction, described above). Inverting the above representation, applying it to plastic strain, and accounting for nonlinearity result in the following form of the flow rule:

$$\frac{d\varepsilon^P_{ij}}{dt} = \lambda' \sigma'_{ij} \tag{13.42}$$

where λ' is a nonlinear fluidity (meaning the inverse of viscosity) of the following general form:

$$\lambda' = \lambda'(\varepsilon_{ij}, T, \varepsilon^P_{ij}, \alpha\eta_{ij}) \tag{13.43}$$

where $\alpha\eta, \eta = 1, \ldots, n$ is a generic representation for a set of n *internal variables,* these variables being necessary to capture the physics of the inelastic deformations.

Note that if sufficient information can be assembled to afford the precise specification of Eq. (13.43), then Eq. (13.42) can be integrated in time, thereby providing the value of the plastic strain tensor for any and all times. Substitution of the resulting value of the plastic strain tensor into Eq. (13.5) will thus complete the constitutive model for an elastoplastic material.

The above simplistic development illustrates the rather ad hoc way that the theory of plasticity was originally developed, and an early generation of the model was termed *the Prandtl–Reuss equations* (Prandtl 1928). This form can be constructed by first differentiating Eq. (13.5) with respect to time and substituting Eq. (13.42) into this equation, thereby resulting in the following stress formulation:

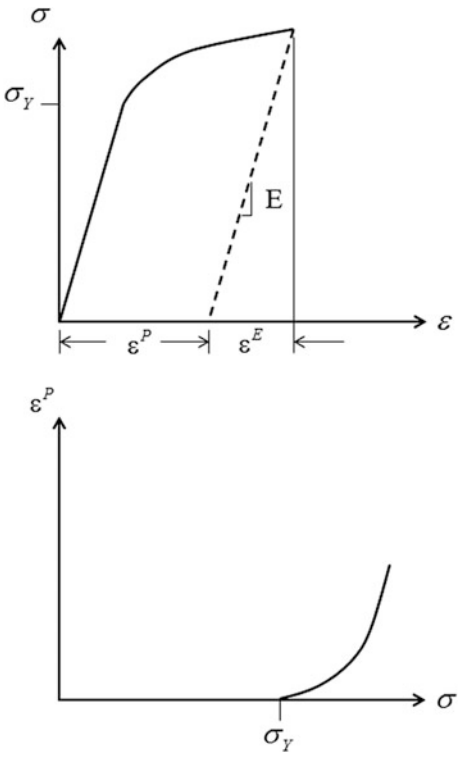


Fig. 13.12 Depiction of the evolution of plastic strain in a uniaxial test

$$\frac{d\varepsilon_{ij}}{dt} = \frac{1}{2\mu}\frac{d\sigma_{ij}}{dt} - \frac{\lambda}{2\mu(3\lambda+2\mu)}\frac{d\sigma_{kk}}{dt}\delta_{ij} + \lambda'\sigma'_{ij} \tag{13.44}$$

The uniaxial simplification of the Prandtl–Reuss equations is demonstrated in Fig. 13.12.

The above equations represent perhaps the first modern attempt to model the multi-dimensional elastoplastic behavior of solids. However, they were proposed for metals subjected to monotonic loading conditions, so that subsequent generalization has been necessary for the purpose of modeling geologic media, especially when subjected to nonmonotonic loading conditions.

Nonetheless, the above model is instructive for the purpose of elucidating how Eq. (13.43) may be deduced from experimental data. To see how this may be accomplished, first square Eq. (13.42) and solve for λ', thus resulting in the following:

$$\lambda' = \frac{3}{2}\frac{d\overline{\varepsilon}^P}{d\overline{\sigma}}\frac{1}{\overline{\sigma}}\frac{d\overline{\sigma}}{dt} \tag{13.45}$$

where *the equivalent plastic strain rate* is defined as follows:

$$\frac{d\overline{\varepsilon}^p}{dt} \equiv \left(\frac{2}{3}\frac{d\varepsilon_{ij}^P}{dt}\frac{d\varepsilon_{ij}^P}{dt}\right)^{1/2} \tag{13.46}$$

and *the equivalent uniaxial stress* is defined as follows:

$$\overline{\sigma} \equiv (3J_2)^{1/2} \tag{13.47}$$

The parameter $d\overline{\varepsilon}^P/d\overline{\sigma}$, called *the equivalent uniaxial plastic strain rate*, can be determined from experimental results as described below.

Note that both the equivalent plastic strain rate and the equivalent uniaxial stress are by definition invariants. It should be apparent that since Eq. (13.44) is a stress formulation, in light of Eq. (13.47), the only unknown in Eq. (13.45) is the leading quantity. This quantity can be obtained from a uniaxial ramp test ($\sigma_{11} \neq 0, \sigma_{ij} = 0$ all other *i, j*), in which, due to the form of flow rule (13.42)

$$\frac{d\varepsilon_{22}^P}{dt} = \frac{d\varepsilon_{33}^P}{dt} = -\frac{1}{2}\frac{d\varepsilon_{11}^P}{dt} \tag{13.48}$$

The above implies that the material is *plastically incompressible, that is*

$$\frac{d\varepsilon_{kk}^P}{dt} = 0 \tag{13.49}$$

Substituting (13.48) into Eqs. (13.45)–(13.47) for the case of a uniaxial test thus results in the following:

$$H' \equiv \frac{d\sigma_{11}}{d\varepsilon_{11}^P} = \frac{d\overline{\sigma}}{d\overline{\varepsilon}^P} \tag{13.50}$$

where *H'* is called *the plastic hardening modulus* because, as shown in Fig. 13.13, it can be obtained directly from a uniaxial test, thereby resulting in the following formula:

$$H' = H'(\overline{\varepsilon}^P) = \frac{EE^T}{E - E^T} \tag{13.51}$$

where E^T is the tangent modulus obtained from the uniaxial test, as shown in Fig. 13.13. Note that it has been assumed that *H'* is a function of the equivalent plastic strain, $\overline{\varepsilon}^P$. Alternatively, it is also often commonly assumed that *H'* may be assumed to be a function of the plastic work, W^P, defined as follows:

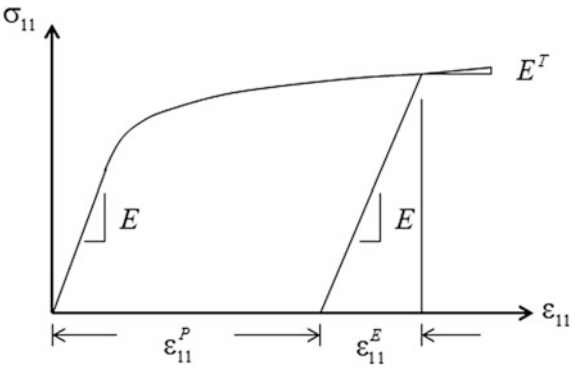


Fig. 13.13 Depiction of experimental parameters required to determine the plastic hardening modulus

$$W^P \equiv \int \sigma_{ij} d\varepsilon_{ij}^P \tag{13.52}$$

The Prandtl–Reuss equations, though powerful, are nonetheless limited. Perhaps the foremost of these limitations is that they are designed for modeling objects loaded monotonically. Furthermore, they are accurate only for materials that are plastically incompressible. Roadway materials are typically not plastically incompressible, so that more advanced models are necessary in order to capture the physically observed response of road way materials. These will be described in the following sections.

13.2.3.2 Drucker's Postulate

In order to accurately capture the elastoplastic behavior of road way materials, a more complex approach is taken to the flow rule, one that is based on the assumption that the material is stable and workhardening. As reported by D.C. Drucker (Drucker 1951), a material is described to be workhardening and stable if the following two conditions are met:

(1) Stable material

$$\frac{d\sigma_{ij}}{dt}\frac{d\varepsilon_{ij}}{dt} > 0 \quad \text{during loading} \tag{13.53a}$$

(2) Workhardening material

$$\Delta\sigma_{ij}\Delta\varepsilon_{ij}^P \geq 0 \quad \text{on completion of loading cycle} \tag{13.53b}$$

The above is called *Drucker's postulate*. The consequences of this postulate were described by Paul M. Naghdi (1924–94) as follows (Naghdi 1960). Consider

an increment of stress, $\Delta\sigma_{ij}$, as a nine-dimensional vector in stress hyperspace. Similarly, consider the nine components of the plastic strain increment, $\Delta\varepsilon_{ij}^{P}$, as a vector in the same hyperspace. Then, it follows from Drucker's postulate that if the material is stable and workhardening

$$\Delta\sigma_{ij}\Delta\varepsilon_{ij}^{P} = \Delta\vec{\sigma}\cdot\Delta\vec{\varepsilon}^{P} = \left|\Delta\vec{\sigma}\right|\left|\Delta\vec{\varepsilon}^{P}\right|\cos\psi \geq 0 \tag{13.54}$$

where

$$\psi \equiv \measuredangle\left(\Delta\vec{\sigma}, \Delta\vec{\varepsilon}^{P}\right) \tag{13.55}$$

Inequality (13.54) implies that

$$-\frac{\pi}{2} \leq \psi \leq \frac{\pi}{2} \tag{13.56}$$

or that the angle between $\Delta\vec{\sigma}$ and $\Delta\vec{\varepsilon}^{P}$ is acute. Consider a geometric depiction of inequality (13.54), as shown in Fig. 13.14. Suppose the yield surface is convex (implying that it is on the interior loading side of the hyperplane A-A at point P) and tangent to hyperplane A-A at point P. Note that for physical reasons, all admissible stress increments that terminate at point P must initiate either within or on the current yield surface, as depicted in the figure.

Inequality (13.54) now leads to the conclusion that

$$\Delta\vec{\varepsilon}^{P} \perp F(\sigma_{ij}) \tag{13.57}$$

It follows that for a stable and workhardening material

$$\frac{d\varepsilon_{ij}^{P}}{dt}(\sigma_{kl}) = \overline{\lambda}(\sigma_{kl})\frac{\partial F(\sigma_{kl})}{\partial\sigma_{ij}} \tag{13.58}$$

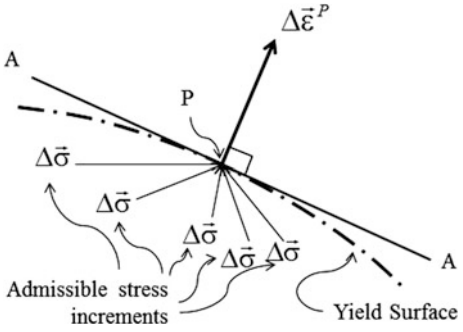


Fig. 13.14 Depiction of stable workhardening material behavior

where $\overline{\lambda}$ is a scalar called *the flow parameter* that describes the length of the plastic strain increment. Note that Eq. (13.58) is a generalization of the flow rule (13.42) deployed in the Prandtl–Reuss equations.

Equation (13.58) is termed *the normality condition* because it implies that for the current value of stress, the increment of plastic strain is normal to the current yield surface in stress hyperspace. Although this result may seem innocuous, the normality condition is a powerful tool in the development of advanced plasticity theories.

To see how this works, consider the von Mises yield criterion, given by Eq. (13.15). Substitution of this equation into (13.58) will result in the following:

$$\frac{d\varepsilon_{ij}^{P}}{dt} = \overline{\lambda}\frac{\partial}{\partial\sigma_{ij}}\left(\sqrt{J_2} - C_{VM}^{Y}\right) = \overline{\lambda}\left[\frac{1}{2}(J_2)^{-1/2}\sigma_{ij}^{'}\right] = \lambda'\sigma_{ij}^{'} \tag{13.59}$$

where, by comparison,

$$\lambda' = \frac{1}{2\sqrt{J_2}}\overline{\lambda} \tag{13.60}$$

The above can be seen to be equivalent to Eq. (13.42), the flow rule for the Prandtl–Reuss equations, thus implying that the von Mises yield criterion is associated with flow rule (13.42). When a flow rule is associated with a given yield criterion, it is termed *an associated flow rule.*

Because the von Mises yield criterion employs the second deviatoric stress invariant, J_2, the Prandtl–Reuss equations are sometimes termed J_2 *plasticity theory*. As will be demonstrated shortly, there are other more advanced theories of plasticity that may be developed using Drucker's postulate and the resulting normality condition.

13.2.3.3 The Flow Rule Associated with the Drucker–Prager Yield Criterion

Similar to the procedure described above for the von Mises yield criterion, the Drucker–Prager yield criterion, given by Eq. (13.21), may be substituted into the normality condition, Eq. (13.58), thereby producing the following result:

$$\frac{d\varepsilon_{ij}^{P}}{dt} = \overline{\lambda}\frac{\partial}{\partial\sigma_{ij}}\left(\sqrt{J_2} + \frac{\mu}{\sqrt{6}}I_1 - C_{DP}^{Y}\right) = \overline{\lambda}\left[\frac{1}{2}(J_2)^{-1/2}\sigma_{ij}^{'} + \frac{\mu}{\sqrt{6}}\delta_{ij}\right] \tag{13.61}$$

The proportionality term, $\overline{\lambda}$, similar to the employed previously for the Prandtl–Reuss equations, can be determined by squaring Eq. (13.61), with the result that

$$\overline{\lambda} = \sqrt{3}\frac{d\overline{\varepsilon}^P}{d\overline{\sigma}}\frac{d\overline{\sigma}}{dt}(1+\mu^2)^{-1/2} \tag{13.62}$$

The determination of the equivalent uniaxial plastic strain rate, $d\overline{\sigma}/d\overline{\varepsilon}^P$, is somewhat more complicated than for a material obeying the von Mises yield criterion. Nonetheless, it can be obtained from a uniaxial or other experimental data (see below).

13.2.3.4 Nonassociated Flow Rules

Not all materials flow in accordance with Drucker's postulate. Indeed, geologic media are known to be both unstable and worksoftening due to microscale dissipation that may either be a result of fracture or interfacial slip. There are two significantly different approaches taken to modeling the macroscale response of these materials. The more advanced of the two involves a procedure termed *multi-scaling*, and this approach will be discussed in Chap. 15. A more expedient, but less physically based approach, that has been used extensively for the purpose of modeling the unstable response of geologic media is to postulate the existence of a function of stress, G, other than the yield criterion as a potential for the flow rule. Since this leads to a flow rule that is not associated with the yield criterion, such flow rules are termed *nonassociated flow rules*. These are described by a flow rule of the general form:

$$\frac{d\varepsilon_{ij}^P}{dt} = \overline{\lambda}\frac{\partial G}{\partial \sigma_{ij}} \tag{13.63}$$

where the precise nature of the stress dependence within the function $G = G(\sigma_{ij})$ is determined via a suite of experiments. This type of flow rule will be considered in further detail in the section on viscoplastic material behavior.

13.2.4 The Workhardening Rule

It is observed in essentially all elastoplastic materials that when the material has been loaded beyond the yield point, any unloading will immediately lead to linear elastic response during unloading, thus implying that the state of stress during plastic loading is always on the yield surface, i.e., never proceeding beyond the current yield surface. Accordingly, in order for the state of stress to remain on the yield surface during plastic loading of a material point it is necessary for the yield surface to undergo some evolution during this process. Historically, phenomenological methodologies have been developed for this purpose. Essentially all procedures toward this end deploy the concept that after initial yielding, the yield

function depends on one or more *internal variables* (also called *hardening variables*), denoted herein by the symbol α (with appropriate subscripts based on the tensorial order of the internal variable in question), so that the yield function is generalized to the following form:

$$F = F(\sigma_{ij}, \varepsilon^P_{ij}, \alpha 1, \alpha 2_{ij}, \ldots) \tag{13.64}$$

Note that the plastic strain tensor, ε^P_{ij}, may be regarded as an internal variable, but is normally listed separately as described above. The number of internal variables required to accurately capture the evolution of the yield surface will depend on the material in question. However, for most materials sufficient accuracy for engineering design purposes will be obtained with two internal variables, one scalar valued and one second-order tensor valued, as shown in Eq. (13.64).

As a consequence of the necessity for the state of stress to remain on the yield surface during plastic loading, it follows from Eq. (13.64) that during plastic loading:

$$\frac{dF}{dt} = \frac{\partial F}{\partial \sigma_{ij}}\frac{d\sigma_{ij}}{dt} + \frac{\partial F}{\partial \varepsilon^P_{ij}}\frac{d\varepsilon^P_{ij}}{dt} + \frac{\partial F}{\partial \alpha 1}\frac{d\alpha 1}{dt} + \frac{\partial F}{\partial \alpha 2_{ij}}\frac{d\alpha 2_{ij}}{dt} + \cdots = 0 \tag{13.65}$$

Because the above ensures that the state of stress remains consistent with the yield surface during plastic loading, it is termed *the consistency condition.* Furthermore, in accordance with Eq. (13.65), it follows that, similar to the evolution law for the plastic strain tensor, *internal variable evolution laws* must be postulated for any additional internal variables required to accurately model the material behavior, and these must be of the following general form:

$$\frac{d\alpha 1}{dt} = f1(\varepsilon_{ij}, \varepsilon^P_{ij}, \alpha 1, \alpha 2_{ij}, \ldots) \tag{13.66}$$

and

$$\frac{d\alpha 2_{ij}}{dt} = f2_{ij}(\varepsilon_{ij}, \varepsilon^P_{ij}, \alpha 1, \alpha 2_{ij}, \ldots) \tag{13.67}$$

Equations (13.66) and (13.67) are often termed *workhardening or hardening rules*, since they are intended to be phenomenological descriptors of the plastic workhardening of the material.

Note that due to Eq. (13.4) the above may be written equivalently in terms of the stress tensor, and this choice is determined by whatever is pragmatic for the material in question. Various forms of the above internal variable evolution laws have been deployed for the yield function (13.64). Some of the more commonly used forms are discussed below.

13.2.4.1 Isotropic Hardening

The simplest method for ensuring that the state of stress remains on the yield surface during loading is the so-called *isotropic hardening* model, in which it is assumed that the initial yield surface expands isotropically about the origin in stress space during plastic loading. As shown in Fig. 13.15 for the case of the von Mises yield criterion, only a single scalar internal variable, called *the isotropic internal variable*, $\alpha 1$, is required, and this scalar may be deployed by simply replacing the material constant C^Y in the yield criteria described above with $\alpha 1$. Because the stress dependence in the yield criterion is unaltered when isotropic hardening is assumed, the flow rule is unaltered from that described in previous sections.

The details regarding isotropic hardening depend on the yield criterion selected, as will be demonstrated below for both the von Mises and Drucker–Prager yield criteria.

Isotropic Hardening with the von Mises Yield Criterion

When the von Mises yield criterion is deployed, yield criterion (13.15) is now modified to the following:

$$F_{VM} \equiv \sqrt{J_2} - \alpha 1_{VM} \tag{13.68}$$

where it should be apparent that the initial value of $\alpha 1_{VM}$, termed $\alpha 1^I_{VM}$, is, in accordance with Eq. (13.17), given by the following:

$$\alpha 1^I_{VM} = C^Y_{VM} = \frac{\sigma^T_Y}{\sqrt{3}} \tag{13.69}$$

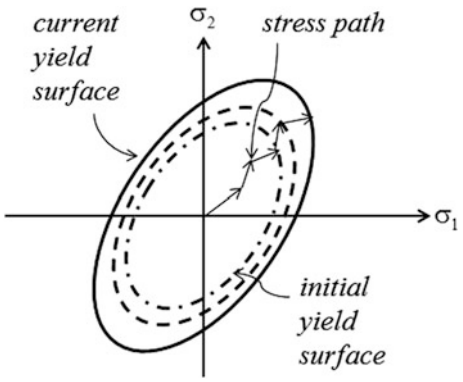


Fig. 13.15 Depiction of an isotropically expanding yield surface in two-dimensional principal stress subspace

Substitution of Eq. (13.68) into consistency condition (13.65) now results in the following evolution law for $\alpha 1$ during plastic loading:

$$\frac{d\alpha 1_{VM}}{dt} = \frac{\partial F_{VM}}{\partial \sigma_{ij}} \frac{d\sigma_{ij}}{dt} = \frac{\partial}{\partial \sigma_{ij}} \left(\sqrt{J_2} - \alpha 1_{VM} \right) \frac{d\sigma_{ij}}{dt} = \frac{\sigma'_{ij}}{2\sqrt{J_2}} \frac{d\sigma_{ij}}{dt} \tag{13.70}$$

The above evolution law will ensure that the state of stress remains on the von Mises yield surface during plastic loading, implying that the yield surface is simultaneously expanding isotropically as the stress state evolves.

13.2.4.2 Isotropic Hardening with the Drucker–Prager Yield Criterion

When the Drucker–Prager yield criterion is deployed, yield criterion (13.21) is now modified to the following:

$$F_{DP} \equiv \sqrt{J_2} + \frac{\mu}{\sqrt{6}} I_1 - \alpha 1_{DP} \tag{13.71}$$

where it should be apparent that the initial value of $\alpha 1_{DP}$, termed $\alpha 1^I_{DP}$, is, in accordance with Eq. (13.22), given by the following:

$$\alpha 1^I_{DP} = C^Y_{DP} = \sigma^S \tag{13.72}$$

Substitution of Eq. (13.71) into consistency condition (13.65) now results in the following evolution law for $\alpha 1$ during plastic loading:

$$\begin{aligned} \frac{d\alpha 1_{DP}}{dt} &= \frac{\partial F_{DP}}{\partial \sigma_{ij}} \frac{d\sigma_{ij}}{dt} = \frac{\partial}{\partial \sigma_{ij}} \left(\sqrt{J_2} + \frac{\mu}{\sqrt{6}} I_1 - \alpha 1_{DP} \right) \frac{d\sigma_{ij}}{dt} \\ &= \left(\frac{\sigma'_{ij}}{2\sqrt{J_2}} + \frac{\mu}{\sqrt{6}} \delta_{ij} \right) \frac{d\sigma_{ij}}{dt} \end{aligned} \tag{13.73}$$

The above evolution law will ensure that the state of stress remains on the Drucker–Prager yield surface during plastic loading, implying that the yield surface is simultaneously expanding isotropically as the stress state evolves.

A similar procedure to that described above for the von Mises and Drucker–Prager yield criteria may be employed for the purpose of constructing the isotropic hardening model for the Tresca and Mohr–Coulomb yield criteria.

When the loading is cycled in tension or compression uniquely, the isotropic hardening model may be sufficiently accurate for engineering purposes. Although it is simple and convenient, it suffers from the fact that it is usually not accurate when the loading is cycled between tension and compression. This is due to the Bauschinger effect (after Johann Bauschinger 1834–93), an observed asymmetry in the yield points in tension and compression caused mainly by the molecular-scale

physics of dislocations in metals, and due to the reflective asymmetry of Coulomb friction in geologic materials. Thus, when the loading sign changes repeatedly, it may be necessary to employ another hardening model, such as that described in the next section.

13.2.4.3 Kinematic Hardening

An alternative to the isotropic hardening model that is sometimes accurate for objects undergoing significant load reversals is the so-called *kinematic hardening model*, in which the yield surface is assumed to translate kinematically without changing shape and/or size during plastic loading, as shown in Fig. 13.16.

This type of genesis of the yield surface may be accommodated by introducing the following transformation:

$$F(\sigma_{ij}, \alpha 2_{ij}) \rightarrow F^{I}(\sigma_{ij} - \alpha 2_{ij}) \tag{13.74}$$

where it can be seen that the second-order tensor-valued *kinematic internal variable*, $\alpha 2_{ij}$, depicts the location of the center of the yield surface in stress space that was originally at the coordinate origin, as shown in Fig. 13.17. *The initial values of all components of this tensor are therefore identically zero*. Note that since the stress dependence within the yield criterion is altered in this hardening model, according to Drucker's postulate, the flow rule must necessarily be altered, as will be described in detail below.

Unlike the case of isotropic hardening, it should be apparent that there are an infinite number of choices for translation of the yield surface that pass through the current state of stress in stress hyperspace. Therefore, the consistency condition alone cannot be utilized to determine the hardening rule for kinematic hardening models. Instead, a generic form must be postulated for the hardening rule. Toward this end, a form that has gained acceptance based on experimental evidence is *Zeigler's rule* (Zeigler 1959), given by the following:

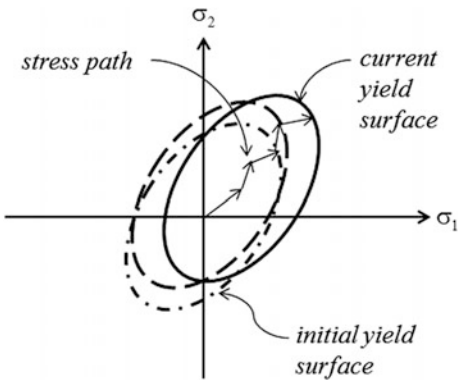


Fig. 13.16 Depiction of a kinematically translating yield surface in two-dimensional principal stress subspace

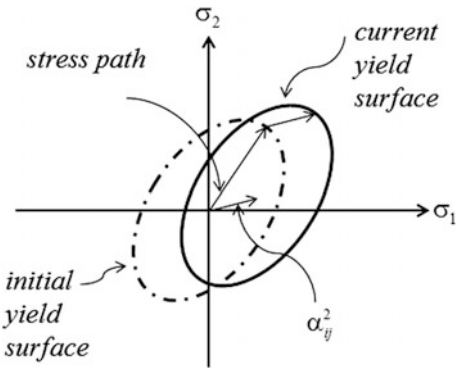


Fig. 13.17 Depiction of the evolution of the yield surface center during kinematic hardening

$$\frac{d\alpha 2_{ij}}{dt} = \overline{\mu}(\sigma_{ij} - \alpha 2_{ij}) \tag{13.75}$$

where $\overline{\mu}$ is a material property called *the kinematic hardening parameter* to be determined experimentally. It can be seen from Eq. (13.75) that Zeigler's rule only assumes the *direction* that the yield surface will translate for any given state. The *magnitude* of that translation during plastic loading, given by $\overline{\mu}$, may be determined by deploying the consistency condition, thus guaranteeing that the current state of stress remains on the yield surface during plastic loading. This may be accomplished by first noting that, as a direct consequence of Eq. (13.74), the following mathematical property of the yield function must hold during plastic loading:

$$\frac{\partial F}{\partial \alpha 2_{ij}} = -\frac{\partial F}{\partial \sigma_{ij}} \tag{13.76}$$

Substituting Eqs. (13.75) and (13.76) into the consistency condition, Eq. (13.65), now results in the following:

$$\frac{\partial F}{\partial \sigma_{ij}}\left[\frac{d\sigma_{ij}}{dt} - \overline{\mu}(\sigma_{ij} - \alpha 2_{ij})\right] = 0 \tag{13.77}$$

Solving the above for the kinematic hardening parameter, $\overline{\mu}$, results in the following:

$$\overline{\mu} = \frac{\frac{\partial F}{\partial \sigma_{ij}}\frac{d\sigma_{ij}}{dt}}{\frac{\partial F}{\partial \sigma_{kl}}(\sigma_{kl} - \alpha 2_{kl})} \tag{13.78}$$

The above equation, together with Eq. (13.75), thus describes the workhardening rule for kinematic hardening.

The specific details regarding kinematic hardening depend on the yield criterion selected, as will be demonstrated below for both the von Mises and Drucker–Prager yield criteria.

Kinematic Hardening with the von Mises Yield Criterion

It follows from Eq. (13.74) that when kinematic hardening is used with the von Mises yield criterion, the initial yield criterion is now replaced with the following:

$$F_{VM} = \sqrt{J_2^{'}} - C_{VM}^{Y} \tag{13.79}$$

where

$$J_2^{'} \equiv \frac{1}{2}\left(\sigma_{ij}^{'} - \alpha 2_{ij}^{'}\right)\left(\sigma_{ij}^{'} - \alpha 2_{ij}^{'}\right) \tag{13.80}$$

and $\alpha 2_{ij}^{'}$ is the deviatoric component of the kinematic hardening variable, defined as follows:

$$\alpha 2_{ij}^{'} \equiv \alpha 2_{ij} - \frac{\alpha 2_{kk}}{3}\delta_{ij} \tag{13.81}$$

Note also that, in view of the modified yield criterion, substitution of Eq. (13.79) into the normality condition, Eq. (13.58), results in the following modified flow rule:

$$\frac{d\varepsilon_{ij}^{P}}{dt} = \frac{1}{2}\bar{\lambda}\frac{1}{\sqrt{J_2^{'}}}\left(\sigma_{ij}^{'} - \alpha 2_{ij}^{'}\right) \tag{13.82}$$

In view of Eq. (13.82), the value of $\bar{\lambda}$ must necessarily be altered from that deployed with isotropic hardening (i.e., Equation 13.45). Accordingly, for the case of kinematic hardening, squaring Eq. (13.82) and solving for $\bar{\lambda}$ will result in the following:

$$\bar{\lambda} = \sqrt{3}\frac{d\bar{\varepsilon}^{P}}{d\bar{\sigma}^{'}}\frac{d\bar{\sigma}^{'}}{dt} \tag{13.83}$$

where

$$\bar{\sigma}^{'} \equiv \left(3J_2^{'}\right)^{1/2} \tag{13.84}$$

The above, together with Eqs. (13.75) and (13.78), describing the evolution of the kinematic hardening variable, $\alpha 2_{ij}$, completes the description of the combined hardening model with the von Mises yield criterion.

Kinematic Hardening with the Drucker–Prager Yield Criterion

For the case of kinematic hardening with the Drucker–Prager model, the yield criterion is modified to the following:

$$F_{DP} = \left[\frac{1}{2}\left(\sigma'_{ij} - \alpha 2'_{ij}\right)\left(\sigma'_{ij} - \alpha 2'_{ij}\right)\right]^{1/2} + \frac{\mu}{\sqrt{6}}(\sigma_{kk} - \alpha 2_{kk}) - C^{Y}_{DP} \tag{13.85}$$

Substituting the above into the normality condition, Eq. (13.58), results in the following flow rule:

$$\begin{aligned}\frac{d\varepsilon^{P}_{ij}}{dt} &= \overline{\lambda}\frac{\partial}{\partial\sigma_{ij}}\left\{\left[\frac{1}{2}\left(\sigma'_{ij} - \alpha 2'_{ij}\right)\left(\sigma'_{ij} - \alpha 2'_{ij}\right)\right]^{1/2} + \frac{\mu}{\sqrt{6}}(\sigma_{kk} - \alpha 2_{kk}) - C^{Y}_{DP}\right\} \\ &= \overline{\lambda}\left\{\frac{1}{2}\left[\frac{1}{2}\left(\sigma'_{kl} - \alpha 2'_{kl}\right)\left(\sigma'_{kl} - \alpha 2'_{kl}\right)\right]^{-1/2}\left(\sigma'_{ij} - \alpha 2'_{ij}\right) + \frac{\mu}{\sqrt{6}}\delta_{ij}\right\}\end{aligned} \tag{13.86}$$

As in the case of kinematic hardening used with the von Mises yield criterion, the value of $\overline{\lambda}$ must necessarily be altered. For the case of kinematic hardening used with the Drucker–Prager yield criterion, Eq. (13.86) is squared and solved for $\overline{\lambda}$, thereby resulting in the following:

$$\overline{\lambda} = \sqrt{3}\frac{d\overline{\varepsilon}^{P}}{d\overline{\sigma}'}\frac{d\overline{\sigma}'}{dt}(1 + \mu^2)^{-1/2} \tag{13.87}$$

The above procedure may also be utilized with the Tresca and Mohr–Coulomb yield criteria to create a kinematic hardening model for these yield surfaces.

The above, together with Eqs. (13.75) and (13.78), describing the evolution of the kinematic hardening variable, $\alpha 2_{ij}$, completes the description of the kinematic hardening model with the Drucker–Prager yield criterion.

13.2.4.4 Combined Hardening

It should be apparent that both the isotropic and kinematic hardening models are single-internal-variable models. As such, both are often found to be too simplistic to accurately reflect the physics of dissipative processes that occur in inelastic media. For example, when a typical elastoplastic material is subjected to uniaxial reverse

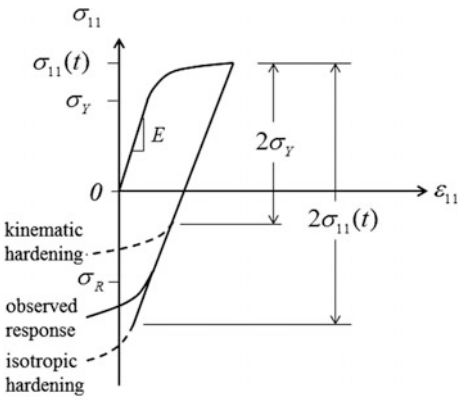


Fig. 13.18 Comparison of isotropic and kinematic hardening models to a uniaxial test on a typical elastoplastic material

loading, as shown in Fig. 13.18, neither isotropic nor kinematic hardening models reflect the material response accurately.

As a result, it is often found to be propitious to employ a two-internal-variable model. Perhaps the simplest means of accomplishing this is to deploy both the isotropic and kinematic hardening models simultaneously, thereby producing both translation and expansion of the yield surface in stress space during plastic loading. This type of model has been found to be considerably more accurate than either of the above two models, especially for cases involving significant load reversals.

In order to implement combined hardening, it should be apparent that the yield criterion must be modified to the following form during plastic loading:

$$F(\sigma_{ij} - \alpha 2_{ij}) - \alpha 1 = 0 \tag{13.88}$$

The consistency condition therefore takes on the following form:

$$\frac{\partial F}{\partial \sigma_{ij}}\frac{d\sigma_{ij}}{dt} + \frac{\partial F}{\partial \alpha 2_{ij}}\frac{d\alpha 2_{ij}}{dt} - \frac{d\alpha 1}{dt} = 0 \tag{13.89}$$

Now consider Fig. 13.18 once again. It should be apparent from the figure that when the uniaxial loading is cycled from tension to compression, re-yielding will occur in compression with the stress value $\sigma_{11} = \sigma_R$, and this value can be determined from a cyclic uniaxial test. Now suppose that *the hardening ratio*, β, is defined as follows:

$$\beta \equiv \frac{[\sigma_{11}(t) - \sigma_R - 2\sigma_Y]}{[2\sigma_{11}(t) - 2\sigma_Y]} \tag{13.90}$$

where $\sigma_{11}(t)$ is the peak applied uniaxial tensile stress, as shown in Fig. 13.18. It is apparent from Eq. (13.89) that:

$$\begin{aligned} \beta = 1 &\Rightarrow \quad \textit{isotropic hardening} \\ 0 < \beta < 1 &\Rightarrow \textit{combined hardening} \\ \beta = 0 &\Rightarrow \quad \textit{kinematic hardening} \end{aligned} \tag{13.91}$$

The hardening ratio will be utilized to construct the evolution law for $\alpha 1$ in the following sections.

It should be apparent that Eq. (13.76) still holds in the case of combined hardening. In addition, it is assumed that in the case of combined hardening, the generic form for the evolution of the kinematic hardening variable, $\alpha 2_{ij}$, remains unchanged, that is—Zeigler's rule, given by Eq. (13.75)—is utilized for combined hardening. Thus, substituting Eqs. (13.75) and (13.76) into the combined hardening consistency condition Eq. (13.89) results in the following equation:

$$\overline{\mu} = \frac{\frac{\partial F}{\partial \sigma_{ij}} \frac{d\sigma_{ij}}{dt} - \frac{d\alpha 1}{dt}}{\frac{\partial F}{\partial \sigma_{kl}} (\sigma_{kl} - \alpha 2_{kl})} \tag{13.92}$$

The above equation, together with Eq. (13.75) thus describes the workhardening rule for $\alpha 2_{ij}$ for the case of combined hardening.

The specific details regarding combined hardening depend on the yield criterion selected, as will be demonstrated below for both the von Mises and Drucker–Prager yield criteria.

Combined Hardening with the von Mises Yield Criterion

It follows from Eq. (13.74) that when combined hardening is used with the von Mises yield criterion, the initial yield criterion is now replaced with the following:

$$F_{VM} = \sqrt{J_2'} - \alpha 1_{VM} \tag{13.93}$$

where J_2' is as defined in Eq. (13.80).

Note also that, in view of the modified yield criterion, substitution of Eq. (13.90) into the normality condition, Eq. (13.58), results in the following modified flow rule:

$$\frac{d\varepsilon_{ij}^P}{dt} = \frac{1}{2}\overline{\lambda}\frac{1}{\sqrt{J_2'}}\left(\sigma_{ij}' - \alpha 2_{ij}'\right) \tag{13.94}$$

The above can be seen to be identical in the form to that employed in the case of kinematic hardening with the von Mises yield criterion (i.e., Eq. 13.82). However, because the hardening now includes expansion of the yield surface, the normality condition will produce different results from that resulting when only kinematic

hardening is deployed. This is reflected via the current value of the kinematic hardening variable, $\alpha 2'_{ij}$, in the above flow rule. Note also that the form of the scalar $\overline{\lambda}$ is also unchanged from that obtained in the kinematic hardening case, but the value of $\overline{\lambda}$ is altered via the current value of $\alpha 2'_{ij}$.

The evolution law for the isotropic hardening variable is obtained by multiplying isotropic hardening evolution law (13.76) by the hardening ratio, β, thus resulting in the following:

$$\frac{d\alpha 1_{VM}}{dt} = \beta \frac{\sigma'_{ij}}{2\sqrt{J_2}} \frac{d\sigma_{ij}}{dt} \tag{13.95}$$

The above, together with Eqs. (13.75) and (13.92) describing the evolution of the kinematic hardening variable, $\alpha 2_{ij}$, completes the description of the combined hardening model with the von Mises yield criterion.

Combined Hardening with the Drucker–Prager Yield Criterion

For the case of combined hardening with the Drucker–Prager model, the yield criterion is modified to the following:

$$F_{DP} = \left[\frac{1}{2}\left(\sigma'_{ij} - \alpha 2'_{ij}\right)\left(\sigma'_{ij} - \alpha 2'_{ij}\right)\right]^{1/2} + \frac{\mu}{\sqrt{6}}(\sigma_{kk} - \alpha 2_{kk}) - \alpha 1_{DP} \tag{13.96}$$

Substituting the above into the normality condition, Eq. (13.58), results in the following flow rule:

$$\frac{d\varepsilon^P_{ij}}{dt} = \overline{\lambda}\left\{\frac{1}{2}\left[\frac{1}{2}\left(\sigma'_{kl} - \alpha^{2'}_{kl}\right)\left(\sigma'_{kl} - \alpha^{2'}_{kl}\right)\right]^{-1/2}\left(\sigma'_{ij} - \alpha^{2'}_{ij}\right) + \frac{\mu}{\sqrt{6}}\delta_{ij}\right\} \tag{13.97}$$

As in the case of combined hardening with the von Mises yield criterion, the form of the flow rule for combined hardening with the Drucker–Prager yield criterion is identical to that obtained for kinematic hardening. However, the actual values obtained for the plastic strain rate tensor are altered by the fact that the current components of the kinematic hardening variable, $\alpha 2'_{ij}$, are modified due to the expansion of the yield surface during plastic loading. Note also that the form of the scalar $\overline{\lambda}$ is also unchanged from that obtained in the kinematic hardening case, but that the value of $\overline{\lambda}$ is altered via the current value of $\alpha 2'_{ij}$.

The evolution law for the isotropic hardening variable is obtained by multiplying Eq. (13.73), the isotropic hardening law, by the hardening ratio,β, thus resulting in the following:

$$\frac{d\alpha 1_{DP}}{dt} = \beta\left(\frac{\sigma'_{ij}}{2\sqrt{J_2}} + \frac{\mu}{\sqrt{6}}\delta_{ij}\right)\frac{d\sigma_{ij}}{dt} \tag{13.98}$$

The above, together with Eqs. (13.75) and (13.92) describing the evolution of the kinematic hardening variable, $\alpha 2_{ij}$, completes the description of the combined hardening model with the Drucker–Prager yield criterion.

13.3 The Elastoplastic Initial Boundary Value Problem

The model that is produced by adjoining the stress–elastic constitutive Eq. (13.5) with the yield criterion, the flow, and the workhardening rule described in the previous sections is called *the elastoplastic initial boundary value problem.* The construction of this model begins with Tables 9.6 and 9.7, wherein the list of unknowns and available equations is described in detail. The reader will recall that Table 9.6 listed 25 unknown variables to be predicted as functions of position, $\vec{x}$, and time, t, in the continuous body $V + S$. In this chapter, the plastic strain tensor has been introduced (thereby increasing the list of unknowns to 31), so that the list of unknowns is now as follows:

Displacement vector	u_i
Strain tensor	ε_{ij}
Plastic strain tensor	ε^P_{ij}
Stress tensor	σ_{ij}
Body force per unit mass	f_i
Internal energy per unit mass	u
Heat flux vector	q_i
Heat source per unit mass	r
Entropy per unit mass	s
Temperature	T

The prediction of these variables is afforded with the use of the 10 equations and one inequality shown in Table 9.7, plus the stress–elastic strain constitutive Eq. (13.5), the yield criterion, the flow rule, and the workhardening rule.

Recall from Chap. 10 that when dealing with an elastoplastic object the body is by definition *at constant temperature* in both time and space, thus obviating the necessity to include temperature, T, in the model. Since the temperature is constant, the heat flux vector, $\vec{q}$, and the heat source term, r, must necessarily be identically zero, and it follows that for the elastoplastic case the internal energy, u, can be treated as an ancillary variable due to the fact that the conservation of energy simplifies to a means of calculating the internal energy once the stress field, $\sigma_{ij}(\vec{x}, t)$, strain field, $\varepsilon_{ij}(\vec{x}, t)$, and plastic strain field, $\varepsilon^P_{ij}(\vec{x}, t)$, are predicted with the model.

In addition, the entropy, s, although not constant, will be shown below to be an ancillary variable. Given that the body force vector, f_i, can be predetermined using Newton's gravitational law when this is the source of the body force, this reduces the above total of 31 variables to the following 21 primary variables: $u_i, \varepsilon_{ij}, \varepsilon_{ij}^P$ and σ_{ij}. Accordingly, the quasi-static elastoplastic initial boundary value problem may be now stated in Table 13.1.

Analytic methods for solving the elastoplastic IBVP can be found in numerous excellent texts on the subject. Unfortunately, the road way problem is sufficiently complex that analytic methods cannot be used to model elastoplastic road ways. Therefore, in this text methods for solving elastoplastic IBVPs will be confined to computational methods, to be described in Chap. 14.

13.4 Multi-dimensional Viscoplasticity

Sometimes the base geologic material is observed to exhibit significant rate dependence, and this is often due to a significant amount of moisture within the base material. Due to the physical nature of the base material, the response will also demonstrate nonlinearity that is similar to that observed in base material that is rate independent. Accordingly, it is oftentimes found that only single nonlinear viscosity is required to accurately capture the response of the material. As a result, it is deemed unnecessary to employ a nonlinear viscoelastic constitutive model to account for the rate dependence, since such an approach (see Chap. 12) utilizes multiple viscosities.

Instead, it has been found that it is often sufficiently accurate to employ an extended form of the plasticity models described previously in this chapter, with the viscosity $(1/\lambda')$ formulated in such a way as to exhibit rate dependence. This type of constitutive model, termed *viscoplasticity*, is therefore described by the following set of equations:

$$\sigma_{ij} = 2\mu(\varepsilon_{ij} - \varepsilon_{ij}^I) + \lambda\delta_{ij}(\varepsilon_{kk} - \varepsilon_{kk}^I) \tag{13.99}$$

where ε_{ij}^I is called *the inelastic strain tensor*. Note the similarity between Eq. (13.99) and elastoplastic Eq. (13.5). Clearly, the only difference is that the permanent strain, ε_{ij}^I, is generalized in such a way as to permit rate dependence. Accordingly, it is predicted with the use of a set of evolution laws that describe the behavior of the material as follows:

$$\dot{\varepsilon}_{ij}^I = \dot{\lambda}^I \frac{\partial G}{\partial \sigma_{ij}} \tag{13.100}$$

Table 13.1 Quasi-static elastoplastic initial boundary value problem

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S), f_i(\text{in } V+S)$

Geometry: $\vec{x}\ (\text{on } S),\ n_i\ (\text{on } S)$

Material Properties: $\rho, \lambda, \mu, \alpha$ *plus the coefficients appearing in the internal variable evolution laws*

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}, \varepsilon^P_{ij}, \alpha 1, \alpha 2_{ij}$ = **28 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = \lambda(\varepsilon_{kk} - \varepsilon^P_{kk})\delta_{ij} + 2\mu(\varepsilon_{ij} - \varepsilon^P_{ij})$	6
$\frac{d\varepsilon^P_{ij}}{dt} = \overline{\lambda}\frac{\partial F}{\partial \sigma_{ij}}, F = F(\sigma_{kl}, \alpha 1, \alpha^2_{kl})$	6
$\frac{d\alpha 1}{dt} = f1(\varepsilon_{ij}, \varepsilon^P_{ij}, \alpha 1, \alpha 2_{ij}, \ldots)$	1
$\frac{d\alpha 2_{ij}}{dt} = f2_{ij}(\varepsilon_{kl}, \varepsilon^P_{kl}, \alpha 1, \alpha 2_{kl}, \ldots)$	6
Total	**28**

Initial Conditions:

Displacements:	$u_i(\vec{x}, 0) = \hat{u}_i$	$\forall \vec{x} \in V + S$
Stresses:	$\sigma_{ij}(\vec{x}, 0) = \hat{\sigma}_{ij}$	$\forall \vec{x} \in V + S$
Strains:	$\varepsilon_{ij}(\vec{x}, 0) = \hat{\varepsilon}_{ij}$	$\forall \vec{x} \in V + S$

(continued)

Table 13.5 (continued)

Plastic strain tensor:	$\varepsilon_{ij}^{P}(\vec{x},0)=0$	$\forall \vec{x} \in V+S$
Isotropic hardening variable:	$\alpha 1(\vec{x},0)=\hat{\alpha}1$	$\forall \vec{x} \in V+S$
Kinematic hardening variable:	$\alpha 2_{ij}(\vec{x},0)=0$	$\forall \vec{x} \in V+S$
Boundary Conditions:		
Tractions:	$t_i=\hat{t}_i$ on S_1	
Displacements:	$u_i=\hat{u}_i$ on S_2	$S=S_1+S_2$

where $\dot{\lambda}^I$ is a history-dependent function of the state variables of the general form:

$$\dot{\lambda}^I = \dot{\lambda}^I(\sigma_{ij}, \varepsilon_{ij}^I, \alpha 1, \alpha 2_{ij}) \tag{13.101}$$

where $\alpha 1$ is a scalar-valued internal variable, given by an equation of the general form:

$$\frac{d\alpha 1}{dt} = f_1(\sigma_{ij}, \varepsilon_{ij}^I, \alpha 1, \alpha 2_{ij}) \tag{13.102}$$

and $\alpha 2_{ij}$ is a second-order tensor-valued internal variable, given by an equation of the general form:

$$\frac{d\alpha 2_{ij}}{dt} = f_{2ij}(\sigma_{ij}, \varepsilon_{ij}^I, \alpha 1, \alpha 2_{ij}) \tag{13.103}$$

In addition, G is a function of the stress that is nonassociative (see above). There are various forms proposed in the literature for Eqs. (13.100)–(13.103), depending on the material used in the construction of the road way (Cela 1998; Elaskar et al. 2000; Masad 2005; Tashman 2005), and this type of material model is appropriate for use with geologic base materials, especially soils with a significant amount of moisture.

It is to be emphasized that viscoplastic material models are quite challenging in two different ways. First of all, they require a difficult and time-consuming suite of laboratory experiments in order to completely characterize the material properties contained within the model. Secondly, because of the nature of the material behavior observed in rate-dependent geologic media, the resulting set of equations is mathematically *stiff*, which is a terminology used to describe nonlinear differential equations that are unstable when solved numerically with traditional

computational algorithms (Gear 1971, 1981). Thus, special care must be taken when employing this type of material model, as the effort may not justify the added accuracy. Although this type of constitutive model can be incorporated into a finite element model so that predictions of the stresses, strains, and displacements can be made as functions of time and space in the road way, it will not be covered in further detail within this text.

13.5 Multi-dimensional Thermoviscoplasticity

When the temperature in the object of interest varies spatially and/or with time, the above elastoplastic and viscoplastic models must necessarily be generalized. Although several different types of material models have presented in this chapter, for the purpose of introducing temperature dependence, it is propitious to group all of these constitutive models under the generic title *thermoviscoplasticity*.

In this case, a thermoviscoplastic material is defined to be one in which the stress, σ_{ij}, is a single-valued function of the strain, ε_{kl}, the plastic (or inelastic) strain, ε_{kl}^{P}, and the temperature, T, as follows:

$$\sigma_{ij}(\vec{x},t) = \underset{ij}{\overset{\tau=t}{E}}_{\tau=0}\left[\varepsilon_{kl}(\vec{x},\tau), \varepsilon_{kl}^{P}(\vec{x},\tau), T(\vec{x},\tau)\right] \tag{13.104}$$

In addition, the plastic strain is determined via a set of *internal variable evolution laws* of the following generic form:

$$\frac{d\alpha\eta_{ij}}{dt} = f\eta_{ij}(\varepsilon_{kl}, \varepsilon_{kl}^{P}, T, \gamma_k, \alpha\eta_{kl}), \quad \eta = 1, \ldots, n \tag{13.105}$$

where the number of internal variables is assigned the generic value n. Although only two internal variables have been described in the current chapter, the number of internal variables will ultimately be determined by what is pragmatic for modeling the physical mechanisms of dissipation for the material in question. For simplicity, in the following discussion, it will be assumed that there are only two internal variables, $\alpha 1$ (a scalar-valued internal variable), and $\alpha 2_{ij}$ (a second-order tensor-valued internal variable).

The above equations are adjoined with the following (from Table 9.5):

(1) strain-displacement equations:

$$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{13.106}$$

(2) equations of motion:

$$\sigma_{ji,j} + \rho f_i = \rho \frac{d^2 u_i}{dt^2} \tag{13.107}$$

(3) conservation of energy:

$$\rho \frac{du^I}{dt} = \sigma_{ij} \frac{d\varepsilon_{ij}}{dt} - q_{i,i} + \rho r \tag{13.108}$$

(4) entropy production inequality:

$$\rho \frac{ds}{dt} + \frac{\partial}{\partial x_i}\left(\frac{q_i}{T}\right) - \frac{\rho r}{T} \geq 0 \tag{13.109}$$

Similar to that introduced for thermoelastic media in Chap. 11, the temperature gradient, γ_i, as described in Eq. (11.67), is defined as follows:

$$\gamma_i \equiv \frac{\partial T}{\partial x_i} \tag{13.110}$$

The list of 41 unknowns for a thermoviscoplastic continuum is thus as follows:

Displacement vector	u_i
Strain tensor	ε_{ij}
Plastic strain tensor	ε_{ij}^P
Scalar internal variable	$\alpha 1$
Second-order tensor internal variable	$\alpha 2_{ij}$
Stress tensor	σ_{ij}
Body force per unit mass	f_i
Internal energy per unit mass	u^I
Heat flux vector	q_i
Heat source per unit mass	r
Entropy per unit mass	s
Temperature	T
Temperature gradient	γ_i

The prediction of the above variables is afforded with the use of the 13 equations and one inequality (13.106)–(13.110) listed above.

The thermoviscoplasticity problem is considerably more complicated than the viscoplasticity problem, and this is due at least in part to the coupling between mechanics and thermodynamics. This complicating feature will be elucidated in the following section.

13.5.1 Thermodynamic Constraints on Thermoviscoplastic Material Behavior

The development of thermodynamic constraints on thermoviscoplastic materials will follow the method developed by Coleman and Noll in Chap. 11, but will be extended for the case of materials with internal variables (Coleman and Gurtin 1967; Kratochvil and Dillon 1970; Allen 1991). The challenge is to solve the above set of equations in terms of the independent variables $\vec{x}$ and t, implying that a solution is sought for the 41 unknowns $u_i, \varepsilon_{ij}, \varepsilon^P_{ij}, \sigma_{ij}, \alpha 1, \alpha 2_{ij}, f_i, u^I, T, \gamma_i, q_i, r$ and s as functions of $\vec{x}$ and t. The solution procedure is described in Table 13.2.

Table 13.2 The procedure for obtaining thermodynamic constraints on thermoviscoplastic media

Step 1: Note that normally the body force per unit mass, $f_i = f_i(\vec{x}, t)$, can be specified (by using Newton's gravitational law) as well as the heat source, $r = r(\vec{x}, t)$. However, in this case specify a priori instead $u_i = u_i(\vec{x}, t)$, and
$T = T(\vec{x}, t) \quad \forall \vec{x}, t \in V + S \Rightarrow$ **37 remaining unknowns**
Step 2: Use Eq. (13.106) to solve for $\varepsilon_{ij} = \varepsilon_{ij}(\vec{x}, t) \Rightarrow$ **31 remaining unknowns**
Step 3: Use Eq. (13.110) to solve for $\gamma_i = \gamma_i(\vec{x}, t) \Rightarrow$ **28 remaining unknowns**
Step 4: Assume that there exist 24 constitutive equations for a thermoviscoplastic material as follows):

$$\sigma_{ij}(\vec{x}, t) = \sigma_{ij}\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t)\right) \tag{13.111}$$

$$u^I(\vec{x}, t) = u^I\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t)\right) \tag{13.112}$$

$$q_i(\vec{x}, t) = q_i\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t)\right) \tag{13.113}$$

$$s(\vec{x}, t) = s\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t)\right) \tag{13.114}$$

$$\frac{d\varepsilon^P_{ij}}{dt} = \Omega^P_{ij}\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t), \dot{\varepsilon}^P_{kl}(\vec{x}, t)\right) \tag{13.115}$$

$$\frac{d\alpha 1}{dt} = \Omega 1\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t), \dot{\varepsilon}^P_{kl}(\vec{x}, t)\right) \tag{13.116}$$

$$\frac{d\alpha 2_{ij}}{dt} = \Omega 2_{ij}\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t), \dot{\varepsilon}^P_{kl}(\vec{x}, t)\right) \tag{13.117}$$

where, according to the principle of equipresence (Truesdell and Toupin 1954), since the temperature gradient is included in one constitutive equation, it must be included in all of them. Note that the above relate the kinetic variables to the kinematic variables (which have been determined above). Use (13.111)–(13.117) to solve for $\sigma_{ij} = \sigma_{ij}(\vec{x}, t), u^I = u^I(\vec{x}, t), q_i = q_i(\vec{x}, t), s = s(\vec{x}, t), \varepsilon^P_{ij} = \varepsilon^P_{ij}(\vec{x}, t), \alpha 1 = \alpha 1(\vec{x}, t), \alpha 2_{ij} = \alpha 2_{ij}(\vec{x}, t) \Rightarrow$ **4 remaining unknowns**
Step 5: Use Eq. (13.107) to solve for $f_i = f_i(\vec{x}, t) \Rightarrow$ **1 remaining unknown**
Step 6: Use Eq. (13.108) to solve for $r = r(\vec{x}, t) \Rightarrow$ **0 remaining unknowns**
Step 7: Recall that the Helmholtz free energy is defined as follows:

$$h \equiv u^I - Ts \tag{13.118}$$

Thus, as a result of constitutive Eqs. (13.112) and (13.114), it follows that

$$h(\vec{x}, t) = h\left(\varepsilon_{kl}(\vec{x}, t), \varepsilon^P_{kl}(\vec{x}, t), T(\vec{x}, t), \gamma_k(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{kl}(\vec{x}, t)\right) \tag{13.119}$$

Step 8: Applying the chain rule of differentiation to the above equation and substituting this result into entropy production inequality (13.109) thus results in the following:

(continued)

Table 13.3 (continued)

$$
\begin{aligned}
&- \left[\rho \frac{\partial h}{\partial \varepsilon_{ij}}(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl}) - \sigma_{ij}(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\varepsilon_{ij}}{dt} \\
&- \left[\rho \frac{\partial h}{\partial T}(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl}) + \rho s(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{dT}{dt} \\
&- \left[\rho \frac{\partial h}{\partial \gamma_i}(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\gamma_i}{dT} \\
&- \left[\rho \frac{\partial h}{\partial \varepsilon_{ij}^P}(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\varepsilon_{ij}^P}{dt} \\
&- \left[\rho \frac{\partial h}{\partial \alpha 1}(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\alpha 1}{dt} \\
&- \left[\rho \frac{\partial h}{\partial \alpha 2_{kl}}(\varepsilon_{kl}, \varepsilon_{kl}^P, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\alpha 2_{kl}}{dt} - \frac{q_i \gamma_i}{T} \geq 0
\end{aligned}
\tag{13.120}
$$

Step 9: Now note that the quantities inside the brackets in inequality (13.120) are independent of the quantities outside the brackets, and the quantities outside the brackets are independent of one another, so that the above produces the following constraints on the allowable material behavior for a thermoelastic material:

$$\sigma_{ij} = \rho \frac{\partial h}{\partial \varepsilon_{ij}} \tag{13.121}$$

$$s = -\frac{\partial h}{\partial T} \tag{13.122}$$

$$\frac{\partial h}{\partial \gamma_i} = 0 \tag{13.123}$$

Step 10: It follows from Eq. (13.123) together with (13.112) and (13.114) that

$$h = h(\varepsilon_{kl}, \varepsilon_{kl}^p, T, \alpha 1, \alpha 2_{kl}) \tag{13.124}$$

A special case of Eq. (13.124) is found to be sufficient for most materials, given by the following:

$$h = h(\varepsilon_{kl}, \varepsilon_{kl}^p, T) \tag{13.125}$$

Equation (13.125) is now expanded in a Taylor series as follows:

$$
h = \frac{1}{\rho}\left(\begin{array}{l} h^{000} + h_{ij}^{100}\varepsilon_{ij} + + h_{ij}^{010}\varepsilon_{ij}^P + h^{001}\theta + h_{ijkl}^{110}\varepsilon_{ij}\varepsilon_{kl}^P \\ + h_{ij}^{101}\varepsilon_{ij}\theta + h_{ij}^{011}\varepsilon_{ij}^P\theta + \frac{1}{2}h_{ijkl}^{200}\varepsilon_{ij}\varepsilon_{kl} + \frac{1}{2}h_{ijkl}^{020}\varepsilon_{ij}^P\varepsilon_{kl}^P + \frac{1}{2}h^{002}\theta^2 + \ldots \end{array}\right) \tag{13.126}
$$

where $h^{000}, h_{ij}^{100}, h_{ij}^{010}, h^{001}, h_{ijkl}^{110}, h_{ij}^{101}, h_{ij}^{011}, h_{ijkl}^{200}, h_{ijkl}^{020}$ and h^{002} are material constants and recall that

$$\theta \equiv T - T_R \tag{13.127}$$

Note that T_R is defined to be the reference temperature at which no strain is observed when the object is unstressed.

Step 11: Substitute Eq. (13.126) into Eq. (13.121) to obtain the following:

$$\sigma_{ij} = h_{ij}^{100} + h_{ijkl}^{110}\varepsilon_{kl}^P + h_{ij}^{101}\theta + h_{ijkl}^{200}\varepsilon_{kl} \tag{13.128}$$

Experimental evidence suggests that during isothermal process, the plastic strain produces negligible free energy, so that

$$\frac{\partial h}{\partial \varepsilon_{ij}^P} = -\frac{\partial h}{\partial \varepsilon_{ij}} \tag{13.129}$$

Substituting the above approximation into Eq. (13.128) and introducing the nomenclature for material constants introduced in Chap. 11 now results in the following equivalent form of Eq. (13.128):

$$\sigma_{ij} = E_{ijkl}(\varepsilon_{kl} - \varepsilon_{kl}^P - \alpha_{kl}\theta) \tag{13.130}$$

where, as defined in Chap. 11, E_{ijkl} is the elastic modulus tensor and α_{kl} is the thermal expansion tensor.

(continued)

Table 13.3 (continued)

In the isotropic case, Eq. (13.130) reduces to the following (see Chap. 11):The above equations, together with the flow

$$\sigma_{ij} = \lambda(\varepsilon_{kk} - \varepsilon^P_{kk})\delta_{ij} + 2\mu(\varepsilon_{ij} - \varepsilon^P_{ij}) - \beta\delta_{ij}\theta \tag{13.131}$$

The above equations, together with the flow rule and the internal variable evolution laws described previously in this chapter, will define the constitutive model for a thermoviscoplastic material. For example, differentiating equation (13.131) with respect to time and introducing flow rule (13.42) will result in the Prandtl–Reuss equations.

Step 12: As in the case of the thermoelastic material, expand the heat flux vector in terms of the temperature gradient to obtain:

$$q_i(\varepsilon_{kl}, \varepsilon^P_{kl}, T, \gamma_k, \alpha 1, \alpha 2_{kl}) = -k_{ij}(\varepsilon_{kl}, \varepsilon^P_{kl}, T, \alpha 1, \alpha 2_{kl})\gamma_j \tag{13.132}$$

where k_{ij} is the thermal conductivity tensor. The above is an extension of Fou*rier's law of heat conduction* (Fourier 1822), in which it can be seen that the thermal conductivity tensor is not necessarily constant. However, in most practical circumstances it is found that it is indeed a rather weak function of the state variables listed above.

Step 14: Substituting Eqs. (13.121)–(13.123) into inequality (13.120) will result in the following:

$$\begin{aligned} &- \left[\rho \frac{\partial h}{\partial \varepsilon^P_{ij}}(\varepsilon_{kl}, \varepsilon^P_{kl}, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\varepsilon^P_{ij}}{dt} \\ &- \left[\rho \frac{\partial h}{\partial \alpha 1}(\varepsilon_{kl}, \varepsilon^P_{kl}, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\alpha 1}{dt} \\ &- \left[\rho \frac{\partial h}{\partial \alpha 2_{kl}}(\varepsilon_{kl}, \varepsilon^P_{kl}, T, \gamma_k, \alpha 1, \alpha 2_{kl})\right] \frac{d\alpha 2_{kl}}{dt} - \frac{q_i \gamma_i}{T} \geq 0 \end{aligned} \tag{13.133}$$

Inequality (13.133) indicates that entropy is generated whenever the internal state is changing, in addition to the entropy generated by flux of heat. In addition, it provides constraints on the allowable form of the thermoviscoplastic constitutive model.

This completes the construction of thermodynamic constraints on thermoviscoplastic media, thus demonstrating that it is possible to solve for all of the unknowns using the thermoviscoplastic constitutive model, while at the same time obtaining powerful constraints on the constitutive behavior of thermoviscoplastic media

13.5.2 The Thermoviscoplastic Initial Boundary Value Problem

The results of the procedure described in Table 13.2 may now be utilized to simplify the thermoviscoplastic initial boundary value problem. To accomplish this, first note that the temperature, T, appears in the thermoviscoplastic constitutive Eq. (13.130). As such, it will be necessary to treat it as a primary variable that must be adjoined to the twenty-eight primary variables $u_i, \sigma_{ij}, \varepsilon_{ij}, \varepsilon^P_{ij}, \alpha 1, \alpha 2_{ij}$ required to solve the thermoviscoplastic initial boundary value problem. The means whereby this new primary unknown is predicted is the conservation of energy Eq. (13.108). This equation may be manipulated so as to remove the secondary unknowns u^I, s, h, γ_i by first employing Eq. (13.118) to solve for the internal energy. The time derivative of the resulting equation may then be taken, and this result is substituted into Eq. (13.108) to obtain the following:

$$\rho\left(\frac{\partial h}{\partial t}+T\frac{\partial s}{\partial t}+s\frac{\partial T}{dt}\right)=\sigma_{ij}\frac{d\varepsilon_{ij}}{dt}-q_{i,i}+\rho r \tag{13.134}$$

Now, Eq. (13.125) may be utilized to take the time derivative of the Helmholtz free energy, thus altering Eq. (13.134) to the following form:

$$-\left(\rho\frac{\partial h}{\partial\varepsilon_{ij}}-\sigma_{ij}\right)\frac{d\varepsilon_{ij}}{dt}+\rho\left(\frac{\partial h}{\partial T}+s\right)\frac{dT}{dt}+\rho T\frac{ds}{dt}+\rho\frac{\partial h}{\partial\varepsilon_{ij}^{P}}\frac{d\varepsilon_{ij}^{P}}{dt}=-q_{i,i}+\rho r \tag{13.135}$$

Using Eqs. (13.121) and (13.122), it can be seen that the two parenthetical terms in Eq. (13.135) are identically zero.

Now, note from Eqs. (13.122) and (13.125) that the time derivative of the entropy may be written as follows:

$$\frac{ds}{dt}=\frac{\partial s}{\partial\varepsilon_{ij}}\frac{d\varepsilon_{ij}}{dt}+\frac{\partial s}{\partial T}\frac{dT}{dt}+\frac{\partial s}{\partial\varepsilon_{ij}^{P}}\frac{d\varepsilon_{ij}^{P}}{dt} \tag{13.136}$$

Substituting Eq. (13.136) into Eq. (13.135) now results in the following:

$$\rho T\left(\frac{\partial s}{\partial\varepsilon_{ij}}\frac{d\varepsilon_{ij}}{dt}+\frac{\partial s}{\partial T}\frac{dT}{dt}+\frac{\partial s}{\partial\varepsilon_{ij}^{P}}\frac{d\varepsilon_{ij}^{P}}{dt}\right)+\rho\frac{\partial h}{\partial\varepsilon_{ij}^{P}}\frac{d\varepsilon_{ij}^{P}}{dt}=-q_{i,i}+\rho r \tag{13.137}$$

Substituting Eq. (13.122) and (13.126) into the above results in the following form of the conservation of energy for case of isotropic material:

$$\rho c_v\frac{dT}{dt}=\left(kT_{,i}\right)_{,i}+\rho r-\alpha(3\lambda+2\mu)T\left(\frac{d\varepsilon_{kk}}{dt}-\frac{d\varepsilon_{kk}^{P}}{dt}\right)+\sigma_{ij}\frac{d\varepsilon_{ij}^{P}}{dt} \tag{13.138}$$

where the nomenclature introduced in Chap. 11 has been used to describe the material properties.

The above equation now has the secondary variables removed, so that it may be used to obtain the temperature field, $T(\vec{x},t)$.

13.5.2.1 Two-Way Coupled Thermoviscoplasticity

Unfortunately, Eq. (13.138) also contains the dilatational components of both the strain tensor, ε_{kk}, and the plastic strain tensor, ε_{kk}^{P}. This implies that the mechanics and the thermodynamics are *coupled*, meaning that neither can be solved separately from one another in a thermoviscoplastic object. This coupling in thermoviscoplastic media can be substantially larger than that encountered in thermoelastic

Table 13.3 Two-way coupled linear thermoviscoplastic initial boundary value problem

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S), f_i(\text{in } V+S), r(\text{in } V+S)$

Geometry: $\vec{x}(\text{on } S),\ n_i(\text{on } S)$

Material Properties: $\rho, \lambda, \mu, \alpha, k, c_v$ *plus all material properties appearing in internal variable evolution laws*

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}, \varepsilon_{ij}^P, \alpha 1, \alpha 2_{ij}, T$ **= 29 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = \lambda(\varepsilon_{kk} - \varepsilon_{kk}^P)\delta_{ij} + 2\mu(\varepsilon_{ij} - \varepsilon_{ij}^P) - \alpha(3\lambda + 2\mu)\theta\delta_{ij}$	6
$\rho c_v \frac{dT}{dt} = (kT_{,i})_{,i} + \rho r - \alpha(3\lambda + 2\mu)T\frac{d\varepsilon_{kk}}{dt}$	1
$\frac{d\varepsilon_{ij}^P}{dt} = \bar{\lambda}\frac{\partial F}{\partial \sigma_{ij}}, F = F(\sigma_{ij}, \alpha 1, \alpha_{ij}^2)$	6
$\frac{d\alpha 1}{dt} = f1(\varepsilon_{ij}, \varepsilon_{ij}^P, \alpha 1, \alpha 2_{ij}, \ldots)$	1
$\frac{d\alpha 2_{ij}}{dt} = f2_{ij}(\varepsilon_{ij}, \varepsilon_{ij}^P, \alpha 1, \alpha 2_{ij}, \ldots)$	6
Total	**29**

Initial Conditions:

Displacements:	$u_i(\vec{x}, 0) = \hat{u}_i$	$\forall \vec{x} \in V + S$
Stresses:	$\sigma_{ij}(\vec{x}, 0) = \hat{\sigma}_{ij}$	$\forall \vec{x} \in V + S$
Strains:	$\varepsilon_{ij}(\vec{x}, 0) = \hat{\varepsilon}_{ij}$	$\forall \vec{x} \in V + S$
Temperature:	$T(\vec{x}, 0) = \hat{T}$	$\forall \vec{x} \in V + S$
Plastic strain tensor:	$\varepsilon_{ij}^P(\vec{x}, 0) = 0$	$\forall \vec{x} \in V + S$

(continued)

Table 13.3 (continued)

Isotropic hardening variable:	$\alpha 1(\vec{x},0)=\hat{\alpha}1$	$\forall \vec{x} \in V+S$
Kinematic hardening variable:	$\alpha 2_{ij}(\vec{x},0)=0$	$\forall \vec{x} \in V+S$
Boundary Conditions:		
Tractions:	$t_i=\hat{t}_i \quad on \quad S_1$	
Displacements:	$u_i=\hat{u}_i \quad on \quad S_2$	$S=S_1+S_2$
Temperature:	$T=\hat{T} \quad on \quad S_3$	
Heat Flux:	$q_i=\hat{q}_i \quad on \quad S_4$	$S=S_3+S_4$

media due to the inclusion of the last term in Eq. (13.138), called *the rate of plastic work*.

This then results in the so-called *coupled thermoviscoplastic initial boundary value problem*, thus implying that when temperature change occurs in a thermoviscoplastic object, the deformation changes, *and* vice versa, that is, mechanical loading induces temperature change. This initial boundary value problem is described in Table 13.3 for the isotropic case.

13.5.2.2 One-Way Coupled Thermoviscoplasticity

The initial boundary value problem described in Table 13.3 can be quite challenging to solve due to the thermoviscoplastic coupling. Fortunately, this coupling is in many cases quite small, i.e., the terms containing both the strain tensor and the plastic strain tensor produce negligible heating. This assumption appears to be accurate for the purpose of modeling flexible pavement structures. Therefore, this will be assumed to be the case throughout the remainder of this text, thus resulting in the following simplification of Eq. (13.138):

$$\rho c_v \frac{dT}{dt} = (kT_{,i})_{,i} + \rho r \tag{13.139}$$

It can be seen that the above is an uncoupled form of the conservation of energy, in which the temperature is the single unknown field variable, thus earning this simplification the name *uncoupled heat transfer*. It is in fact the theory first developed by Fourier (Fourier 1822). In this case, the procedure described in Table 13.3 simplifies to two separate problems. First, the temperature field, $T(\vec{x},t)$

Table 13.4 Part 1: procedure for solving for the temperature in uncoupled thermoviscoplasticity

Part 1: Solve for $T = T(\vec{x}, t)$

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: r

Geometry: $\vec{x}\,(\text{on } S),\ n_i\,(\text{on } S)$

Material Properties: ρ, k, c_v

Unknowns: T = 1 unknown

Field Equation:

	No. of Equations
$\rho c_v \dfrac{dT}{dt} = (kT_{,i})_{,i} + \rho r$	1
Total	**1**

Initial Conditions:

Temperature: $T(\vec{x}, 0) = \hat{T} \quad \forall \vec{x} \in V + S$

Boundary Conditions:

Temperature: $T = \hat{T} \quad on \quad S_3$

Heat Flux: $q_i = \hat{q}_i \quad on \quad S_4 \quad S = S_3 + S_4$

is predicted, subject to the thermodynamic initial and boundary conditions described in Table 13.3. Thereafter, the temperature may be deployed as a known quantity to predict the remaining twenty-eight unknowns, $u_i(\vec{x}, t), \sigma_{ij}(\vec{x}, t), \varepsilon_{ij}(\vec{x}, t), \varepsilon^P_{ij}(\vec{x}, t), \alpha 1(\vec{x}, t), \alpha 2_{ij}(\vec{x}, t)$. The special case of an initially isotropic object subjected to quasi-static loading is described in Tables 13.4 and 13.5.

13.6 Methods for Modeling Cracking

It should go without saying that any attempt to model road way performance will be unsuccessful unless it accounts for the cracking that develops within the road way over time (see Chap. 8). Toward this end, there are two markedly different methods that have developed within the road way modeling community. The most

Table 13.5 Part 2: procedure for solving for stresses, strains, and displacements in uncoupled thermoviscoplasticity

PART 2: Solve for $u_i = u_i(\vec{x},t), \sigma_{ij} = \sigma_{ij}(\vec{x},t), \varepsilon_{ij} = \varepsilon_{ij}(\vec{x},t),$
$\varepsilon_{ij}^P = \varepsilon_{ij}^P(\vec{x},t), \alpha 1 = \alpha 1(\vec{x},t), \alpha 2_{ij} = \alpha 2_{ij}(\vec{x},t)$

Independent Variables: $\vec{x}, t$

Known Inputs:

Loads: $\hat{t}_i(\text{on } S), f_i(\text{in } V+S)$

Geometry: $\vec{x}\,(\text{on } S),\ n_i\,(\text{on } S)$

Material Properties: $\rho, \lambda, \mu, \alpha$ *plus the coefficients appearing in the internal variable evolution laws*

Unknowns: $u_i, \varepsilon_{ij}, \sigma_{ij}, \varepsilon_{ij}^P, \alpha 1, \alpha 2_{ij}$ = **28 unknowns**

Field Equations:

	No. of Equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\sigma_{ij} = \lambda(\varepsilon_{kk} - \varepsilon_{kk}^I)\delta_{ij} + 2\mu(\varepsilon_{ij} - \varepsilon_{ij}^I) - \alpha(3\lambda + 2\mu)\theta\delta_{ij}$	6
$\frac{d\varepsilon_{ij}^I}{dt} = f_{ij}(\sigma_{kl}, \alpha 1, \alpha_{kl}^2)$	6
$\frac{d\alpha 1}{dt} = f1(\varepsilon_{ij}, \varepsilon_{ij}^I, \alpha 1, \alpha 2_{ij}, \ldots)$	1
$\frac{d\alpha 2_{ij}}{dt} = f2_{ij}(\varepsilon_{kl}, \varepsilon_{kl}^I, \alpha 1, \alpha 2_{kl}, \ldots)$	6
Total	**28**

Initial Conditions:

Displacements: $u_i(\vec{x},0) = \hat{u}_i \quad \forall \vec{x} \in V+S$

(continued)

Table 13.5 (continued)

Stresses:	$\sigma_{ij}(\vec{x},0)=\hat{\sigma}_{ij}$	$\forall \vec{x} \in V+S$
Strains:	$\varepsilon_{ij}(\vec{x},0)=\hat{\varepsilon}_{ij}$	$\forall \vec{x} \in V+S$
Inelastic strain tensor:	$\varepsilon^{I}_{ij}(\vec{x},0)=0$	$\forall \vec{x} \in V+S$
Isotropic hardening variable:	$\alpha 1(\vec{x},0)=\hat{\alpha}1$	$\forall \vec{x} \in V+S$
Kinematic hardening variable:	$\alpha 2_{ij}(\vec{x},0)=0$	$\forall \vec{x} \in V+S$
Boundary Conditions:		
Tractions:	$t_i=\hat{t}_i \quad on \quad S_1$	
Displacements:	$u_i=\hat{u}_i \quad on \quad S_2 \quad S=S_1+S_2$	

commonly utilized approach employs a method termed *continuum damage mechanics*. This approach is both approximate and phenomenological in nature, thereby leading to some inaccuracy in the predictive results. On the other hand, it can be deployed with a certain degree of efficiency that is generally not possible with the alternative approach, termed *fracture mechanics*. The mechanics, as well as the pros and cons of each of these approaches, are discussed in the following two sections.

13.6.1 Damage Mechanics

In this approach, the effects of evolutionary cracking are accounted for in the constitutive equations via a phenomenological damage parameter, D, that is a continuous function of spatial coordinates within the road way. Sometimes termed *continuum damage mechanics,* in this approach the cracks themselves are not modeled explicitly, but instead are accounted for via their effects on the global-scale constitutive equations. Accordingly, cracks within the road way are not modeled explicitly, thereby limiting the applicability to cracks that are small compared to the length scale of the road way (termed herein as microcracks).

Thus, for example, the elastic modulus is now written in the following form:

$$E(\vec{x},t)=[1-D(\vec{x},t)]E^{0}(\vec{x}) \tag{13.140}$$

where E^0 is the initial undamaged elastic modulus and the damage parameter, D, is modeled via a *damage evolution law* of the generic form:

$$\frac{dD(\vec{x},t)}{dt} = f^D\left(\varepsilon_{ij}(\vec{x},t), D(\vec{x},t), \ldots\right) \tag{13.141}$$

As such, D represents the locally averaged effect of microcracking within a statistically homogeneous representative volume element of material on the locally averaged effective modulus within that RVE. Since microcracking can and indeed does vary spatially within the road way, D will necessarily vary spatially due to the coordinate location dependence of the state variables included on the right-hand side of evolutionary Eq. (13.141). Note that due to the form of Eq. (13.140)

$$D(\vec{x}, t=0) = 0 \quad \forall \vec{x} \in V + S \tag{13.142}$$

It should be apparent from Eq. (13.141) that the value of D can be determined experimentally in the laboratory by performing tests on specimens that induce statistically homogeneous states of microcracking, measuring the resulting reduction in the modulus, E, and utilizing the experimental results to characterize the exact nature of damage evolution law (13.141) for the material in question. Additionally, the nonlinear viscoelastic model described in Chap. 12 can be extended to include damage in a similar manner to Eq. (13.140) (Park et al. 1996; Park and Schapery 1997; Schapery 1999).

This approach has inherent difficulties. First, one must assume a phenomenologically based mathematical form for the damage parameter, D. This is not known a priori. Second, this approach requires that the experiments be performed on AC pavement rather than at the level of the constituents, so that laboratory experiments must be performed for each expected aggregate volume fraction and shape distribution, making it costly to use. Third, the numerical algorithm required to deploy this within a finite element code is daunting. Finally, because this approach does not predict macrocracks, it cannot be used to model large-scale phenomena such as alligator and thermal cracking (see Chaps. 1 and 8).

13.6.2 Fracture Mechanics

As described in the previous section, the approach utilized within the cognomen continuum damage mechanics implies that the damage is not modeled explicitly, but rather in a locally homogenized way within the global-scale constitutive equations. By contrast, the approach taken with fracture mechanics is to model each crack explicitly as the production of new surface area due to the applied loads and environmental effects.

Whereas the subject of mechanics is a very old and rich one (Allen 2014), the development of a cogent theoretical framework for the purpose of predicting crack

growth in solids has developed only within the last century. Rooted within older theories of yielding and failure, such as that discussed earlier in this chapter, predictive fracture methodologies are founded within the concept that some threshold value of mechanical variables must be exceeded in order for a crack to propagate within a solid.

Perhaps the first modern attempt to do this was proposed by English aeronautical engineer Arnold A. Griffith (1893–1963), who hypothesized that in order for a crack to propagate within a solid a necessary and sufficient condition is given by the following (Griffith 1921):

$$G \geq G_C \tag{13.143}$$

where G is *the available energy release rate* (meaning—available energy per unit length of crack extension), and G_C is a material property termed *the critical energy release rate*. The ramifications of such a postulate are that if the critical energy release rate, G_C, can be measured experimentally, then inequality (13.143) can be utilized to predict when a crack will propagate, so long as the available energy release rate, G, can be accurately predicted.

While Griffith's criterion is a logical one that has been shown via experimental results to be accurate for a range of materials, it can be somewhat complicated to deploy. This is due to the fact that the available energy release rate, G, may in fact be quite difficult to predict. In fact, over the course of the twentieth century, it developed that a useful means of predicting G for many materials is via the deployment of continuum mechanics, the approach laid out in the preceding chapters of this text.

Unfortunately, utilizing continuum mechanics for such a prediction has turned out to be quite cumbersome. This is at least in part due to the fact deploying continuum mechanics to predict G in elastic media generally leads to a null predicted energy release rate unless the tip of the crack in question is assumed to be mathematically sharp (meaning—the radius of the crack tip is infinitesimally small, an assumption that is not possible in reality), thereby leading to the prediction that the stress state at the tip of the crack is infinitely large (termed singular). Of course, this is also not possible in reality. Nonetheless, it has been found that these two errors somehow manage to nullify one another in such a way that Griffith's criterion is amazingly accurate for certain materials. It has therefore become the most widely utilized criterion for predicting crack growth.

The Griffith criterion is oftentimes inaccurate for the purpose of modeling crack growth in viscoelastic solids (Knauss 1970). Accordingly, various attempts have been made to "patch up" the Griffith criterion so that it may be used to model crack growth in viscoelastic solids (Schapery 1975, 1984; Christensen 1981), but these efforts have in some cases been less than perfect. This may be due at least in part to the fact that in viscoelastic media the time-dependent nature of the material behavior obviates the possibility of a mathematically sharp crack tip, thereby rendering the approach used in so-called "linear elastic fracture mechanics" (wherein the stress field is singular at the crack tip) dubious at best.

Acting on this fact, A. Dugdale proposed the existence of a set of cohesive forces acting on the crack faces and trailing from the tip of a crack (Dugdale 1960), and G. Barenblatt proposed that such a *cohesive zone* could be deployed to remove the stress singularity at the crack tip (Barenblatt 1962). The relationship between Griffith's model and the Dugdale-Barenblatt approach was perhaps first discussed in detail by Willis (1967).

A cohesive zone model can be seen to obviate the physical contradictions mentioned above that are inherent in the Griffith model, that is, no assumption need be made that the crack tip is mathematically sharp, nor is it necessary for the stresses at the crack tip to be singular. Accordingly, this approach to predicting crack growth has been extended (Needleman 1987) and gained increased attention (Tvergaard 1990).

In these recent models, it is hypothesized (on the basis of molecular-scale physics) that the faces of the crack tip are subjected to self-equilibrating tractions using a cubic relation of the following generic form:

$$\begin{aligned} t_n^C &= k_n\left[a_0 + a_1 u_n^C + a_2 (u_n^C)^2 + a_3 (u_n^C)^3\right] \\ t_s^C &= k_s\left[b_0 + b_1 u_s^C + b_2 (u_s^C)^2 + b_3 (u_s^C)^3\right] \end{aligned} \tag{13.144}$$

where k_n and k_s are material constants, for the normal (n) and shear (s) components of the crack face traction vector, $\vec{t}^C$. In addition, the remaining coefficients are fit so that the following conditions are met for the normal component of traction, $\vec{t}^C$, as a function of the cohesive zone opening displacement vector, $\vec{u}^C$:

$$t_n^C(u_n^C = 0) = 0,\ t_n^C(u_n^C = u_n^{\max}) = 0,\ \frac{dt_n^C}{du_n^C}(t_n^{\max}) = 0,\ \frac{dt_n^C}{du_n^C}(u_n^C = u_n^{\max}) = 0 \tag{13.145}$$

and the coefficients for the shear component of crack face traction, t_s^C, are determined in a similar fashion. It can be seen from the above conditions that the crack face traction-displacement formulation may be represented graphically as shown in Fig. 13.19.

Unfortunately, the above formulation leads to some difficulty when deployed within computational schemes such as the finite element method because they contain no length scale, thereby leading to mesh refinement issues (de Borst 2003).

Accordingly, generalizations of Eq. (13.144) have been proposed that include a length scale (Bazant and Li 1997; Allen and Searcy 2001a, b). In this text, a viscoelastic cohesive zone model will be utilized that is of the following generic nonlinear viscoelastic form:

$$t_i^C(x_k^C, t) = t_i^C\left\{u_k^C, \alpha^C\right\} \quad \in S^C \tag{13.146}$$

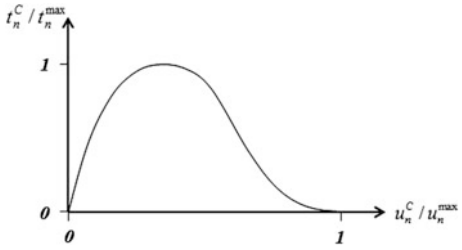


Fig. 13.19 Graphical depiction of the normal traction versus normal displacement relation in the Needleman cohesive zone model

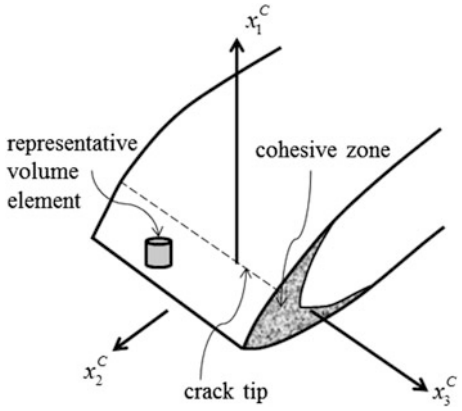


Fig. 13.20 Depiction of a crack with an embedded cohesive zone

where, as shown in Fig. 13.20, S^C is the surface of the cohesive zone ahead of the crack tip and t_i^C are the components of the cohesive zone tractions in the local Cartesian coordinates of the cohesive zone, x_k^C. In addition, α^C is a scalar-valued internal variable termed *the cohesive zone damage variable*. The purpose of Eq. (13.146) is to introduce a physically based mathematical representation of the cohesive zone tractions that is nonmonotonic, such that the cohesive zone forces initially increase with increasing opening of the cohesive zone displacement vector, $\vec{u}^C$, but as the damage within the cohesive zone increases, the cohesive zone tractions eventually descend to zero, thereby signaling crack advancement.

To see how the above concept may be utilized, consider the fibrillated viscoelastic cohesive zone depicted in Fig. 13.21. The micromechanical depiction of the viscoelastic cohesive zone shown in the figure may be modeled analytically, and

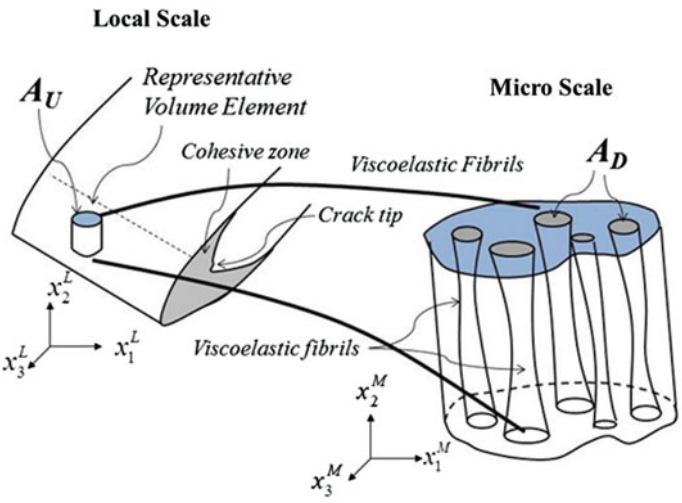


Fig. 13.21 Depiction of a fibrillated viscoelastic cohesive zone

this result can be homogenized within the RVE shown to produce the following traction-displacement relation (Allen and Searcy 2000):

$$t_i(t) = \frac{u_i^C}{\lambda^C}(1 - \alpha^C(t)) \int_0^t E^C(t-\tau)\frac{\partial \lambda^C}{\partial \tau} d\tau \tag{13.147}$$

where $E^C(t)$ is the undamaged relaxation modulus within the cohesive zone (normally taken to be equivalent to the relaxation modulus of the surrounding material), and λ^C is the Euclidean norm of the nondimensionalized cohesive zone opening displacement vector, given by the following:

$$\lambda^C \equiv \left[\left(\frac{u_n^C}{u_n^*}\right)^2 + \left(\frac{u_r^C}{u_r^*}\right)^2 + \left(\frac{u_s^C}{u_s^*}\right)^2 \right]^{1/2} \tag{13.148}$$

where u_i^* are material constants that account for the three fracture modes.

It can be shown from the micromechanical analysis of the cohesive zone depicted in Fig. 13.21 that the cohesive zone damage parameter, α^C, is given by the following:

$$\alpha^C \equiv \frac{A_U^C - A_D^C}{A_U^C} \tag{13.149}$$

where A_U^C is the initial undamaged cross-sectional area of the cohesive zone in the plane of the zone, and A_D^C is the cross-sectional area of the intact material within the cohesive zone in the damaged state. Note that when the cross-sectional area of the

damaged material, A_D^C, goes to zero, $\alpha^C \rightarrow 1$, and the tractions in Eq. (13.147) go to zero, thereby signifying propagation of the crack (see Fig. 13.20).

While Eq. (13.149) gives a micromechanical explanation of the mechanics of the cohesive zone damage parameter, in practice it is found that it can be modeled more practically with a *damage evolution law* of the following form:

$$\begin{array}{ll} \frac{d\alpha^C}{dt} = \alpha_1^C \left(\dot{\bar{\lambda}}^C\right)^{n^C} & \dot{\bar{\lambda}}^C \geq 0 \, and \, \alpha^C \leq 1 \\ \frac{d\alpha^C}{dt} = 0 & \dot{\bar{\lambda}}^C \leq 0 \, or \, \alpha^C = 1 \end{array} \tag{13.150}$$

where α_1^C and n^C are cohesive zone material constants that are obtained from experiments (Kim et al. 2010).

Figure 13.22 depicts the generic form that the traction-displacement relation predicts for a range of constant cohesive zone opening displacement rates. Note that both the cohesive zone response and the work of separation (the area under the traction-displacement curve) within the cohesive zone are rate dependent, commensurate with the observed fracture behavior of viscoelastic media, including asphalt.

This approach to modeling cracking has the advantage that the material properties that are utilized are at the constituent level of the mastic, so that one need not perform costly laboratory experiments for each possible choice of mastic, aggregate volume fraction, and aggregate shape distribution. Thus, it becomes possible to include these parameters as controllable design parameters, unlike the single-scale phenomenological approach described in the previous section. Furthermore, deploying a cohesive zone model can be quite attractive when modeling the road way computationally, as will be demonstrated in Chap. 15.

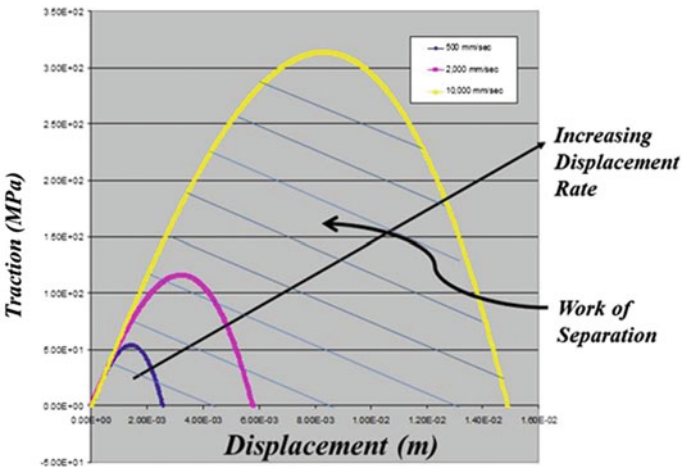


Fig. 13.22 Predicted traction–displacement relation for a viscoelastic cohesive zone loaded at increasing rates

This approach also has some disadvantages. First, it requires that the model be verified against experimental data obtained for different aggregate volume fractions. Second, because it uses a homogenization principle to construct the damage-dependent thermoviscoelastic-moisture constitutive equations for the asphalt concrete, it necessarily results in some approximation in the constitutive model for the asphalt concrete.

13.7 Summary

This chapter has presented an overview of the development of the theories of plasticity, viscoplasticity, and thermoviscoplasticity. In addition, two methodologies that can be utilized to account for the effects of progressive cracking have been briefly reviewed. An understanding of these models is essential to the ability to predict the response of flexible pavements containing asphalt.

13.8 Problems

Problem 13.1 Given: A soft metallic material, such as soldering wire.
Required:

(a) Devise a mechanical apparatus for reproducing Da Vinci's test described in Chap. 2.
(b) Use your apparatus to perform tests on the wire, slowly increasing the load until the relation between the load and displacement becomes nonlinear.
(c) Once the behavior becomes nonlinear, remove the load incrementally until the wire is completely unloaded.
(d) Perform sufficient tests on the wire to determine whether the material is elastoplastic.
(e) Plot the results of your tests on a stress–strain diagram.

Problem 13.2 Given: The strain energy, W, contained within a linear elastic material is given by the following:

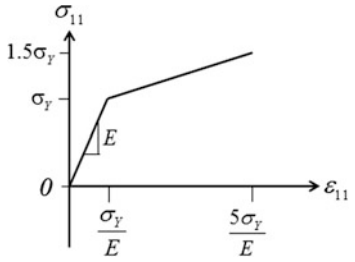
$$W = \frac{1}{2}\sigma_{ij}\varepsilon_{ij}$$

Required: Prove that for isotropic media

$$W = f_1(I_1) + f_2(J_2)$$

Problem 13.3

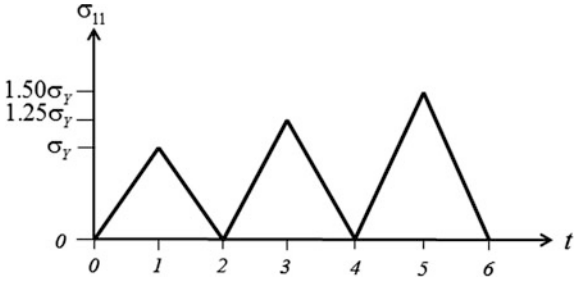
Given: The material with uniaxial rate-independent stress–strain behavior shown below.



Required:

(a) Construct a graph of H' vs. $\bar{\varepsilon}^P$.

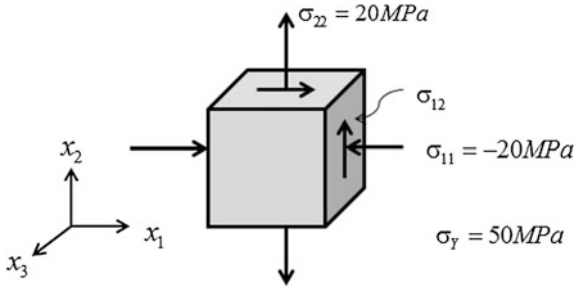
(b) Construct a graph of σ_{11} vs. ε_{11} for the following uniaxial stress input using isotropic hardening.



(c) Construct the genesis of the yield surface in 2D principal stress space.

Problem 13.4

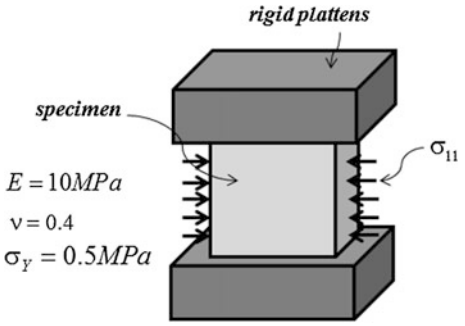
Given: The specimen shown below is loaded such that the stresses are spatially homogeneous.



Required: Determine the value of σ_{12} that will initiate yielding using the following yield criteria:

(a) Von Mises.
(b) Tresca.
(c) Drucker–Prager ($\mu = 0.05$).

Problem 13.5
Given: The experiment shown below is constructed such that the stress and strain are spatially homogeneous and the specimen does not deform in the x_2 coordinate direction.



Required: Predict the values of σ_{11} and ε_{11} at initial yield using

(a) the von Mises yield criterion and
(b) the Drucker–Prager yield criterion ($\mu = 0.05$).

Problem 13.6
Given: Drucker's postulate.

Required: Write out the expression for the plastic strain rate in terms of the stress tensor for materials obeying the following yield criteria:

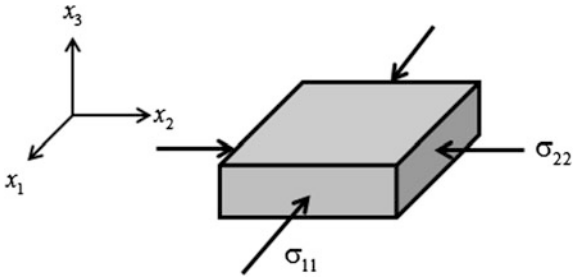
(a) Tresca.
(b) Mohr–Coulomb.
(c) Drucker–Prager.

Problem 13.7
Given: A material is known to yield according to the Drucker–Prager yield criterion.

Required:

(a) Construct the stress formulation for isotropic hardening.
(b) Construct the stress formulation for kinematic hardening.
(c) Construct the stress formulation for combined hardening.

Problem 13.8
Given: The specimen shown below is loaded in biaxial compression such that the stress state is spatially homogeneous and $\sigma_{11} = \sigma_{22} (\sigma_{23} = \sigma_{13} = \sigma_{12} = \sigma_{33} = 0)$.



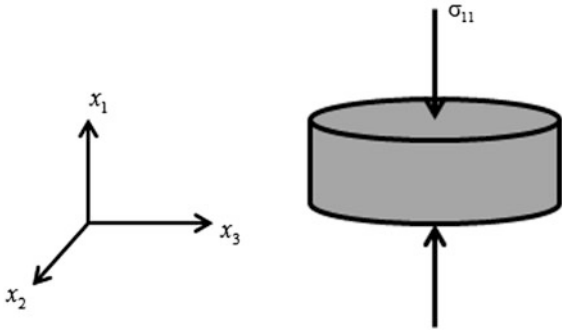
Required: Using the von Mises yield criterion

(a) Construct a depiction of the yield surface in 2D principal stress space at the point at which $\sigma_{11} = \sigma_{22} = -2\sigma_Y$ assuming (1) isotropic hardening and (2) kinematic hardening.
(b) Suppose $\sigma_{11} = -2\sigma_Y$ and σ_{22} is slowly reversed from compression to tension. Calculate the value of σ_{22} at which re-yield will occur assuming (1) isotropic hardening and (2) kinematic hardening.

Problem 13.9
Given: The Drucker–Prager yield criterion.

Required: Construct a nonassociated flow rule by assuming that the plastic strain increment is normal to the component of the gradient of the part of the yield function that is independent of hydrostatic pressure

Problem 13.10 Given: The specimen shown below is loaded axially such that $\sigma_{11} = -\sigma_Y$. Thereafter, the axial stress is held constant and a lateral pressure, p, is applied such that $\sigma_{22} = \sigma_{33} = p$, and this pressure is gradually increased.



Required: Using the Drucker–Prager yield criterion ($\mu = 0.05$) and isotropic hardening, predict and plot p vs. ε_{11}, assuming the uniaxial stress–strain curve is as shown in Problem 13.3.

References

L'Académie Française. (1699). Histoire de l'Académie Royale des Sciences avec les Mémoires de Mathématique et de Physique, 206.

Allen, D. (1991). A review of the theory of thermomechanical coupling in inelastic solids. *Applied Mechanics Reviews, 44,* 361.

Allen, D., & Searcy, C. (2000). Numerical aspects of a micromechanical model of a cohesive zone. *Journal of Reinforced Plastics and Composites, 19,* 240–248.

Allen, D., & Searcy, C. (2001a). A micromechanical model for a viscoelastic cohesive zone. *International Journal of Fracture, 107,* 159–176.

Allen, D., & Searcy, C. (2001b). A micromechanically based model for predicting dynamic damage evolution in ductile polymers. *Mechanics of Materials, 33,* 177–184.

Allen, D. (2014). *How mechanics shaped the modern world*. Springer.

Barenblatt, G. (1962). The mathematical theory of equilibrium crack in brittle fracture. *Advances in Applied Mechanics, 7,* 55.

Bazant, Z. & Li, Y. (1997). Cohesive crack with rate-dependent opening and viscoelasticity: I. mathematical model and scaling. *International Journal of Fracture*, *86*, 247.

Cela, J. (1998). Analysis of reinforced concrete structures subjected to dynamic loads with a viscoplastic Drucker-Prager model. *Applied Mathematical Modelling, 22,* 495.

Christensen, R. (1981). A theory of crack growth in viscoelastic materials. *Engineering Fracture Mechanics, 14,* 215.

Coleman, B., & Gurtin, M. (1967). Thermodynamics with internal state variables. *The Journal of Chemical Physics, 47,* 597.

Coulomb, C. (1776). Essai sur une application des regles des maximis et minimis a quelques problemes de statique relatifs, a la architecture. *Mem. Acad. Roy. Div. Sav. 7*, 343.

de Borst, R. (2003). Numerical aspects of cohesive-zone models. *Engineering Fracture Mechanics, 70,* 1743.

Drucker, D. (1951). A more fundamental approach to plastic stress-strain relations. In *Proceedings of the 1st National Congress of Applied Mechanics* (p. 487). ASME.

Dugdale, D. (1960). Yielding of steel sheets containing slits. *Journal of the Mechanics and Physics of Solids, 8*(2), 100–104.

Elaskar, S., Godoy, L., Gray, D., & Stiles, J. (2000). A viscoplastic approach to model the flow of granular solids. *Int. J. Solids Structures, 37,* 2185.

Fourier, J. (1822). *Théorie analytique de la chaleur*. Firmin Didot.

Gear, C. (1971). *Numerical initial-value problems in ordinary differential equations*. Prentice-Hall.

Gear, C. (1981). Numerical solution of ordinary differential equations: Is there anything left to do? *SIAM Review, 23,* 10.

Griffith, A. (1921). The phenomena of rupture and flow in solids. *Philosophical Transactions of the Royal Society, A221,* 163.

Hill R (1950) *The Mathematical theory of plasticity*. The Oxford University Press.

Kim, Y., Aragão, F., & Allen, D. (2010). Damage modeling of bituminous mixtures considering mixture microstructure, viscoelasticity, and cohesive zone fracture. *Canadian Journal of Civil Engineering, 37,* 1125.

Knauss, W. (1970). Delayed failure-the Griffith problem for linearly viscoelastic materials. *Int. J. Fracture Mech., 6,* 7.

Kratochvil, J., & Dillon, O. (1970). Thermodynamics of crystalline elastic-visco-plastic materials. *Journal of Applied Physics, 41,* 1470.

Lubliner, J. (2008). *Plasticity Theory*. Courier Dover.

Masad, E., Tashman, L., Little, D., & Zbib, H. (2005). Viscoplastic modeling of asphalt mixes with the effects of anisotropy, damage and aggregate characteristics. *Mechanics of Materials, 37,* 1242.

Mises, R. (1913). *Mechanik der festen körper im plastisch deformablen zustand, Göttinger Nachr* (p. 582). Mathematisch-Physikalische Klasse.

Naghdi, P. (1960). Stress-strain relations in plasticity and thermoplasticity. In E. Lee & P. Symonds (Eds.), *Plasticity* (p. 121). Pergamon.

Needleman, A. (1987). A continuum model for void nucleation by inclusion debonding. *Journal of Applied Mechanics, 54,* 525.

Negahban, M. (2012). The *mechanical and thermodynamical theory of plasticity*. CRC Press.

Park, S., Kim, R., & Schapery, R. (1996). A viscoelastic continuum damage model and its application to uniaxial behavior of asphalt concrete. *Mechanics of Materials, 24,* 241.

Park, S., & Schapery, R. (1997). A viscoelastic constitutive model for particulate composites with growing damage. *The International Journal of Solids and Structures, 34,* 931.

Prager, W. (1959). *An introduction to plasticity*. Addison-Wesley.

Prandtl, L. (1928), Ein gedankenmodell zur kinetischen theorie der festern körper, Z. Angew. *Mathematics and Mechanics*, *8*, 85

Schapery, R. (1975). A theory of crack initiation and growth in viscoelastic media I. theoretical development. *International Journal of Fracture*, *11*, 141.

Schapery, R. (1984). Correspondence principles and a generalized J integral for large deformation and fracture analysis of viscoelastic media. *International Journal of Fracture, 25,* 195.

Schapery, R. (1999). Nonlinear viscoelastic and viscoplastic constitutive equations with growing damage. *International Journal of Fracture, 97,* 33.

Tashman, L., Masad, E., Little, D., & Zbib, H. (2005). A microstructure-based viscoplastic model for asphalt concrete. *International Journal of Plasticity, 21,* 1659.

Timoshenko, S. (1953). *History of strength of materials.* McGraw-Hill.

Tresca, H. (1864). *Mémoire sur l'écoulement des corps solides soumis à de fortes pressions* (vol. 59, p. 754). Paris: C.R. Academy of Sciences.

Truesdell, C., & Toupin, R. (1954). The classical field theories, encyclopedia of physics. *Springer, 3,* 1.

Tvergaard, V. (1990). Effect of fibre debonding in a whisker-reinforced metal. *Materials Science and Engineering, 125,* 203.

Willis, J. (1967). A comparison of the fracture criteria of Griffith and Barenblatt. *Journal of the Mechanics and Physics of Solids, 42,* 1397.

Zeigler, H. (1959). A modification of Prager's hardening rule. *Quarterly of Applied Mathematics, 17,* 55.

Chapter 14
Computational Methods for Roadway Analysis and Design

14.1 Introduction

It should now be apparent to the reader that a flexible road way is a quite complex infrastructure. As shown in Fig. 14.1, typical road ways are complicated by four unavoidable factors: loads, geometry, material properties, and environmental conditions. The loads may depend on vehicular weight, tire configuration, tire pressure, tire pressure distribution, and number of loading cycles. The geometry of the road way is affected by terrain features, number of lanes, culvert design, layer depths, and aggregate shape and size, not to mention the geometrics and chemistry associated with additives such as fines (see Chap. 5). The material properties of the road way constituents are invariably dictated by cost and availability of materials. Finally, environmental conditions vary not only spatially across our planet, but they also vary temporally at any given point over the life span of a given road way infrastructure due to the statistical nature of the Earth's climatology.

As detailed in the previous chapters (see Chaps. 10 through 13), the material constitution within each layer can further complicate the problem, as some layers may behave quite differently from others. The previous four chapters focused on the development of models for predicting the behavior of each layer of the road way. Indeed, as shown in Fig. 14.2, each layer of the road way may necessarily have to be modeled with a different constitutive model, thereby resulting in the formulation of an exceedingly intricate initial boundary value problem. As an added complication, environmental conditions introduce a level of complexity to the problem that may further obviate the ability to accurately predict pavement performance (see Chap. 8).

Until fairly recently such impediments would have rendered the ability to predict road way performance scientifically untenable, this despite the rapid expansion in our scientific understanding of the physics and chemistry of road way materials (see Chaps. 2 through 8). However, two simultaneous developments during the latter half of the twentieth century have made the deployment of sound scientific principles within road way models more tenable. First and foremost has been the development

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_14

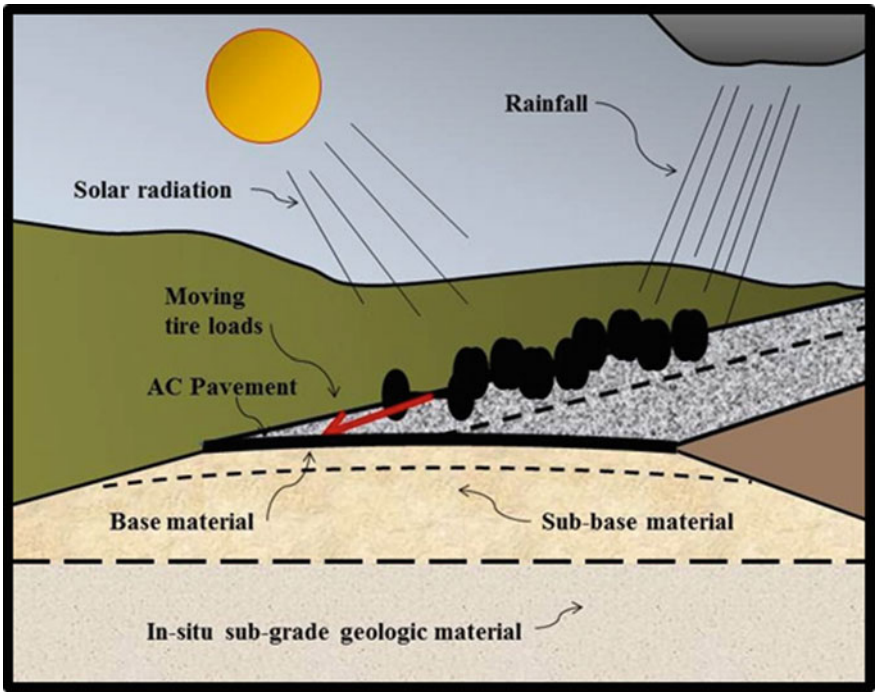


Fig. 14.1 Depiction of a typical roadway subjected to complex loading condition

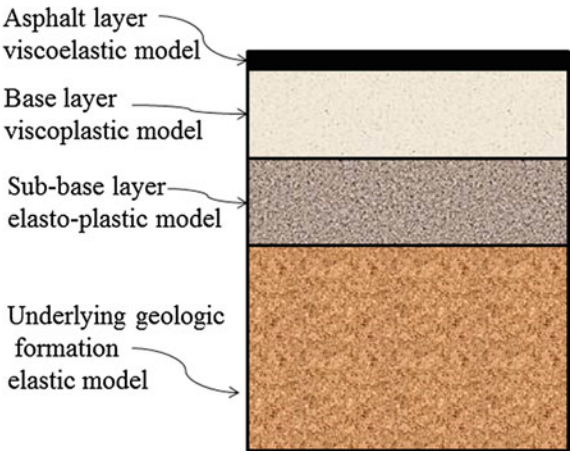


Fig. 14.2 Depiction of layered structure with associated material models in a typical roadway

of the digital computer. Concomitantly, the finite element method has arisen. These two essential innovations have been necessary in order to develop computational algorithms for the purpose of deploying sound scientific principles for the purpose of modeling all of the complexities encountered in typical flexible pavements.

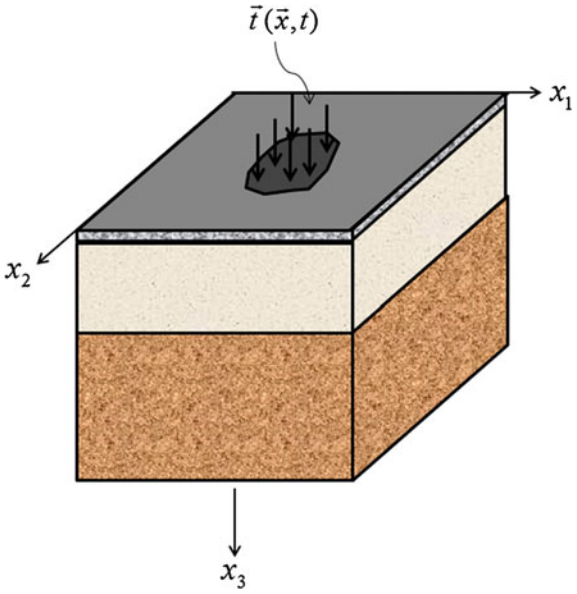


Fig. 14.3 Depiction of a typical tire loading applied to the roadway surface

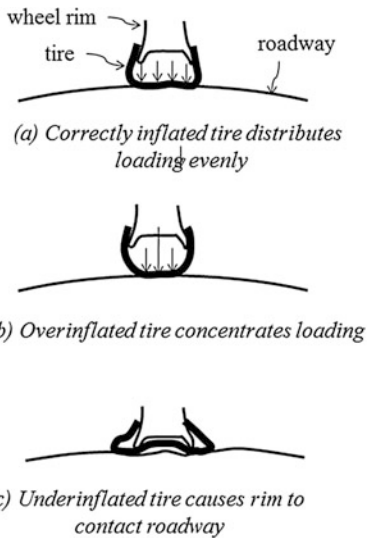


Fig. 14.4 Depictions of various tire inflation configurations

As shown in Fig. 14.3, the typical road way is subjected to traction boundary conditions that vary both in time and spatial coordinates. As demonstrated in Fig. 14.4, the applied loads are normally of traction type because the road way is much stiffer than the tires that come in contact with the road way (see Chap. 11).

If the tire is overinflated, the boundary conditions will remain of traction type because the tire is nonetheless much more compliant than the asphalt concrete even

when overinflated. In this case, despite the fact that the axle loading remains unchanged, the reduction in the contact area due to overinflation can increase the tire tractions to levels that can rapidly damage the road way. This is a major reason that overinflated tires are to be avoided on road ways.

It is possible for the boundary conditions applied to the road way to be of displacement type, such as in the case when tires are significantly underinflated. As also shown in Fig. 14.4, when the rim of a wheel comes into contact with the surface of the road way, the rim will likely be sufficiently stiffer than the road way as to require the road way surface to conform to the shape of the wheel rim where the two are in contact with one another, thereby resulting in stress concentrations at the point of contact (and most likely resulting in rapid damage to the road way surface). This is one important reason that drivers should avoid driving on underinflated (or flat!) tires.

A significant point of this discussion is to clarify that the loads are essentially repetitive in nature and, except for the stochastic effects associated with differences in tires, they may be assumed for practical purposes to be of constant amplitude. When the *input* loads are, in fact, of constant amplitude, the *output* variables are nevertheless not of constant amplitude, for if they were, the road way would last forever. Instead, the amplitudes of the output variables are observed to slowly evolve with the number of loading cycles, as shown schematically in Fig. 14.5.

This unavoidable fact leads one to the conclusion that all output variables in any useful road way model must be dependent on the entire loading and environmental history sustained by the road way. Because of this fact, it becomes necessary for the finite element procedure to employ what is termed *a time-stepping algorithm*. This procedure does nothing more than utilize the finite element method to apply an *increment* of traction and calculate the *increment* in all of the output variables, and then step forward to the succeeding time step, repeating this process recursively until failure of the road way is predicted by the algorithm.

In this chapter, the essentials of the finite element method will be elucidated for the specific problems described above. Because the finite element method in and of itself is a quite broad and complicated subject, not to mention the fact that it is well

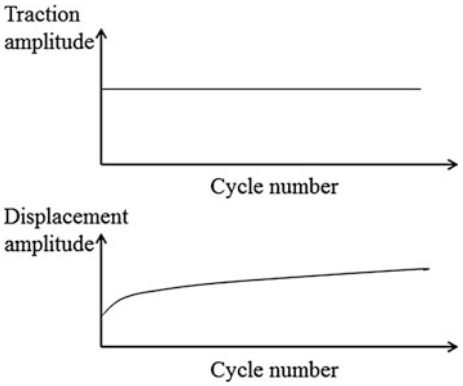


Fig. 14.5 Depiction of the evolution of roadway surface residual displacement as a function of loading cycle number for constant amplitude input traction

documented in the literature, the discussion herein will be limited to what the authors consider to be the minimum information necessary for a robust understanding of this modern methodology that is rapidly becoming the tool of choice for road way design. The interested reader will find a veritable cornucopia of excellent texts on this subject elsewhere (Reddy 2005; Oden 2006; Oden and Reddy 2011; Zienkiewicz and Taylor 2013).

This chapter will be divided into three parts. In the first part, the mathematics of the finite element method will be laid out in as straightforward a process as possible. In the second part, procedures will be developed for implementing both material and fracture models into finite element algorithms. In the third and final part, some typical road way problems will be addressed using the finite element method, thereby demonstrating the enormous power of this computational tool for both analyzing and designing pavements and road ways.

14.2 Fundamentals of the Finite Element Method

The finite element method is a systematic approximate computational methodology for predicting the stresses, strains, and displacements as functions of both time and spatial coordinates in an object of arbitrary shape subjected to external loads and environmental effects. When the proper mathematical equations are utilized to formulate the algorithm, the degree of approximation is limited only by the capacity of the computer used to perform the analysis. However, especially in the case of road ways, the mathematical model is limited by the accuracy of the constitutive model deployed for each layer of the road way. And it should be noted that this limitation is not caused by utilizing the finite element method, but is rather due to the inaccuracies that are inherent in any constitutive model.

Although it has not been the case until recently, there are now numerous commercially available finite element programs that can be used for the purpose of modeling the mechanical response of road ways. These include NASTRAN (McNeal-Schwendler Corporation 2014), ADINA (ADINA System 2011), SAP (Computers and Engineering Software and Consulting, 2015), and Abaqus FEA (Dassault Systèmes 2011). Of these, the last is the most commonly used (to these authors' collective knowledge) for modeling road ways at this writing. There are also a number of open-source finite element codes available on the World Wide Web. Most of these codes are written in either FORTRAN or C++, or an amalgamation of the two. They all function in a similar way, so that selection of the optimum product is a user issue. However, where these codes differ markedly is in the ease of use of their preprocessors and post-processors. These are the software packages that are accessed by the user for the purpose of inputting the information necessary to perform the analysis (such as mesh generation) and outputting the predicted results in a convenient format for road way design purposes.

Because there are so many finite element packages available today, it would seem that no further discussion is warranted on the subject within this text.

However, the authors have chosen to cover the subject in some detail for a number of reasons. First and foremost of these is that the analysis of road ways is unique in the following aspects: the materials that comprise the road way; the geometry of road ways; the loadings applied to road ways; and the environmental conditions that road ways are subjected to. As such, any commercially available code that is not specifically designed to model road ways will oftentimes be cumbersome for the user. Second, although the typical user will not need to have in-depth knowledge of the underlying mathematics and software required to construct a finite element code, there are times when such knowledge is necessary. For these two reasons, we have determined that a synoptic (though rigorous) coverage of the finite element method is warranted within this text. That coverage is contained within the following sections.

The finite element method is essentially nothing more than a computational method for solving sets of partial differential equations, wherein the derivatives are taken with respect to spatial coordinates. As such, the procedure results in some spatial error in approximation of the predicted variables. Fortunately, there is a fairly simple procedure for reducing this approximation-induced error called *mesh refinement*, and this technique will be described further on within the text.

The finite element method was born in the middle third of the twentieth century, when engineers were attempting to develop experimentally inspired methods for designing aircraft and bridges. Simultaneously, mathematicians such as Richard Courant (1888–1972) were developing mathematically inspired techniques that proceeded from a theoretical viewpoint. The two methods converged in the late 1950s, and the term "finite element" was attributed to Ray Clough (1920-2016) in 1960.

The finite element method consists of casting the governing field equations, such as those described in Tables 14.1 and 14.2, into what mathematicians call *a dual formulation*. This means that the governing differential equations are replaced with a more convenient formulation that is (almost) mathematically equivalent, meaning that it will render the same solution for the unknown variables as do the governing differential equations. As described in Chap. 9, the above equations must hold at each and every point in the object of interest. Toward this end, consider a subset of the object in question called *a finite element*, as shown in Fig. 14.6.

Note that the generic finite element may be of arbitrary shape (but is normally not) and must lie wholly within the object of interest. Consistent with the terminology utilized in Table 14.1, the volume and surface of the object of interest are denoted $V + S$, and the volume and surface of the generic finite element are denoted $V^e + S^e$.

The finite element method proceeds by first attempting to satisfy the governing field equations within $V^e + S^e$. Once this step has been completed, the results for the generic finite element are assembled together with other generic elements within the object in such a way that the geometry of the assembled generic elements (called *the finite element mesh)* approximates the geometry of the actual object as closely as possible, and this process is termed *mesh assembly.*

Table 14.1 Review of the one-way coupled heat transfer problem

Independent variables: $\vec{x}, t$	
Known inputs	
Loads	$\hat{q}_i(onS_1), \hat{T}(onS_2), r(inV+S)$
Geometry	$\vec{x}$(on S), n_i(on S)
Material properties	$\rho, k_{ij}, c_v \forall \vec{x} \in V+S$
Unknowns: T, q_i, γ_i = 7 unknowns	
Field equations	No. of equations
$\gamma_i = \frac{\partial T}{\partial x_i}$	3
$\rho c_v \frac{\partial T}{\partial t} = -q_{i,i} + \rho r$	1
$q_i = -k_{ij}\gamma_j$	3
Total	7
Boundary conditions	
Heat flux	$q_i = \hat{q}_i \quad on \quad S_1$
Temperature	$T = \hat{T} \quad on \quad S_2$
Initial conditions	
Temperature	$T(t=0) = known \quad \forall \vec{x} \in V+S$
Temperature gradient	$\gamma_i(t=0) = known \quad \forall \vec{x} \in V+S$
Heat flux	$q_i(t=0) = known \quad \forall \vec{x} \in V+S$

Table 14.2 Review of the quasi-static linear elastic IBVP

Independent variables: $\vec{x}, t$	
Known inputs	
Loads	$\hat{t}_i(on\ S_1), \hat{u}_i(on\ S_2), f_i(in\ V+S)$
Geometry	$\vec{x}$(on S), n_i(on S)
Material properties	$\rho, E_{ijkl}, \alpha, \beta\ \forall \vec{x} \in V+S$
Unknowns $u_i, \varepsilon_{ij}, \sigma_{ij}$ = 15 unknowns	
Field equations	No. of equations
$\varepsilon_{ij} \equiv \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	6
$\sigma_{ji,j} + \rho f_i = 0$	3
$\Delta\sigma_{ij} = C_{ijkl}\Delta\varepsilon_{kl}$	6
Total	15
Boundary conditions	
Tractions	$t_i = \hat{t}_i \quad on \quad S_1$
Displacements	$u_i = \hat{u}_i \quad on \quad S_2$
Initial conditions	
Displacements	$u_i(t=0) = known \quad \forall \vec{x} \in V+S$
Strains	$\varepsilon_{ij}(t=0) = known \quad \forall \vec{x} \in V+S$
Stresses	$\sigma_{ij}(t=0) = known \quad \forall \vec{x} \in V+S$

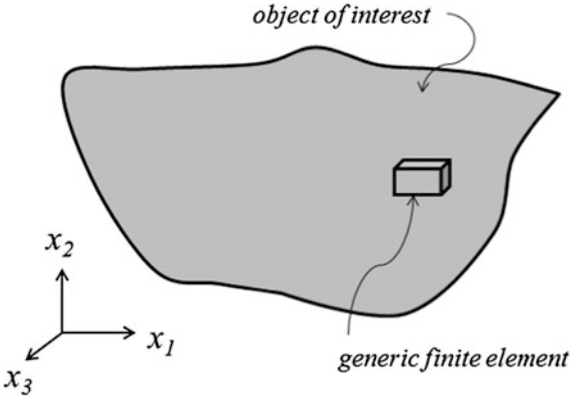


Fig. 14.6 Depiction of a generic finite element within the object of interest

Methods for creating finite element platforms for the purpose of predicting the temperature (called heat transfer) and moisture (called mass transfer) as functions of time and space will be described in Sect. 14.2.1 and 14.2.2, and for the mechanics of deformation in Sect. 14.2.3 and 14.2.4 below.

14.2.1 Construction of the Heat Transfer and Moisture Finite Element Platforms

As described in Chaps. 10–13, while the mechanics of the road way depends on the temperature and moisture in the road way, the reverse is not true; that is, the temperature and moisture fields may in most circumstances be modeled independently of the mechanics of the road way. Therefore, these should be predicted as functions of time and space before performing the mechanics analysis.

As described in Chap. 11, the generic mathematical form of both the heat transfer and the moisture distribution equations is essentially identical, and this form is generally called a diffusion model because it is mathematically parabolic in nature. In this section, the procedure for developing a scalar-valued diffusion finite element model will be described. Toward this end, the generic dependent variable utilized will be the temperature, $T = T(\vec{x}, t)$. However, it should be apparent that when the moisture distribution is to be predicted, the identical field equations are used, so that the finite element algorithm is identical, thereby producing a computational algorithm also capable of predicting the moisture distribution, $M = M(\vec{x}, t)$. These two variables then become required inputs to the mechanics finite element model to be described in the next section.

For those who have already been exposed to the mathematical framework of the finite element method, this section may be skipped. For those who wish to delve further into the finite element method in this section, we begin with a review of the one-way coupled heat transfer initial boundary value problem lain out in Chap. 11 in Table 14.1.

The IBVP described in Table 14.1 is a relatively straightforward one, but it is nonetheless sufficiently complex that analytic solutions are oftentimes not possible, and this is due to complex road way geometry, heterogeneity of the material properties, and complex heating at the boundary of the road way. Fortunately, the finite element method may be used to obtain accurate predictions of the temporal and spatial distribution of temperature within the road way, and this predicted result can then be fed as input to the finite element model for the mechanics of the road way, to be described in the next section.

To see how a dual formulation for the heat transfer problem may be constructed, it is customary to begin with the conservation of energy equation in Table 14.1, given by:

$$\rho c_v \frac{\partial T}{\partial t} = -q_{i,i} + \rho r \tag{14.1}$$

Equation (14.1) is now multiplied by a variation in an admissible temperature field, termed δT, and integrated over the volume of the finite element, V^e, as follows:

$$\int_{V^e} \left(\rho c_v \frac{\partial T}{\partial t} + q_{i,i} - \rho r \right) \delta T \, dV = 0 \tag{14.2}$$

The mathematical complexities associated with the replacement of Eq. (14.1) with Eq. (14.2), thereby resulting in a dual formulation, are profound. The reader is referred to the literature on the finite element method to ascertain these complexities. Suffice it to say herein that, for practical purposes, any set of dependent variables that satisfies the so-called *variational equation* (14.2) also satisfies Eq. (14.1), and this is true for essentially any bounded choice of admissible temperature field, δT.

As will be shown below, there is a choice for δT that is both accurate and convenient, and this choice is almost always utilized with the finite element method. For now, suffice it to say that Eq. (14.1) is now replaced with Eq. (14.2), and this new equation becomes a replacement representation of conservation of energy within the generic element $V^e + S^e$.

Now consider the deployment of the product rule and the divergence theorem [Eq. (9.31)] in the following way:

$$\int_{V^e} (q_i \delta T)_{,i} dV = \int_{V^e} \left(q_{i,i} \delta T + q_i \delta T_{,i} \right) dV = \int_{S^e} q_i \delta T n_i dS \tag{14.3}$$

Rearranging the above and substituting into Eq. (14.2) now results in the following:

$$-\int_{V^e} q_i \delta T_{,i} dV = \int_{V^e} \left(-\rho c_v \frac{\partial T}{\partial t} + \rho r\right) \delta T dV - \int_{S^e} q_i \delta T n_i dS \tag{14.4}$$

From Table 14.1, the heat flux–temperature gradient equation is given by:

$$\gamma_i = T_{,i} \tag{14.5}$$

Therefore, substituting (14.5) into (14.4) results in the following:

$$-\int_{V^e} q_i \delta \gamma_i dV = \int_{V^e} \left(-\rho c_v \frac{\partial T}{\partial t} + \rho r\right) \delta T dV - \int_{S^e} q_i \delta T n_i dS \tag{14.6}$$

Note that (with the exception of the time derivative of the temperature) the known terms have been moved to the right-hand side of the equation.

Now recall the constitutive equation relating the temperature gradient to the heat flux:

$$q_i = -k_{ij} \gamma_j \tag{14.7}$$

Substituting Eq. (14.7) into Eq. (14.6) and rearranging terms thus results in the following:

$$\delta U^e \equiv \int_{V^e} \rho c_v \frac{\partial T}{\partial t} \delta T dV + \int_{V^e} k_{ij} \gamma_j \delta \gamma_i dV - \int_{V^e} \rho r \delta T dV + \int_{S^e} q_i \delta T n_i dS = 0 \tag{14.8}$$

where δU^e is called the variational energy within the element.

Equation (14.8) is now a complete variational form of the heat transfer equation that is ready for deployment with the finite element method, that is, it is a single equation in terms of the primary variable, T.

14.2.2 Construction of the Finite Element Heat Transfer Equations for a Single Element

In order to obtain a solution to the variational form given by Eq. (14.8), it is necessary to *assume* the spatial dependence within the finite element in the primary dependent variable in the equation, T. This is accomplished by assuming that

$$T(\vec{x}, t) = T^{\alpha}(t) h^{\alpha}(\vec{x}) \quad \alpha = 1, \ldots, n \tag{14.9}$$

where n is defined to be the number of *nodes* within the element, $h^{\alpha}(\vec{x})$ are called *shape functions*, and u_i^{α} are termed *the nodal displacements*. The mathematical form

of the shape functions will be determined by the specific element chosen for the analysis, and this subject will be taken up in detail in a later section. For now, it is expedient to proceed with the implementation of Eq. (14.9) into the variational form given by Eq. (14.8).

Equation (14.9) is first substituted into Eq. (14.5), thereby resulting in the following:

$$\gamma_i = T^{\alpha} h^{\alpha}_{,i} \tag{14.10}$$

Substituting Eqs. (14.9) and (14.10) into Eq. (14.8) now results in the following:

$$\left(\int_{V^e} \rho c_v h^{\alpha} h^{\beta} \frac{\partial T^{\alpha}}{\partial t} dV + \int_{V^e} k_{ij} T^{\alpha} h^{\alpha}_{,j} h^{\beta}_{,i} dV - \int_{V^e} \rho r h^{\beta} dV + \int_{S^e} q_i n_i h^{\beta} dS \right) \delta T^{\beta} = 0 \tag{14.11}$$

Because the variation in the temperature is arbitrary, it need not be zero, leading to the result that the entire parenthetical term in Eq. (14.11) must be zero. Accordingly, this term may be written as the following set of equations in terms of the nodal temperature matrix $\{T\}$ for a generic element:

$$[M^e] \left\{ \frac{\partial T}{\partial t} \right\} + [K^e]\{T\} - \{F^e\} = 0 \tag{14.12}$$

where α and β range from one to the number of nodes in the element. In addition,

$$M^e_{\alpha\beta} \equiv \int_{V^e} \rho c_v h^{\alpha} h^{\beta} dV \tag{14.13}$$

$$K^e_{\alpha\beta} \equiv \int_{V^e} k_{ij} h^{\beta}_{,j} h^{\alpha}_{,i} dV \tag{14.14}$$

and

$$F^e_{\beta} \equiv \int_{V^e} \rho r h^{\beta} dV - \int_{S^e} q_i n_i h^{\beta} dS \tag{14.15}$$

The above equations are a set of ordinary differential equations in time. As such, the time dependence in the equations can be accounted for via a standard time-differencing approximation given by the following:

$$\theta \frac{dT}{dt}(t+\Delta t) + (1-\theta)\frac{dT}{dt} \cong \frac{T(t+\Delta t) - T(t)}{\Delta t} \quad 0 \le \theta \le 1 \tag{14.16}$$

where the choice of θ results in one of the following (Reddy 2005):

$$\theta = \left\{ \begin{array}{c} 0 - \textit{forward difference scheme} \\ 1/2 - \textit{Crank} - \textit{Nicolson scheme} \\ 2/3 - \textit{Galerkin scheme} \\ 1 - \textit{backward difference scheme} \end{array} \right\} \tag{14.17}$$

The accuracy of the approximation in time will be determined by the choice of θ and the size of the increment in time, Δt. For most problems involving road ways, a sufficiently accurate solution may be obtained by simply performing successive analyses with decreasing time increments until successive solutions are equivalent to one another to the desired degree of accuracy.

Substituting Eq. (14.16) into Eq. (14.12) will yield the following element equations for the heat transfer problem:

$$\left[\hat{K}^e\right]\{T(t+\Delta t\} = \left\{\hat{F}^e\right\} \tag{14.18}$$

where

$$\left[\hat{K}^e\right] \equiv ([M^e] + \theta \Delta t[K^e]) \tag{14.19}$$

and

$$\left\{\hat{F}^e\right\} \equiv ([M^e] - (1-\theta)\Delta t[K^e]) + \Delta t(\theta\{F^e(t+\Delta t\} + (1-\theta)\{F^e(t)\}) \tag{14.20}$$

Thus, it is possible to construct a set of algebraic equations for an arbitrary element that replace the variational form of the heat transfer model. This set of equations then must be assembled with the equations for all of the other elements within the object to form a set of *global finite element equations*, and the procedure for assembling these equations will be described in a later section. This then completes the construction of the finite element model for the heat transfer problem.

Note also that, due to the identical mathematical form of the moisture model to the heat transfer model, the same finite element code may be deployed to predict the moisture distribution within the road way, so long as the appropriate moisture properties and boundary conditions are employed within the finite element algorithm.

14.2.3 Construction of the Mechanics Finite Element Platform

Although the finite element method may be deployed for a broad range of problems in applied physics, it was inspired by the need to perform accurate stress analyses of linear elastic objects. For this reason, and also because significant portions of road ways can be modeled accurately using linear elasticity theory, this is the problem that will be addressed in detail herein for the purpose of demonstrating the mechanics finite element method to the reader. Applications to inelastic materials will be discussed briefly thereafter. For those who have already had an initial course in finite elements, this section may be skipped. For those who wish to delve further into the finite element method described within this section, we begin with a review of the quasi-static linear elastic initial boundary value problem, laid out in Table 11.1, and reproduced above as Table 14.2.

For the road way engineer, the solution to the above boundary value problem poses a significant challenge, and the crux of this challenge has to do with geometry —the geometric shape of the road way, the shape of the aggregate(s) deployed, and the shape of the loading applied to road ways by vehicles passing over it. These issues, each one complicated by geometry, are what led engineers and scientists to develop the finite element method in the twentieth century.

Although the formulation of the boundary value problem described in Table 14.2 was essentially completed by 1822 (Maugin 2014), practitioners invariably found it nearly impossible to solve the 15 equations in 15 unknowns (stresses, strains, and displacements) whenever the object in question was of complex shape. And as economics drove the industrial revolution, the shapes of structural parts became increasingly complicated. This complexity is never more apparent than in today's road ways. Thus, the finite element method is ideal for the purpose of analyzing and designing road ways. To wit, if the deflections of the road way surface can be predicted accurately in time, then failure by rutting can potentially be avoided in many cases. Furthermore, if the spatial and temporal distribution of stresses in the road way can be predicted, then perhaps even road way cracking can be substantially mitigated (to be discussed later in this chapter).

To see how the dual formulation for the mechanics problem works, it is customary to begin with the equilibrium equations in Table 14.2, given by:

$$\sigma_{ji,j} + \rho f_i = 0 \tag{14.21}$$

The variational form of Eq. (14.21) will be constructed in the next section.

14.2.4 Construction of an Incrementalized Variational Form of the Mechanics Field Equations

Because the typical road way is subjected to a cyclic loading history, it is customary to utilize the model to predict the response to a succession of *increments* in the applied loads. This necessitates the incrementalization of the governing field equations. To accomplish this task, the tractions on the surface of the road way are incrementalized as follows:

$$\Delta t_i \equiv t_i^{t+\Delta t} - t_i^t \tag{14.22}$$

where the superscripts denote the time at the start of the step, t, and the time at the end of the step, $t + \Delta t$. The increment in the applied tractions will necessarily result in an increment in the output variables, which are defined as follows:

$$\begin{aligned} \Delta u_i &\equiv u_i^{t+\Delta t} - u_i^t \\ \Delta \varepsilon_{ij} &\equiv \varepsilon_{ij}^{t+\Delta t} - \varepsilon_{ij}^t \\ \Delta \sigma_{ij} &\equiv \sigma_{ij}^{t+\Delta t} - \sigma_{ij}^t \end{aligned} \tag{14.23}$$

Equilibrium equation (14.21) is now defined to be applied at time $t + \Delta t$, thus resulting in the following:

$$\sigma_{ji,j}^{t+\Delta t} + \rho f_i^{t+\Delta t} = 0 \tag{14.24}$$

Note that on the surface of the Earth, the body force does not normally change with time, but we have nevertheless noted it as such for purposes of generality.

Equation (14.24) is now multiplied by a variation in an admissible displacement field, termed δu_i, and integrated over the volume of the finite element, V^e, as follows:

$$\delta U^e = \int_{V^e} \left(\sigma_{ji,j}^{t+\Delta t} + \rho f_i^{t+\Delta t} \right) \delta u_i dV = 0 \tag{14.25}$$

The mathematical complexities associated with the replacement of Eq. (14.24) with Eq. (14.25), thereby resulting in a dual formulation, are profound. The reader is referred to the literature on the finite element method to ascertain these complexities (Reddy 2005; Oden 2006; Oden and Reddy 2011). Suffice it to say herein that, for practical purposes, any set of dependent variables that satisfies the variational equation (14.25) also satisfies Eq. (14.24), and this is true for essentially any bounded choice of admissible displacement field, δu_i .

As will be shown below, there is a choice for δu_i that is both accurate and convenient, and this choice is almost always utilized with the finite element method. For now, suffice it to say that Eq. (14.24) is now replaced with Eq. (14.25), and this new equation becomes a replacement representation of equilibrium within the element $V^e + S^e$.

Now consider the deployment of the product rule and the divergence theorem, Eq. (9.31), in the following way:

$$\int_{V^e} \frac{\partial}{\partial x_j}\left(\sigma_{ji}^{t+\Delta t}\delta u_i\right)dV = \int_{V^e}\left(\sigma_{ji,j}^{t+\Delta t}\delta u_i + \sigma_{ji}^{t+\Delta t}\delta u_{i,j}\right)dV = \int_{S^e}\sigma_{ji}^{t+\Delta t}\delta u_i n_j dS \tag{14.26}$$

Rearranging the above and substituting into Eq. (14.25) results in the following:

$$\int_{V^e}\sigma_{ji}^{t+\Delta t}\delta u_{i,j}dV = \int_{V^e}\rho f_i^{t+\Delta t}\delta u_i dV + \int_{S^e} t_i^{t+\Delta t}\delta u_i dS \tag{14.27}$$

Now suppose that the term containing the gradient of the displacement is decomposed in the following way:

$$u_{i,j} = \varepsilon_{ij} + \omega_{ij} \tag{14.28}$$

where ε_{ij} is the strain tensor, defined by Eq. (9.40), and ω_{ij}, called *the rotation tensor*, is defined as follows:

$$\omega_{ij} \equiv \frac{1}{2}\left(u_{i,j} - u_{j,i}\right) \tag{14.29}$$

Next, recall that the stress tensor is symmetric due to conservation of angular momentum, Eq. (9.85), so that, when Eq. (14.28) is substituted into Eq. (14.27), it is found that

$$\sigma_{ji}\omega_{ij} = 0 \tag{14.30}$$

Note that the rotation tensor is antisymmetric $(\omega_{ij} = -\omega_{ji})$. It can be shown that when any second-order symmetric tensor is multiplied by a second-order antisymmetric tensor, the result is necessarily zero, so that Eq. (14.27) simplifies to the following:

$$\int_{V^e}\sigma_{ij}^{t+\Delta t}\delta\varepsilon_{ij}^{t+\Delta t}dV = \int_{V^e}\rho f_i^{t+\Delta t}\delta u_i dV + \int_{S^e} t_i^{t+\Delta t}\delta u_i dS \tag{14.31}$$

Next, substitute definitions (14.23) into Eq. (14.31), thereby resulting in the following:

$$\int_{V^e}\left(\sigma_{ij}^{t} + \Delta\sigma_{ij}\right)\delta\left(\varepsilon_{ij}^{t} + \Delta\varepsilon_{ij}\right)dV = \int_{V^e}\rho f_i^{t+\Delta t}\delta u_i dV + \int_{S^e} t_i^{t+\Delta t}\delta u_i dS \tag{14.32}$$

It is assumed that within the time-stepping scheme, the quantity ε_{ij}^{t} is known a priori, so that no variation is taken with respect to this variable, thereby simplifying Eq. (14.32) to the following:

$$\int_{V^e} \Delta\sigma_{ij}\delta\Delta\varepsilon_{ij}dV = \int_{V^e} \rho f_i^{t+\Delta t}\delta u_i dV + \int_{S^e} t_i^{t+\Delta t}\delta u_i dS - \int_{V^e} \sigma_{ij}^{t}\delta\Delta\varepsilon_{ij}dV \quad (14.33)$$

Note that the known terms have been moved to the right-hand side of the above equation.

Now suppose that the constitutive relationship can be written in the following incremental form:

$$\Delta\sigma_{ij} = C_{ijkl}^{t}\Delta\varepsilon_{kl} \quad (14.34)$$

Substituting Eq. (14.34) into Eq. (14.33) thus results in the following:

$$\int_{V^e} C_{ijkl}^{t}\Delta\varepsilon_{ij}\delta\Delta\varepsilon_{ij}dV = \int_{V^e} \rho f_i^{t+\Delta t}\delta u_i dV + \int_{S^e} t_i^{t+\Delta t}\delta u_i dS - \int_{V^e} \sigma_{ij}^{t}\delta\Delta\varepsilon_{ij}dV \quad (14.35)$$

Equation (14.35) is now a complete variational form of the equilibrium equations that is ready for deployment within the finite element method. However, for purposes of computer implementation, it is customary to write the above in a more convenient form using *Voigt notation*, defined as follows:

$$\begin{aligned} &\sigma_1 \equiv \sigma_{11}, \sigma_2 \equiv \sigma_{22}, \sigma_3 \equiv \sigma_{33}, \sigma_4 \equiv \sigma_{23}, \sigma_5 \equiv \sigma_{13}, \sigma_6 \equiv \sigma_{12} \\ &\varepsilon_1 \equiv \varepsilon_{11}, \varepsilon_2 \equiv \varepsilon_{22}, \varepsilon_3 \equiv \varepsilon_{33}, \varepsilon_4 \equiv 2\varepsilon_{23}, \varepsilon_5 \equiv 2\varepsilon_{13}, \varepsilon_6 \equiv 2\varepsilon_{12} \end{aligned} \quad (14.36)$$

such that Eq. (14.34) can be written as follows:

$$\Delta\sigma_i = C_{ij}^{t}\Delta\varepsilon_j \quad (14.37)$$

Implementation of Eqs. (14.36) and (14.37) into Eq. (14.35) results in the following matrix form of the variational principle:

$$\begin{aligned} \int_{V^e} \underbrace{[\Delta\varepsilon]}_{1X6}\underbrace{[C^t]}_{6X6}\underbrace{[\delta\Delta\varepsilon]}_{6X1} dV = &\int_{V^e} \rho \underbrace{[f^{t+\Delta t}]}_{1X3}\underbrace{[\delta\Delta u]}_{3X1} dV + \int_{S^e} \underbrace{[t^{t+\Delta t}]}_{1X3}\underbrace{[\delta\Delta u]}_{3X1} dS \\ &- \int_{V^e} \underbrace{[\sigma^t]}_{1X6}\underbrace{[\delta\Delta\varepsilon]}_{6X1} dV \end{aligned} \quad (14.38)$$

The above variational form is now ready for the finite element implementation process.

14.2.5 Construction of the Finite Element Mechanics Equations for a Single Element

In order to obtain a solution with the variational form given by Eq. (14.38), it is necessary to *assume* within a generic element the spatial dependence in the primary dependent variable in the equation, Δu_i. This is accomplished by assuming that

$$\Delta u_i(\vec{x},t) = \Delta u_i^{\alpha}(t) h_i^{\alpha}(\vec{x}) \quad \alpha = 1,\ldots,n, \text{no sum on i} \tag{14.39}$$

where n is defined to be the number of *nodes* within the element, and u_i^{α} are termed *the nodal displacements*. Note that the *shape functions,* $h_i^{\alpha}(\vec{x})$, may be identical to those deployed for the heat transfer and moisture analyses, except that insomuch as the dependent variable in the mechanics problem is vector-valued, the shape functions must also be vector-valued, thereby introducing substantial mathematical complexity into the mechanics problem.

The mathematical form of the shape functions will be determined by the specific element chosen for the analysis, and this subject will be taken up in detail in a later section. For now, it is expedient to proceed with the implementation of Eq. (14.39) into the variational form.

To accomplish this, Eq. (14.39) is first substituted into strain-displacement Eq. (9.40), thereby resulting in the following:

$$\underbrace{[\Delta\varepsilon]}_{6X1} = \underbrace{[B]}_{6X3n}\underbrace{[\Delta\bar{u}]}_{3nX1} \tag{14.40}$$

where

$$\underbrace{[\Delta\bar{u}]^T}_{1X3n} \equiv \left[\Delta u_1^1 \quad \Delta u_2^1 \quad \Delta u_3^1 \quad \Delta u_1^2 \quad \Delta u_2^2 \quad \Delta u_3^2 \quad \ldots \quad \Delta u_1^n \quad \Delta u_2^n \quad \Delta u_3^n\right] \tag{14.41}$$

and the matrix $[B]$ is given by the following:

$$\underbrace{[B]}_{6X3n} = \begin{bmatrix} \frac{\partial h_1^1}{\partial x_1} & 0 & 0 & \frac{\partial h_1^2}{\partial x_1} & 0 & 0 & \ldots & \frac{\partial h_1^n}{\partial x_1} & 0 & 0 \\ 0 & \frac{\partial h_2^1}{\partial x_2} & 0 & 0 & \frac{\partial h_2^2}{\partial x_2} & 0 & \ldots & 0 & \frac{\partial h_2^n}{\partial x_2} & 0 \\ 0 & 0 & \frac{\partial h_3^1}{\partial x_3} & 0 & 0 & \frac{\partial h_3^2}{\partial x_3} & \ldots & 0 & 0 & \frac{\partial h_3^n}{\partial x_3} \\ 0 & \frac{\partial h_2^1}{\partial x_3} & \frac{\partial h_3^1}{\partial x_2} & 0 & \frac{\partial h_2^2}{\partial x_3} & \frac{\partial h_3^2}{\partial x_2} & \ldots & 0 & \frac{\partial h_2^n}{\partial x_3} & \frac{\partial h_3^n}{\partial x_2} \\ \frac{\partial h_1^1}{\partial x_3} & 0 & \frac{\partial h_3^1}{\partial x_1} & \frac{\partial h_1^2}{\partial x_3} & 0 & \frac{\partial h_3^2}{\partial x_1} & \ldots & \frac{\partial h_1^n}{\partial x_3} & 0 & \frac{\partial h_3^n}{\partial x_1} \\ \frac{\partial h_1^1}{\partial x_2} & \frac{\partial h_2^1}{\partial x_1} & 0 & \frac{\partial h_1^2}{\partial x_2} & \frac{\partial h_2^2}{\partial x_1} & 0 & \ldots & \frac{\partial h_1^n}{\partial x_2} & \frac{\partial h_2^n}{\partial x_1} & 0 \end{bmatrix} \tag{14.42}$$

Furthermore,

$$\underbrace{[\delta\Delta u]}_{3X1} = \underbrace{[H]}_{3X3n}\underbrace{[\delta\Delta\bar{u}]}_{3nX1} \tag{14.43}$$

where

$$\underbrace{[H]}_{3X3n} = \begin{bmatrix} h_1^1 & 0 & 0 & h_1^2 & 0 & 0 & \dots & h_1^n & 0 & 0 \\ 0 & h_2^1 & 0 & 0 & h_2^2 & 0 & \dots & 0 & h_2^n & 0 \\ 0 & 0 & h_3^1 & 0 & 0 & h_3^2 & \dots & 0 & 0 & h_3^n \end{bmatrix} \tag{14.44}$$

Substituting Eqs. (14.40) and (14.43) into variational principle (14.38) and factoring out the variational terms results in the following set of element equations for a generic element:

$$\underbrace{[K^e]}_{3nX3n}\underbrace{[\Delta\bar{u}]}_{3nX1} = \underbrace{[F^e(t+\Delta t)]}_{3nX1} - \underbrace{[R^e(t)]}_{3nX1} \tag{14.45}$$

where

$$\underbrace{[K^e]}_{3nX3n} = \int_{V^e} \underbrace{[B]}_{3nX6}^{T}\underbrace{[C^t]}_{6X6}\underbrace{[B]}_{6X3n}dV \tag{14.46}$$

$$\underbrace{[F^e(t+\Delta t)]}_{1X3n}^{T} = \int_{V^e} \rho\underbrace{\left[\vec{f}(t+\Delta t)\right]}_{1X3}\underbrace{[H]}_{3X3n}dV + \int_{S^e}\underbrace{[\vec{t}(t+\Delta t)]}_{1X3}\underbrace{[H]}_{3X3n}ds \tag{14.47}$$

and

$$\underbrace{[R^e(t)]}_{1X3n}^{T} = \int_{V^e}\underbrace{[\sigma(t)]}_{1X6}^{T}\underbrace{[B]}_{6X3n}dV \tag{14.48}$$

This completes the construction of the discretized finite element equations for a single generic element within a linear elastic object under quasi-static conditions. Note that they are now a set of algebraic equations, rather than the differential equations listed in Table 14.2. Note also that due to the fact that the tangent modulus matrix, $[C^t]$ may, in fact, be nonconstant if the material is not linear elastic, the equations may be nonlinear, and this issue will be discussed in further detail below. It remains both to choose an appropriate element and to perform the assembly process, to be described in the following sections.

14.2.6 Choosing an Appropriate Element

It should be apparent that accuracy of the model demands that the assemblage of elements used to simulate the shape of the actual object to be analyzed be as geometrically similar to the object as possible. Toward this end, it is found that irregular-shaped elements are often the most propitious choice. On the other hand, from the standpoint of mathematical difficulty, it is desirable for the element to be as simple as possible. Accordingly, there is no unique answer for the choice of an element and, in fact, numerous elements have been developed for use with the finite element method over the past half century. For the purpose of demonstrating the method as simply as possible, the simplest possible choice is a three-noded triangular element (in two dimensions), as shown in Fig. 14.7. A four-noded quadrilateral element is also mathematically convenient, but is not necessarily convenient for simulating the shape of irregular-shaped objects such as aggregate.

Although higher order triangles and quadrilateral elements are possible, the mathematical complexity that comes with the deployment of these elements tends to obfuscate the purpose herein, which is demonstrative in nature. Although the road way is quite rightfully modeled as a three-dimensional object, for the purpose herein, a two-dimensional demonstration should lend clarity to the development of the element shape functions (See the discussion in Chap. 11 for an explanation regarding the reduction of a three-dimensional boundary value problem to two dimensions).

Toward this end, consider the triangular element shown in Fig. 14.8. Note that two Cartesian coordinate systems are shown: both the global coordinate system (x_1, x_2) and a local coordinate system (x_1^L, x_2^L). The reason for this is a matter of convenience, as will be shown shortly. Note also that the vertices of the triangle are termed nodes. This element is therefore called a three-node triangular element. Suppose for the moment that the displacements of the three nodes $\left(u_1^1, u_2^1, u_1^2, u_2^2, u_1^3, u_2^3\right)$ are somehow known. For this case, the value of n in

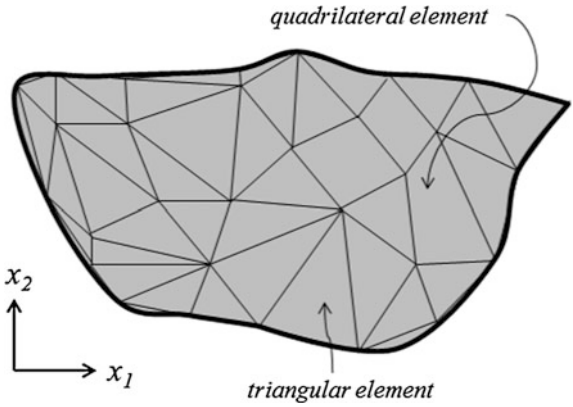


Fig. 14.7 Cartoon depicting a simplified finite element mesh deploying triangular and quadrilateral elements to represent a two-dimensional object

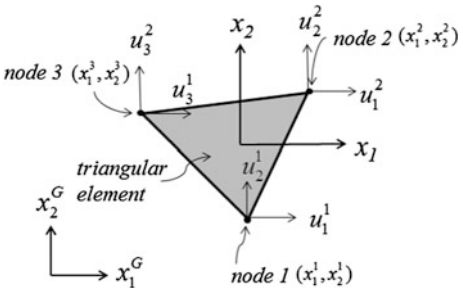


Fig. 14.8 Depiction of a two-dimensional three-node triangular element

Eq. (14.39) is 3, thereby implying that there can be only six unknown coefficients in the shape functions h_i^α. Supposing that the same shape functions are employed for both the x_1 and x_2 coordinate directions, this implies that the lowest order displacement field within the triangular element must be of the following form:

$$\begin{aligned} u_1(x_1, x_2) &= c_1 + c_2 x_1 + c_3 x_2 \\ u_2(x_1, x_2) &= c_4 + c_5 x_1 + c_6 x_2 \end{aligned} \tag{14.49}$$

Since the nodal displacements are (for the moment) assumed to be known, the above implies that:

$$\begin{aligned} u_1(x_1^1, x_2^1) &= u_1^1 = c_1 + c_2 x_1^1 + c_3 x_2^1 \\ u_1(x_1^2, x_2^2) &= u_1^2 = c_1 + c_2 x_1^2 + c_3 x_2^2 \\ u_1(x_1^3, x_2^3) &= u_1^3 = c_1 + c_2 x_1^3 + c_3 x_2^3 \\ u_2(x_1^1, x_2^1) &= u_2^1 = c_4 + c_5 x_1^1 + c_6 x_2^1 \\ u_2(x_1^2, x_2^2) &= u_2^2 = c_4 + c_5 x_1^2 + c_6 x_2^2 \\ u_2(x_1^3, x_2^3) &= u_2^3 = c_4 + c_5 x_1^3 + c_6 x_2^3 \end{aligned} \tag{14.50}$$

The above is two independent sets of three equations in the six unknowns $c_1, c_2, c_3, c_4, c_5, c_6$. Solving for these six unknowns and substituting them back into Eq. (14.49) will yield the following:

$$u_i(x_1, x_2) = u_i^\alpha h \quad \text{no sum on i} \tag{14.51}$$

where

$$\begin{aligned} h_i^1 &= \frac{1}{2A_e}\left[\left(x_1^2 x_2^3 - x_1^3 x_2^2\right) + \left(x_2^2 - x_2^3\right)x_1 + \left(x_1^3 - x_1^2\right)x_2\right] \\ h_i^2 &= \frac{1}{2A_e}\left[\left(x_1^3 x_2^1 - x_1^1 x_2^3\right) + \left(x_2^3 - x_2^1\right)x_1 + \left(x_1^1 - x_1^3\right)x_2\right] \\ h_i^3 &= \frac{1}{2A_e}\left[\left(x_1^1 x_2^2 - x_1^2 x_2^1\right) + \left(x_2^1 - x_2^2\right)x_1 + \left(x_1^2 - x_1^1\right)x_2\right] \end{aligned} \tag{14.52}$$

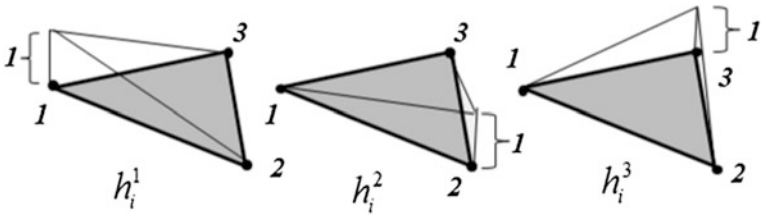


Fig. 14.9 Graphic depiction of the shape functions for a two-dimensional three-node triangular element

and A^e is the area of the element, given by

$$A^e = \left(x_1^2 x_2^3 - x_1^3 x_2^2\right) + \left(x_1^3 x_2^1 - x_1^1 x_2^3\right) + \left(x_1^1 x_2^2 - x_1^2 x_2^1\right) \tag{14.53}$$

Note that

$$h_i^\alpha(x_1^\beta, x_2^\beta) = \delta_{\alpha\beta} \tag{14.54}$$

and

$$\sum_{\alpha=1}^{3} h_i^\alpha = 1 \tag{14.55}$$

Equations (14.54) and (14.55) imply that Eq. (14.50) is satisfied. Figure 14.9 shows a graphic depiction of the three shape functions for the three-node triangular element.

While the mathematics can become quite cumbersome, the concept behind the derivation of the shape function for more advanced elements is conceptually similar to that deployed herein to obtain the shape functions for the two-dimensional three-node triangular element. For this reason, and because this subject is covered in detail in many texts on the subject of finite elements, discussion of these more advanced types of elements will not be covered herein. Suffice it to say that when the shape functions have been determined, the calculation of the terms in Eqs. (14.46)–(14.48) is a straightforward process. This then completes the development of the finite element equations (14.45) for a typical element.

14.2.7 Assembly of the Global Mechanics Finite Element Equations

The development to this point has focused on the construction of the finite element equations for a single generic element, denoted with superscripts *e*. In order to

construct the global solution, it is necessary to assemble the equations for all of the elements comprising *the finite element mesh* representing the object being modeled. This may be accomplished by summing the governing equations for all of the elements where, in view of Eq. (14.25):

$$\delta E^T = \sum_{e=1}^{N} \delta E^e = 0 \tag{14.56}$$

where N is the total number of elements, δE^T is the variation in the total mechanical energy in the object, and δE^e is the variation in the mechanical energy in each element.

The element equations derived in the previous section may now be substituted into the above to obtain the following variational form of equilibrium in the object of interest:

$$\sum_{e=1}^{N} \delta E^e = \sum_{e=1}^{N} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} K_{ij}^e u_j^e - \sum_{i=1}^{n} F_i^e \right] \delta u_i^e = 0 \tag{14.57}$$

where n is the number of nodes in each element. The correspondence between element node numbers and global node numbers may be established by constructing the so-called B*oolean connectivity matrix*, [MB] (Reddy 2005), a matrix of size *NXn* where the rows of [MB] are constructed by assigning the elements of the *ith* row to be the successive global numbers of the nodes of the *ith* element, with *i* ranging from 1 to the global number of elements (note that these should be numbered in counterclockwise order to ensure that the global stiffness matrix is positive definite). The Boolean connectivity matrix may then be utilized to establish a correspondence between the global displacement components, u_i^G, and the element displacement components, u_j^e, via the observation that the variation in the total energy in the object, δE^T, may be written in terms of global displacements as follows:

$$\delta E^T = \sum_{i=1}^{N^G} \left[\sum_{j=1}^{N^G} K_{ij}^G u_j^G - F_i^G \right] \delta u_i^G = 0 \tag{14.58}$$

where N^G is the global number of degrees of freedom.

Equating (14.57) and (14.58) will now produce an equivalence that, with the use of the Boolean connectivity matrix, will provide a means of assembling the global equations. In addition, it will be found that interelement forces cancel one another out (due to Newton's third law), and boundary conditions may be implemented to the global equations in a straightforward manner (Reddy 2005), thereby resulting in global equations of the following form:

$$\underbrace{\left[K^G\right]}_{N^G X N^G}\underbrace{\left\{u^G\right\}}_{N^G X 1} = \underbrace{\left\{F^G\right\}}_{N^G X 1} \tag{14.59}$$

As pointed out previously in this chapter, the road way problem will normally be nonlinear, in this case Eq. (14.59) will be utilized in a time-stepping procedure to predict successive *increments* in the global displacement array, $\{\Delta u^G\}$. The means for obtaining these incremental displacements in time will be discussed in the next section.

14.2.8 Accounting for Nonlinearity with Newton Iteration

The hygro-thermomechanical response of road ways is in general nonlinear. A simple means of observing this reality is to apply a load to the road way surface and measure the resulting surface deflection. Doubling the load usually will not double the displacement, thereby demonstrating the nonlinearity of the road way (see Chap. 9). From a mathematical standpoint, this nonlinearity results from the nonlinear character of the mathematical equations utilized to model the road way. Specifically, the following are mathematically nonlinear:

- Elastoplastic constitutive models
- Viscoplastic constitutive models
- Fracture models, including cohesive zone models

It will be shown in due course that one or more of the above models must be implemented into road way models in order to provide any hope of making accurate predictions of road way performance. Therefore, it can be stated that, in general, *all useful road way performance models are inherently mathematically nonlinear.*

In the case of computationally based road way models, these nonlinearities may all be accounted for by deploying a time-stepping algorithm that decomposes the load history into small increments of load in time, together with a nonlinear iterative scheme designed to obtain convergence of the nonlinear equations on each increment of time. The former of these—incrementing the loading in time—can normally be chosen by the simple process of subdividing the time step in successive computational simulations until a small enough time step is deployed such that the predicted results are unaltered.

The second part—deployment of a nonlinear iterative scheme—is the subject of this section. While many iterative techniques are possible, for the sake of simplicity only the Newton iteration method will be described herein. The authors have found this method to be sufficiently accurate for the purpose of modeling inherent nonlinearities in typical computational road way models. As mentioned above, this nonlinearity is, in the case of road ways, usually reflected within the bulk constitutive equations, or the fracture model, or both.

For the sake of simplicity, it will be assumed in this discussion that the nonlinearity is reflected via the constitutive equations, implying that Eq. (14.37) is nonlinear, that is, the modulus tensor C^t_{ijkl} is not constant in time. Such would be the case for an elastoplastic or viscoplastic material as described below. In this case, the incremental algorithm described above will necessarily lead to inherent error in the predicted increments in the state variables due to the fact that during the time increment Δt, the modulus tensor is altered from C^t_{ijkl} at the beginning of the increment to $C^{t+\Delta t}_{ijkl}$ at the end of the increment. Thus, using only the former value in the finite element algorithm will necessarily lead to an erroneous result.

It should be self-evident that the magnitude of the error induced by this approximation will depend both on the size of the time increment and the severity of the inherent nonlinearity. While there are methods for predicting acceptable time increments, this issue will be forgone herein in the interest of expediency. Suffice it to say that when the iterative algorithm fails to converge, it is likely due to having chosen a time increment that is too large.

In order to account for the error induced by the approximation represented by Eq. (14.37), it is propitious to employ the *Newton method of iteration*, so-named because it was apparently first employed by Isaac Newton in 1669. Amplifying on Newton's method, Joseph Raphson (c. 1648–c. 1715) published a more general methodology in 1690, thereby resulting in the oft-used name *Newton–Raphson method.*

As described in the previous sections, the finite element method reduces a set of partial differential equations to a set of algebraic equations. When the governing differential equations are mathematically nonlinear (see Chap. 9), the resulting finite element discretized equations will also be nonlinear, but in this case they are termed *algebraically nonlinear*. The Newton–Raphson iterative method is designed for just such a circumstance—solving sets of nonlinear algebraic equations.

To see how this may be applied to the road way problem, consider the global finite element equations, given by Eq. (14.59). Due to the nonlinearity in these equations, when the resulting displacements, strains and stresses are calculated, they will be somewhat in error. This error can be captured by constructing the following global equations:

$$\{f(t+\Delta t)\}_i \equiv \left\{F^G(t+\Delta t)\right\} - \left\{R^G(t+\Delta t)\right\}_i \tag{14.60}$$

where the subscript i denotes the ith iteration, such that

$$\left\{R^G(t+\Delta t)\right\}_i \equiv \int_V \left[(\sigma(t+\Delta t))_i\right]^T [B] dV \tag{14.61}$$

is termed the global reaction matrix. In the above, it should be noted that the predicted stresses on the ith iteration will usually be incorrect due to the nonlinearity in the equations, thereby leading to a non-null matrix $\{f\}_i$. When the correct

solution is obtained, $\{f\}$ should in fact be null. Thus, expanding $\{f\}$ in a Taylor series will result in the following:

$$\{0\} = \{f(t+\Delta t)\} = \{f(t+\Delta t)\}_i + \left[\frac{\partial f}{\partial u^G}\right]_{i+1} \{\Delta\Delta u^G\}_{i+1} + \cdots \tag{14.62}$$

where $\{\Delta\Delta u^G\}_{i+1}$ is the estimated error in the displacements due to the approximation introduced in Eq. (14.37). However, note that

$$\left[\frac{\partial f}{\partial u^G}\right]_{i+1} = -\left[K^G\right]_{i+1} \tag{14.63}$$

where $[K^G]_{i+1}$ is the global stiffness matrix on the i + 1th iteration. Thus, substituting Eq. (14.63) into Eq. (14.62) results in the following:

$$\left[K^G\right]_{i+1}\left\{\Delta\Delta u^G\right\}_{j+1} = \{f(t+\Delta t)\}_i = \left\{F^G(t+\Delta t)\right\} - \left\{R^G(t+\Delta t)\right\}_i \tag{14.64}$$

It can be seen that the above equation now presents a recursive relation for improving the estimated displacement field, implying that as follows:

$$\left\{u^G(t+\Delta t)\right\} = \left\{u^G(t)\right\} + \sum_{i=1}^{n} \left\{\Delta\Delta u^G\right\}_i \tag{14.65}$$

The solution procedure is then to repeat Eqs. (14.64) and (14.65) until the most recently calculated displacement sub-increment is negligibly small, such as given by the following:

$$\frac{\left\|\left\{\Delta\Delta u^G\right\}_n\right\|}{\left\|\sum_{i=1}^{n}\left\{\Delta\Delta u^G\right\}_i\right\|} \leq \delta_{AL} \tag{14.66}$$

thereby determining the number of iterations, n, where δ_{AL} is a small number chosen so as to ensure sufficient engineering accuracy for the problem at hand, and the symbol $\|\|$ denotes the Euclidean norm of the enclosed parameter.

Although the above procedure will nearly always converge for road way problems when the time step is sufficiently small, it does suffer from one significant shortcoming—the fact that Eq. (14.63) requires that the global stiffness must be reconstructed on each iteration. Since this step is oftentimes computationally intensive, it is possible to instead use a somewhat less accurate approximation, such that Eq. (14.63) is modified to the following:

$$[K^t]\left\{\Delta\Delta u^G\right\}_{j+1} = \{f(t+\Delta t)\}_i = \left\{F^G(t+\Delta t)\right\} - \left\{R^G(t+\Delta t)\right\}_i \tag{14.67}$$

where $[K^t]$ is the global stiffness matrix from the previous time step. This approximation will both reduce allowable size of the time step for convergence and increase

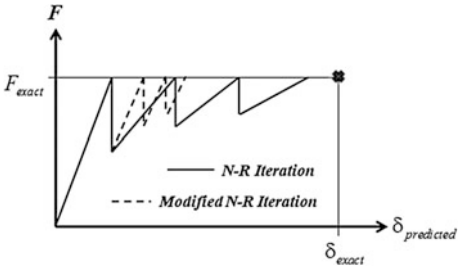


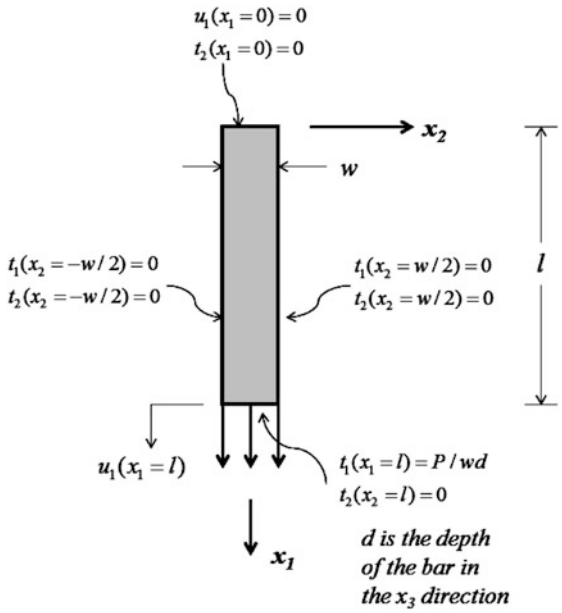
Fig. 14.10 Cartoon demonstrating the comparative convergence rates for the Newton–Raphson and modified Newton–Raphson methods

the required number of iterations to obtain acceptable accuracy, but it nonetheless will oftentimes lead to increased computational efficiency since the global stiffness matrix need not be recalculated during the iterative process. This iterative procedure is termed *the Modified Newton–Raphson method.* Figure 14.10 illustrates the difference between the Newton–Raphson and the modified Newton–Raphson method.

The following example problem can be used as a simple means of verifying the Newton–Raphson and modified Newton–Raphson algorithms within a finite element code.

Example Problem 14.1

Given: A uniaxial bar is subjected to a monotonically increasing loading $P = P_0 t$, as shown below. In addition, the bar is composed of a nonlinear elastic material with modulus $E = E_0[1 - k_0(\varepsilon_{11})^2]$.



$$E_0 = 10^5, k_0 = 10^5, d = 1, w = 1, l = 10, P_0 = 100$$

Required:

(a) Derive the incremental Newton–Raphson equations for estimating the displacement increment at the end of the bar, $\Delta u(L)$

(b) Use the result obtained in (a) to estimate the displacement $\Delta u(L)$ at time t = 0.01 with both the Newton–Raphson and the modified Newton–Raphson approaches

Solution:

(a) According to the Newton–Raphson method, the displacement increment is estimated using the following formula:

$$\Delta u_1(l)_i = \frac{\left[P(t=1) - (\sigma_{11})_i wd\right]}{K_i} \tag{14.68}$$

where the subscript implies the iteration number. In addition,

$$K_i = \frac{E_0[1 - k_0(\varepsilon_{11})_i^2]wd}{l} \tag{14.69}$$

Also

$$(\varepsilon_{11})_i = \frac{\sum_{j=1}^{i} \Delta u_j}{l} \tag{14.70}$$

$$(\sigma_{11})_i = E_0[1 - k_0(\varepsilon_{11})_i^2](\varepsilon_{11})_i \tag{14.71}$$

(b) Predicted results are shown in Table 14.3 using the above formulation, where convergence using the Newton–Raphson method is obtained in 6 iterations, whereas convergence using the modified Newton–Raphson method is obtained in 7 iterations when convergence criterion (14.66) is used with $\delta_{AL} = 5.0 \cdot 10^{-7}$. Finally, the predicted stress as a function of iteration number is shown graphically in Fig. 14.11, where it should be noted that the exact axial stress is $\sigma_{11} = 100$ and the exact axial strain is $\varepsilon_{11} = 0.001153$.

14.3 Implementation of Constitutive and Fracture Models to a Mechanics Finite Element Code

14.3.1 Implementation of Plasticity

Recall from Chap. 13 the Drucker–Prager yield criterion with combined hardening, given by the following:

Table 14.3 Predicted results for example problem 14.1

Iteration No.	N-R			M N-R		
	$u_1(l)_i$	$(\varepsilon_{11})_i$	$(\sigma_{11})_i$	$u_1(l)_i$	$(\varepsilon_{11})_i$	$(\sigma_{11})_i$
1	0.01000	0.00100	90.00000	0.01000	0.00100	90
2	0.01111	0.00111	97.39369	0.01100	0.0011	96.69
3	0.01141	0.00114	99.23610	0.01133	0.00113	98.76195
4	0.01150	0.00115	99.76875	0.01145	0.00115	99.51791
5	0.01152	0.00115	99.92932	0.01150	0.00115	99.80943
6	0.01153	0.00115	99.97833	0.01152	0.00115	99.92423
7				0.01153	0.00115	99.9698

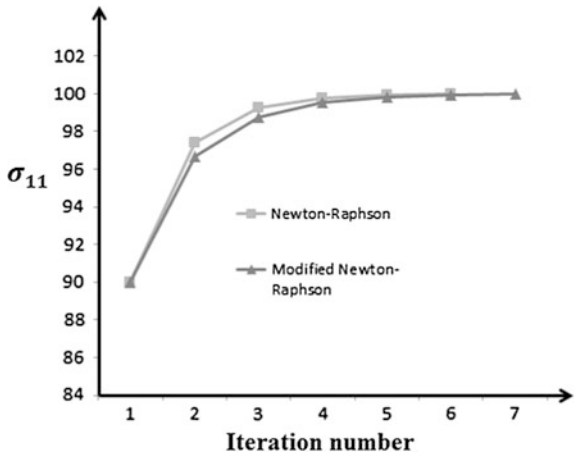


Fig. 14.11 Graphical depiction of predicted stress in example problem 14.1

$$F_{DP} = \sqrt{J_2'} + \frac{\mu^2}{\sqrt{6}} I_1' - \alpha 1_{DP} \tag{14.72}$$

where

$$J_2' \equiv \frac{1}{2}\left(\sigma_{ij}' - \alpha 2_{ij}'\right)\left(\sigma_{ij}' - \alpha 2_{ij}'\right) \tag{14.73}$$

in which

$$\alpha 2_{ij}' \equiv \alpha 2_{ij} - \frac{\alpha 2_{kk}}{3}\delta_{ij} \tag{14.74}$$

and

$$I_1' \equiv \sigma_{kk} - \alpha 2_{kk} \tag{14.75}$$

The consistency condition takes on the following form:

$$\frac{\partial F_{DP}}{\partial \sigma_{ij}} \frac{d\sigma_{ij}}{dt} + \frac{\partial F_{DP}}{\partial \alpha 2_{ij}} \frac{d\alpha 2_{ij}}{dt} + \frac{\partial F_{DP}}{\partial \alpha 1_{DP}} \frac{d\alpha 1_{DP}}{dt} = 0 \tag{14.76}$$

The evolution law for the isotropic hardening variable is given by:

$$\frac{d\alpha 1_{DP}}{dt} = \beta \left(\frac{\sigma_{ij}'}{2\sqrt{J_2}} + \frac{\mu}{\sqrt{6}} \delta_{ij} \right) \frac{d\sigma_{ij}}{dt} \tag{14.77}$$

In addition, Zeigler's rule for combined hardening is given by:

$$\frac{d\alpha 2_{ij}}{dt} = \bar{\mu}(\sigma_{ij} - \alpha 2_{ij}) \tag{14.78}$$

Substituting (14.78) into (14.76) thus results in the following:

$$\bar{\mu} = \frac{\frac{\partial F_{DP}}{\partial \sigma_{ij}} \frac{d\sigma_{ij}}{dt} + \frac{\partial F_{DP}}{\partial \alpha 1} \frac{d\alpha 1_{DP}}{dt}}{\frac{\partial F_{DP}}{\partial \sigma_{kl}} (\sigma_{kl} - \alpha 2_{kl})} \tag{14.79}$$

Recall also the normality condition, given by:

$$\frac{d\varepsilon_{ij}^P}{dt}(\sigma_{kl}) = \bar{\lambda}(\sigma_{kl}) \frac{\partial F_{DP}(\sigma_{kl})}{\partial \sigma_{ij}} \tag{14.80}$$

Substituting (14.72) into (14.80) thus results in the following:

$$\frac{d\varepsilon_{ij}^P}{dt} = \bar{\lambda} \left\{ \frac{1}{2\sqrt{J_2'}} \left(\sigma_{ij}' - \alpha_{ij}^{2'} \right) + \frac{\mu}{\sqrt{6}} \delta_{ij} \right\} \tag{14.81}$$

Squaring (14.81) results in the following:

$$\bar{\lambda} = \sqrt{3} \frac{d\bar{\epsilon}^P}{d\bar{\sigma}'} \frac{d\bar{\sigma}'}{dt} (1 + \mu^2)^{-1/2} \tag{14.82}$$

where

$$\bar{\sigma}' \equiv \left(3J_2' \right)^{1/2} \tag{14.83}$$

in which

$$\frac{d\bar{\epsilon}^p}{dt} \equiv \left(\frac{2}{3}\frac{d\varepsilon_{ij}^P}{dt}\frac{d\epsilon_{ij}^P}{dt}\right)^{1/2} \tag{14.84}$$

The above equations are adjoined with the stress–elastic strain relation, given by:

$$\sigma_{ij} = E_{ijkl}(\varepsilon_{kl} - \varepsilon_{kl}^P) \tag{14.85}$$

The above set of equations provides a complete description of the Drucker–Prager model, that is, the model consists of the following four components:

1. Stress–elastic strain relation—Eq. (14.85)
2. Yield criterion—Eq. (14.72)
3. Flow rule—Eq. (14.81)
4. Workhardening rules—Eqs. (14.77) through (14.79)

Unfortunately, the above formulation is not particularly convenient for implementation within the finite element method. This is due to the fact that the finite element method utilizes formulation that assumes that the stress–strain relation is of the general form:

$$\Delta\sigma_{ij} = C_{ijkl}\Delta\varepsilon_{kl} \tag{14.86}$$

where the symbol Δ implies a finite increment of the quantity appended, and the tensor C_{ijkl} is termed *the tangent modulus tensor*. Within the context of the finite element method, this would signify the incremental output due to an increment in the input boundary conditions. The formulation given by Eq. (14.86) normally results in the most computationally efficient method of deploying the finite element method. As such, it is propitious to reformulate the Drucker–Prager model in a way that is consistent with Eq. (14.86). The challenge is therefore to determine the exact form of the tangent modulus tensor.

In order to accomplish this restructuring, it is first necessary to write the differential form of Eq. (14.85) as follows:

$$\frac{d\sigma_{ij}}{dt} = E_{ijkl}\left(\frac{d\epsilon_{kl}}{dt} - \frac{d\epsilon_{kl}^P}{dt}\right) \tag{14.87}$$

where it should be apparent that the time derivatives may be replaced with the incremental symbol Δ. Thus, in order to cast Eq. (14.87) in a form that is consistent with Eq. (14.86), it is necessary to deal with the terms containing the time

derivative of the plastic strain tensor. This may be accomplished in the following way.

First, substitute Eq. (14.80) into Eq. (14.87) to obtain the following:

$$\frac{d\sigma_{ij}}{dt} = E_{ijkl}\left(\frac{d\varepsilon_{kl}}{dt} - \bar{\lambda}\frac{\partial F_{DP}}{\partial \sigma_{kl}}\right) \tag{14.88}$$

Substituting the above into Eq. (14.76) thus results in:

$$\frac{\partial F_{DP}}{\partial \sigma_{ij}}\left[E_{ijkl}\left(\frac{d\varepsilon_{kl}}{dt} - \bar{\lambda}\frac{\partial F_{DP}}{\partial \sigma_{kl}}\right)\right] + \frac{\partial F_{DP}}{\partial \alpha 2_{ij}}\frac{d\alpha 2_{ij}}{dt} + \frac{\partial F_{DP}}{\partial \alpha 1_{DP}}\frac{d\alpha 1_{DP}}{dt} = 0 \tag{14.89}$$

Now recall that due to the form of the Drucker–Prager yield criterion

$$\frac{\partial F_{DP}}{\partial \alpha 2_{ij}} = -\frac{\partial F_{DP}}{\partial \sigma_{ij}} \tag{14.90}$$

And, it can also be seen from Eq. (14.72) that

$$\frac{\partial F_{DP}}{\partial \alpha 1_{DP}} = -1 \tag{14.91}$$

so that Eq. (14.89) may be written as follows:

$$\frac{\partial F_{DP}}{\partial \sigma_{ij}}\left[E_{ijkl}\left(\frac{d\varepsilon_{kl}}{dt} - \bar{\lambda}\frac{\partial F_{DP}}{\partial \sigma_{kl}}\right)\right] - \frac{\partial F_{DP}}{\partial \sigma_{ij}}\frac{d\alpha 2_{ij}}{dt} - \frac{d\alpha 1_{DP}}{dt} = 0 \tag{14.92}$$

Substituting Eq. (14.80) into the third term above results in the following:

$$\frac{\partial F_{DP}}{\partial \sigma_{ij}}\left[E_{ijkl}\left(\frac{d\varepsilon_{kl}}{dt} - \bar{\lambda}\frac{\partial F_{DP}}{\partial \sigma_{kl}}\right)\right] - \frac{1}{\bar{\lambda}}\frac{d\varepsilon_{ij}^{P}}{dt}\frac{d\alpha 2_{ij}}{dt} - \frac{d\alpha 1_{DP}}{dt} = 0 \tag{14.93}$$

The above equation can be written equivalently as follows:

$$\frac{\partial F_{DP}}{\partial \sigma_{ij}}\left[E_{ijkl}\frac{d\varepsilon_{kl}}{dt}\right] = \bar{\lambda}\left[E_{ijkl}\frac{\partial F_{DP}}{\partial \sigma_{ij}}\frac{\partial F_{DP}}{\partial \sigma_{kl}} + \frac{1}{\bar{\lambda}^2}\frac{d\varepsilon_{ij}^{P}}{dt}\frac{d\alpha 2_{ij}}{dt} + \frac{1}{\bar{\lambda}}\frac{d\alpha 1_{DP}}{dt}\right] \tag{14.94}$$

where the reason for the form of the last term will become apparent below. Solving Eq. (14.94) for $\bar{\lambda}$ now results in:

$$\bar{\lambda} = \frac{\dfrac{\partial F_{DP}}{\partial \sigma_{tu}} E_{tuvw} \dfrac{d\varepsilon_{vw}}{dt}}{\dfrac{\partial F_{DP}}{\partial \sigma_{pq}} E_{pqrs} \dfrac{\partial F_{DP}}{\partial \sigma_{rs}} + \dfrac{1}{\bar{\lambda}^2} \dfrac{d\varepsilon_{ij}^P}{dt} \dfrac{d\alpha 2_{ij}}{dt} + \dfrac{1}{\bar{\lambda}} \dfrac{d\alpha 1_{DP}}{dt}} \tag{14.95}$$

Now define c such that

$$\left(\frac{d\sigma_{ij}}{dt} - c \frac{d\varepsilon_{ij}^P}{dt} \right) \frac{\partial F_{DP}}{\partial \sigma_{ij}} = 0 \tag{14.96}$$

The above, together with the normality condition and the consistency condition, implies that (Allen 1980):

$$c \frac{\partial F_{DP}}{\partial \sigma_{ij}} \frac{\partial F_{DP}}{\partial \sigma_{ij}} = \frac{1}{\bar{\lambda}} \frac{\partial F_{DP}}{\partial \sigma_{ij}} \frac{d\sigma_{ij}}{dt} = \frac{1}{\bar{\lambda}^2} \frac{d\varepsilon_{ij}^P}{dt} \frac{d\alpha 2_{ij}}{dt} + \frac{1}{\bar{\lambda}} \frac{d\alpha 1_{DP}}{dt} \tag{14.97}$$

Substituting the above into Eq. (14.95) now results in the following:

$$\bar{\lambda} = \frac{\dfrac{\partial F_{DP}}{\partial \sigma_{tu}} E_{tuvw} \dfrac{d\varepsilon_{vw}}{dt}}{\dfrac{\partial F_{DP}}{\partial \sigma_{pq}} E_{pqrs} \dfrac{\partial F_{DP}}{\partial \sigma_{rs}} + c \dfrac{\partial F_{DP}}{\partial \sigma_{pq}} \dfrac{\partial F_{DP}}{\partial \sigma_{pq}}} \tag{14.98}$$

Substituting Eq. (14.98) into Eq. (14.88) now gives:

$$\frac{d\sigma_{ij}}{dt} = C_{ijkl} \frac{d\varepsilon_{kl}}{dt} \tag{14.99}$$

where, by inspection

$$C_{ijkl} = E_{ijkl} - \frac{E_{ijmn} \dfrac{\partial F_{DP}}{\partial \sigma_{mn}} E_{tukl} \dfrac{\partial F_{DP}}{\partial \sigma_{tu}}}{c \dfrac{\partial F_{DP}}{\partial \sigma_{pq}} \dfrac{\partial F_{DP}}{\partial \sigma_{pq}} + \dfrac{\partial F_{DP}}{\partial \sigma_{pq}} E_{pqrs} \dfrac{\partial F_{DP}}{\partial \sigma_{rs}}} \tag{14.100}$$

It can be seen that the incremental constitutive equation is now in the desired form given by Eq. (14.86).

In order to determine the unknown coefficient c in the above equation, Eq. (14.80) is substituted into Eq. (14.96), thereby resulting in the following:

$$c = \frac{d\sigma_{ij} d\varepsilon_{ij}^P}{d\varepsilon_{kl}^P d\varepsilon_{kl}^P} \tag{14.101}$$

Since the material is elastically isotropic, during a uniaxial test ($\sigma_{11} \neq 0$, $\forall\, other\ \sigma_{ij} = 0$) the above simplifies to the following:

$$c = \frac{\frac{d\sigma_{11}}{dt}\left(\frac{d\varepsilon_{11}}{dt} - \frac{1}{E}\frac{d\sigma_{11}}{dt}\right)}{\left(\frac{d\varepsilon_{11}}{dt} - \frac{1}{E}\frac{d\sigma_{11}}{dt}\right)^2 + 2\left(\frac{d\varepsilon_{22}}{dt} + \frac{\nu}{E}\frac{d\sigma_{11}}{dt}\right)^2} \tag{14.102}$$

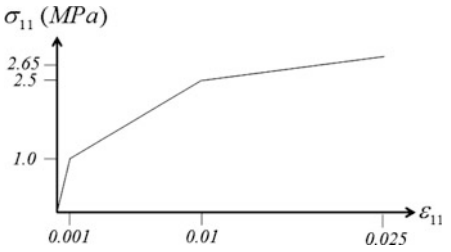
Alternatively, c may be obtained from a hydrostatic compression test ($\sigma_{11} = \sigma_{22} = \sigma_{33} = -p$), in which:

$$c = \frac{-1}{3}\frac{dp}{dt}\left(\frac{1}{3}\frac{d\varepsilon_{kk}}{dt} + 3E\frac{dp}{dt}\right)^{-1} \tag{14.103}$$

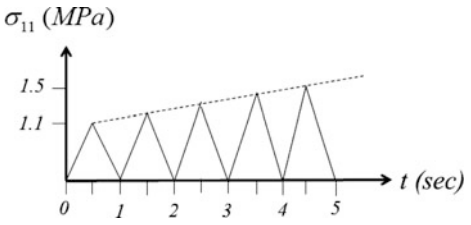
The above incrementalized equations may be supplemented with the radial return algorithm (Lubliner 1990) for the purpose of transitioning from elastic to elastoplastic response.

Example Problem 14.2

Given: A uniaxial bar is known to have material properties as follows: $E = 10^3$ MPa, $\nu = 0.3$, $\beta = 1.0$, $\mu = 0.0$, $\sigma_Y = 1.0$ MPa. The uniaxial stress–strain curve is as follows:



Required: Use the Drucker–Prager model to predict the response of the material to the following cyclic loading history.



Solution: When the plasticity routine is functioning properly, the output of a cyclic uniaxial simulation should appear like that shown in Fig. 14.12.

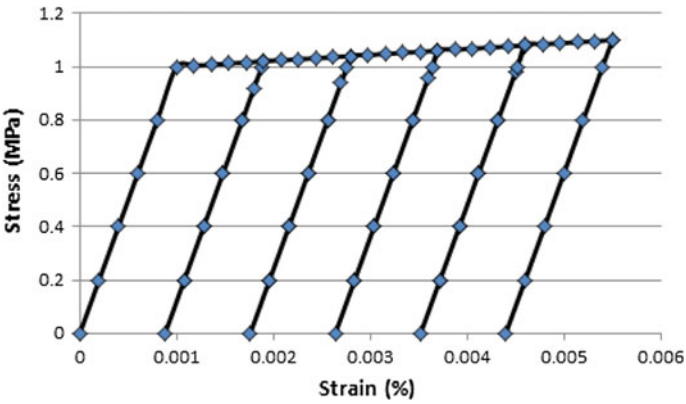


Fig. 14.12 Depiction of the predicted response of an elastoplastic uniaxial bar subjected to cyclic loading described in example problem 14.2

14.3.2 Implementation of Viscoelasticity

In order to utilize the finite element method to solve linear viscoelastic IBVPs, it is necessary to convert the integral formulation of the constitutive equations to an incremental form that may be deployed within a time-marching scheme. Accordingly, an incremental formulation will be constructed from the convolution integral representation given by Eq. 12.38. The formulation contained herein follows that reported in Zocher et al. (1997). For simplicity, only the isotropic case is presented here. As a further simplification, it will be assumed herein that Poisson's ratio is constant in time. The reader interested in the formulation for more generally anisotropic media is referred to the above reference.

Consider the isotropic form of the stress formulation for linear viscoelastic media, given by the following:

$$\sigma_i = f_{ij} \int_0^t E(t-\tau) \frac{\partial \varepsilon_j}{\partial \tau} d\tau \tag{14.104}$$

where for convenience single-subscripted notation has been employed for the stress, σ_i, and strain, ε_j and the range on 1 and j, as well as the values of the coefficients of $f_{ij} = f_{ij}(\nu)$ will depend on the dimensionality of the problem. For example, for the three-dimensional case

$$[f_{ij}] \equiv \frac{1}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{(1-2\nu)}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{(1-2\nu)}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{(1-2\nu)}{2} \end{bmatrix} \tag{14.105}$$

Now suppose that the stresses and strain are known at time t_k. Suppose furthermore that Eq. (14.104) is to be integrated for the purpose of calculating $\sigma_i(t_k + \Delta t)$ for a generic increment of time, Δt, given $\varepsilon_j(t)$, $t \le t_k + \Delta t$. Applying Eq. (14.104) for this circumstance results in the following:

$$\begin{aligned} \sigma_i(t_k + \Delta t) &= f_{ij} \int_0^{t_k + \Delta t} E(t_k + \Delta t - \tau) \frac{\partial \varepsilon_j}{\partial \tau} d\tau \\ &= f_{ij} \int_0^{t_k} E(t_k + \Delta t - \tau) \frac{\partial \varepsilon_j}{\partial \tau} d\tau + f_{ij} \int_{t_k}^{t_k + \Delta t} E(t_k + \Delta t - \tau) \frac{\partial \varepsilon_j}{\partial \tau} d\tau \end{aligned} \tag{14.106}$$

Now define the following:

$$\begin{aligned} \Delta E \equiv E(t_k + \Delta t - \tau) - E(t_k - \tau) \Rightarrow \\ E(t_k + \Delta t - \tau) = E(t_k - \tau) + \Delta E \end{aligned} \tag{14.107}$$

Substituting Eq. (14.107) into Eq. (14.106) now results in the following:

$$\begin{aligned} \sigma_i(t_k + \Delta t) &= f_{ij} \int_0^{t_k} E(t_k - \tau) \frac{\partial \varepsilon_j}{\partial \tau} d\tau \\ &+ f_{ij} \int_0^{t_k} \Delta E \frac{\partial \varepsilon_j}{\partial \tau} d\tau + f_{ij} \int_{t_k}^{t_k + \Delta t} E(t_k + \Delta t - \tau) \frac{\partial \varepsilon_j}{\partial \tau} d\tau \end{aligned} \tag{14.108}$$

The first term on the right-hand side of Eq. (14.108) can be seen to be the known state of stress at the beginning of the time increment, $\sigma_i(t_k)$, so that Eq. (14.108) may be rewritten as follows:

$$\Delta\sigma_i \equiv \sigma_i(t_k + \Delta t) - \sigma_i(t_k) = f_{ij} \int_{t_k}^{t_k + \Delta t} E(t_k + \Delta t - \tau) \frac{\partial \varepsilon_j}{\partial \tau} d\tau + \Delta\sigma_i^R \tag{14.109}$$

where the residual stress term is defined as follows:

$$\Delta\sigma_i^R \equiv f_{ij} \int_0^{t_k} \Delta E \frac{\partial \varepsilon_j}{\partial \tau} d\tau \tag{14.110}$$

The above provides a means of predicting the stress increment for a given increment of strain during the time increment, Δt. For example, consider the case wherein the relaxation modulus, $E(t)$, is fit with a Prony series (with M terms) of the following form:

$$E(t) = E_\infty + \sum_{m=1}^{M} E_m e^{-\frac{t}{\tau_m}} \tag{14.111}$$

Suppose also that it can assumed that during the time increment of interest the components of the rate of strain, $\dot{\varepsilon}_j \equiv \varepsilon_j^0$, may be assumed to be approximately constant, that is:

$$\begin{aligned} &\varepsilon_j(t) = \varepsilon_j(t_k) + \varepsilon_j^0 \cdot (\tau - t_k) H(\tau - t_k) \Rightarrow \\ &\frac{\partial \varepsilon_j}{\partial \tau} = \varepsilon_j^0 H(\tau - t_k) + \varepsilon_j^0 \cdot (\tau - t_k)\delta(\tau - t_k) \quad t_k \le \tau \le t_k + \Delta t \end{aligned} \tag{14.112}$$

Substituting Eqs. (14.111) and (14.112) into Eq. (14.109) in closed form therefore results in the following:

$$\Delta\sigma_i = f_{ij} E' \Delta\varepsilon_j + \Delta\sigma_i^R \tag{14.113}$$

where it should be apparent that $\Delta\varepsilon_j = \varepsilon_j^0 \Delta t$, and

$$E' = E_\infty + \frac{1}{\Delta t} \sum_{m=1}^{M} E_m \tau_m (1 - e^{-\frac{\Delta t}{\tau_m}}) \tag{14.114}$$

Now, consider the incremental residual stress term, $\Delta\sigma_i^R$, defined in Eq. (14.110). Substitution of Eq. (14.111) into Eq. (14.107) and this result into Eq. (14.110) will result in the following:

$$\Delta\sigma_i^R = -f_{ij} \sum_{m=1}^{M} (1 - C_1^m) S_j^m \tag{14.115}$$

where

$$C_1^m \equiv e^{-\frac{\Delta t}{\tau_m}} \tag{14.116}$$

and

$$S_j^m \equiv \int_0^{t_k} E_m e^{-\frac{(t_k-\tau)}{\tau_m}} \frac{\partial \varepsilon_j}{\partial \tau} d\tau \tag{14.117}$$

Equation (14.117) may be replaced by a recursive formula by first writing Eq. (14.117) in the following equivalent form:

$$S_j^m = \int_0^{t_k-\Delta t} E_m e^{-(t_k-\tau)/\tau_m} \frac{\partial \varepsilon_j}{\partial \tau} d\tau + \int_{t_k-\Delta t}^{t_k} E_m e^{-(t_k-\tau)/\tau_m} \frac{\partial \varepsilon_j}{\partial \tau} d\tau \tag{14.118}$$

The above may be written in the following recursive form (Zocher et al. 1997):

$$S_j^m(t_k) = C_1^m S_j^m(t_k - \Delta t) + \Delta S_j^m \tag{14.119}$$

where

$$\Delta S_j^m \equiv \int_{t_k-\Delta t}^{t_k} E_m e^{-(t_k-\tau)/\tau_m} \frac{\partial \varepsilon_j}{\partial \tau} d\tau \tag{14.120}$$

Integrating Eq. (14.120) will result in the following:

$$\Delta S_j^m = C_2^m \varepsilon_j^0(t_k - \Delta t) \tag{14.121}$$

where $\varepsilon_j^0(t_k - \Delta t)$ is the rate of strain for the previous time step, and

$$C_2^m \equiv E_m \tau_m (1 - e^{-\frac{\Delta t}{\tau_m}}) \tag{14.122}$$

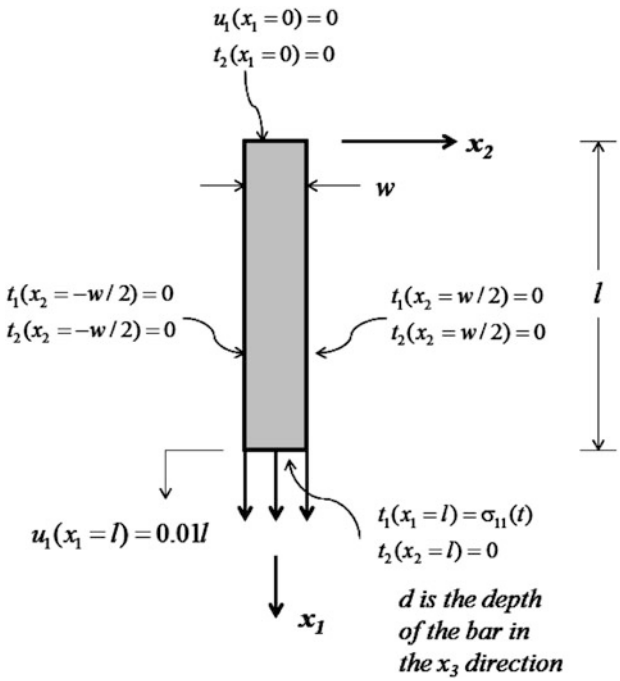
Equation (14.113) may now be utilized within the time-stepping finite element algorithm to increment the stresses for an isotropic linear viscoelastic material.

Example Problem 14.3

Given: A linear viscoelastic uniaxial bar is subjected to a uniaxial relaxation test such that $\varepsilon(t) = 0.01H(t)$, as shown below. In addition, the material behavior of the bar is modeled with a standard linear solid, given by the following:

$$E(t) = E_\infty + E_1 e^{-t/\tau_1}$$

where $E_\infty = 10^5$ Pa, $E_1 = 10^5$ Pa, $\tau_1 = 10$ s



Required:

(a) Construct a one-dimensional incremental algorithm for the purpose of determining the predicted stress, $\sigma_{11}^P = \sigma_{11}^P(t)$
(b) Use the algorithm to calculate the predicted stress up to the time t = 60 s.
(c) Compare the predicted stress, $\sigma_{11}^P(t)$, to the exact stress, $\sigma_{11}^E(t)$.

Solution:

(a) Reduction in the incremental algorithm to uniaxial form results in the following recursive formulation:

$$
\begin{aligned}
&\textit{Initialize}\\
&C_0 = e^{-\Delta t/\tau_1}\\
&E' = E_\infty + \tfrac{1}{\Delta t}[E_1\tau_1(1 - C_0)]\\
&C_1 = 1 - C_0\\
&C_2 = E_1\tau_1 C_1\\
&S = 0\\
&\textit{Increment}\\
&t = t + \Delta t\\
&\Delta\sigma_{11}^R = -C_1 S(t - \Delta t)\\
&\Delta\sigma_{11}^P = E'\Delta\varepsilon_{11} + \Delta\sigma_{11}^R\\
&\sigma_{11}^P(t + \Delta t) = \sigma_{11}^P(t) + \Delta\sigma_{11}^P\\
&\Delta S = C_2\dot{\varepsilon}(t - \Delta t)\\
&S(t) = C_0 S(t - \Delta t) + \Delta S
\end{aligned}
$$

Table 14.4 Predicted results for example problem 6.3

time (s)	σ_{11}^{P}	σ_{11}^{P}	σ_{11}^{E}
	$\Delta t = 12$ s	$\Delta t = 6$ s	
0	2000	2000	2000
6		1752	1548.8
12	1582.3	1412.7	1301.2
18		1226.5	1165.3
24	1175.4	1124.3	1090.7
30		1068.2	1049.8
36	1052.8	1037.4	1027.3
42		1020.5	1015
48	1015.9	1011.3	1008.2
54		1006.2	1004.5
60	1004.8	1003.4	1002.5

(b) The above incremental equations may be used to predict the stress versus time for an arbitrary value of Δt, where it should be noted that after the initial increment of time $\Delta\varepsilon_{11} = \dot{\varepsilon} = 0$. Results for two different values of Δt are shown in Table 14.4.

(c) The exact solution is obtained by integrating the convolution integral as follows:

$$\sigma_{11}^{E}(t) = \int_0^t E(t-\tau)\frac{\partial(0.01H(\tau)}{\partial\tau}d\tau$$
$$= \int_0^t E(t-\tau)0.01\delta(\tau)d\tau$$

Applying the sifting property, Eq. (9.28) thus gives the following:

$$\sigma_{11}^{E}(t) = E(t-0) = E_\infty + E_1 e^{-t/\tau_1}$$

Predicted results using the incremental algorithm are compared with the exact result in Table 14.4, wherein it can be seen that the incremental algorithm converges rapidly as the time step, Δt, is decreased. This is demonstrated graphically in Fig. 14.13.

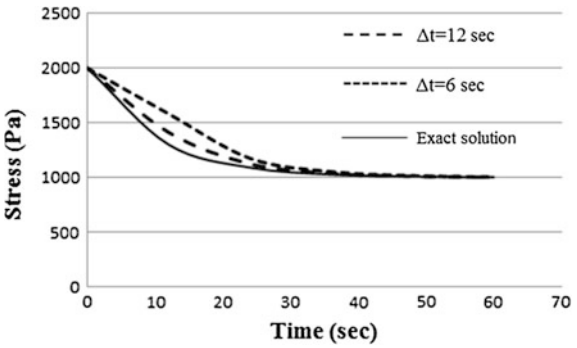


Fig. 14.13 Graphical comparison of predicted stress as a function of time step

14.3.3 Implementation of a Cohesive Zone Model

The implementation of a cohesive zone model within the finite element framework will allow for the prediction of cracking on any length scale desired. It is necessary only to incrementalize the cohesive zone model to be deployed within the FE code. Toward this end, the computational implementation of the nonlinear viscoelastic cohesive zone model introduced in Chap. 13 will be described herein as a generic introduction to this subject.

The technique for integrating the previously introduced viscoelastic cohesive zone model is essentially the same as that utilized to integrate the viscoelasticity model within the bulk material described in Chap. 12. This can be accomplished by first referring to Eqs. (13.147) and (13.150), reproduced here for simplicity as follows:

$$t_i(t) = \frac{1}{\lambda^C}\frac{u_i^C}{u_i^*}\left(1-\alpha^C(t)\right)\int_0^t E^C(t-\tau)\frac{\partial \lambda^C}{\partial \tau}d\tau \tag{14.123}$$

$$\begin{aligned} \frac{d\alpha^C}{dt} &= \alpha_1^C\left(\dot{\lambda}^C\right)^{n^C} \quad \dot{\lambda}^C \geq 0 \; and \; \alpha^C \leq 1 \\ \frac{d\alpha^C}{dt} &= 0 \qquad\qquad\quad \dot{\lambda}^C \leq 0 \; or \; \alpha^C = 1 \end{aligned} \tag{14.124}$$

The above equations may be incremented in time (Allen and Searcy 2000). To see how this is accomplished, note that, given the variable at time t, it is desired to predict the variables at time $t+\Delta t$. Therefore, first define the increment in the viscoelastic relaxation modulus as follows:

$$\begin{aligned} &\Delta E^C \equiv E^C(t+\Delta t-\tau) - E^C(t-\tau) \Rightarrow \\ &E^C(t+\Delta t-\tau) = E^C(t-\tau) + \Delta E^C \end{aligned} \tag{14.125}$$

Furthermore, let

$$\Delta\bar{\lambda}^C \equiv \bar{\lambda}^C(t+\Delta t) - \bar{\lambda}^C(t) \Rightarrow \bar{\lambda}^C(t+\Delta t) = \bar{\lambda}^C(t) + \Delta\bar{\lambda}^C \tag{14.126}$$

$$\Delta u_i^C \equiv u_i^C(t+\Delta t) - u_i^C(t) \Rightarrow u_i^C(t+\Delta t) = u_i^C(t) + \Delta u_i^C \tag{14.127}$$

$$\Delta\alpha^C \equiv \alpha^C(t+\Delta t) - \alpha^C(t) \Rightarrow \alpha^C(t+\Delta t) = \alpha^C(t) + \Delta\alpha^C \tag{14.128}$$

In order to accommodate the integration of the integral equation over the time increment, it is assumed that the crack opening displacement rate is constant during the time increment, that is:

$$\bar{\lambda}^C(\tau) = \bar{\lambda}^C(t) + \dot{\bar{\lambda}}_0^C(\tau - t)H(\tau - t) \quad t \le \tau \le t + \Delta t \tag{14.129}$$

where $\dot{\bar{\lambda}}_0^C$ is the constant displacement rate. In order to integrate Eq. (14.124) in time, it is assumed that the time derivative may be approximated as follows:

$$\frac{d\alpha^C}{dt} \cong \alpha_1^C \left[\frac{\bar{\lambda}^C(t) + \bar{\lambda}^C(t+\Delta t)}{2}\right]^{n^C} \tag{14.130}$$

Thus, for small time steps

$$\alpha^C(t+\Delta t) \cong \alpha^C(t) + \frac{d\alpha^C}{dt} \cdot \Delta t \tag{14.131}$$

The above approximations allow the formulation of the following recursive time-stepping algorithm:

$$\Delta t_i^C = \frac{u_i^C(t) + \Delta u_i^C}{\delta_i^C} \frac{[1 - \alpha^C(t+\Delta t)]}{\bar{\lambda}^C + \Delta\bar{\lambda}^C} \cdot E^C(\Delta t) \cdot \dot{\bar{\lambda}}_0^C \Delta t + \left(\Delta t_i^C\right)^R \tag{14.132}$$

where

$$\begin{aligned}\left(\Delta t_i^C\right)^R = &\left\{\frac{\left[u_i^C(t) + \Delta u_i^C\right][1 - \alpha^C(t+\Delta t)]}{\delta_i^C\left[\bar{\lambda}^C(t) + \Delta\bar{\lambda}^C\right]}\right\} \cdot \left[E_\infty^C \bar{\lambda}^C(t) + \sum_{j=1}^{p^C} \sigma_j^C(t)\right] \\ &- \left\{\frac{u_i^C(t)[1 - \alpha^C(t)]}{\delta_i^C\left[\bar{\lambda}^C(t) + \Delta\bar{\lambda}^C\right]}\right\} \cdot \left[E_\infty^C \bar{\lambda}^C(t) + \sum_{j=1}^{p^C} \sigma_j^C(t)\right] \\ &+ \left\{\frac{\left[u_i^C(t) + \Delta u_i^C\right][[1 - \alpha^C(t+\Delta t)]]}{\delta_i^C\left[\bar{\lambda}^C(t) + \Delta\bar{\lambda}^C\right]}\right\} \cdot \left[-\sum_{j=1}^{p^C}\left[1 - e^{-E_j^C \Delta t/\eta_j^C}\right]\sigma_j^C(t)\right]\end{aligned} \tag{14.133}$$

where it is assumed that the cohesive zone relaxation modulus is modeled with a Prony series containing p^C terms, thereby resulting in

$$E^C(\Delta t) \equiv E_\infty^C + \frac{1}{\Delta t}\sum_{j=1}^{p^C} \eta_j^C \left(1 - e^{-E_j^C \Delta t/\eta_j^C}\right) \tag{14.134}$$

and

$$\sigma_j^C(t) = e^{-E_j^C \Delta t/\eta_j^C} \cdot \left[\sigma_j^C(t - \Delta t)\right] + \eta_j^C \dot{\bar{\lambda}}^C \left[1 - e^{-E_j^C \Delta t/\eta_j^C}\right] \tag{14.135}$$

The above completes the development of a recursive relation for incrementing the cohesive zone tractions in time. They are then implemented to the FE algorithm by placing equal and opposite tractions according to the above equations on the faces of the cohesive zones, wherever they are to be implemented within the algorithm.

This then may be utilized to predict the evolution of cracks within the road way, whereby the development of damage within the cohesive zones will result in the cohesive zone tractions going to zero where the cohesive zones fail, in the process producing predicted crack extension within the road way.

14.4 Summary

In this chapter, a brief overview has been given of the computational methods required to perform finite element analyses of asphalt mixtures and asphaltic road ways. It has been shown that the finite element method may be deployed to predict both temperature and moisture within the road way as functions of time and spatial coordinates, and these may then be utilized, together with input loads on the road way, to predict the mechanical response, including stresses, strains, and displacements of the road way.

Newton iteration has been discussed as a means of accounting for nonlinearities due to plasticity, viscoelasticity, viscoplasticity, and evolving cracks within the road way. The finite element method has then been utilized to demonstrate the power of computational mechanics in predicting the performance of road ways as a function of global, local, and even microscale input variables, including tire pressure distribution, fillers, fines, aggregate, base material and compaction, and even environmental effects.

As tools for computational analysis, both micromechanics and multi-scaling have been discussed in some detail, with emphasis placed on how designers can account for the effects of controllable parameters such as mix volume fractions (see Chap. 7) and pavement thickness to improve road way performance (see Chap. 8), and particular attention has been paid to how these parameters can affect road way

rutting and cracking (see Chaps. 1 and 8). The tools discussed in this chapter will therefore serve the designer well in improving performance of road ways of the future.

14.5 Problems

Problem 14.1
Given: The governing equations for the temperature and moisture field problems
Required:

(a) Describe a one-to-one relation between the input loads, geometry, and material properties for these two field problems
(b) Describe a one-to-one relation between the output variables for these two field problems.

Problem 14.2
Given: A material is known to behave uniaxially according to the equation $\sigma = E\varepsilon$ where $E \equiv E_0 \cos(\varepsilon/\varepsilon_0)$, $E_0 = 10^5$, $\varepsilon_0 = \pi/0.02$
Required: Predict $\sigma = \sigma(\varepsilon)$ on the range $0 \leq \varepsilon \leq 0.01$ using Newton iteration and compare to the exact solution.

Problem 14.3
Given: The algorithm for integrating linear viscoelastic constitutive equations described in Sect. 14.3.2.
Required:

(a) Reduce the algorithm to a one-dimensional form
(b) Use the algorithm constructed in part (a) to model a ramp test for a material point subjected to a constant strain rate of $\dot{\varepsilon} = 0.0001$, where the material may be modeled with a standard linear solid with material properties $E_\infty = 10^5$ Pa, $E_1 = 10^5$ Pa, $\tau_1 = 10$ s.
(c) By comparing your predictions to the exact answer, determine the minimum time step that will produce identical results to three significant digits up to the time t = 10 s.

References

Allen, D. (1980) A note on the combined isotropic-kinematic work hardening rule. *International Journal for Numerical Methods in Engineering*, 15, 1724–1728.
ADINA System. (2011). Encyclopedia of mathematics.
Allen, D., & Searcy, (2000). Numerical aspects of a micromechanical model for a cohesive zone. *Journal of Reinforced Plastics Composites, 19,* 240.

Computers and Engineering Software and Consulting. (2015). SAP2000 v11 Manuals.
Dassault Systèmes. (2011). ABAQUS 6.9 User Documentation, SIMULIA.
Lubliner, J. (1990). *Plasticity theory*. MacMillan.
Maugin, G. (2014). *Continuum mechanics through the eighteenth and nineteenth centuries*. Springer.
McNeal-Schwendler Corporation. (2014). MSC Nastran Release Overview, MSC.
Oden, J. (2006). *Finite elements of nonlinear continua*. Dover.
Oden, J., & Reddy, J. (2011). *An introduction to the mathematical theory of finite elements*. Dover.
Reddy, J. (2005). *An introduction to the finite element method* (3rd ed.). McGraw-Hill.
Zienkiewicz, O., & Taylor, R. (2013). The *finite element method* (7th ed.). Elsevier.
Zocher, M., Allen, D., & Groves, S. (1997). A three dimensional finite element formulation for thermoviscoelastic orthotropic media. *International Journal for Numerical Methods in Engineering, 40,* 2267.

Chapter 15
Computational Modeling Applications

15.1 Introduction

In the previous chapter, numerical methods were introduced for predicting the response of asphalt mixtures and flexible pavements. In this chapter, the power of computational methods will be demonstrated via a number of example problems, and these methods will be utilized to make computational predictions of pavement performance. Much of what appears in this chapter was previously published by the first two authors and co-workers, and is reprinted herein by permission of the publisher Taylor & Francis Ltd. (Allen et al. 2017a, b, c).

15.2 Computational Techniques for Road way Design and Analysis Using the Finite Element Method

15.2.1 Computational Micromechanics

As discussed in Chap. 11, it is oftentimes impractical to perform a detailed analysis of an infrastructural component that accounts for material inhomogeneities within the structure. This may be due to cost and/or time limitations, but oftentimes the enormous number of inhomogeneities can render a detailed analysis that accounts for each inclusion computationally impractical even with a supercomputer. For example, consider a practical rule of thumb—each inclusion requires on the order of one hundred constant strain triangular elements in order to obtain accurate predictions of the distributions of stresses and strains within the inclusion. Thus, a structural component such as that shown in Fig. 15.1 will necessarily require on the order of hundreds of thousands of elements (or more!) to accurately account for the material inhomogeneities within the component. It should therefore be apparent that such analyses are oftentimes computationally impractical.

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0_15

Fig. 15.1 A typical specimen of asphalt concrete

Alternatively, it is oftentimes necessary to perform multiple analyses on the selfsame structural component, each on a different length scale. For example, when focusing on the global length scale (the scale of the structural component), an analysis may be performed wherein the component is assumed to be statistically homogeneous, as described in Chap. 11. For purposes of this analysis, it may be necessary to perform a micromechanics analysis in order to predict the homogenized material properties to be deployed in the global scale analysis, as also described in Chap. 11. The resulting homogenized properties may then be utilized to perform a global scale analysis as a means of designing the structural component.

In this problem, the micromechanics approach is first employed to obtain the homogenized material properties (see Chap. 11), and for this purpose, the micromechanical analysis may be performed computationally using the finite element method described in the previous section. Toward this end, consider the finite element meshes shown in Fig. 15.2 for three different volume fractions of aggregate, wherein the aggregate and binder are treated as separate materials, but the material properties in each finite element are nonetheless spatially homogeneous. Although the meshes shown are two-dimensional, research suggests that at least for the case of macroscopically isotropic materials, a two-dimensional analysis is sufficient for this purpose (Rodin 1996).

In order to predict macroscopic properties that are representative of the macroscopically homogenized material, a minimum size of the sample is required. Research suggests that the minimum size in the absence of microcracking depends on the shape, spacing, angularity, orientation distribution, and volume fraction of the aggregate (Helms et al. 1999). For typical asphalt mixes, the minimum number of aggregate grains is normally around 20–30, thereby guiding the meshes shown in Fig. 15.2. While more refined meshes may be deployed, they will normally require more computational time without producing significantly improved estimates of the homogenized properties.

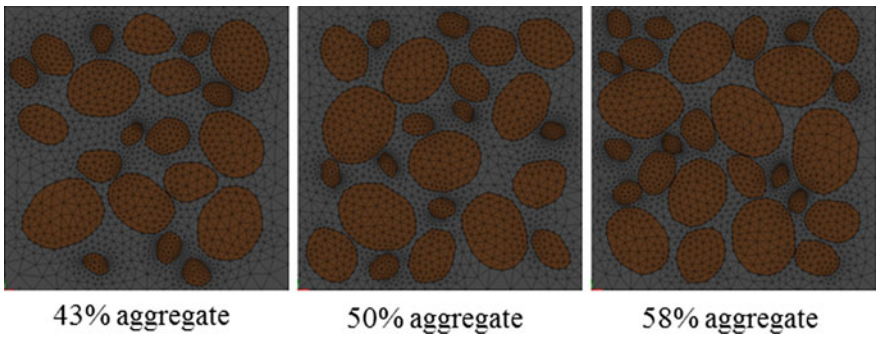


Fig. 15.2 Finite element meshes for three different volume fractions of aggregate imbedded within a cementitious binder

In order to ascertain whether this is the case, the analysis is performed on the mesh, and the higher order terms in Eq. (11.60) may be checked. If these terms are small compared to the average state variables within the mesh, then the predictions made with the mesh will produce macroscopically averaged properties that are representative of the homogenized material, and the mesh is called a *representative volume element* (RVE). Alternatively, one can perform analyses with successively more refined finite element meshes and calculate the predicted homogenized properties obtained with each analysis. When the predicted properties converge to the desired degree of accuracy, the mesh is sufficiently refined.

The RVE is then subjected to spatially homogeneous tractions (or displacements) on the boundary, and the resulting stresses and strains within the object are volume averaged using Eqs. (11.54) and (11.59) to obtain the average material properties for the global scale. When the component is macroscopically isotropic, this process will be quite simple. However, should either of the constituents be orthotropic, or should the inclusions take on some preferred orientation, then the resulting global scale properties will be macroscopically anisotropic to some degree, and the number of analyses necessary to characterize the macroscopic properties can be cumbersome. Fortunately, in road way materials the macroscopic response may usually be assumed to be isotropic, and in this case a single analysis is sufficient to obtain the macroscopic material properties.

Assuming the material may be modeled as macroscopically isotropic, the homogenized properties may be obtained from a single uniaxial simulation in which the RVE is subjected to loading in any one direction, but is free to contract normal to the loading direction. For example, if both constituents are linear elastic, the results of the simulation can be used to calculate the average stress, $\bar{\sigma}_{11}$, in the loading direction (x_1) using Eq. (11.54), and the average strain in the loading ($\bar{\varepsilon}_{11}$) and transverse directions ($\bar{\varepsilon}_{22}$) using Eq. (11.59). These results may then be utilized to predict the homogenized Young's modulus and Poisson's ratio as follows:

$$\overline{E} = \frac{\bar{\sigma}_{11}}{\bar{\varepsilon}_{11}} \tag{15.1}$$

and

$$\bar{\nu} = -\frac{\bar{\varepsilon}_{22}}{\bar{\varepsilon}_{11}} \tag{15.2}$$

The above may be contrasted to the following bounds, obtained by assuming parallel assemblages of the constituents.

Upper (Voigt) bound:

$$\overline{E} \simeq V_A E_A + V_B E_B \tag{15.3}$$

Lower (Reuss) bound:

$$\overline{E} = \frac{E_A E_B}{V_A E_B + V_B E_A} \tag{15.4}$$

where V_A and V_B are the volume fractions of the two constituents. As an example, the RVEs shown in Fig. 15.2 have been modeled using the finite element method, with elastic properties representative of granite aggregate ($E_A = 5.5\,\text{GPa}, \nu_A = 0.3$) and asphalt binder ($E_B = 0.10\,\text{GPa}, \nu_B = 0.4$). A plot of the axial stress for the case wherein the volume fractions of aggregate and binder are equivalent is shown in Fig. 15.3. It is shown in the figure that the stress concentrates within the aggregate.

Predicted results for the homogenized Young's modulus are shown in Fig. 15.4 for three different volume fractions of aggregate. It can be seen that the micromechanical analysis using the finite element method produces predicted results that are significantly different from the upper (not shown in the figure because it is off the diagram) and lower bounds, thereby supporting the importance of performing micromechanical analyses, especially computationally.

15.2.2 Simulating the Resilient Modulus Test

The methods developed within this textbook can be utilized to simulate a variety of standard pavement tests such as the resilient modulus test (AASHTO 2007), and although this test is normally performed on unbound materials, the predicted results obtained for asphalt can nonetheless be useful for design purposes.

In this test, a cylindrical test specimen is first subjected to a spatially and temporally constant hydrostatic pressure, p, as shown in Fig. 15.5. The specimen is subsequently subjected to a cyclic pressure, σ_d, termed the deviator stress, in the axial direction, as shown in Fig. 15.6. The output axial strain due to the deviator

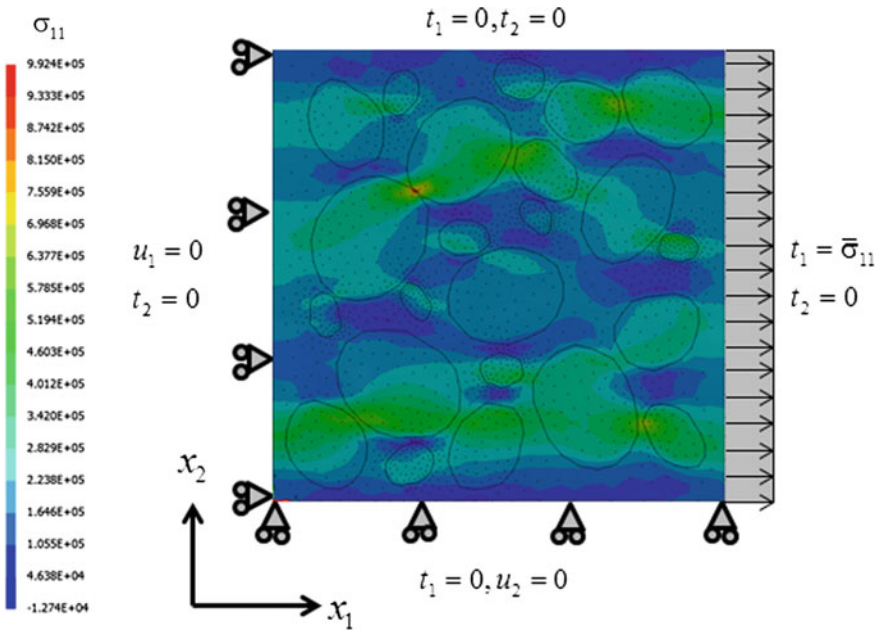


Fig. 15.3 Depiction of the axial stress in an RVE containing 43% volume fraction of aggregate

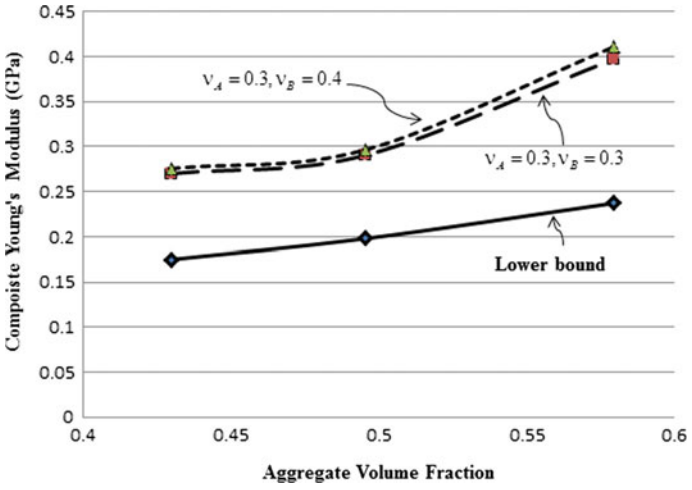


Fig. 15.4 Comparison of computed average Young's modulus to lower bound for various volume fractions (note that ν_B is Poisson's ratio for the asphalt binder)

stress is recorded, as also shown in Fig. 15.7. A cross-plot of the deviator stress versus the axial strain is shown in Fig. 15.8, wherein the recoverable axial strain, ε_r, is shown on the seventh loading cycle.

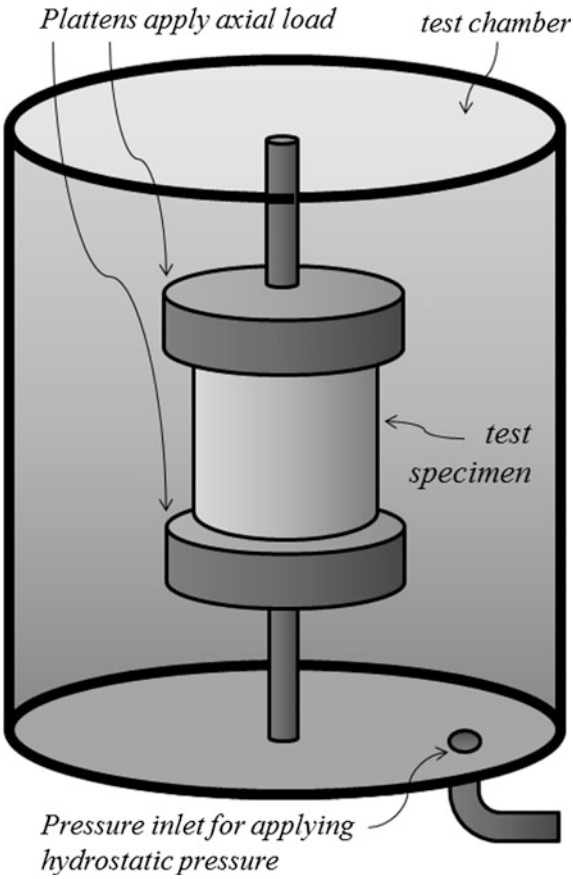


Fig. 15.5 Depiction of a resilient modulus test

The resilient modulus obtained from the test is defined as follows:

$$M_R \equiv \frac{\sigma_d}{\varepsilon_r} \tag{15.5}$$

It is shown in Fig. 15.7 that the resilient modulus varies from cycle to cycle, but it normally shakes down to a constant value after a number of loading cycles, and this is the value that is intended for use as a design tool (AASHTO 2007). Experiments on typical base materials indicate that the resilient modulus depends on the loading rate, the confining pressure, and the deviator stress (Huang 2004). Therefore, it is customary to utilize loading rates defined by the standard, and a series of costly experiments are performed with varying values of both the confining pressure and the deviator stress, and graphs showing these dependences are produced (Huang 2004). Oftentimes, the number of tests can run to 25 or more, thereby expending considerable time and resources for the purpose of evaluating the resilient modulus.

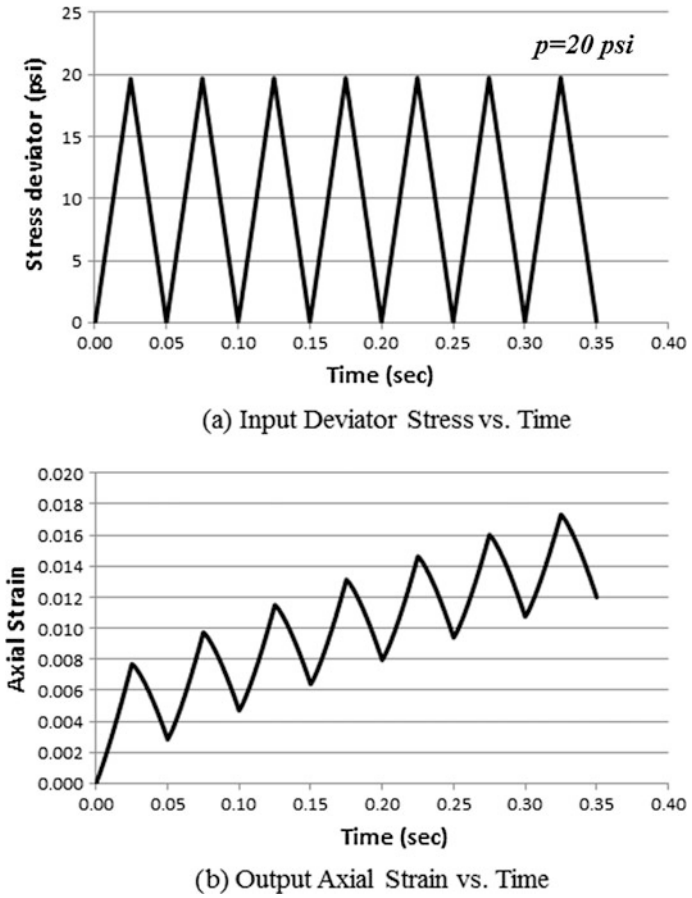


Fig. 15.6 Typical input deviator stress and output axial strain versus time from a resilient modulus test

A question comes to mind: Is all this costly testing really necessary? As a qualitative answer to this question, suppose that the computational mechanics methodology described within this textbook is employed to simulate the resilient modulus test. For example, suppose that the expanding multi-scaling approach developed above is utilized to predict the linear viscoelastic relaxation modulus of an asphaltic concrete. The reader will recall that this process produces predicted values of the relaxation modulus as a function of the volume fractions of mineral filler, fines, and aggregate. Accordingly, should the results of this technique be utilized to predict the resilient modulus, it might be possible to predict the resilient modulus for a given asphalt binder as a function of these same design variables, namely volume fractions of mineral filler, fines, and aggregate. This approach has been employed herein to model the resilient modulus using the predicted linear viscoelastic properties produced above.

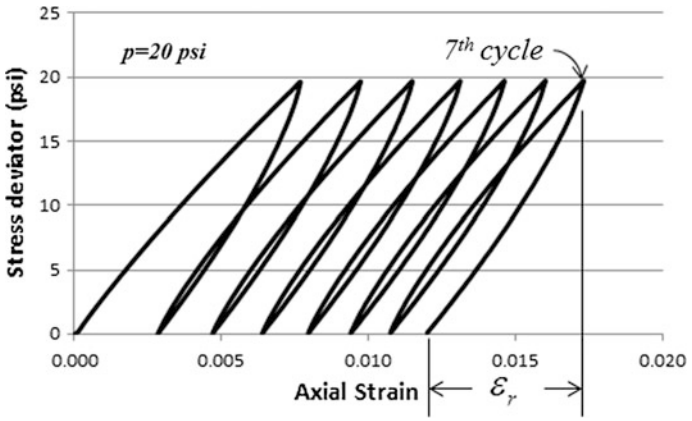


Fig. 15.7 Cross-plot deviator stress versus axial strain from the resilient modulus test shown in Fig. 15.6

The following results are reported in SI units, whereas MEPDG uses the English system of units. The predicted cross-plot stress deviator versus axial strain curves are shown in Fig. 15.8 for cases B2-F2-A1 B2-F2-A2 (see Table 15.5). Similar predictions can be made for other values of the stress deviator, thereby leading to predictions of the resilient modulus versus the deviator stress for various aggregate volume fractions, as shown in Fig. 15.9. It is notable that the predicted change in magnitude of the resilient modulus is considerably less than that observed in actual experiments, and this is presumably due to the fact that a linear viscoelastic material model has been employed, and no damage accumulation has been included within the simulations. Including these two energy dissipation mechanisms would presumably improve the accuracy of the simulations.

15.2.3 Multi-scaling

As shown in Fig. 15.10, examination of the materials utilized in road way construction will reveal that there are in fact multiple length scales within these materials (see Chap. 2–7), essentially all of them resulting from the geometry of the various materials added to the asphalt binder. For example, the typical asphalt mastic is constructed by adding microfine mineral fillers such as lime or sand, and the length scale of these particles typically ranges from 10^{-7} to 10^{-5} m (see Chap. 5). For notational purposes, this will be denoted as *the binder length scale*, and variables described on this length scale will be superscripted with the letter B. Note that this length scale is nonetheless sufficiently large compared to the molecular length scale (approximately 10^{-10} m), so that continuum mechanics may be deployed with reasonable accuracy.

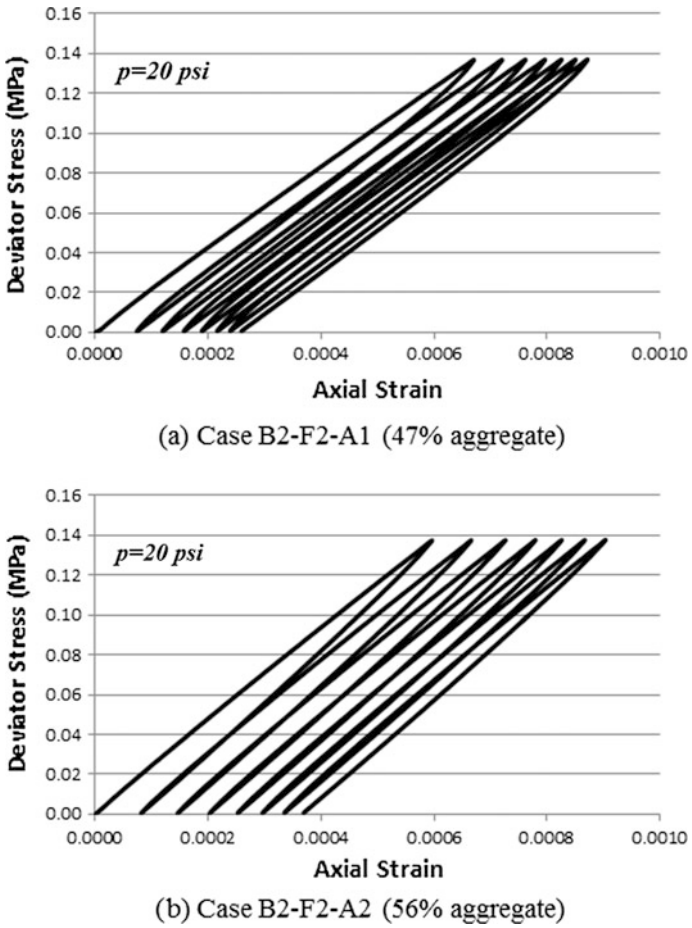


Fig. 15.8 Predicted cross-plot deviator stress versus axial strain curves for two different aggregate volume fractions

The next larger length scale pertaining to road ways is the scale of fines, which ranges from about 75×10^{-6} m to about 2×10^{-4} m. For notational purposes, this will be denoted as *the fines length scale*, and variables described on this length scale will be superscripted with the letter F.

As described in the section in Chap. 11 on micromechanics, there is also a length scale associated with the aggregate, which typically ranges from 10^{-3} to 5×10^{-2} m. This scale will be denoted as *the aggregate length scale*, and variables described on this length scale will be superscripted with the letter A.

Finally, the length scale of the road way generally runs from 3×10^{-1} to 2×10^{1} meters. This scale will be denoted as *the road way length scale*, and variables described on this length scale will be superscripted with the letter R,

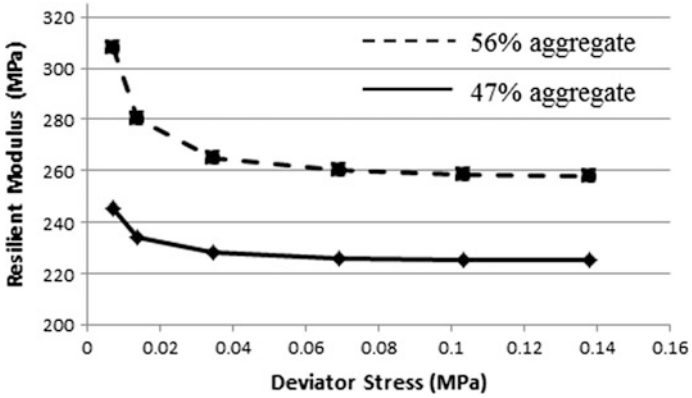


Fig. 15.9 Predicted resilient modulus versus deviator stress for two different aggregate volume fractions

thereby leading to the multi-scale superscripted notation given by the binder length scale, l^B, the fines length scale, l^F, the aggregate length scale, l^A, and the road way length scale, l^R, and variables on these length scales will be superscripted accordingly, (B, F, A, R).

Road way engineers are well aware that physical phenomena on all four of these length scales can contribute to failure of road ways (see Chap. 8). Therefore, significant research has been reported in recent years that attempts to account for these effects. Although many design approaches account for the effects of local and microscale physics phenomenologically (AASHTO 2007), researchers are turning increasingly to more physically based methods of accounting for these effects.

Perhaps the most robust approach would be to model each and every inclusion, void, crack, and particle within the entire road way as a separate homogeneous material embedded within the road way. Common sense dictating that such a complicated IBVP would not avail a closed-form analytic solution, the logical candidate for a solution technique of such a complex problem would be the finite element method. However, given that each inclusion requires a minimum of 10–20 elements, and that a typical section of road way is composed of literally hundreds of thousands of inclusions, the analyst is led to the conclusion that such a highly refined finite element mesh would necessarily contain millions to tens of millions of degrees of freedom. At the speed that today's computers run, such calculations for any given time would typically require the order of perhaps a CPU week or more. Add to that the necessity to deploy a time marching scheme for the purpose of predicting the degradation of the road way over time, and the calculation could amplify to as much as *CPU centuries*. Thus, this approach will perhaps lead to a predictive tool that will take longer to run than the current life of road ways—an approach that is clearly untenable, at least at the time of this writing.

This conundrum is not unique to the structural mechanics community. The pace of scientific developments has led to numerous theories in recent times that cannot

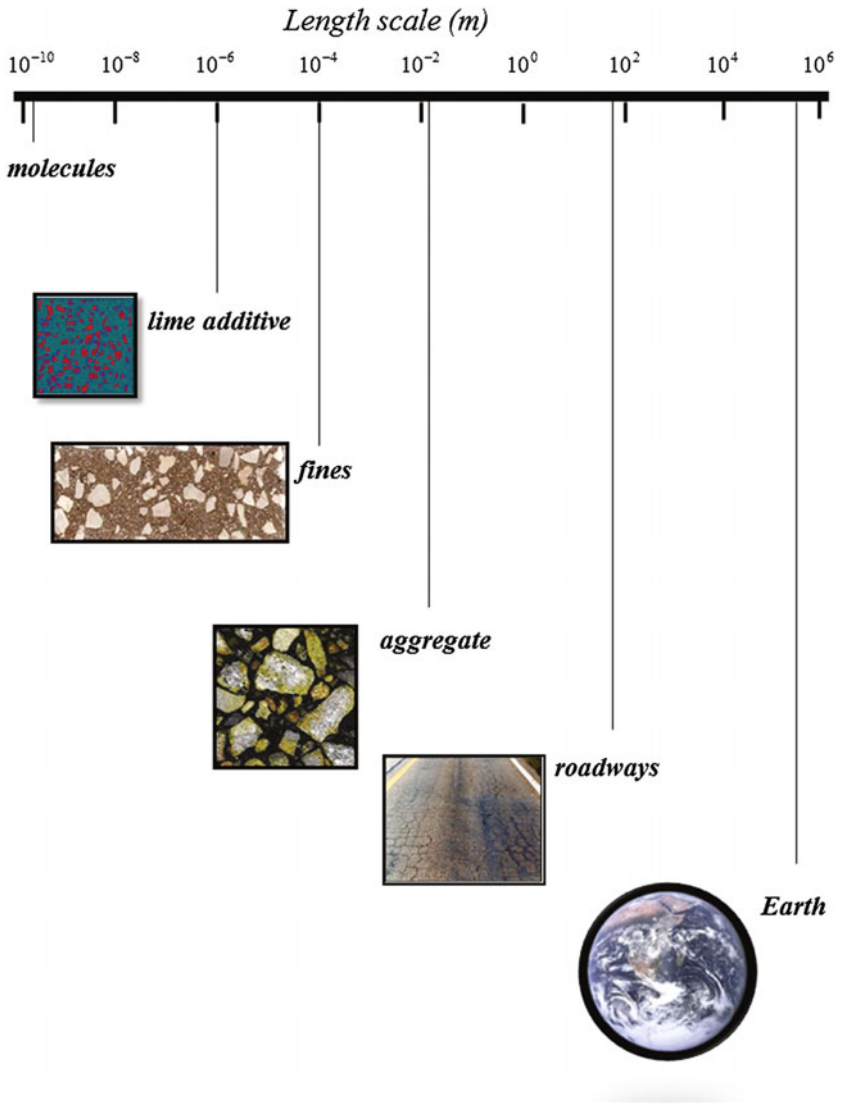


Fig. 15.10 Depiction of length scales relative to roadway infrastructure

be deployed using today's computational capabilities, including global plate tectonics, weather, and the biomechanics of the human body (Allen 2014).

Fortunately, there is Moore's law (Moore 1965), which states that computer capacity will double every eighteen months, and this law has been shown to be accurate for more than a century. Should it continue to be accurate for the foreseeable future, it is quite possible that road ways can be modeled in such a precise way in the next century. Perhaps it will even be possible to model each and every

molecule within the road way at some point in the future, but this is simply not possible at the time of this writing.

Until then, it will be expedient to model the physics and chemistry on each length scale separately and link them together using homogenization techniques recursively, as described in the previous section on micromechanics. In this text, this recursive process of linking length scales will be termed *multi-scaling*.

15.2.3.1 Expanding Multi-scaling

When the process of multi-scaling is deployed on successively increasing length scales, together with intervening homogenization of material properties, it will be termed *expanding multi-scaling*. Note that, due to the fact that the analysis on the smaller scale does not depend on the analysis on the larger scale, this process is mathematically one-way coupled, thereby potentially introducing some approximation within the model. This approach will yield accurate predictions under the following conditions:

(1) The length scales described above are widely separated ($l^B \ll l^F \ll l^A \ll l^R$);
(2) The statistical geometry of the inclusions, including the locally averaged size distribution, orientation, and shape, may be assumed to be spatially invariant throughout the road way;
(3) The material properties of all of the constituents within the road way may be assumed to be statistically spatially homogeneous within the road way; and
(4) Any chemical changes occurring during mixing are accounted for within the model.

The expanding multi-scaling procedure commences with the smallest length scale, l^B, wherein a micromechanical analysis is performed, as described in the previous section. The results of this analysis are then homogenized to obtain averaged material properties for the purpose of performing an analysis on the next larger length scale. This process is then repeated on the next larger length scale, l^F, and thereafter on the aggregate length scale, l^A, thereby arriving at an analysis at the road way length scale, l^R.

As mentioned above, this process is not only computationally much simpler than a full-scale analysis that accounts for every inclusion on a single length scale, but it also produces mathematically accurate results so long as the above assumptions hold within the road way (Helms et al. 1999; Allen 2002; Souza et al. 2008).

Consider the multi-scale example with volume fractions of constituents at each length scale shown in Table 15.1. For example, the makeup of the asphalt concrete for the B2-F2-A2 case is shown in Table 15.2. Note that in this analysis, the binder at the smallest length scale, l^B, is termed *bitumen* or *binder*, whereas at the fines length scale, l^F, the homogenized binder is termed *modified binder*. Furthermore, at the aggregate length scale, l^A, the homogenized binder is termed *mastic*, and at the road way length scale, l^R, the homogenized binder is termed *asphalt concrete*.

Table 15.1 Depiction of geometric parameters for constituents at each length scale in a sample asphalt concrete

Binder length scale

Case B1

Material⟶	*Bitumen*	*Filler*	*Interphase*	*Total*
Volume Fraction	*0.941*	*0.039*	*0.02*	*1.00*

Case B2

Material⟶	*Bitumen*	*Filler*	*Interphase*	*Total*
Volume Fraction	*0.867*	*0.102*	*0.031*	*1.00*

Case B3

Material⟶	*Bitumen*	*Filler*	*Interphase*	*Total*
Volume Fraction	*0.801*	*0.161*	*0.038*	*1.00*

Fines length scale

Case F1

Material⟶	*Binder*	*Fines*	*Total*
Volume Fraction	*0.565*	*0.435*	*1.00*

Case F2

Material⟶	*Binder*	*Fines*	*Total*
Volume Fraction	*0.48*	*0.52*	*1.00*

Case F3

Material⟶	*Binder*	*Fines*	*Total*
Volume Fraction	*0.405*	*0.595*	*1.00*

Aggregate length scale

Case A1

Material⟶	*Mastic*	*Aggregate*	*Voids*	*Total*
Volume Fraction	*0.48*	*0.47*	*0.05*	*1.00*

Case A2

Material⟶	*Mastic*	*Aggregate*	*Voids*	*Total*
Volume Fraction	*0.39*	*0.56*	*0.05*	*1.00*

Case A3

Material⟶	*Mastic*	*Aggregate*	*Voids*	*Total*
Volume Fraction	*0.30*	*0.65*	*0.05*	*1.00*

Table 15.2 Depiction of the asphalt concrete makeup for the B2-F2-A2 case described in Table 15.5

Asphalt concrete makeup on roadway length scale							
Case B2-F2-A2							
Material⟶	***Bitumen***	***Filler***	***Interphase***	***Fines***	***Aggregate***	***Voids***	***Total***
Vol. Fraction	***0.162***	***0.019***	***0.006***	***0.203***	***0.56***	***0.05***	***1.00***
Mass Fraction	***0.071***	***0.017***	***0.004***	***0.242***	***0.666***	***0***	***1.00***

Table 15.3 Elastic properties of typical roadway fillers

Filler⟶	Lime	Interphase	Fines	Aggregate
Young's modulus (GPa)	20.0	10.0	50.0	50.0
Poisson's ratio	0.3	0.3	0.3	0.3

Note also that three different geometric configurations are considered on each length scale, thus leading to three different analyses to be performed on each length scale, the intent being to determine the sensitivity of the road way performance to variations in *the primary design variables: filler volume fraction, fines volume fraction, aggregate volume fraction, and asphalt concrete depth* (see Chap. 8).

On the binder length scale, it is assumed that bitumen is embedded with a microfine mineral filler such as lime (see Chap. 5), and that this process produces an interphase between the filler and the bitumen. Both the filler and the interphase are assumed to be isotropic elastic, with properties shown in Table 15.3. Three different volume fractions of filler are considered, with finite element meshes shown in Fig. 15.11. Note that in this case, because the filler is a relatively small volume fraction, it is possible to analyze a single particle (only a quarter of which is modeled due to symmetry), called *a unit cell*. Research has shown that the homogenization process yields comparable results to more computationally intensive RVE results when the particles represent a small volume fraction (Helms et al. 1999).

The analyses of these three unit cells are utilized, together with the homogenization methodology described in the previous section, to construct the homogenized material properties on the next length scale, l^F. The binder is assumed to be linear viscoelastic, with coefficients in the Prony series describing the relaxation modulus shown in Table 15.4 (Note that these were obtained by multiplying the shear relaxation modulus obtained in Example Problem 12.3 by using $E = 2G(1+\nu)$, where $\nu = 0.4$). Note that $\mu(t)$ must be converted to $E(t)$ using Eqs. (11.46) and (11.48).

Homogenization of the resulting solution using Eq. (15.1) leads to the predicted relaxation data shown in Fig. 15.12 (Allen et al. 2017a). These curves are then fit using the approach described in Chap. 12, and the resulting relaxation modulus for each volume fraction of filler is utilized to perform the analysis on the next larger length scale.

The homogenized binder properties are next utilized to perform the analyses on the fines length scale, l^F, with three different volume fractions of fines, as shown in

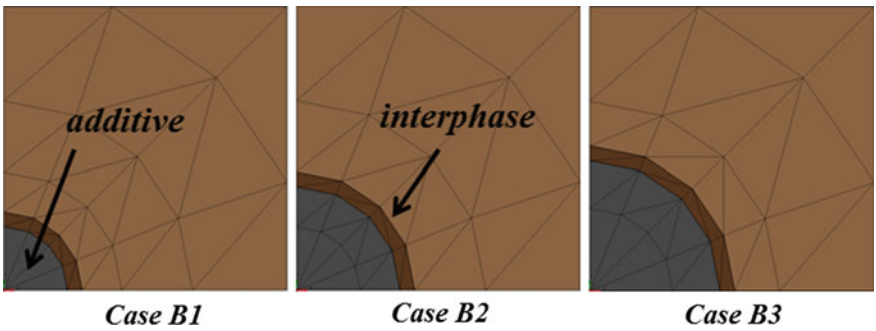


Fig. 15.11 Depiction of three different geometries for filler within bitumen

Table 15.4 Prony series coefficients used to model the relaxation modulus of the binder

E_∞(MPa)	E_1(MPa)	E_2(MPa)	E_3(MPa)	E_4(MPa)	E_5(MPa)	τ_1(s)	τ_2(s)	τ_3(s)	τ_4(s)	τ_5(s)
2.8	11.0	25.0	56.0	10.6	2.6	1.E-3	1.E-2	1.E-1	1.E0	1.E1

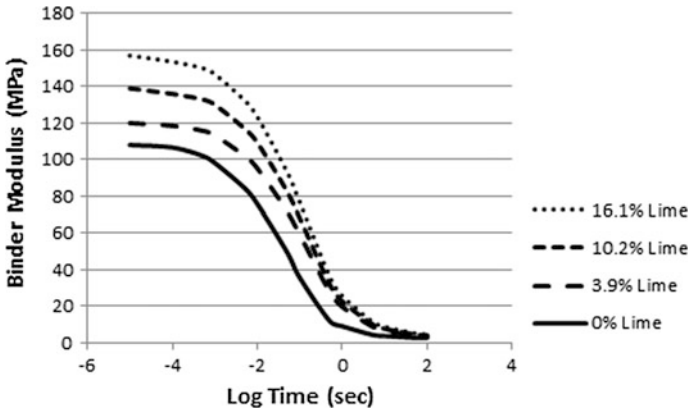


Fig. 15.12 Predicted relaxation modulus for analyses performed on the cases shown in Fig. 15.11 (Allen et al. 2017a)

Fig. 15.13. The process of homogenization is repeated, thereby resulting in the mastic relaxation modulus for three different volume fractions of fines, as shown in Fig. 15.14 (Allen et al. 2017a).

The mastic relaxation modulus is then utilized to perform the analyses for the RVE at the aggregate length scale, l^G, with three different volume fractions of aggregate (see Chap. 3), as shown in Fig. 15.15. Note that on this length scale, it is also assumed that approximately 4% of air voids are included in the analyses. The analyses and resulting homogenization process result in Fig. 15.16 (Allen et al. 2017a), with predicted relaxation modulus for three different volume fractions of aggregate within the asphalt concrete shown in Table 15.5 (Allen et al. 2017a).

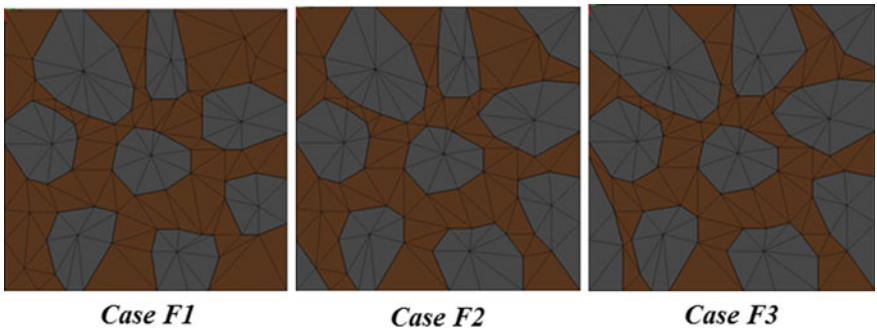


Fig. 15.13 Depiction of three different geometries for binder with fines added

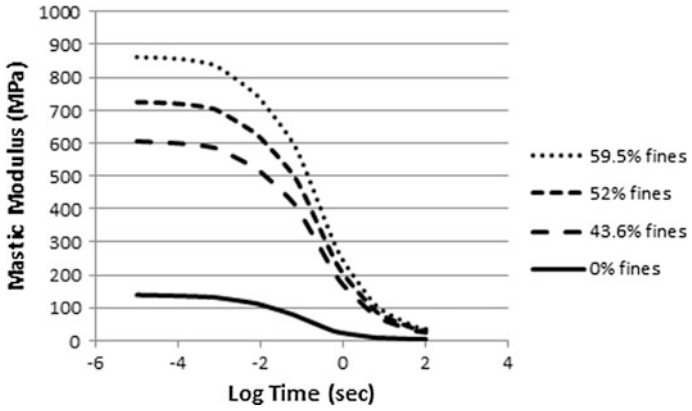


Fig. 15.14 Predicted relaxation modulus for analyses performed on the cases shown in Fig. 15.13 (Allen et al. 2017a)

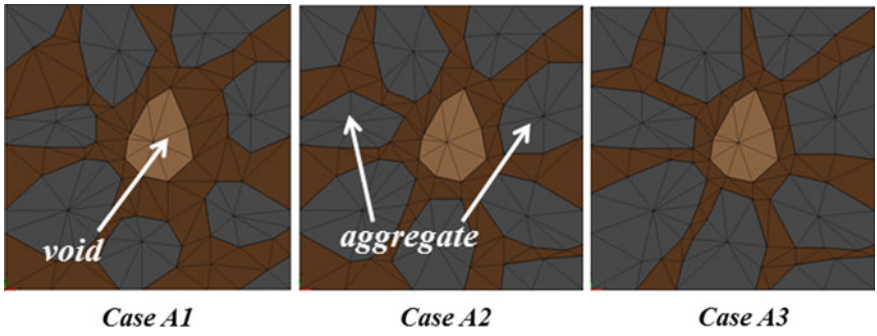


Fig. 15.15 Depiction of three different geometries for mastic with aggregate and voids added

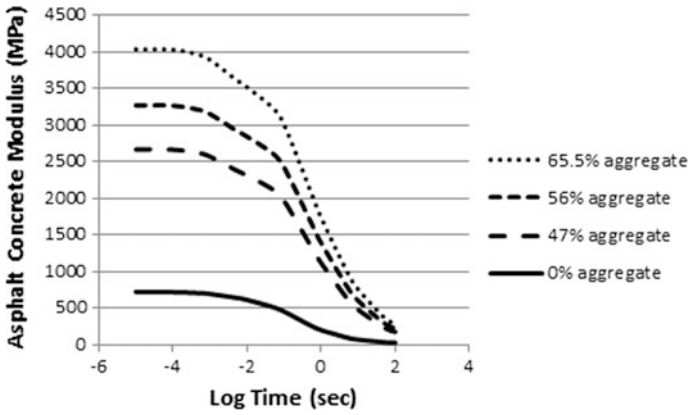


Fig. 15.16 Predicted relaxation output for analyses performed on the cases shown in Fig. 15.15 (Allen et al. 2017a)

The three previous scales of analysis now result in a homogenized set of road way length scale asphalt concrete relaxation moduli as functions of fillers, fines, and aggregate volume fractions (see Chaps. 8, 10, and 12). Figure 15.17 demonstrates the mechanical effect on the relaxation modulus produced by each of the fillers (Allen et al. 2017a). As shown in the figure, the properties of the asphalt concrete are increased multifold by introducing (usually) inexpensive geologic particulates. Note that this procedure could also be utilized for other constituents, such as void volume fraction and cracking (to be described later in this chapter).

The asphalt concrete relaxation modulus may now be utilized to perform analyses at the road way length scale, l^R. Toward this end, Fig. 15.18 depicts a typical finite element mesh constructed for the purpose of modeling the performance of a cross section of the road way. This particular two-dimensional mesh contains 3,162 degrees of freedom and 3,000 constant strain triangles. The analysis is carried out assuming plane strain conditions in the direction of travel (normal to the plane of the depiction).

As described in Table 15.6, three analyses are performed at this length scale, each with a different thickness of the asphalt concrete layer, while holding all other design variables constant. If all of the design variables (filler volume fraction, fines volume fraction, aggregate volume fraction, and road way thickness) are considered, this results in a matrix of $3^4 = 81$ possible road way configurations, thereby demonstrating the open-ended nature of the road way design problem.

For convenience, only a few of the road way scale results are shown herein. Each road way configuration is subjected to five successive cycles of road way traction (for a typical truck tire). Figure 15.19 shows the peak predicted stress component σ_{11} as a function of location in the road way for the case B2-F2-A1 (Allen et al. 2017a). Figure 15.20 shows the predicted plastic strain component ε_{11}^P after five loading cycles for the B2-F2-A1 case (Allen et al. 2017a).

As a measure of road way performance, maximum residual deformations of the road way surface are predicted as a function of loading cycle number. Figure 15.21

Table 15.5 Homogenized relaxation modulus for asphalt concrete (Allen et al. 2017a)

Property ⟶	E_∞(MPa)	E_1(MPa)	E_2(MPa)	E_3(MPa)	E_4(MPa)	E_5(MPa)	τ_1(s)	τ_2(s)	τ_3(s)	τ_4(s)	τ_5(s)
Case↓											
B2-F2-A1	78.8	64.5	0.8	75.3	48.4	07.5	0.00144	0.01443	0.1443	1.443	14.43
B2-F2-A2	25	30.2	0.7	140.3	20.8	58.2	0.00144	0.01443	0.1443	1.443	14.43
B2-F2-A3	82.8	40.2	0.8	430	137.5	51.6	0.00144	0.01443	0.1443	1.443	14.43

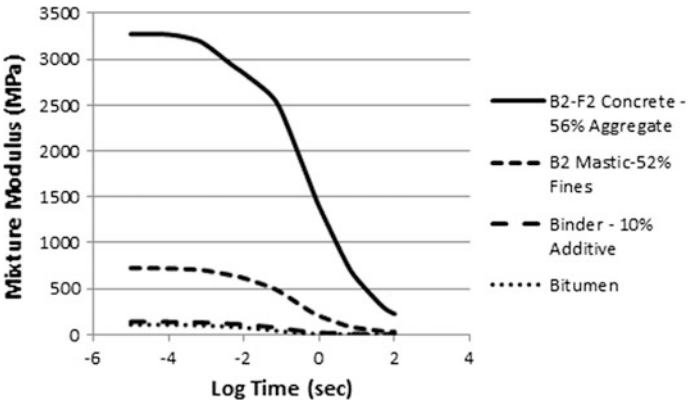


Fig. 15.17 Predicted effect of introducing fillers to bitumen on the asphalt concrete relaxation modulus (Allen et al. 2017a)

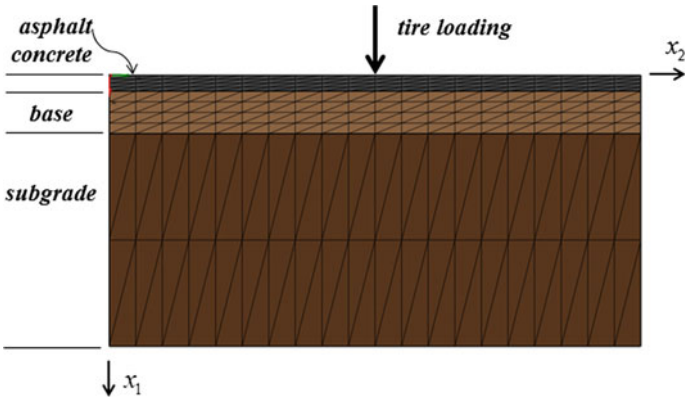


Fig. 15.18 Depiction of finite element mesh utilized to predict roadway performance

Table 15.6 Analyses performed at the roadway length scale

Case R1			
Layer Type ⟶	*Asphalt Concrete*	*Base*	*Subgrade*
Depth	*0.05 m*	*0.4 m*	*2.0 m*
Case R2			
Layer Type⟶	*Asphalt Concrete*	*Base*	*Subgrade*
Depth	*0.10 m*	*0.4 m*	*2.0 m*
Case R3			
Layer Type⟶	*Asphalt Concrete*	*Base*	*Subgrade*
Depth	*0.15 m*	*0.4 m*	*2.0 m*

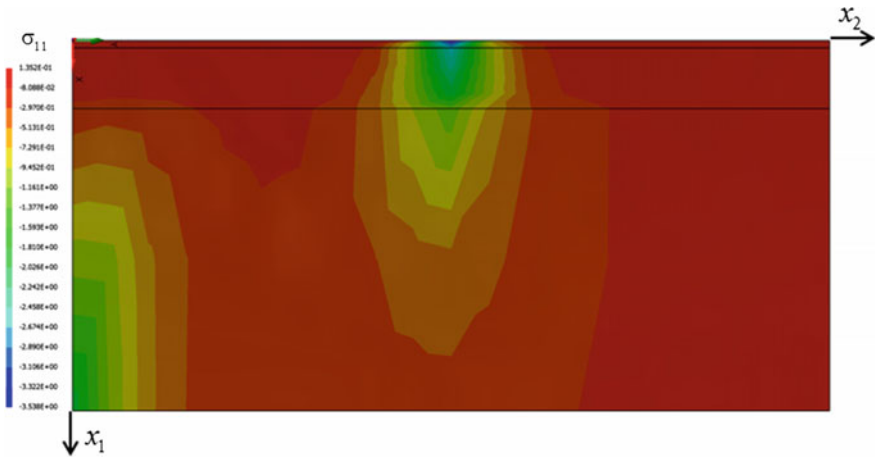


Fig. 15.19 Predicted peak stress σ_{11} (Case B2-F2-A2-R1) (Allen et al. 2017a)

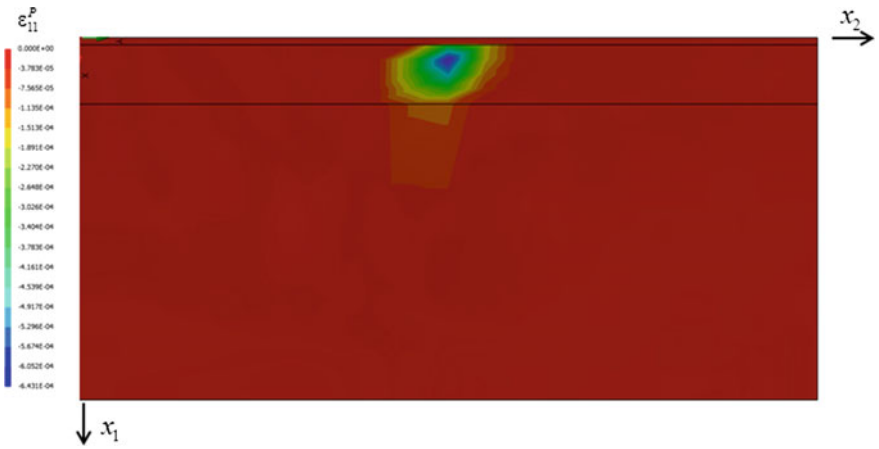


Fig. 15.20 Predicted residual plastic strain component ε_{11}^{P} after five loading cycles (Case B2-F2-A2-R1) (Allen et al. 2017a)

shows the predicted effect of varying aggregate volume fraction on permanent deformation (an aggregate length scale design variable) (Allen et al. 2017a). Similarly, Fig. 15.22 shows the maximum permanent surface deformation plotted as a function of cycle number for three different thicknesses of asphalt concrete (a road way length scale design variable) (Allen et al. 2017a).

As a final example, consider the case wherein the base material is compacted in such a way as to increase the initial yield point of the base material (without increasing the elastic modulus). As shown in Fig. 15.23, while the effect of further compaction of the base material is not quite as pronounced on the residual deformation as is the effect of increasing asphalt concrete thickness, it is nevertheless detectable (Allen et al. 2017a).

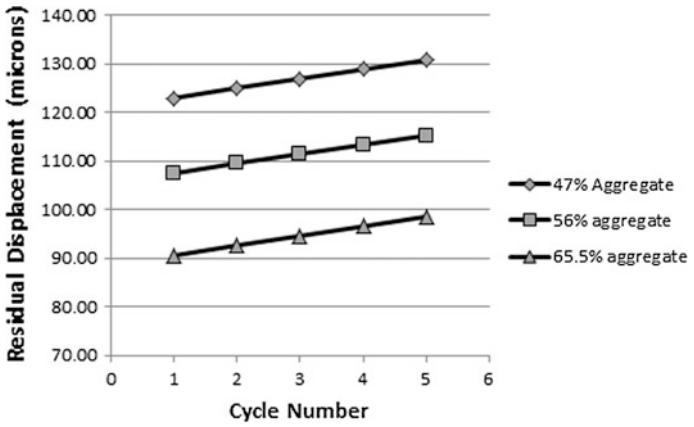


Fig. 15.21 Predicted maximum residual deformation versus cycle number for three different aggregate volume fractions. (Cases B2-F2-A1-R2, B2-F2-A2-R2, and B2-F2-A3-R2) (Allen et al. 2017a)

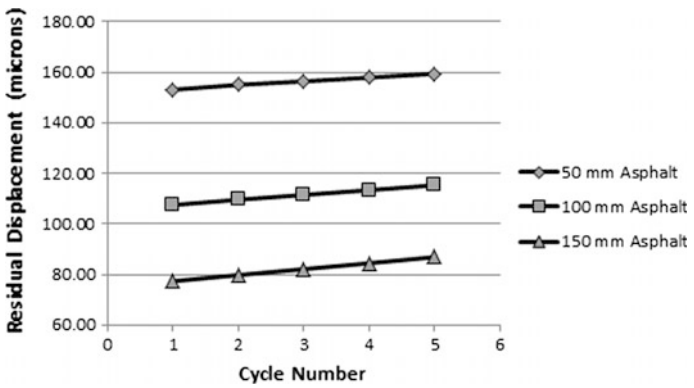


Fig. 15.22 Predicted maximum residual deformation versus cycle number for three different asphalt concrete thicknesses (Allen et al. 2017a)

Whether this effect is significant will depend on a cost analysis, but it is often found that improving the properties of the base material is less costly than increasing the thickness of the surface layer (see Chap. 8).

It should now be apparent that the expanding multi-scale procedure leads to a road way analysis tool that can be utilized to assess the cost-effectiveness of microfine mineral fillers, millimeter-scale fines, centimeter-scale aggregate, meter-scale road way layers, and even base treatments (see Chaps. 2–8). Note that cracking has not been included within the present analyses, nor has the effect of environment been accounted for. Thus, while the analysis contained herein is only a demonstration of the capability of multi-scaling, it should be apparent that it can be used as a powerful tool in the road way design process. The effects of both cracking and environment will be considered in later sections.

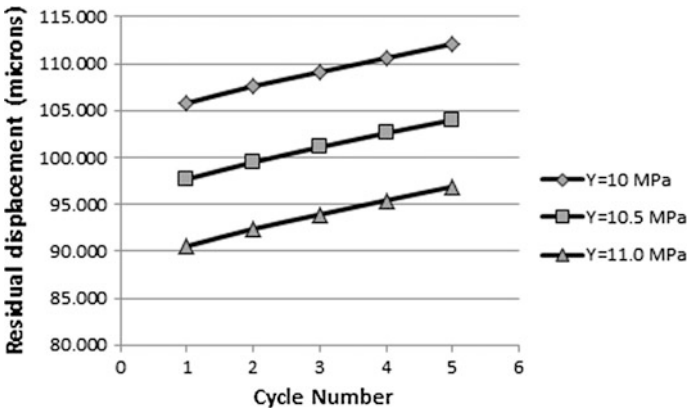


Fig. 15.23 Predicted maximum residual deformation versus cycle number for three different base yield points (denoted as Y in the figure) (Allen et al. 2017a)

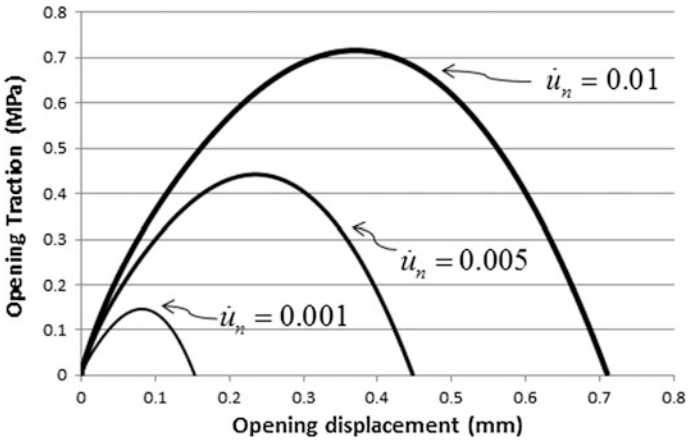


Fig. 15.24 Depiction of normal component of cohesive zone model for three different opening displacement rates (Allen et al. 2017b)

15.2.3.2 Modeling Evolving Cracks in the Road way

The computational algorithm for incrementing the cohesive zone model introduced in Chap. 14 may be used to predict crack growth in the road way. As an example, Fig. 15.24 depicts the output cohesive zone normal traction component versus displacement for three different loading rates of the binder properties shown in Table 15.4 $(\delta_1^C = 1.0, \alpha_1^C = 0.025, n^C = 0.50)$, thereby illustrating the rate-dependent fracture behavior of asphalt binder (Allen et al. 2017b).

Figure 15.25 shows the predicted results for the same cohesive zone as that depicted in Fig. 15.24, but this time subjected to a cyclic loading in displacement

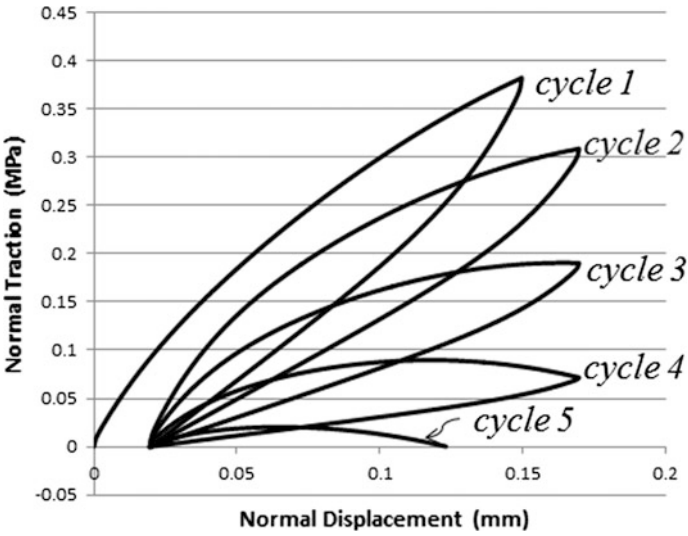


Fig. 15.25 Predicted normal traction versus displacement in a viscoelastic cohesive zone subject to cyclic loading (Allen et al. 2017b)

control (Allen 2017b). As shown in the figure, the fracture toughness of the cohesive zone is not only rate dependent, but it is also history dependent. As shown in the figure, the cohesive zone fails on the fifth cycle, and at a load level that is lower than on the previous cycles, thereby illustrating the history dependence of the fracture toughness in the viscoelastic cohesive model.

Next, consider the case of a compact tension specimen of mastic (Case B2-F2) subjected to opposing tip displacements that are slowly increased monotonically in time, as shown in Fig. 15.26. Note that the mastic has been embedded with a precrack, as shown in the figure. The viscoelastic properties of the mastic are as shown in Table 15.7, and the finite element mesh utilized for the analysis is shown in Fig. 15.27, with the distribution of the vertical component of normal stress (σ_{11}) just before propagation of the crack shown in the figure (Allen et al. 2017b). Figure 15.28 shows the crack length versus tip displacement for three different displacement rates, demonstrating the rate dependence of the fracture model (Allen et al. 2017b). Finally, Fig. 15.29 depicts the evolution of the normal component of traction within the cohesive zone as the crack tip advances (Allen et al. 2017b).

Note that in the example problem described above, the test has been simulated in displacement control. By contrast, consider the case wherein the specimen shown in Fig. 15.26 is subjected to load control, with the tip loads increasing at constant rate. This results in a slightly different prediction of the traction curve as the crack propagates, as shown in Fig. 15.30 (Allen et al. 2017b).

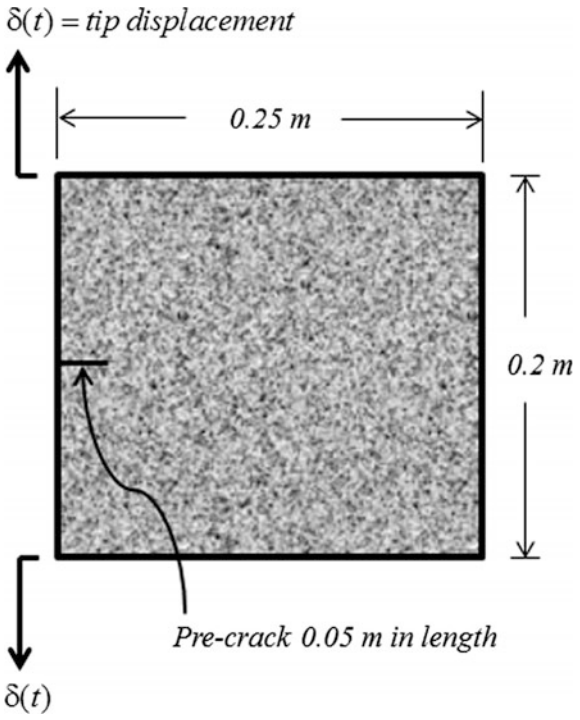


Fig. 15.26 Depiction of a precracked specimen of mastic subjected to opposing tractions

15.2.3.3 Contracting Multi-scaling

Now consider the reverse of the contracting multi-scaling process, one in which the analysis is performed on successively smaller length scales. For example, in some cases, the designer may determine that the global scale homogenized analysis leads to the prediction of locally excessive stresses, strains, or displacements within the global component. In such a case, it is possible to utilize the output tractions from the global analysis to perform a more detailed study of a portion of the structure, one wherein the material heterogeneity is included in the analysis.

As an example, consider once again the road way shown in Fig. 15.18. As shown in Fig. 15.31, suppose that it is determined that as a result of the global scale analysis (denoted with superscripts G) that a more detailed analysis is performed on the local scale (denoted with superscripts L). Using this approach, a more highly refined finite element mesh may be constructed for the purpose of determining the local effects of the distribution of tire traction on the road way.

Figure 15.32 shows the predicted initial components of surface displacement due to the tire loading, whereas Figs. 15.33 and 15.34 show plots of the initial in-plane normal components of stress (Allen et al. 2017b).

Figure 15.35 shows the predicted components of stress just below the surface of the road way as functions of the horizontal coordinate x_2^L (Allen et al. 2017b). Note

Table 15.7 Mastic properties used to model the specimen shown in Fig. 15.26

Bulk Material Properties ($\nu = \mathbf{0.4}$)										
E_∞(MPa)	E_1(MPa)	E_2(MPa)	E_3(MPa)	E_4(MPa)	E_5(MPa)	τ_1(s)	τ_2(s)	τ_3(s)	τ_4(s)	τ_5(s)
29.7	117.4	3.36	335.8	184.5	87.0	0.1443E-2	0.1443E-1	0.14427E0	0.14427E1	0.14427E2
Cohesive zone material properties										
E_∞^C(MPa)	E_1^C(MPa)	E_2^C(MPa)	E_3^C(MPa)	E_4^C(MPa)	E_5^C(MPa)	τ_1^C(s)	τ_2^C(s)	τ_3^C(s)	τ_4^C(s)	τ_5^C(s)
29.7	117.4	3.36	335.8	184.5	87.0	0.1443E-2	0.1443E-1	0.1443E0	0.1443E1	0.1443E2
$\delta_1^C = 1.0, \delta_2^C = 2.5, \alpha_1^C = 20.0,\ n^C = 0.5$										

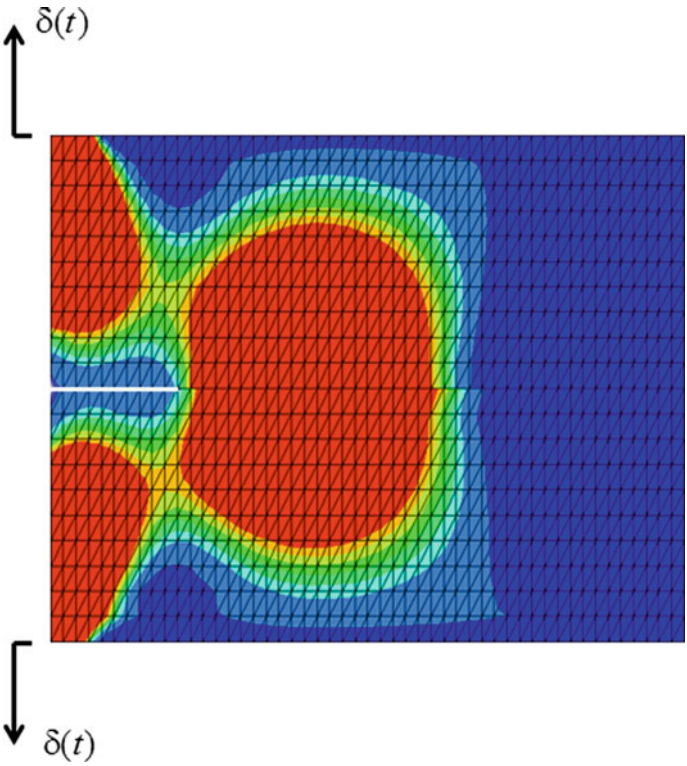


Fig. 15.27 Finite element mesh and distribution of stress component σ_{11} prior to propagation of the crack tip (Allen et al. 2017b)

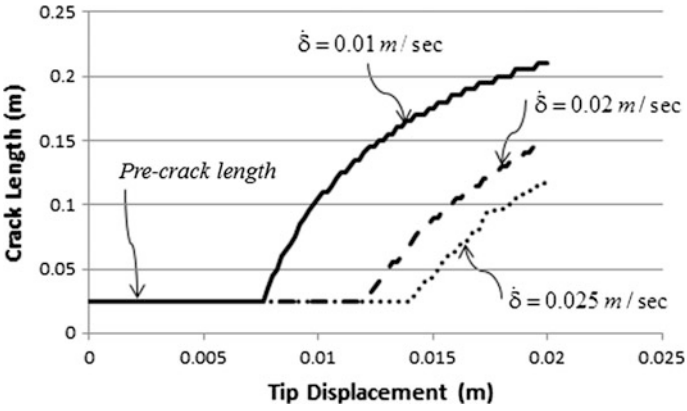


Fig. 15.28 Predicted crack growth rate within the specimen shown in Fig. 15.26 for three different loading rates (Allen et al. 2017b)

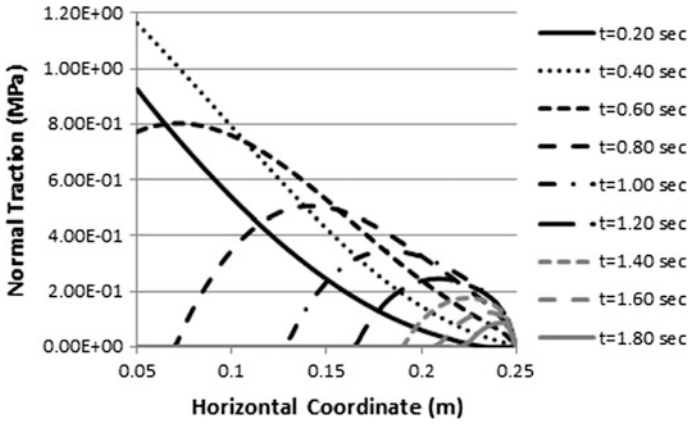


Fig. 15.29 Normal component of traction on the faces of the cohesive zone ahead of the crack tip as the crack advances (displacement control) (Allen et al. 2017b)

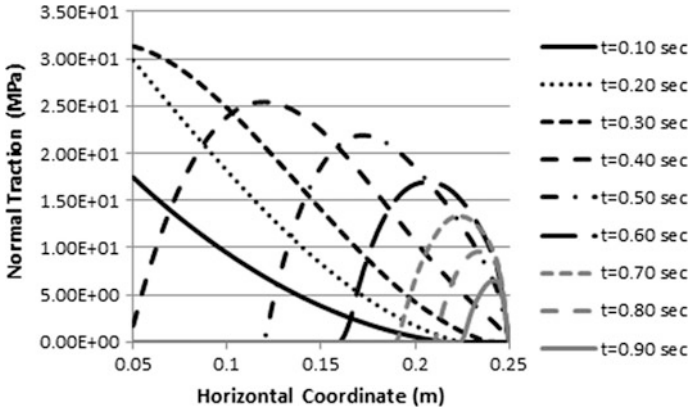


Fig. 15.30 Normal component of traction on the faces of the cohesive zone ahead of the crack as the crack advances (load control) (Allen et al. 2017b)

that in Fig. 15.35, the shear stress has been plotted both on the vertical plane, denoted σ_{12} , and the 45° plane, denoted $\sigma_{12}(45^{\circ})$. This is due to the fact that it is sometimes observed that rutting cracks form at an angle with respect to the road way surface. However, as shown in the figure, in the present analysis, the shearing stress on the vertical plane exceeds the shearing stress on the 45° plane. Furthermore, the three normal components of stress are all comparable to one another beneath the tire load, signifying that the state of stress is nearly hydrostatic compression, a circumstance wherein cracks cannot grow. On the basis of these observations, it is anticipated that if cracking does occur, it will be due to the vertical component of shearing stress, σ_{12} (termed Mode II cracking), just inside the edge of the tire loading. Thus, a third analysis is now performed in which a

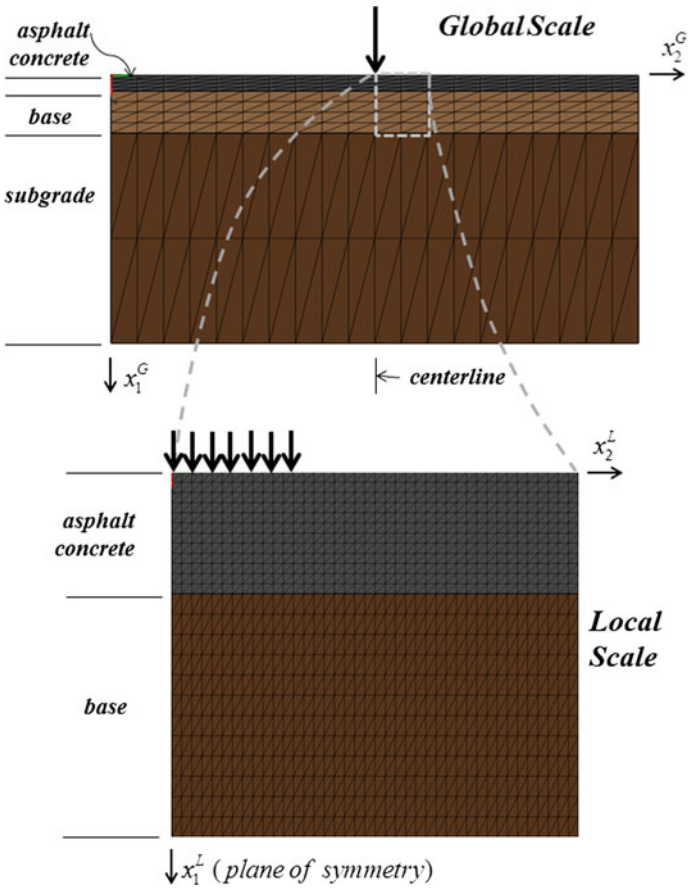


Fig. 15.31 Depiction of a local segment of the roadway selected for a more refined analysis

cohesive zone is embedded on the vertical plane at a distance of 0.015 m from the centerline of the tire loading, as shown in Fig. 15.36.

The analysis predicts that the crack initiates at the road way surface and propagates downward, with the crack length versus tire pressure shown in Fig. 15.37 (Allen et al. 2017b). Figures 15.38, 15.39, and 15.40 show the predicted contours of the in-plane components of stress after the crack has propagated to the bottom of the asphalt layer, wherein it can be seen that the road way surface beyond the crack is essentially unloaded, thereby further concentrating the stresses beneath the tire loading (Allen et al. 2017b).

Figure 15.41 shows the evolution of the cohesive zone shearing traction as the tire pressure is increased monotonically (Allen et al. 2017b). Note that once the crack is initiated, very little increase in the tire pressure is required to propagate the vertical crack through the entire asphalt concrete layer, implying that the crack is very nearly unstable.

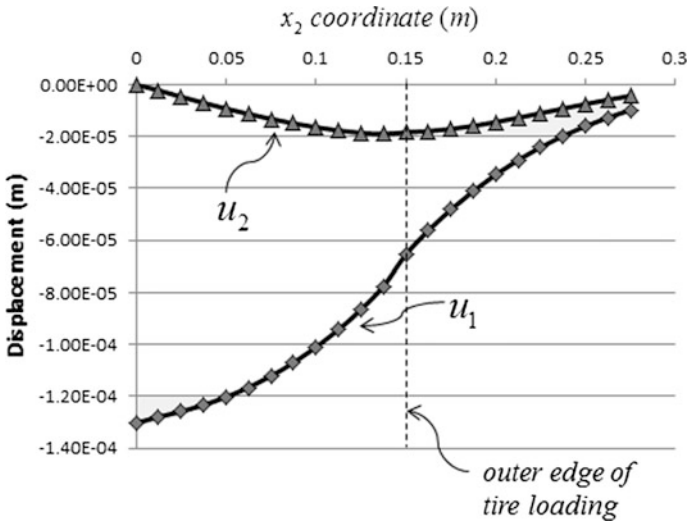


Fig. 15.32 Predicted initial surface displacement components (Allen et al. 2017b)

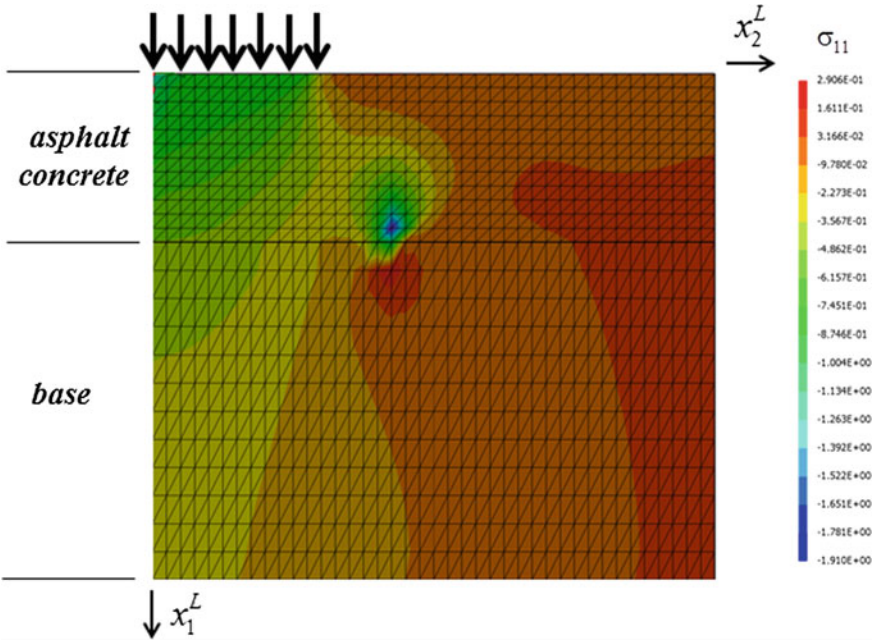


Fig. 15.33 Predicted initial normal stress distribution in the vertical direction (Allen et al. 2017b)

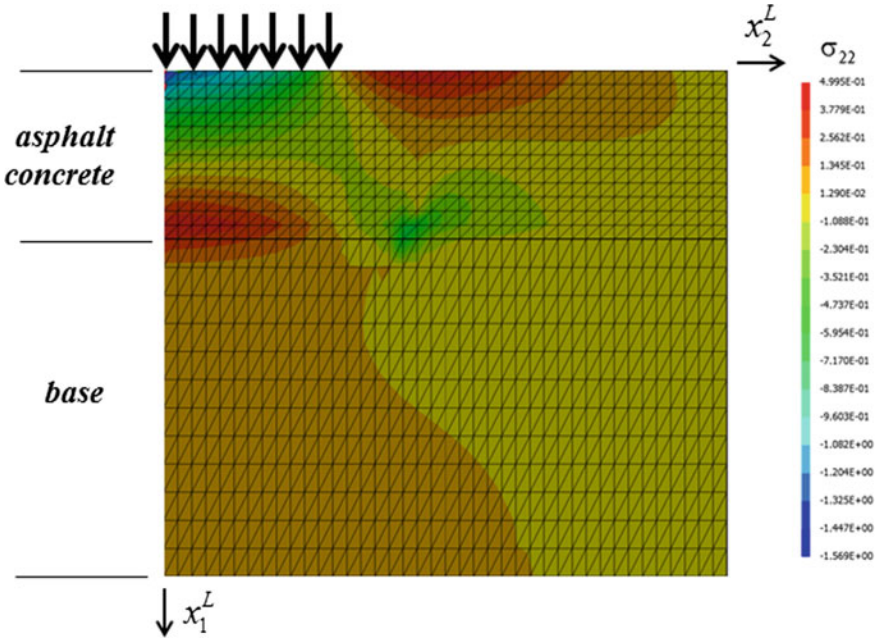


Fig. 15.34 Predicted initial normal stress distribution in the horizontal direction (Allen et al. 2017b)

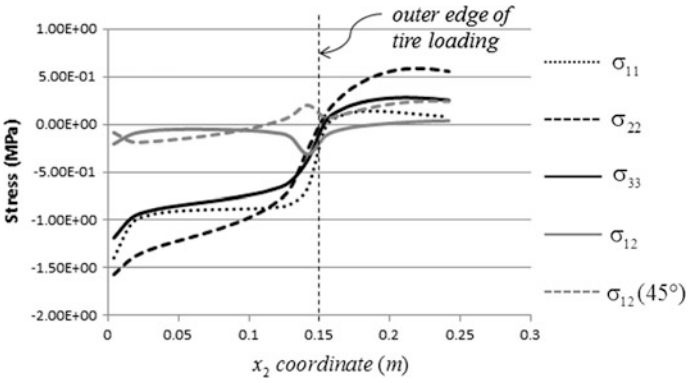


Fig. 15.35 Predicted initial stress components beneath the roadway surface (Allen et al. 2017b)

The process of downward multi-scaling can be continued with decreasing length scales if it is desirable to investigate the physics of the cracking process in greater detail. For example, the results of the analysis at the local scale can be utilized to perform an analysis at the aggregate length scale, as depicted in Fig. 15.42.

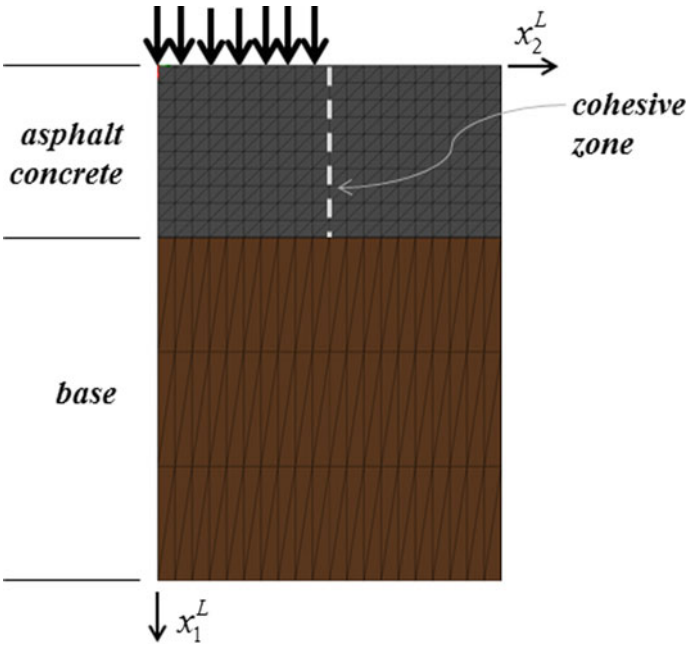


Fig. 15.36 Finite element mesh of roadway cutout showing embedded cohesive zone

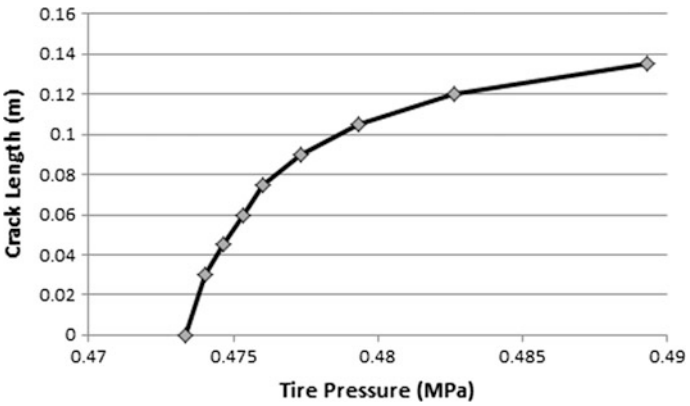


Fig. 15.37 Predicted crack length versus tire pressure for configuration depicted in Fig. 15.36 (Allen et al. 2017b)

Figure 15.43 depicts a comparison of the total crack surface area versus monotonically increasing horizontally applied normal tractions and vertically applied shear tractions (Allen et al. 2017b). As shown in the figure, cracks are predicted to grow much more rapidly due to shearing tractions, indicating that

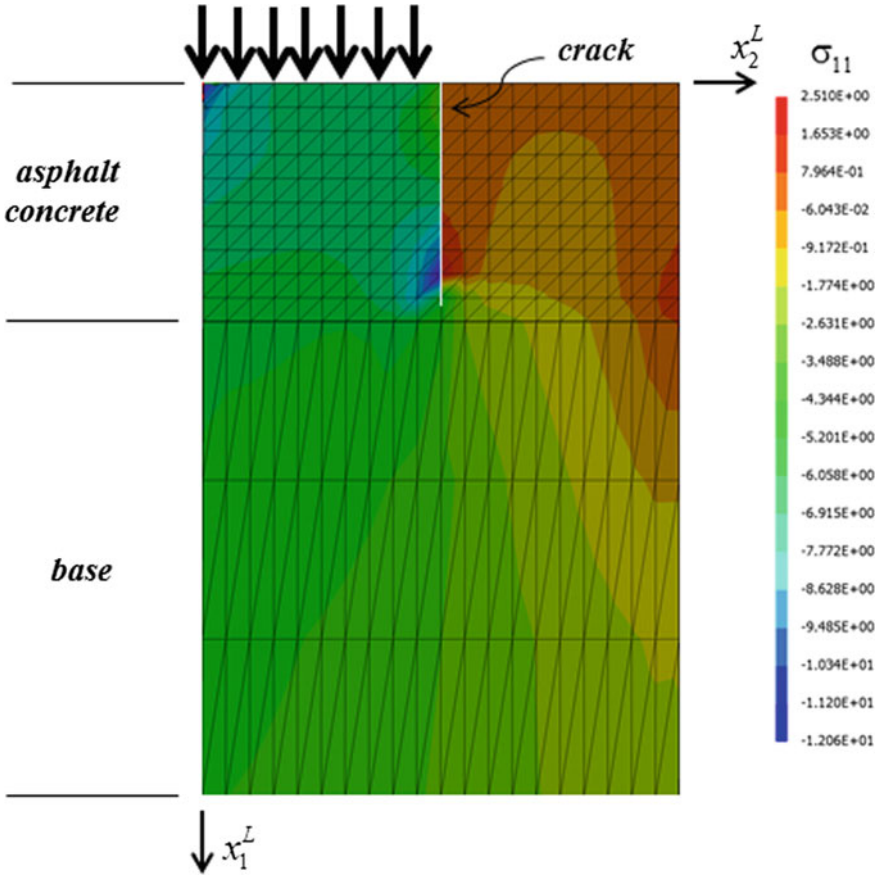


Fig. 15.38 Contour showing predicted stress component σ_{11} in cracked roadway (Allen et al. 2017b)

cracking is more likely to occur in the present case due to shear stresses resulting from the tire pressure applied at the road way surface.

The predicted crack growth for the shear traction case is further detailed in Fig. 15.44, wherein the evolution of the distribution of cracking due to monotonically increasing shearing traction is shown in successive diagrams of the RVE (Allen et al. 2017b).

15.2.3.4 Two-Way Coupled Multi-scaling

Road ways are often observed to undergo cracking on multiple length scales, and these cracks can number into the hundreds or even thousands per cubic meter of the road way (see Chaps. 1 and 8). Because cracks cause stress concentrations, each

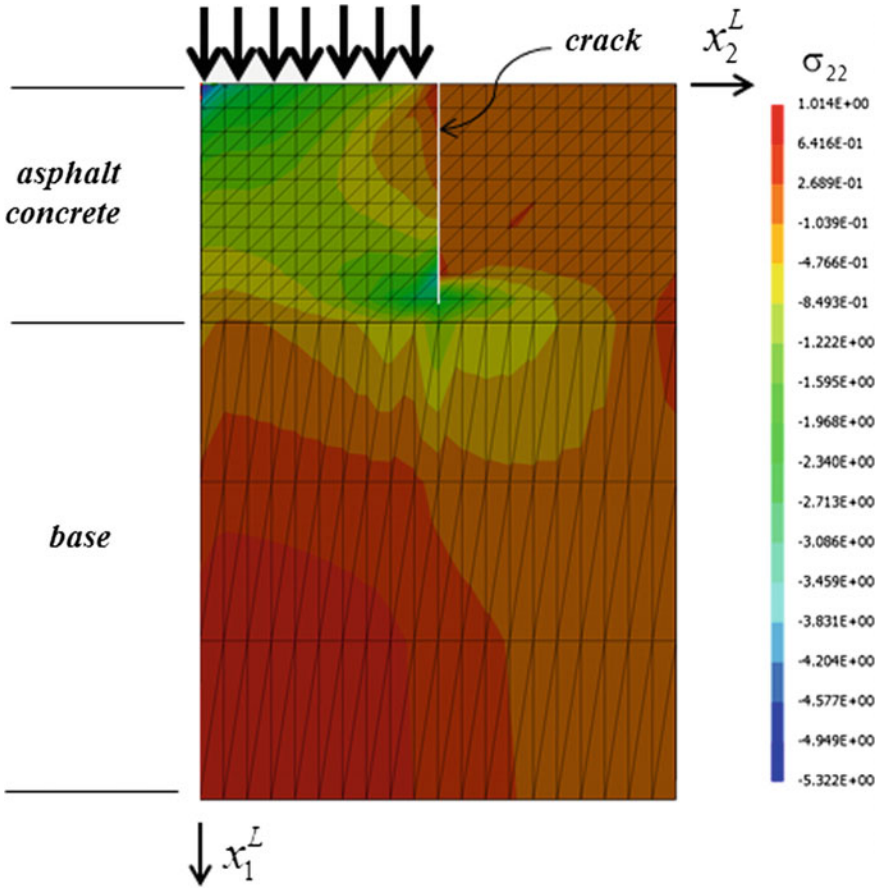


Fig. 15.39 Contour showing predicted stress component σ_{22} in cracked roadway (Allen et al. 2017b)

crack will necessitate the deployment of hundreds or even thousands of finite elements to accurately predict the stresses near each crack tip. Thus, when there are numerous cracks present, it can become computationally untenable to account for the effects of these cracks on pavement performance using a single scale finite element mesh. Fortunately, the length scales associated with road way cracking tend to congregate around the same length scales as those associated with the geometric length scales of the road way such as the filler length scale, l^F, the aggregate length scale, l^A, and the road way length scale, l^A, and quite often these length scales are widely separated, that is,

$$l^F \ll l^A \ll l^R \tag{15.6}$$

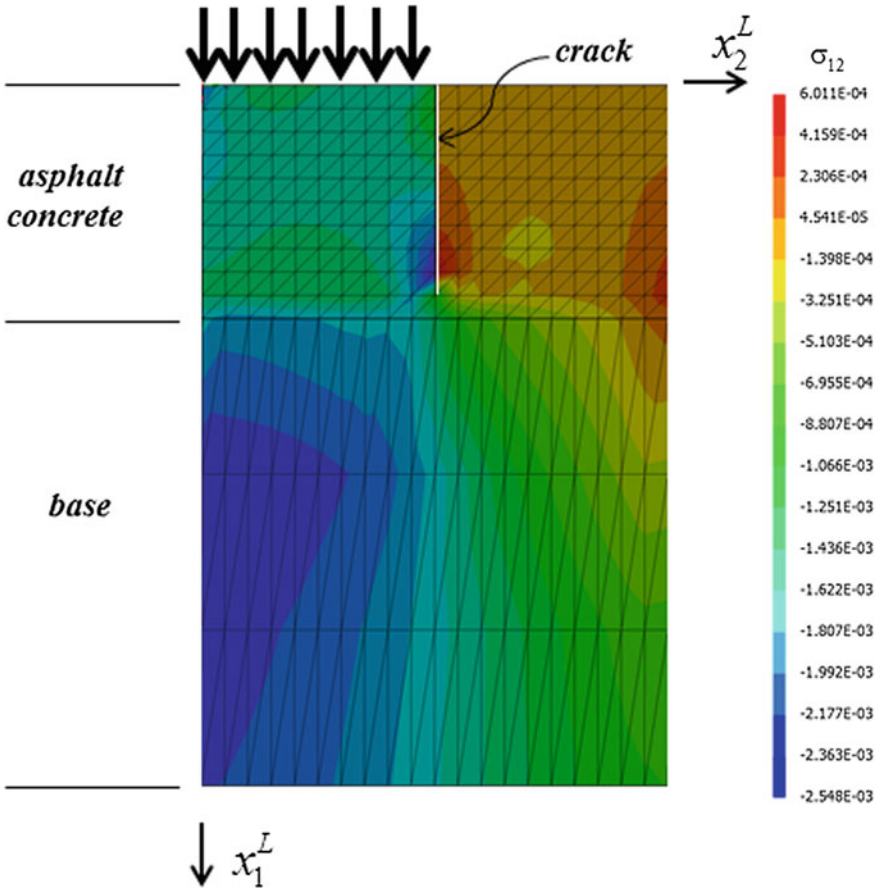


Fig. 15.40 Contour showing predicted stress component σ_{12} in cracked roadway (Allen et al. 2017b)

Accordingly, as in the cases of expanding and contracting multi-scaling, analyses can be performed on the different length scales and coupled via homogenization techniques.

However, whereas the various layers of the road way may be assumed to be initially spatially homogeneous on length scales that are large compared to the various fillers, this assumption will not necessarily remain accurate when cracks form and propagate within the road way. This is due to the fact that cracks tend to concentrate at spatial locations within the road way where tire loads are applied, as shown in Fig. 15.45.

Due to the development of these spatially concentrated cracks, the structural stiffness of the road way will vary significantly as a function of both spatial coordinates and time, thereby leading to still further concentration of the failure of the road way. In a mathematical sense, this implies that not only does the global

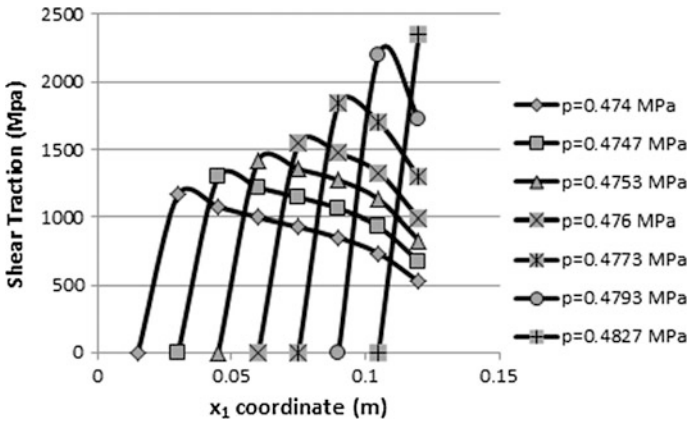


Fig. 15.41 Depiction of the predicted cohesive zone shearing traction versus coordinate location for increasing tire pressure (Allen et al. 2017b)

scale analysis of the road way affect the local scale analysis (as described above), but the local scale analysis also affects the global scale analysis, and the means whereby this two-way dependence is accounted for is termed herein as *two-way coupled multi-scaling*.

The process of performing two-way coupled multi-scaling is shown pictorially in Fig. 15.46, where it should be understood that the typical RVE depicted on the local scale is one of many RVEs that are distributed throughout the global finite element mesh, with one local RVE per global finite element, and these local RVEs are implanted within the global mesh wherever evolutionary cracking is anticipated to occur *on the local scale*, l^L, as compared to the global scale, l^G, and the microscale, l^M, the length scale on which the cohesive zone model is formulated. Note that the local scale can be any length scale, so long as it is small compared to the length scale of the road way, l^R, and large compared to the microscale. Note also that cohesive zones may be implanted within the global mesh as well as the local mesh, so that global scale cracking can be simulated simultaneously with local scale cracking. Thus, it is conceivable that this multi-scaling process could be deployed on several different length scales simultaneously (Allen 2002, Allen and Searcy 2006), but for demonstrative purposes, it is only utilized herein on two length scales, that is, in this text, we consider only the case wherein

$$l^L \equiv l^A \text{ and } l^G \equiv l^R \tag{15.7}$$

In order to affect this two-way coupling between the local and global length scales, the modeling process utilizes a time stepping procedure as described at the beginning of this chapter. On the first time step, it is assumed that the road way is undamaged, and a global analysis is performed for an increment of loading, wherein it is assumed that the road way is initially spatially (layered) homogeneous at the

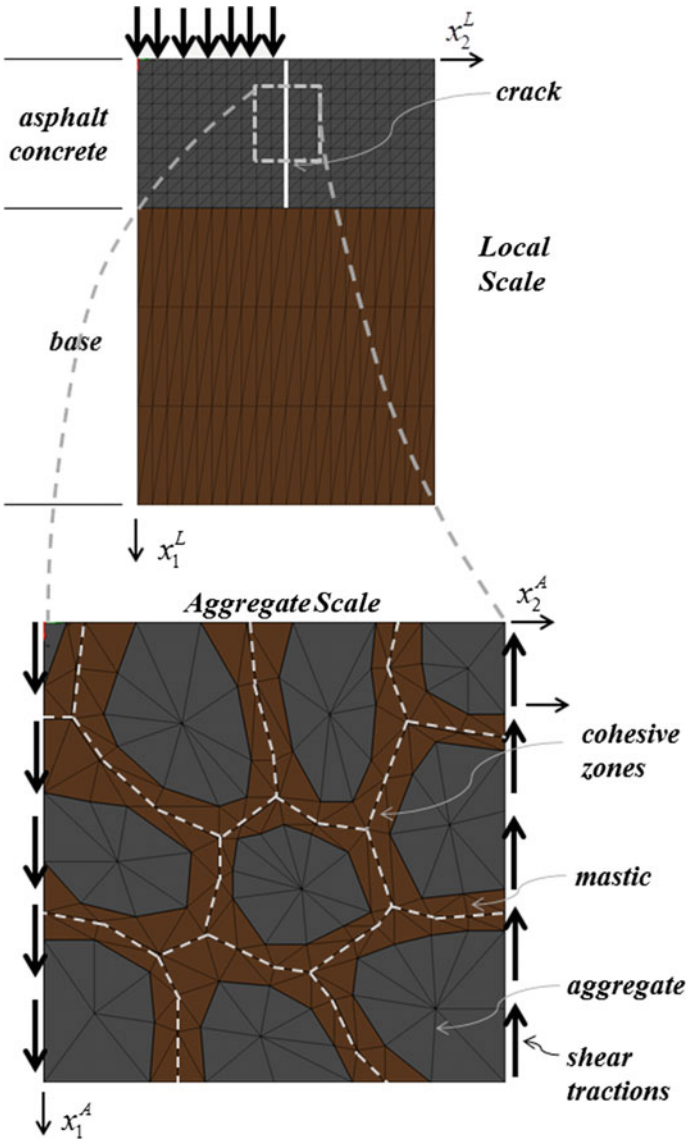


Fig. 15.42 Depiction of downward multi-scaling from the local length scale to the aggregate length scale

road way length scale. The globally calculated stresses obtained from this first loading increment are then applied to each and every local scale RVE, thereby producing spatially varying predicted states in the evolutionary cracking in each RVE, with greater cracking occurring in those RVEs that undergo greater global stresses. This process causes the local analyses to be dependent on (coupled to) the

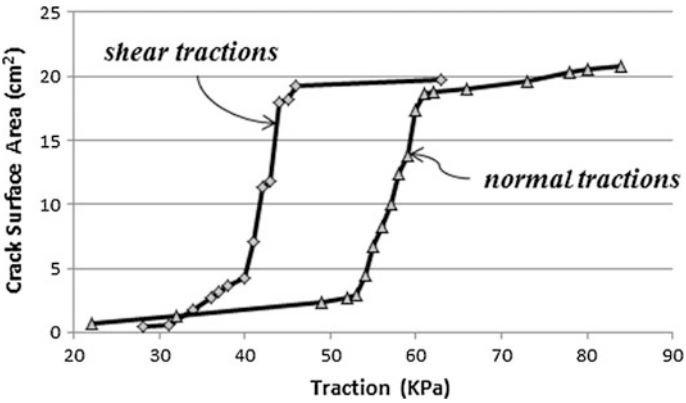


Fig. 15.43 Comparison of predicted crack growth due to normal and shear tractions applied to the RVE shown in Fig. 15.37 (Allen et al. 2017b)

global analyses due to the spatial variation in the global scale stresses. Once the local analysis has been performed within each local RVE on the given time step, the predicted results can be utilized to construct *the homogenized tangent modulus* within each RVE, and this modulus can be seen to vary spatially due to the different states of cracking predicted within each RVE.

In order to determine the homogenized tangent modulus within each RVE, consider a generic RVE with cracks, as shown in Fig. 15.47.

The volume, V^L, of the RVE is composed of mastic volume, V_M^L, aggregate volume, V_A^L, and crack volume, V_C^L, and the surface, S^L, of the RVE is composed of the exterior surface, S_E^L and crack surface, S_C^L. Because there are cracks within the RVE, the procedure for determining averaged properties within the RVE must be modified somewhat from that discussed previously. Whereas the volume average of these variables is utilized when the RVE is undamaged, this is not strictly correct when there are cracks present. This is due to the fact that stress is defined to be a force per unit area acting on a *surface*, so that the proper means of calculating the stress and strain are to perform a *boundary average* of these variables on the exterior surface of the RVE (Costanzo et al. 1996). Accordingly, the surface-averaged incremental stress is defined as follows:

$$\Delta\Sigma_{ij}^G \equiv \frac{1}{2V^L}\int_{S_E^L}\left(\Delta\sigma_{ik}^L n_k^L x_j^L + \Delta\sigma_{jk}^L n_k^L x_i^L\right)dS \tag{15.8}$$

where n_k^L are the components of the unit outer normal vector. Similarly, the exterior surface-averaged incremental strain within the RVE is defined as follows (note that the symbol E is the Greek symbol capital epsilon rather than E, the Latin symbol, which is utilized for the modulus tensor):

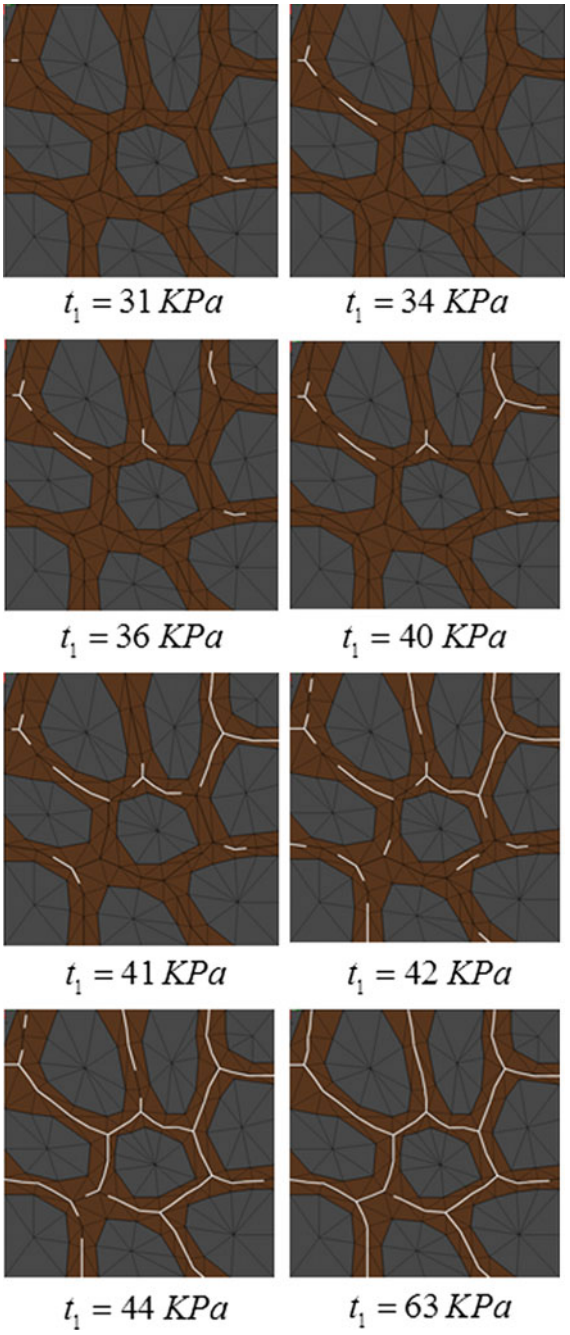


Fig. 15.44 Predicted evolution of cracking due to shear traction t_1 (Allen et al. 2017b)

Fig. 15.45 Photograph showing typical alligator cracking along wheel path within the roadway (*Source* Road Science, a Division of ArrMaz)

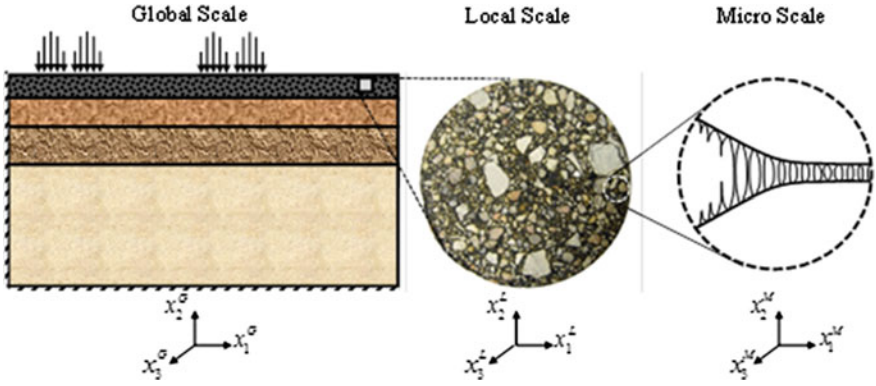


Fig. 15.46 Depiction of the two-way coupled multi-scale problem

$$\Delta \mathrm{E}_{ij}^{G} \equiv \frac{1}{2V^{L}} \int\limits_{S_{E}^{L}} \left(\Delta u_{i}^{L} n_{j}^{L} + \Delta u_{j}^{L} n_{i}^{L} \right) dS \tag{15.9}$$

Since the cracks are either traction-free or self-equilibrating, it can be shown that under quasi-static conditions (Costanzo et al. 1996):

$$\Delta \Sigma_{ij}^{G} = \Delta \bar{\sigma}_{ij}^{L} = \frac{1}{V^{L}} \int\limits_{V} \sigma_{ij}^{L} dV \tag{15.10}$$

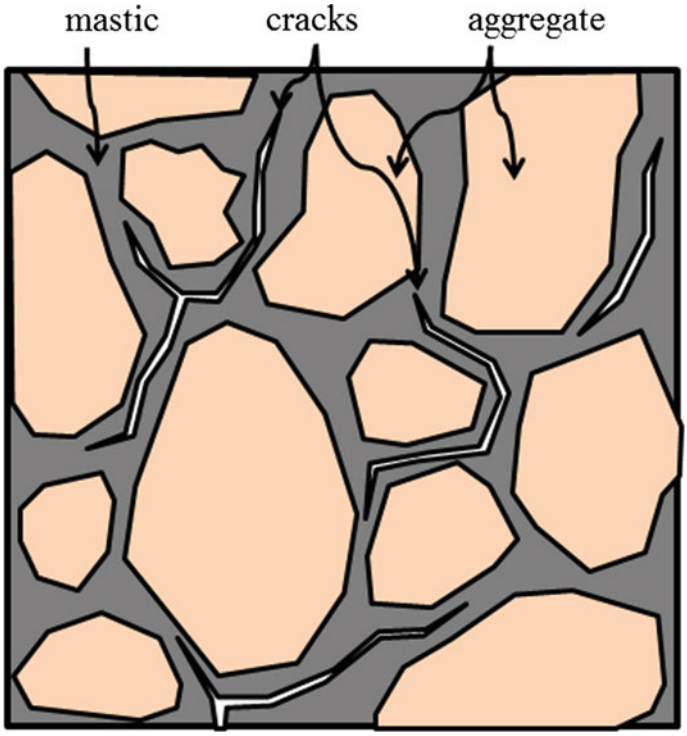


Fig. 15.47 Generic RVE with cracks

It should be noted that Eq. (15.10) is less cumbersome to calculate for a generic RVE than is Eq. (15.8), and this is the procedure recommended herein (Allen et al. 2017c).

As for Eq. (15.9), it may also be shown to be equivalent to the following (Costanzo et al. 1996, Allen and Yoon 1998, Allen 2002, Allen et al. 2017c):

$$\Delta \mathrm{E}_{ij}^{G} = \Delta \bar{\varepsilon}_{ij}^{L} - \frac{1}{2V^{L}} \int_{S_C^L} \left(\Delta u_i^C n_j^C + \Delta u_j^C n_i^C \right) dS \tag{15.11}$$

where $\bar{\varepsilon}_{ij}^{L}$ is the volume-averaged strain within the RVE, Δu_i^C are the components of the increment in the displacement vector on the crack faces, and n_i^C are the components of the unit outer normal to the crack faces. It should be apparent that either (15.9) or (15.10) can be utilized to calculate $\Delta \mathrm{E}_{ij}^{G}$. For the case wherein the cracks are all stationary at the end of the current load increment, it can be shown that (Allen et al. 2017c):

$$\Delta\Sigma_{ij}^{G} = C_{ijkl}^{G}\Delta\mathrm{E}_{kl}^{G} \tag{15.12}$$

where C_{ijkl}^{G} is the tangent modulus tensor for the current loading increment to be determined from the analysis of the RVE. Note that if the cracks are not stationary, the above equation will in general overestimate the value of the tangent modulus tensor. In the case of road ways, wherein cracks generally propagate in a stable way over many loading cycles, this overestimate may be inconsequential.

As described in Chap. 11, due to symmetry of the locally averaged stress and strain tensors, as well as thermodynamic constraints, there are at most twenty-one unique coefficients in C_{ijkl}^{G}. Furthermore, if it can be approximated as orthotropic, there are only nine unique coefficients in C_{ijkl}^{G}. These can be determined by applying small one-dimensional shear and normal displacement increments to the RVE in each of the coordinate directions at the end of each time step, thereby resulting in the calculation of the nine unknown components of C_{ijkl}^{G}. However, this is a rather cumbersome process that may not be necessary for the purposes of engineering accuracy.

As a simplification, when the damaged RVE may be assumed to be reasonably isotropic, the inverse of Eq. (15.12) simplifies to the following form on each time step (Allen et al. 2017c):

$$\Delta\mathrm{E}_{ij}^{G} = D_{1}^{G}\Delta\Sigma_{kk}^{G}\delta_{ij} + 2D_{2}^{G}\Delta\Sigma_{ij}^{G} \tag{15.13}$$

where D_{1}^{G} and D_{2}^{G} are time-dependent compliances that may be determined directly from the incremental tractions applied to the RVE during the time step. This can be accomplished as follows: First note that in Eq. (15.13), shearing is uncoupled from dilatation, so that the incremental coefficient D_{2}^{G} may be calculated using the following formula, obtained by setting $i = 1$ and $j = 2$ in Eq. (15.13):

$$D_{2}^{G} = \frac{\Delta\mathrm{E}_{12}^{G}}{2\Delta\Sigma_{12}^{G}} \tag{15.14}$$

Similarly, the incremental coefficient D_{1}^{G} may now be determined by contracting Eq. (15.13), thereby resulting in the following formula:

$$D_{1}^{G} = \frac{\Delta\mathrm{E}_{kk}^{G}}{3\Delta\Sigma_{kk}^{G}} - \frac{2}{3}D_{2}^{G} \tag{15.15}$$

from which D_{1}^{G} may be calculated.

In order to obtain the instantaneous tangent modulus tensor, Eq. (15.13) must now be inverted, thereby resulting in the following:

$$\Delta\Sigma_{ij}^{G} = C_{1}^{G}\Delta\mathrm{E}_{kk}^{G}\delta_{ij} + 2C_{2}^{G}\Delta\mathrm{E}_{ij}^{G} \tag{15.16}$$

where

$$C_1^G = -\frac{D_1^G}{2D_2^G(3D_1^G + 2D_2^G)},\ 2C_2^G = \frac{1}{2D_2^G} \tag{15.17}$$

For purposes of the implementation to a finite element program, it may be simpler to write Eq. (15.16) in matrix form using Voigt notation (see Chap. 11) which is as follows (Allen et al. 2017c):

$$\begin{Bmatrix} \Delta\Sigma_1 \\ \Delta\Sigma_2 \\ \Delta\Sigma_3 \\ \Delta\Sigma_4 \\ \Delta\Sigma_5 \\ \Delta\Sigma_6 \end{Bmatrix} = \begin{bmatrix} C_1 + 2C_2 & C_1 & C_1 & 0 & 0 & 0 \\ C_1 & C_1 + 2C_2 & C_1 & 0 & 0 & 0 \\ C_1 & C_1 & C_1 + 2C_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & C_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & C_2 \end{bmatrix} \begin{Bmatrix} \Delta E_1 \\ \Delta E_2 \\ \Delta E_3 \\ \Delta E_4 \\ \Delta E_5 \\ \Delta E_6 \end{Bmatrix} \tag{15.18}$$

Under plane strain conditions, Eq. (15.18) reduces to the following:

$$\begin{Bmatrix} \Delta\Sigma_1 \\ \Delta\Sigma_2 \\ \Delta\Sigma_6 \end{Bmatrix} = \begin{bmatrix} C_1 + 2C_2 & C_1 & 0 \\ C_1 & C_1 + 2C_2 & 0 \\ 0 & 0 & C_2 \end{bmatrix} \begin{Bmatrix} \Delta E_1 \\ \Delta E_2 \\ \Delta E_6 \end{Bmatrix} \tag{15.19}$$

$$\Delta\Sigma_3 = C_1(\Delta E_1 + \Delta E_2)$$

Also, under plane stress conditions, Eq. (15.18) reduces to the following:

$$\begin{Bmatrix} \Delta\Sigma_1 \\ \Delta\Sigma_2 \\ \Delta\Sigma_6 \end{Bmatrix} = \begin{bmatrix} \frac{(D_1 + D_2)}{4(D_1D_2 + D_2^2)} & \frac{-D_1}{4(D_1D_2 + D_2^2)} & 0 \\ \frac{-D_1}{4(D_1D_2 + D_2^2)} & \frac{(D_1 + D_2)}{4(D_1D_2 + D_2^2)} & 0 \\ 0 & 0 & \frac{1}{4D_2} \end{bmatrix} \begin{Bmatrix} \Delta E_1 \\ \Delta E_2 \\ \Delta E_6 \end{Bmatrix} \tag{15.20}$$

$$\Delta E_3 = D_1(\Delta\Sigma_1 + \Delta\Sigma_2)$$

The spatially varying tangent modulus is then utilized within the corresponding global finite element on the subsequent time step, thereby causing the global analysis to be dependent on (coupled to) the local analyses. Applying this procedure repeatedly within a time marching scheme produces the desired two-way coupled multi-scaling algorithm, as described by the flowchart shown in Fig. 15.48.

There are certain shortcomings to performing two-way coupled analyses, including the following:

- The procedure is both mathematically and computationally complex;
- Modeling long-term response of the road ways requires considerable computer time;
- There are currently no readily available user-friendly multi-scale finite element codes available for public use;

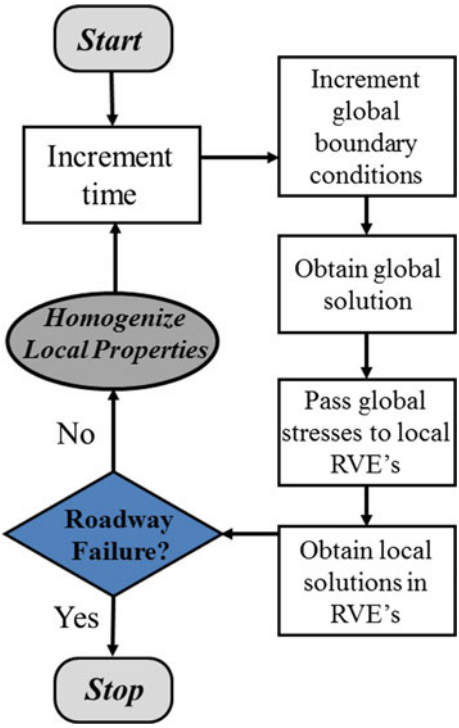


Fig. 15.48 Flowchart describing the two-way coupled multi-scaling algorithm

- The procedure may produce spurious results when deployed by novices; and
- Simpler one-way coupled analyses may be sufficient for many design purposes.

On the other hand, the two-way coupled approach is gaining rapid interest due to the following advantages:

- Design variables on widely differing length scales can be accounted for directly in the design process;
- Multiple analyses such as those described for expanding and contracting multi-scaling are obviated; and
- Material properties are required only at the constituent scale, thereby decreasing laboratory experiment costs substantially.

Because two-way coupled algorithms are not generally available in either open-source or commercially available finite element codes at the time of this writing, it is necessary to develop one's own multi-scale algorithm. Accordingly, the authors and co-workers have over a period of several years developed a two-way coupled multi-scale finite element code computer code. This code currently contains the following capabilities for the purpose of modeling and designing flexible pavements:

- A user-friendly preprocessor for generating both global and local finite element meshes,

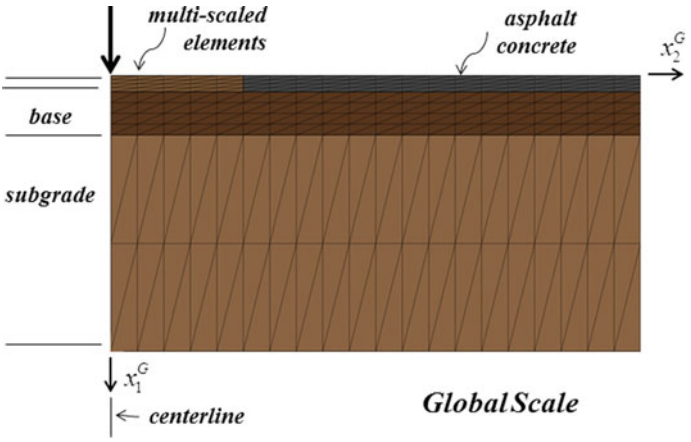


Fig. 15.49 Global finite element mesh utilized in two-way coupled multi-scale example

- Both two- and three-dimensional formulations,
- Both quasi-static and dynamic (explicit) formulations,
- Some post-processing capabilities, and
- A library of material models for various layers within a road way, including plasticity, viscoplasticity, and viscoelasticity.

The authors and co-workers have previously published extensively on the algorithm that is utilized in this section (Zocher et al. 1997; Allen and Yoon 1998; Allen and Searcy 2000; Phillips et al. 1999; Allen 2002; Soares et al. 2008; Allen and Searcy 2006; Souza et al. 2008; Souza and Allen 2009, 2011a, b), and for that reason, further detail regarding this procedure will be foregone in this text.

As an example of two-way coupled multi-scaling, consider once again the road way depicted in Fig. 15.18. Suppose that it is the intent of the designer to study the onset and evolution of cracking in the asphalt concrete layer. Toward this end, a global finite element mesh is constructed, as shown in Fig. 15.49. Note that for the purposes of computational thrift only the forty elements within the asphalt concrete layer nearest to the load application point are multi-scaled (as depicted in medium brown in Fig. 15.49). Each of these global elements is multi-scaled with the local mesh shown in Fig. 15.50. Material properties utilized throughout the simulation are depicted in Table 15.8.

The multi-scaled finite element mesh is subjected to a monotonically increasing point load applied at the road way centerline, as shown in Fig. 15.49, and this load is applied at a constant rate of 0.25 MN/s. Figure 15.51 shows the evolution of cracking in RVEs 1, 3, 5, and 7, the four topmost local meshes closest to the load application point, as a function of the applied load (Allen et al. 2017c). As shown in the figure, crack growth at the local scale decreases with distance from the point of load application. For example, no damage is predicted in RVE 9 (not shown in the figure).

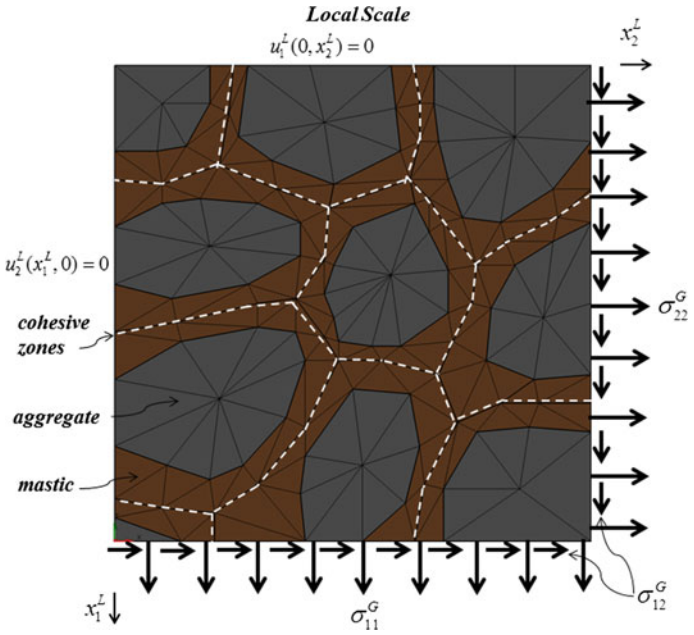


Fig. 15.50 Local finite element mesh utilized in two-way coupled multi-scale example

Figure 15.52 shows the predicted specific crack density versus applied force for the four RVEs shown in Fig. 15.51 (Allen et al. 2017c), where specific crack density is the predicted total crack length within the RVE at any point in time divided by the maximum cracking allowed within the RVE mesh (0.1825 cm). This example demonstrates the ability of the two-way coupled multi-scaling approach to model damage as a function of spatial and temporal coordinates.

Due to the fact that multiple finite element meshes are deployed simultaneously, as well as the necessity to account for nonlinearity via a time marching scheme, the multi-scaling approach described herein is computationally intensive, especially when modeling the response of road ways to cyclic loading conditions. When doing so, it may be necessary to utilize supercomputers to predict road way life, especially when modeling the road way in three dimensions. This difficulty notwithstanding, significant insight can nonetheless be gained for the purpose of designing road ways with the use of an ordinary desktop computer, and this can be accomplished by modifying the design variables one by one and predicting the resulting road way response for the first few loading cycles (Allen et al. 2017c).

In order to demonstrate how this design procedure may be employed within the two-way coupled multi-scaling approach, consider two cases, one wherein the base layer initial yield point is 6.0 MPA, and a second wherein the base layer is further consolidated so as to increase the initial yield point by 10%, all other design parameters held unchanged. As shown in Fig. 15.53, the maximum residual deformation (directly beneath the applied load) is plotted versus loading cycle

Table 15.8 Material properties utilized in two-way coupled multi-scale example

Global material properties										
1. Linear Elastic Asphalt Concrete Properties (not including multi-scaled elements)										
E = 1.5 GPa, ν ***= 0.36 (layer thickness = 0.15 m)***										
2. Elastoplastic Base Layer Properties										
E = 2.0 GPa, ν ***= 0.35,*** σ_Y^C ***= 6.0 MPa,*** μ ***= 0, H' = 1080 MPa,*** β ***= 1.0 (layer thickness = 0.4 m)***										
3. Linear Elastic Subgrade Layer Properties										
E = 2.0 GPa, ν ***= 0.35 (layer thickness = 2.0 m)***										
Local Material Properties (59% volume fraction of aggregate)										
1. **Linear Elastic Aggregate Properties**										
E = 2.0 ***GPa,*** ν **= 0.30**										
2. **Linear Elastic Asphalt Binder properties**										
E = 0.757 GPa, ν **= 0.40**										
3. **Cohesive Zone Properties**										
E_∞^C (GPa)	E_1^C (GPa)	E_2^C (GPa)	E_3^C (GPa)	E_4^C (GPa)	E_5^C (GPa)	τ_1^C (s)	τ_2^C (s)	τ_3^C (s)	τ_4^C (s)	τ_5^C (s)
2970	11740	336	33580	18450	8700	0.1443E-2	0.1443E-1	0.1443E0	0.1443E1	0.1443E2
$\delta_1^C = 1.0, \delta_2^C = 2.5, \alpha_1^C = 0.50,\ n^C = 0.025$										

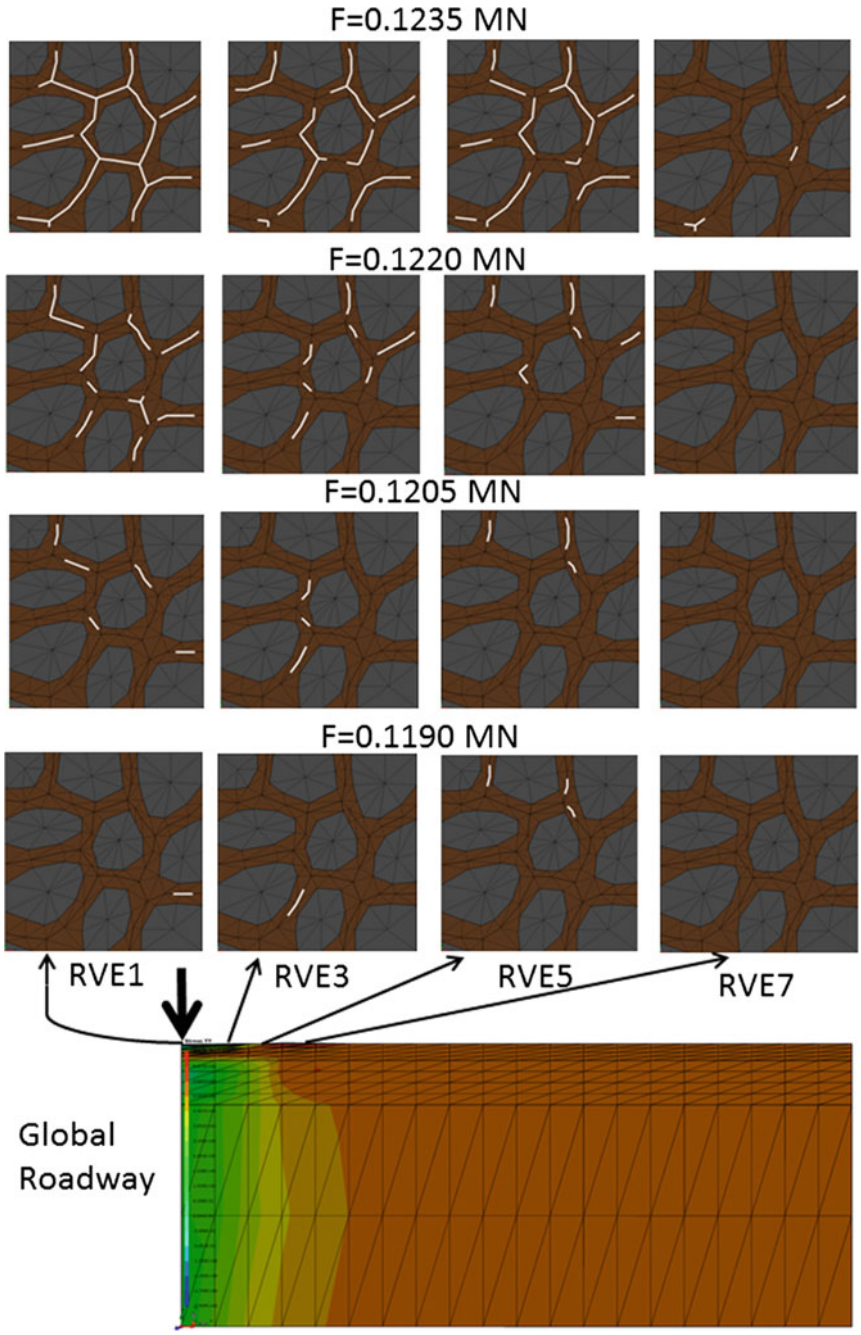


Fig. 15.51 Graphical depiction of predicted crack growth in selected RVEs as a function of applied force (Note that the stress component σ_{11}^{G} is shown for the global roadway) (Allen et al. 2017c)

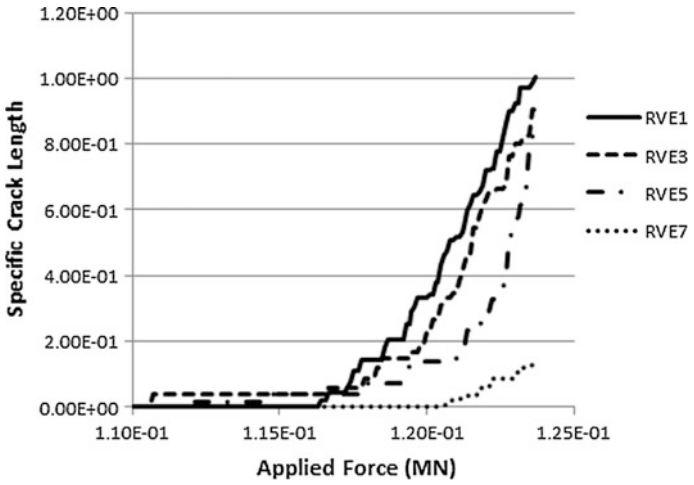


Fig. 15.52 Predicted specific crack density in selected RVEs as a function of applied force (Allen et al. 2017c)

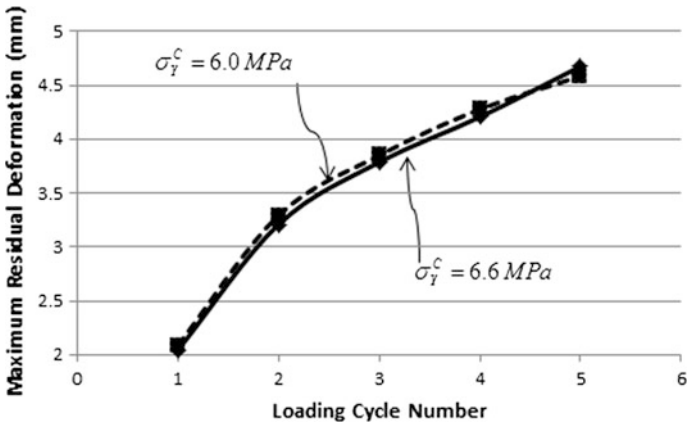


Fig. 15.53 Predicted maximum residual deformation as a function of loading cycle number for two different base layer yield points (Allen et al. 2017c)

number for the two different cases (Allen et al. 2017c). As shown in the figure, the predicted residual deformation for the weaker base material is predicted to be larger than the stronger base material for the first four cycles of loading, but on the fifth loading cycle, the two (somewhat counterintuitively) reverse, with the predicted residual deformation to be larger for the case wherein the base layer is further consolidated, thereby demonstrating the complex nature of road way design. In this design scenario, as shown in Fig. 15.54, the multi-scaling algorithm predicts that cracking in RVE3 (where the cracking is most pronounced) will occur in the asphalt

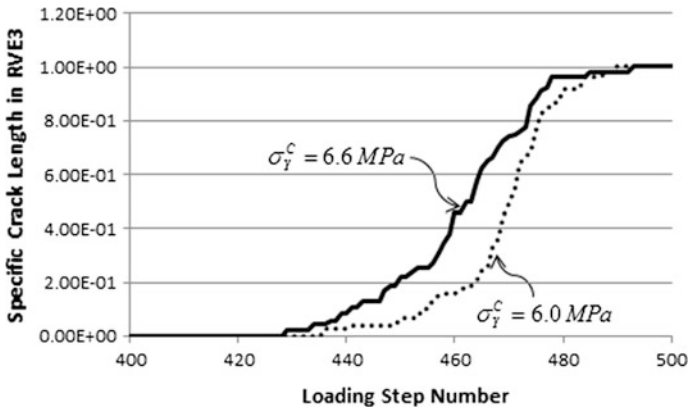


Fig. 15.54 Predicted crack length as a function of loading step number for two different base yield points (Allen et al. 2017c)

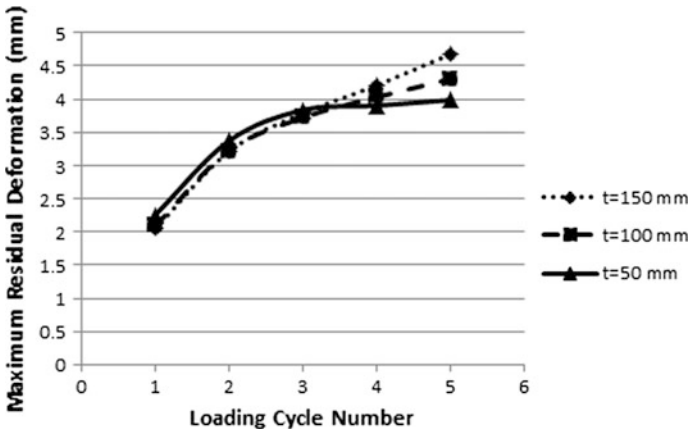


Fig. 15.55 Predicted residual deformation as a function of loading cycle for three different asphalt concrete thicknesses (Allen et al. 2017c)

concrete slightly sooner if the base layer is further consolidated, thereby enhancing residual deformations in the road way (Allen et al. 2017c).

Now consider the case wherein the same road way configuration is subjected to cyclic loading with time. The maximum residual deformation versus loading cycle number is now predicted for three different asphalt concrete thicknesses. As shown in Fig. 15.55, the multi-scale analysis results in smaller predicted residual deformations in the thicker pavement on the first three loading cycles (Allen et al. 2017c). However, on the fourth and fifth loading cycles, the predicted trend is once again reversed, with the thinner pavement undergoing smaller residual deformations.

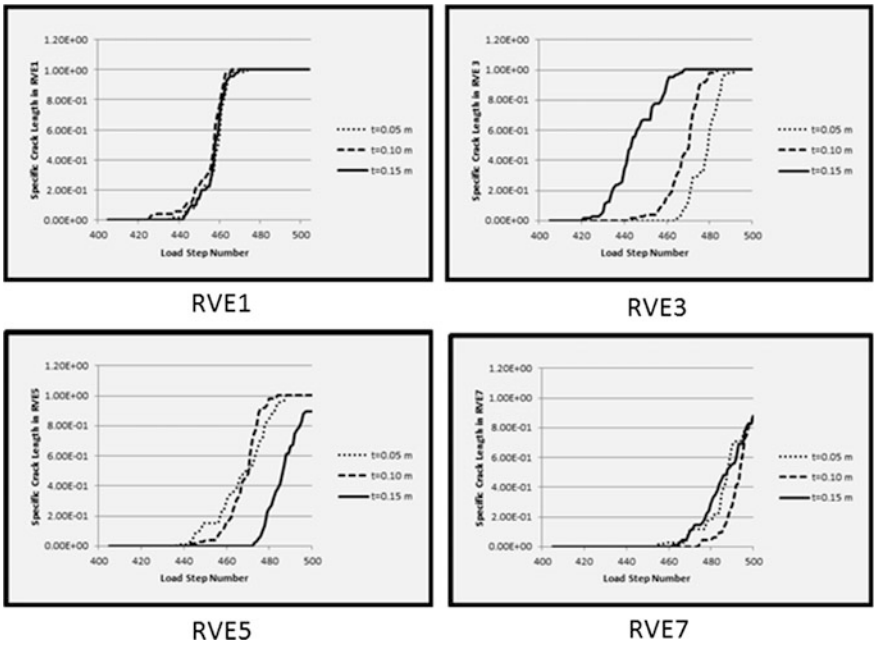


Fig. 15.56 Predicted RVE cracking as a function of load step number for three different pavement thicknesses (Allen et al. 2017c)

As in the previous design scenario, this counterintuitive result may be traced to the fact that the predicted residual deformations over the first three loading cycles are caused solely by plastic deformations in the base layer. As shown in Fig. 15.56, significant cracking is predicted thereafter, and while this cracking is comparable in RVEs 1, 5, and 7, cracking occurs much more rapidly in the thicker asphalt layer in RVE 3, thereby inducing larger residual deformation in the thicker pavement than in the thinner pavement (Allen et al. 2017c).

The previous two examples demonstrated the ability of the two-way coupled multi-scaling approach to predict the effects of pavement thickness and base yield point on road way performance, both of which are global scale design variables. Suppose, however, that the designer is more interested in determining the effects of local scale variables on road way performance. For example, suppose the examples described above are performed once again, but in this case, the volume fraction of aggregate is varied, all other design variables held fixed. Figure 15.57 shows predicted residual deformation versus loading cycle number for three different volume fractions of aggregate (Allen et al. 2017c). As shown in the figure, there is significantly less residual deformation for the pavement with 59% volume fraction of aggregate, but little difference between the 49% and 54% aggregate volume fraction cases.

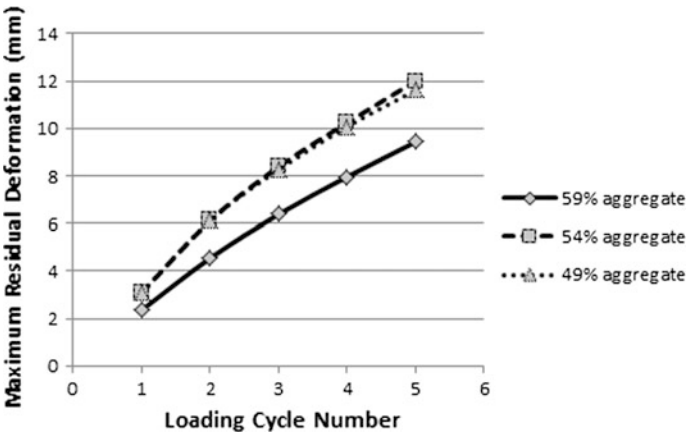


Fig. 15.57 Predicted residual deformation as a function of loading cycle number for three different volume fractions of aggregate (Allen et al. 2017c)

As shown in Fig. 15.58, the evolution of pavement cracking occurs in all locations later in the 59% aggregate volume fraction concrete than in the other two cases, thereby providing evidence that higher volume fraction concrete provides superior road way performance for this design scenario (Allen et al. 2017c).

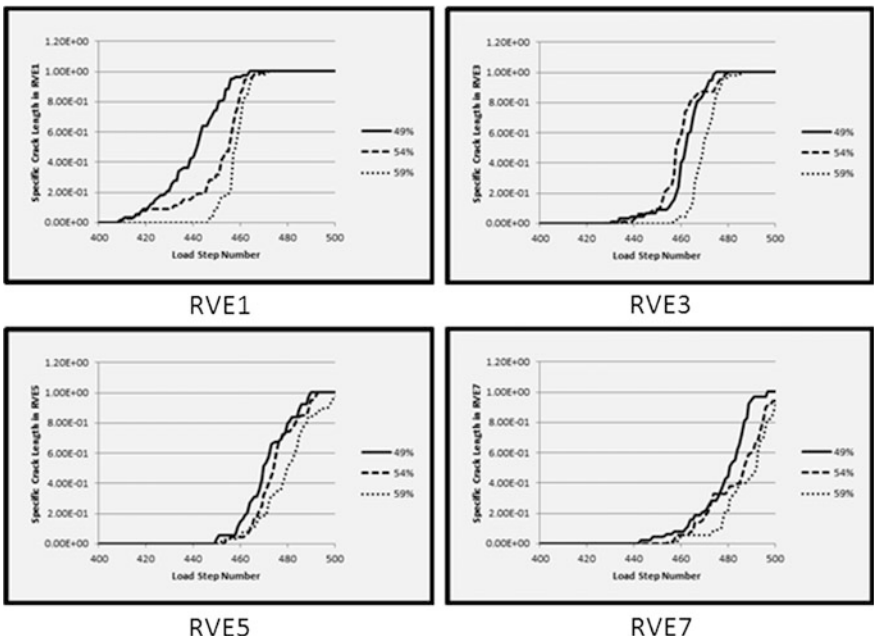


Fig. 15.58 Predicted pavement cracking as a function of load step number for three different volume fractions of aggregate (Allen et al. 2017c)

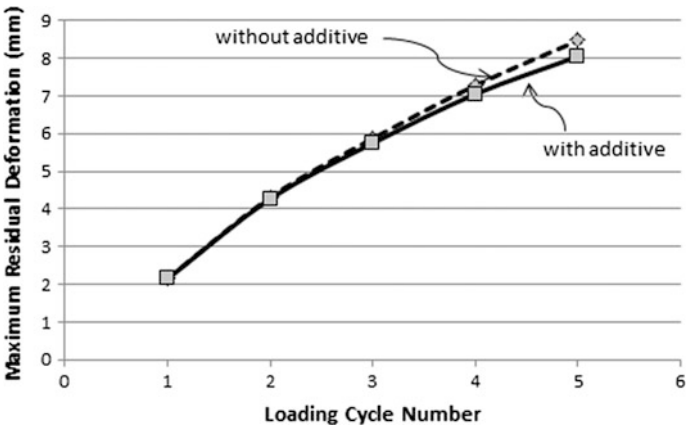


Fig. 15.59 Predicted maximum residual deformation as a function of loading cycle number for two different cohesive zone relaxation moduli (Allen et al. 2017c)

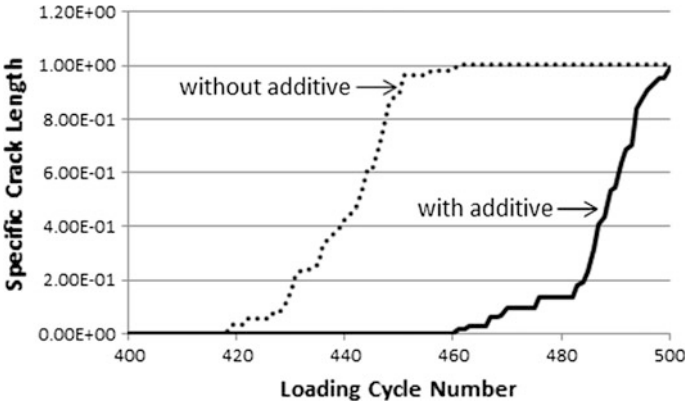


Fig. 15.60 Comparison of cracking as a function of loading step in RVE1 for treated and untreated asphalt concrete (Allen et al. 2017c)

As a final demonstration of the power of the two-way coupled multi-scaling approach for designing road ways, consider the case wherein the designer has access to an asphalt filler that produces a 10% increase in the cohesive zone properties. Predicted maximum residual deformation versus cycle number is shown for two cases in Fig. 15.59, and while the results are not dramatically different, there is a clear trend toward improved performance with the filler included in the asphalt concrete (Allen et al. 2017c).

As shown in Fig. 15.60, this improved performance is once again traceable to the evolution of cracking, wherein it is found that the untreated asphalt cracks significantly sooner than does the treated asphalt (Allen et al. 2017c).

While the results shown herein may not hold true for other road way configurations, they demonstrate the power of the multi-scaling approach for the purpose of improving pavement design.

15.3 Summary

In this chapter, several different computational methods have been demonstrated for the purpose of modeling asphalt mixtures and flexible pavements. Specifically, the following computational methodologies have been deployed herein:

(1) Micromechanics,
(2) Expanding multi-scaling,
(3) Contracting multi-scaling, and
(4) Two-way coupled multi-scaling.

Each of the above approaches has its place in the modeling of flexible pavements.

15.4 Problems

Problem 15.1
Given: An isotropic linear elastic aggregate is embedded in an isotropic linear elastic binder with properties of the constituents given as follows:

$$E_A = 10\text{GPa}, \nu_A = 0.3$$
$$E_B = 1.0\text{GPa}, \nu_B = 0.4$$

Required: Using the finite element code of your choice

(a) Construct finite element meshes for three different volume fractions of aggregate given by $V_A = 0.40, 0.50, 0.60$ and a volume fraction of voids given by $V_V = 0.05$ in all three cases.

(b) Use your meshes to estimate both the Young's modulus and Poisson's ratio of the asphalt concrete by the finite element method for the three different volume fractions of aggregate.

(c) Plot your results on two graphs showing Young's modulus and Poisson's ratio versus volume fraction of aggregate and compare to the Voigt and Reuss bounds.

References

Allen, D., & Yoon, C. (1998). Homogenization techniques for thermoviscoelastic solids containing cracks. *The International Journal of Solids Structures, 35,* 4035.

Allen, D., & Searcy (2000) Numerical aspects of a micromechanical model for a cohesive zone. *The Journal of Reinforced Plastics Composites, 19,* 240.

Allen, D. (2002). Homogenization principles and their application to continuum damage mechanics. *Composites Science and Technology, 61,* 2223.

Allen, D., & Searcy, C. (2006). A model for predicting the evolution of multiple cracks on multiple length scales in viscoelastic composites. *The Journal of Materials Science, 41,* 6501.

Allen, D. (2014). *How mechanics shaped the modern world*. Springer.

Allen, D., Little, D., Soares, R., & Berthelot, C. (2017a). Multi-scale computational model for design of flexible pavement—part I: Expanding multi-scaling, *The International Journal on Pavement Engineering, 18*, 309.

Allen, D., Little, D., Soares, R., & Berthelot, C. (2017b). Multi-scale computational model for design of flexible pavement—part II: Contracting multi-scaling, *The International Journal on Pavement Engineering, 18*, 321.

Allen, D., Little, D., Soares, R., & Berthelot, C. (2017c). Multi-scale computational model for design of flexible pavement—part III: Two-way coupled multi-scaling, *The International Journal on Pavement Engineering, 18*, 335.

Costanzo, F., Boyd, J., & Allen, D. (1996). Micromechanics and homogenization of inelastic composite materials with growing cracks. *Journal of the Mechanics and Physics of Solids, 44,* 333.

Helms, K., Allen, D., & Hurtado, L. (1999). A model for predicting grain boundary cracking in polycrystalline viscoplastic materials including scale effects. *The International Journal of Fracture, 95,* 175.

Huang, Y. (2004). *Pavement analysis and design* (2nd ed.). Prentice Hall.

Moore, G. (1965). Cramming more components onto integrated circuits, *Electronics Magazine, 4*.

Phillips, M., Yoon, C., & Allen, D. (1999). A computational model for predicting damage evolution in laminated composite plates. *Journal of Engineering Materials and Technology*, *21*, 436.

Rodin, G. (1996). Eshelby's inclusion problem for polygons and polyhedra. *Journal of the Mechanics and Physics of Solids, 44,* 1977.

Soares, R., Kim, Y., & Allen, D. (2008). Multiscale computational modeling for predicting evolution of damage in asphaltic pavements. In A. Loizos, T. Scarpas & I. Al-Qadi (Eds.) *Pavement cracking mechanisms, modeling, detection, testing and case histories* (p. 599). CRC Press.

Souza, F., Allen, D., & Kim, Y. (2008). Multiscale model for predicting damage evolution in composites due to impact loading. *Composites Science and Technology, 68,* 2624.

Souza, F., & Allen, D. (2009). Model for predicting multiscale crack growth due to impact in heterogeneous viscoelastic solids. *Mechanics of Composite Materials, 45,* 145.

Souza, F., & Allen, D. (2011a). Modeling failure of heterogeneous viscoelastic solids under dynamic/impact loading due to multiple evolving cracks using a multiscale model. *Mechanics of Time-Dependent Materials, 14,* 125.

Souza, F., & Allen, D. (2011b). Multiscale modeling of impact on heterogeneous viscoelastic solids containing evolving microcracks. *The International Journal for Numerical Methods in Engineering, 82,* 464.

Standard Method of Test for Determining the Resilient Modulus of soils and Aggregate Materials. (2007). American Association of State Highway and Transportation Officials, AASHTO Report No. T307.

Zocher, M., Allen, D., & Groves, S. (1997). A three dimensional finite element formulation for thermoviscoelastic orthotropic media. *International Journal for Numerical Methods in Engineering, 40,* 2267.

Index

D.N. Little et al., *Modeling and Design of Flexible Pavements and Materials*,
https://doi.org/10.1007/978-3-319-58443-0