Smart Sensors, Measurement and Instrumentation 18

Johannes Lambrechts
Saurabh Sinha

Microsensing Networks for Sustainable Cities

Springer

Smart Sensors, Measurement and Instrumentation

Volume 18

Series editor

Subhas Chandra Mukhopadhyay
School of Engineering and Advanced Technology (SEAT)
Massey University (Manawatu)
Palmerston North
New Zealand
e-mail: S.C.Mukhopadhyay@massey.ac.nz

More information about this series at http://www.springer.com/series/10617

Johannes Lambrechts · Saurabh Sinha

Microsensing Networks for Sustainable Cities

Springer

Johannes Lambrechts
Engineering and Built Environment
University of Johannesburg
Johannesburg
South Africa

Saurabh Sinha
Engineering and Built Environment
University of Johannesburg
Johannesburg
South Africa

ISSN 2194-8402 ISSN 2194-8410 (electronic)
Smart Sensors, Measurement and Instrumentation
ISBN 978-3-319-28357-9 ISBN 978-3-319-28358-6 (eBook)
DOI 10.1007/978-3-319-28358-6

Library of Congress Control Number: 2015960411

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland

Preface

Research, design, implementation, maintenance, and replacement of smart technology in developing countries are very different from those in developed countries. Lack of funding, inadequate government backing for startups, and inequality are considered the most important deterrents to technological progress in developing nations.

Amid milestones such as the invention of the internet, trends towards miniaturization of technologies governed by Moore's law, the increasing popularity of social networking/awareness and the worldwide focus on automating information distribution, smart cities have evolved as central hubs in autonomous and ubiquitous information systems. A central theme of these systems is providing people with up-to-date, real-time, and relevant information about their environment. The range of applications is dramatic, but in a fast-growing urbanization era, people are becoming more mindful of anthropogenic pollution, its numerous sources and its effects on their own health.

Developing countries often rely on international collaboration and in many cases the goodwill of nonprofit organizations to assist in integrating new and sustainable technologies in urban and/or rural areas. This has been accomplished successfully in many cases, but the more serious problem is the long-term operation and maintenance of these systems, which require significant investments in skills development. This book, *Microsensing Networks for Sustainable Cities*, aims to provide content that is relevant to researching technology trends aimed at creating smart and sustainable cities in developing countries. The information provided in this book can be used as initial reference and research guidelines, applicable references to create awareness of obtainable and cost-effective technologies, issues to expect and address, which might not be recorded in publications but found in other sources (commonly in case studies) and consciousness of best practices when researching new and innovative solutions, as well as introductory mathematics. The intended readers of this book will mainly be researchers, research students and prospective practitioners in technology who are committed to identifying, monitoring, and modeling air and water pollution and other applications relevant to smart cities, but more importantly, to sustainability and continual reuse.

Microelectronic component advancements are enabling ubiquitous computing through propriety technologies such as wireless sensor networks and radio-frequency identification to create a global network of connected devices, termed the Internet of Things (IoT). These devices are spatially distributed in cities and across rural and suburban landscapes to monitor the environment. Their small size makes these devices unobtrusive and ideally invisible to the local population. The large drive in mobile computing and especially in cellphones and tablets have invigorated research and development in the field of mobile hardware, low-power devices, sensors, and mobile-friendly operating systems. The synergy of these disciplines is having a ripple effect in developed and developing countries alike and young students are eager to pursue research focused on technology improvements that contribute to the information age. A common denominator in any technology is its energy source. In developing countries, lack of stable and accessible electricity is driving research into renewable and sustainable energy sources.

Various alternative renewable energy sources and techniques of energy harvesting, such as wind energy, hydropower, solar radiation, and energy from mechanical deformation can power, for example, IoT devices. However, there is still a demand for higher voltage supplies (ideally obtained from utilities) to power central storage servers, long-range transceivers, and Internet-connected interface devices. Again, there is a significant difference between developing and developed countries' abilities to provide stable and continuous energy. Economic growth can be severely suppressed by a country's inability to supply its citizens with electricity (only 5.1 % of South Sudan's population have access to electricity); stagnating technological development is inevitable. Such statistics point to major underdeveloped infrastructure and difficult conditions to initiate smart city developments and future sustainability. The countries concerned rely on skills development for their sustainable future and this book aims to provide the necessary baseline to encourage developing countries to initiate new research in the field of microsensing networks for sustainable cities during rapid urbanization.

The final three chapters of this book focus less on technical considerations and mathematical derivations and more on studying planning possibilities for ecologically aware, smart, and sustainable cities in developing countries. This is done by highlighting aspects that have contributed to successful implementation in developed and developing countries.

The authors strongly believe that in order to contribute to sustainability in developing countries, particularly in urban environments, prospective researchers, students, and practitioners must identify and understand the underlying causes and effects of overpopulation, the enabling technologies for smarter cities and possible limitations (technological and economic). The authors have taken it upon themselves to write this book specifically to address these topics and empower developing nations to improve their quality of life through technology.

The authors would like to recognize the research-capacity grant of the Dept. Higher Education and Training, South Africa for sponsoring this work. Furthermore, the authors would like to recognize Dr. Riëtte de Lange, Postgraduate School, University of Johannesburg, South Africa, for her effective administration of this grant.

Contents

Chapter 1
Microsensing Networks for Sustainable Cities: Pollution as a Key Driving Factor

1.1 Introduction

Air and water pollution are inherently associated with increasing rates of urbanization and an increasing population, rapidly moving to cities (UNEP 2015). Urban megacities have increased in number around the globe, and previously branded megacities also deal with an exponential increase in their population. Africa currently (2015) has seven megacities and the growth in population and increasing number of citizens residing in cities (as opposed to bordering countryside, rural and suburban areas) are adversely affecting the environment and hence citizen's health. A drive towards smart cities, cities capable of providing their inhabitants with information about health-related issues, transportation, energy consumption, waste management and various other factors, is also intended to provide information on air and water quality. Advancements in technology have enabled sensing equipment to miniaturize, increasing the number of sensing nodes that can be placed in a city, hence increasing the resolution of informative data to its citizens. Planning for such infrastructure, however, is still relatively new, and many cities, especially in developing nations, desperately require specialists in the field to aid with building new and improved, smart infrastructure. Micro/nano-sensors have improved in robustness, implementation, cost-effectiveness and life-span and these improvements occur frequently. This book includes techniques to integrate these sensors in real-life applications and to group these techniques and provide system integration guidelines, reinforced with theoretical background.

Air and water pollutants, their sources and methods used globally to monitor these parameters are outlined in this chapter to provide background on these issues and what is currently being done in developed and developing countries globally. Low-cost, high-quality sensors are actively researched and are a field within the electrochemical sensor body of knowledge that is expanding owing to these urbanization issues, which have showed large increases in the last 50 years. Implementing the sensors as city-wide nodal distributions through wireless sensor

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_1

networks again requires careful planning, not only for effective use at present, but also to ensure the possibility of future upgrades and improvements to technology. Remote sensing through these networks has many advantages, and real-time monitoring offers the ability of rapid response to potentially harmful events. Individuals have grown accustomed, especially recently, to receive information about their environment *instantly* and almost expect this. Providing this information accurately can be achieved, although striving for fully autonomous infrastructure is a long-term investment, necessitating proper planning.

Mobile devices contain increasingly more sensors with powerful processing techniques and in addition to investigating air and water pollution sensing equipment, this book shows designers how to harness these devices and their processing capabilities to design systems cost-effectively. Energy-harvesting, such as solar-powered device design considerations, is explored in this book. Additional material and resources are developed for open-source mobile operating systems (Android®) and freely available programming languages (Python®) to allow designers to apply the book context easily to practical real-life systems and build new applications.

1.2 Urban Megacities

A city is a permanent location inhabited by humans, independent in its administrative, legal and historical idiosyncrasies. The population of a city can vary greatly; the Vatican City, a micro-city, has a population of approximately 800 (in 2015), and Tokyo, a megacity (a city with a population of more than 10 million people) houses almost 38 million people (2015). Some megacities have been classified as megacities for more than 50 years, such as New York City in the United States of America (USA), which was branded a megacity in the 1950s. New York City is considered a developed city in a developed nation and its infrastructure, and specifically important in this book's context, its waste management systems, are well established. Africa is considered a developing continent and it has seven megacities as of 2015: Nairobi (Kenya), Johannesburg/Pretoria (South Africa), Lagos (Nigeria), Kinshasa/Brazzaville (Democratic Republic of Congo), Cairo (Egypt), Accra (Ghana), and Khartoum (Republic of Sudan) (KPMG 2013). The abundance of natural resources in Africa has resulted in huge interest in African countries especially from countries outside the continent. Africa holds 90 % of the world's platinum, 70 % of its cotton, 50 % of its gold and 30 % of its diamonds, as well as abundant quantities of crude oil in the Gulf of Guinea (KPMG 2013). Africa has a young, growing, vibrant, urbanized population that promises future growth and investment in many of its countries. Investors are being lured to large African cities for lucrative and diverse industry potential. It is projected that by 2016, over 500 million Africans alone will live in urban centers and cities. Africa is more urbanized than India and nearly as urbanized as China. Large projects in Africa, such as the BRICS (Brazil, Russia, India, China, and South Africa) Cable Project to increase global access to high-bandwidth internet, in which 21 African countries are included, the Lesotho Queen Mamohato Memorial Hospital, the Blue Line of the Lagos Rail network, Durban Waste-to-Energy Project, O3b Networks to combine the reach of satellites with the speed of fiber-based internet, and the Ethiopia Djibouti Railway are some of these large projects and investments in Africa (KPMG 2012) Growth in African megacities has been slow owing to socio-political problems, but the political stability in many countries over the past decade has given investors reason to follow through on these projects. This inevitably leads to a growing population and growing urbanization. In the early 1800s less than 3 % of the global population lived in cities; this number exceeded the 50 % mark in 2008 and is expected to grow to 66 % in the next 25 years (UNWUR 2014). The reason for highlighting these points is that pollution, as a result of improper waste management and increasing industrial activity, will have to be dealt with in these developing nations, while also investing in improved systems in developed nations. To avoid, or at least minimize, pollution in the developing nations where multi-billion dollar projects are the driving force towards industrialization, proper systems must be in place to monitor the inevitable distribution of toxic gases and chemicals into the air and water. According to the World Health Organization (WHO) air pollution statistics for PM10 and PM2.5 particles, developing countries have the largest pollution contributions measured in $\mu g\ m^{-3}$. Values obtained between 2009 and 2013 showed that the 10 countries with the highest air pollution were Pakistan, Afghanistan, Bahrain, Senegal, Qatar, Bangladesh, the United Arab Emirates, Mongolia, Egypt and India. These Middle-Eastern and African countries have large populations, but not the largest in the world, which demonstrates that large cities do not necessarily have higher air pollution compared to smaller cities. It depends on the procedures and strategy of management of potential pollution sources, which are not evident in these countries. It is therefore necessary to implement such procedures and educate the inhabitants on the adverse effects on health that pollution presents to humans.

1.3 Smart Cities

A smart city uses currently available technologies to enhance living conditions through access to information about parameters that affect its inhabitants. These parameters include the status of its education and employment, utilities, transportation, energy consumption, health-related issues, water quality, air quality, waste management and any other relevant information that could potentially benefit the community. In a sense, many cities can already be considered smart, depending on the scope of technology and its reach. Reaching the status of a fully autonomous smart city is a long-term process and some cities aim to achieve such status faster, but there is no real quantity that qualifies a city as being smart, or not smart. The global market for striving for smart status is ever increasing with advancements in technology. Some areas of technology, such as miniaturization, have enabled implementation of sensors and actuators that have previously only been achievable in small quantities, as price and performance limitations could not be overcome. Challenges such as climate change, economic restructuring, online retail and entertainment, ageing populations, and public finances can be monitored and addressed using technology, and serve as a major drive towards such use. Three important advantages of using information technologies on city-wide scale are making more efficient use of physical infrastructure through technology-aided design and maintenance, citizen participation through electronic channels, and learning and adapting to changing circumstances in the city. The results can also be shared with a global audience to provide valuable information for growth and development of emerging and developing countries and their megacities. The internet, smart phones, radio-frequency identification (RFID) and wireless communication provide various means of gathering and sharing information between large groups of people and are currently considered the building blocks for smart cities.

Possibly a better description of a smart city is smart planning during urbanization. For example, on a smaller scale, what makes one household smarter than another? Or similarly, what makes one household's living room smarter than another? Is it one additional sensor that measures a parameter that the other does not, a more expensive monitoring system, or smarter planning for future improvements? The third option makes most sense, since there can never be a status where a city has reached its smartest level. However, fast urbanization and more citizens living in close proximity in major cities around the world need smarter city planning for the future. A smart city should provide its citizens with the information they require to make informed choices about their lifestyle, whether these concern transportation, health, retail, or entertainment.

Another possible guide to becoming a smart city is outlined in five steps: set the vision of an efficient and sustainable city, combine hardware and software solutions to improve the efficiency or urban operating systems, integrate operation and information to improve city efficiency, add innovation for a sustainable future and drive collaboration between global and local cities. Again, this does not define a smart city, or add any value in determining the status of the city on the smart-scale, but it offers a holistic approach to improving the quality of life of its inhabitants.

Currently there are a number of cities that employ technologies for the purposes mentioned above with examples such as Beijing, China, which has high-definition enforcement systems, supervision and monitoring for its expressway, closed-circuit television (CCTV) in many areas, traffic flow detection systems, weather detection and guidance systems, and integration in 10 control centers situated at strategic points in the city. In Dallas, Texas, the supervisory control and data acquisition system improves efficiency and resilience to disruptions of the water distribution system, and the integrated mobility management platforms improve the efficiency of multimodal transportation. In Paris, France, a smart grid energy management project at district level, including energy efficiency solutions for buildings and homes, automation systems, renewable installations and infrastructure to charge electric vehicles, contributes to its smart status. Also, in Quito in Ecuador, real-time adaptive traffic control systems covering 600 crossings, 6,700 traffic lights and 1,500 CCTV cameras, weather information for the city airport, city access control and web-based travel information all aim to improve the city's efficiency. In South Africa, construction has started on a new Chinese-financed US $7 billion city in Modderfontein, Gauteng, near Johannesburg. Work on the first set of 300 residential units and some of the roads are already under way. The Chinese firm Shanghai Zendai acquired 1,600 ha of land in Modderfontein in 2013 to build the city. Plans for the city include a central business district, churches, a library, hospital and medical facilities, a sport and international conference center, schools and low-cost housing, among others. The project will house around 30,000 families and create about 200,000 fixed jobs for the local community. These examples show that although each city implements a different vision for improving the life of its citizens, all of the cities (and other cities globally) make use of technology to achieve this goal; its application depends on the issues the city faces and deems important to solve.

During smart city planning, it is necessary to improve and optimize the management of natural resources, such as water and breathable air, for future sustainability through development of technology and business models to manage, treat, and improve the availability and quality of these resources and protect biodiversity. The air in densely populated cities is polluted in part by large industries on its borders, but increasingly by road traffic pollution from fuel combustion. The first step in combating air pollution is gathering information and intelligence on the relevant particles that pollute each area. Air pollution is commonly measured at specific points in fixed locations. Although the technology used at these locations is of high quality and give good and accurate results, it does not report information on nearby locations. Air pollution can vary greatly even over short distances and wind, temperature and humidity affect readings. This means that there is an opportunity for cities to increase the number of locations monitored for air pollution, as it affects all citizens, but not equally. Providing accurate and high-resolution information to its population is considered a step forward towards planning for a smart city. Water quality can also be extensively monitored, treated, and effective preventative measures can be applied if technology and planning are combined and managed. This book focuses on the two main contributors to the decline in human health due to urbanization, namely poor air and water quality resulting from lack of planning.

1.4 Air Pollutants

Air pollution contributes to global warming and has significant health risks for humans and other mammals. The WHO estimates around 4.6 million deaths annually (Rajasegarar et al. 2014) occur as a result of respiratory diseases caused by polluted environments. The air in the atmosphere consists mainly of nitrogen

(78 %), oxygen (21 %), and small proportions of carbon dioxide, CO_2 (0.035 %), water vapor, and noble gases such as argon (0.934 %) (Mackenzie and Mackenzie 1995). Air pollution has a familiar form as a haze visible in the sky, but it can also be invisible, which is arguably more dangerous, since it cannot be purposefully avoided. Burning of fossil fuels such as gasoline, wood or natural gas can produce carbon monoxide (CO) emissions. CO is a colorless, odorless, toxic gas. Sulfur oxides, particularly sulfur dioxide (SO_2), are produced in nature by active volcanoes, but also by human activity such as combustion of coal and gasoline in industrial processes. SO_2 in the presence of a catalyst such as nitrogen dioxide (NO_2) can cause acid rain in the form of H_2SO_4. In the visible spectrum, pollutants such as nitrogen oxides, particularly NO_2, forms a reddish-brown haze with a sharp and distinctive odor. NO_2 is again caused by high-temperature combustion but also in nature by electric discharges from thunderstorms. CO, SO_2, and NO_2 are the predominant concerns of air pollution, but various other harmful pollutants are present in the atmosphere (Chand and Cunningham 1970).

Atmospheric methane (CH_4) is a volatile organic compound (VOC) and other hydrocarbon VOCs can generate ozone (O_3), effectively increasing the lifetime of CH_4 in the atmosphere. CH_4 is not a toxic gas but it can displace oxygen (O_2), causing asphyxia if O_2 content drops below a safe level. Particulate matter (PM) consists of small particles of solid or liquid suspended gas that can occur naturally from active volcanoes, dust storms, of forest fires, but can also be generated by human activity such as burning of fossil fuels in vehicles, power plants, or similar industrial processes. These particles can influence lung function and in severe cases lead to lung cancer. Atoms or molecules with unpaired valence electrons from ionizing radiation, high-temperature reactions, electrical discharges and electrolysis are linked to cardiopulmonary disease. The ozone layer is damaged by chlorofluorocarbons allowing harmful ultraviolet radiation from the sun to reach dangerous levels on the surface of the earth, increasing the probability of skin cancer and damage to sensitive organs such as eyes. A hazardous compound that reacts with oxides of nitrogen and sulfur is ammonia (NH_3), which is often associated with agricultural processes. It is a colorless gas with a distinctive pungent smell, recognizable in household general purpose cleaning solutions. Although not toxic to humans, it can be converted to carbamoyl phosphate and enter the urea cycle. Aquatic animals, however, do not have a mechanism to excrete this from their bodies and therefore it is classified as harmful to the environment. Ferrite gas sensors have proved to be cardinal in measuring and identifying gaseous compounds in the environment; research on this was conducted by Gadkari et al. (2011). Compounds from toxic metals such as mercury (Hg), which is used in commercial products, for example thermometers, and lead (Pb) (Mudgal et al. 2010), used as organic compounds in vehicle fuel additives, can cause a multitude of health risks such as convulsions, coma, renal failure and death, depending on the exposure. Arsenic (As) and cadmium (Cd) are also categorized as toxic heavy metals and can be measured with techniques such as capacitive measurement systems, as described by Baglio (2003).

1.5 Air Pollution Sources

To quantify the dispersion of air pollutants within the atmosphere, mathematical, numerical and measurement techniques are used. Parameters required to model the dispersion of pollutants are identifying the toxic substances, their source and whether it can be modelled as a point, area, or volume source. Source characteristics also include the emission rate of the pollutant, stack height, exit velocity and temperature of the gas, and the stack diameter. In addition, contributing meteorological conditions, such as wind direction and velocity, ambient temperature and atmospheric stability, have to be considered.

The Pollution Standards Index is an index number developed by the USA Environment Protection Agency. Outside of the USA, different countries use different names for their indices such as Air Quality Health Index, Air Pollution Index and Pollutant Standards Index. The index represents the highest concentration of the five most common pollutants in the atmosphere, averaged over a 24 h period. These five pollutants are PM, importantly PM10 (coarse dust particles), which are PM with aerodynamic particles with a diameter smaller than 10 μm but larger than 2.5 μm, CO, SO_2, O_3 and NO_2. According to the WHO, PM affects more people than any other pollutant. The major components of PM are sulfate, nitrates, ammonia, sodium chloride, carbon, mineral dust and water. It consists of a complex mixture of solid and liquid particles of organic and inorganic substances suspended in the air. Particles with aerodynamic diameter smaller than 2.5 μm (fine particles) are dangerous to humans and other mammals, since if these particles are breathed into the lungs, they can reach peripheral regions of the bronchioles, potentially interfere with gas exchange in the lungs and cause severe and long-term damage.

The two largest sources contributing to air pollution are anthropogenic and natural sources. The anthropogenic impact on the environment refers to the impact due to human activity (man-made sources). An informal equation developed to quantify the human impact on the environment is the $I = PAT$ equation, where I is the human **i**mpact, P is the **p**opulation of the region, A is the **a**ffluence (wealth), and T refers to the **t**echnological advancements. Among the many anthropogenic sources, this equation specifically examines the influence of technology in a region on pollution. An increase in population results in increased land use, increased resource use, and hence increased pollution. Proportional to an increase in affluence, higher consumption per capita increases production requirements, again leading to more pollution from industrialization. Although an 'increase' in technology might refer to more efficient and environmentally friendly manufacturing techniques, the efficiency is generally lower than the increase in demand and still contributes negatively to pollution. Higher demand in the energy industry to generate bio-fuels, electricity for households and businesses, the oil shale industry, nuclear power generation and even wind power disrupts the natural environment and leads to climate change. Efficient energy is usually coupled with higher costs and an increase in trade and services and can have a larger impact on the environment during the harvesting process, defeating the purpose of its goal.

A controversial topic that highlights this is the production of hybrid cars aimed at zero emissions from its electrical drivetrain. Manufacturing a hybrid car requires technologically advanced techniques and highly automated assembly lines. This type of manufacturing process requires tremendous inputs of energy, particularly the forging of materials like steel, aluminum, glass and plastic. Interestingly, lightweight vehicles can sometimes be more energy-intensive to build than heavier cars because lighter metals like aluminum are harder to forge than stainless steel. Experts estimate that 10 to 20 % of a vehicle's total lifetime greenhouse gas emissions are released during the manufacturing stage alone (Unnasch et al. 2007). Manufacturing of commercial products is related to harvesting of energy, where cleaning agents, paint, pesticides, paper and pharmaceuticals are prepared in large factories emitting large quantities of CO into the atmosphere. Environmental protection organizations regulate these processes, but this does not eliminate the pollution; it only prolongs the process. Mining, agriculture and transport also affect biodiversity and toxic gas-related illness in humans and other living organisms.

Natural sources of pollution also exist, and not only humans are to blame for polluting the environment, although humans are the only cognitive species that allows its own habitat to be destroyed. Natural sources include volcanic activity that produces sulfur and ash, CO generated by wildfires, the decay of radium in the earth's crust that forms radon, a noble gas that is considered hazardous, methane emitted by animal feces and dust in its natural form in large areas of land where erosion is prominent. Although the exact ratio of anthropogenic versus natural sources of pollution is not known, it can generally be assumed that human activity such as the burning of fossil fuels could tip the balance and lead to climate change faster than it would occur naturally. In Gerlach (2011) the CO_2 emissions from volcanos are estimated at around 0.13–0.44 billion metric tons per year, whereas anthropogenic CO_2 emissions are estimated at 35 billion tons for the same year. This shows that human CO_2 pollution is far more substantial compared to natural causes (only constituting a maximum of 1.26 % of the total emissions. Nations that are identified as contributing to a large extend to the production of these harmful emissions are Pakistan (0.18 billion tons), Kazakhstan (0.25 billion tons), Poland (0.44 billion tons), and South Africa (0.44 billion tons).

The need to quantify and estimate the effects of air pollution by anthropogenic and natural sources is evidently a large driving factor to avoid illness from air pollution. A commonly used model to calculate the dispersion of air pollutant emissions is the complete equation for Gaussian dispersion modeling of continuous, buoyant air pollution plumes (C), and is given by

$$C = \frac{Q}{u}\frac{f}{\sigma_y}\frac{g_1 + g_2 + g_3}{\sigma_z\sqrt{2\pi}} \tag{1.1}$$

where Q is the source pollutant emission rate in g s^{-1}, u is the horizontal wind velocity along the plume centerline in m s^{-1}, f is the crosswind dispersion parameter, σ_y is the horizontal standard deviation of the emission distribution in m, $g = g_1 + g_2 + g_3$ where g is the vertical dispersion parameter, and g_1 is vertical dispersion with no reflections, g_2 is vertical dispersion for reflections from the ground, and g_3 is vertical dispersion for reflection from an inversion aloft. σ_z is the vertical standard deviation of the emission distribution. This equation requires the input of the pollutant plume centerline height above ground level, H, and to determine the plume rise the Briggs equations can be used. Briggs divided air pollution plumes into four general categories, namely cold jet plumes in calm ambient air conditions, cold jet plumes in windy ambient air conditions, and hot buoyant plumes, also in calm ambient and windy conditions. This model is further investigated and applied in this book.

1.6 Water Pollutants

Water pollution is the contamination of natural water bodies by chemical, physical, radioactive or pathogenic microbial substances (Hogan and McGinley 2014). The effects of water pollution increase the mortality rate of aqueous species and terrestrial species drinking water from these water bodies, causing a reduction in biodiversity and extinction of entire ecosystems. Without fresh water of sufficient quantity and quality, sustainable development is not possible. Microbial diseases due to overpopulation lead to millions of deaths per year worldwide. Common examples of chemical pollutants are Hg released by mining activities, NO_x compounds used in agriculture, chlorinated organic compounds from sewage plants and various acids from manufacturing activities. Physical water pollutants generally refer to larger objects that are not necessarily toxic in their natural state but could be harmful to animals when consumed. A smaller contributor to chemical waste is radioactive waste, generally discarded from nuclear power plants. Radioactive waste should ideally not be a reality, but human error and negligent behavior can cause adverse effects from accidental spillage into land masses or more gravely, water bodies. A pathogen is anything that can cause disease, typically an infectious agent such as a virus, bacterium or parasite causing disease in its host. A pathogenic microbe is such an organism that cannot be seen with the naked eye and can cause an infection or disease. Common pathogenic microbes in natural water bodies are usually introduced from untreated sewage or surface runoff (storm water) from intensive livestock grazing. Widespread parasitic and bacterial disease occurs in developing nations where population density, water scarcity and inadequate sewage treatment are prevalent. Microbes can enter the body through respiratory tracts, gastrointestinal tracts, urogenital tracts or skin lesions. If these microbes reach their target in the host body, they can multiply rapidly and avoid attacks by the immune system; resulting in disease and even the death of the host. Many cities in Africa are unable to provide adequate water supply since most of the water is lost in pipe leakage (UNEP 2015). Many of these city dwellers are forced to boil their water or buy expensive bottled water. To effectively treat water pollution, its sources must be understood and defined to enable modeling and estimation of its effects.

1.7 Water Pollution Sources

Two major source categories are identified that cause widespread water pollution: Firstly, point sources, referring to physical sewage or factory waste released into natural water bodies, and secondly, non-point sources that define agricultural and urban storm water runoff.

Point sources, similar to radioactive waste, are sources that are only introduced when human error or negligence occurs and waste from sewage or factories are spilled in natural water bodies. Regulatory bodies aim to minimize these effects and such behavior is punishable by law, usually resulting in the guilty party having to pay a fine and having to take measures to ensure such an event does not occur again.

Non-point sources are, however, more difficult to control, since their pathways to water bodies are usually across land masses in the event of rain or storms washing organic or inorganic chemicals towards natural water systems. Intensive agricultural areas and high-density urban areas can cause inorganic chemical water pollutants such as monopotassium phosphate (KH_2PO_4), a nutrient source fertilizer in the greenhouse trade, and ammonium nitrate (NH_4NO_3), a high-nitrogen level fertilizer, to enter drinking water streams. Erosion control techniques such as contour plowing and crop rotation can prevent these runoff events, but again, a regulatory body is required to ensure these practices are implemented. Automotive products and improper storage and use thereof, such as methanol (CH_3OH) and ethanol (CH_3CH_2OH), gasoline and oil compounds such as octane can also leak from non-point sources into the water systems. The effect of these non-point streams far away from the event can be traced using hydrology transport models. These models mathematically simulate the river or stream flow and calculate the water quality parameters at distant locations. The key component of a hydrological transport model is the surface runoff element, which allows the assessment of sediment, fertilizer, pesticide, and various other chemical pollutants. Such a model should be able to analyze the effects of land use and climate changes on in-stream water quality with consideration of groundwater interactions. Worldwide, these models have been implemented, although each country develops its model differently. Models exist in Australia, China, Japan, Italy, Europe, Canada, Scandinavia and Portugal, where some of these models have evolved to include data on remote sensing and changes in geographical information that can occur during such an event.

The two main sources of water pollution are also divided into six basic categories by the United States Environmental Protection Agency. These categories are: biodegradable waste, consisting predominantly of human and animal waste providing an energy source of organic carbon for bacteria. Plant nutrients such as phosphates and nitrates entering water supplies through sewage and livestock fertilizer runoff. Natural of human-induced thermal pollution (heat) which increases the amount of dissolved oxygen (DO) content and reducing aquatic life diversity. Organic solid matter sediment from non-point sources which clog municipal systems, smother aquatic life, and causing water to become increasingly turbid. Human-made hazardous and toxic materials that are not disposed of properly. Radioactive pollutants from factories, hospitals, and uranium mines that may take many years before it is no longer considered dangerous. In general, it is difficult, sometimes even impossible, to estimate the amount of water pollution that originates in specific parts of the world since many pollutants come from non-point sources.

1.8 Air Pollution Monitoring

Countries worldwide each has a unique way of monitoring air pollution based on criteria set by their air quality standards regulating bodies. A large drive by companies to provide cost-effective, mobile and reliable air quality monitoring equipment is noticeable worldwide. Two main standards exist when monitoring air quality: primary standards that protect against adverse health effects and secondary standards that protect against welfare effects such as damage to farm crops, vegetation or buildings. There are common air pollutants that are considered criteria air pollutants, such as CO, NO_2, SO_2, O_3, PM and Pb. Although these pollutants are monitored in most cases, the procedures to obtain data from each region are different. To inform the general public of the current air quality, an air quality index (AQI) is an indexed number that many government agencies use to quantify air quality. The aim of the AQI is to standardize the ranges of numbers that indicate air quality. This model is used in essence worldwide, but each country implements it differently. Canada, for instance, implements the Air Quality Health Index (AQHI), which provides a number between 1 and 10 + to indicate the level of risk associated with current exposure. The health risks are divided into low (1–3), moderate (4–6), high (7–10), and very high (above 10). Hong Kong replaced the air pollution index with the AQHI in 2013. Its index is also on a scale of 1 to 10 + and considers NO_2, SO_2, O_3, PM10, and PM2.5. The health risk categories are subdivided into low (risk) (1–3), medium (4–6), high (7), very high (8–10), and serious (above 10). Mainland China implements an individual AQI score ranging between (non-linearly) 0 and 300 + for each of the criteria air pollutants over a period of one, eight, or 24 h. The final AQI score is the highest of the six scores with that particular pollutant being the city's major pollutant. India, Singapore, South Korea, the United Kingdom, Europe, and the USA each has its own implementation of the AQI with unique ranges and health indices.

Air pollution monitors range from relatively large stationary systems to cost-effective mobile sensors. Urban, industrial, roadside, and research-based sensors are available, as well as lower-cost solutions that interact with mobile devices such as smartphones to provide real-time air quality measurements for the surroundings. These sensors are advantageous when many users log information on their surroundings to create a well-populated map of air quality. Prototypes of these sensors are in circulation, with prices still relatively high (>US $1000) and could only be mass-produced if the prices are substantially lowered. Commercial air quality sensor products have been the topic of a large amount of research and development in recent years, and have produced products that are available for purchase at competitive prices. Several examples of such products exist, with prices below the US $1000 mark, depending on the number of sensors present and the quality (accuracy) thereof. These devices could potentially replace the traditional expensive, stationary, and complex instruments, although the quality of the measurements is not necessarily trusted by organizations such as the WHO for use in predictions and statistics (WHO 2013). The lower-cost devices are used by academics, industry, communities and individuals and symbolize the future of air-quality monitoring.

The WHO ranked Delhi in India as the world's most polluted city in a survey during May 2014. The survey stated that air quality values that affect the health of the city's inhabitants cannot be fully trusted for various factors. Although air quality sensors—such as the low-cost sensors mentioned previously—are available, the data are inconsistent and do not provide an exact snapshot of current circumstances. PM2.5 sensors are generally not calibrated against each other. The manufacturers also tend to self-certify their products, since currently there is no method to assess the quality of these instruments. The determining factor for the consumer is the price of the product, where the lowest bid is usually chosen above its technical merit. Lack of established standards makes it difficult to certify these products to a single baseline. Even if the readings are accurate, there is currently no way to compare these results officially to determine the impact of the measurements. Data fabrication is also prevalent in many circumstances. Companies that produce substantial amounts of toxic gases and chemicals are required to monitor their footprint and send the data to pollution control boards. Some companies install the equipment only to meet statutory requirements but can manipulate the data to reduce or eliminate penalties or legal action. Finally, air quality measurements are predominantly focused within densely populated urban areas and do not necessarily account for air quality outside these regions. In certain instances, a person walking or biking to work is exposed to more harmful gases compared to driving the same route inside an air-conditioned vehicle. The time of exposure is therefore significantly longer and even in non-urban areas, the effects can be just as harmful.

Providing a larger number of people, especially in developing nations, with the ability to monitor air quality with low-cost devices allows the gathering of additional data in locations not covered by continuous monitoring. Although this is ideal, such an endeavor is not always feasible. In many African countries, poverty and lack of clean drinking water outweigh the perception of clean air. If theoretically a large number of cost-effective devices could be provided to a large portion of a population in a developing nation, the nature of these measurements would require consistent and accurate placement in areas prone to pollution. Dust and aerosol at low levels above the surface of the earth, at levels of between 1 and 2 m, have a significant influence on measurements and do not convey the environmental pollution.

Another city ranked as one of the most polluted, according to the WHO, is Ulaanbaatar, Mongolia. Estimates suggest that one-tenth of deaths in the city are related to particulate air pollution. Extremely harmful concentrations exceeding 200 μg m^{-3} of both PM10 and PM2.5 are recorded in Ulaanbaatar (Davy et al. 2011). An international group of environmental journalists monitored the environment using mobile devices. These devices cost less than US $50 and according to environmental researchers, having the ability to purchase and use many low-cost sensors for the cost of regulatory sensors is tremendously powerful—despite the unproven accuracy of these devices. Monitoring air pollution is much more involved than manufacturers and suppliers of low-cost sensors suggest. Without attention to detail and input from pollution experts, the gathered data are essentially meaningless, since these are not comparable. Similar challenges exist in water pollution monitoring, where lack of standardized measurements and low-cost equipment influence the integrity of gathered data.

1.9 Water Pollution Monitoring

The International Organization for Standardization (ISO) defines monitoring as: "*the programmed process of sampling, measurement and subsequent recording or signaling, or both, of various water characteristics, often with the aim of assessing conformity to specified objectives*". From Bartram and Ballance (1996) this definition can be sub-divided into three separate categories where monitoring is a long-term standardized measurement of the aquatic environment to define trends. Also, surveys are of finite duration; they are intensive programs to measure and observe the quality of the aquatic environment for a specific purpose. Surveillance entails continuous, specific measurements and observations for the purpose of water quality management and operational activities.

Water sampling and analysis initially require a clear and concise objective of the required information that needs to be identified, similar to the medical practice where blood tests are done to determine specific contamination. Water is an absolute necessity for all life but can also be the carrier of many diseases. Infectious water-related diseases can be categorized as waterborne, water-hygiene, water-contact, and water-habitat vector diseases. Waterborne infectious diseases are classified when the pathogen is present in the water and ingested during consumption thereof. Water-hygiene diseases can be reduced by concentrating on high levels of personal, domestic and community hygiene. Almost all waterborne diseases are therefore water-hygiene diseases, among which scabies, skin and eye infections are prevalent. Water contact diseases are transmitted when skin is in contact with pathogen-infested water. Water-habitat vector diseases are transmitted by insect vectors that live near water.

The principal reason for monitoring water quality has traditionally been the need to verify whether the observed water quality is suitable for the intended uses, as described by Bartram and Ballance (1996). Water pollution monitoring and prevention can significantly decrease the cost of polluted water treatments and provide access to clean drinking water in more areas if distributed safely. Several prevention techniques exist, since water pollution originates from various sources. Green infrastructure is a technique that uses vegetation, soils and natural processes to manage water for healthier environments. In highly populated areas, green infrastructure refers to patches of natural land that facilitate habitat, flood protection, cleaner air and ultimately cleaner water. In rural areas, green infrastructure refers to storm water management to simulate natural processes of storing and soaking up large volumes of water. Multivariate statistical methods can also be employed to monitor water pollution, as described by Karami et al. (2012). Certain laws (particularly in the USA) require that responsible jurisdictions establish priority rankings for water bodies that are too polluted or degraded to meet water quality standards and provide a total maximum daily load, a calculation of the maximum amount of a pollutant that a water body can receive and still safely meet water quality standards. Permit programs can play an important role in minimizing the waste and pollution loads released into water bodies.

Somewhat different compared to air pollution, water pollution is generally more visible to humans, as is the act of pollution. Air pollution can come from various gaseous sources, identifiable by sight or smell, but if the gas has no distinct characteristic, or is unknown to the observer, the contaminant can easily be inducted into the atmosphere. Water pollution occurs more abruptly; this does not always happen, but it tends to emanate from negligent discharge of sewage or industrial waste products. For this reason, there is not a single recipe for monitoring and more importantly preventing water pollution, and a trial-and-error approach is sometimes followed. If something bad happened—do not let it happen again. Protection agencies play an extremely important role in advising industries and municipal plants, working from previous experience on how to prevent pollution. Equally important as the techniques employed to monitor pollution, is the hardware equipment, trustworthy and quality components are crucial for accurate and reproducible data.

1.10 Air Quality Sensors

Inherent physical limitations and a large number of possible applications lead to the study an active and complex field, namely gas sensors. Air quality sensing is important; it is used in industrial pollution sensing, healthcare, electronic manufacturing, the automotive industry, environmental studies, and increasingly for personal use. Humans are ever more aware of their surroundings and the effects these could have on their health. Gas sensors (Liu et al. 2012) are not confined to large industries that can afford expensive equipment and monitoring stations, but aim to become more and more affordable to individual users. Price versus the performance of gas sensors plays an important role in their use and availability. Sensors must be sensitive and selective to the target gas and have low power consumption to be integrated into mobile devices and operate for extended times. Many solid-state gases require elevated temperatures of up to 500 °C to be detected, making the power consumption of these devices inevitably high. This also limits the use of complementary metal-oxide semiconductor (CMOS) devices for sensors, as bulk material temperature limitations do not allow for operation at these temperatures (Gardner et al. 2010). Non-standard CMOS materials are also required to generate high temperatures (such as platinum) and are not compatible with CMOS processes.

Several methods for gas sensing are used commercially and for research purposes. The two main categories are based on methods that induce an electrical variation, such as change in resistance, capacitance or frequency of the material, for example CMOS sensors, polymers, carbon nanotubes or moisture-absorbing materials. Different variations, such as optical changes, acoustic methods, gas chromatographs and calorimetric methods are divided into a separate category of sensing types. Again, each category and sensing method has its own advantages and disadvantages based on its sensitivity, selectivity, response time, energy consumption, reversibility, adsorptive capacity, fabrication cost and ease of implementation. Sensors must also have stable and reproducible output over a set period of time. These are important considerations, since it ensures that readings are trustworthy and accurate, failure to incorporate these important characteristics is a major contributing factor to why low-cost sensors have not yet been accepted as standard products to report toxic gas concentrations. Various approaches to improve sensitivity and selectivity are researched, including approaches using dielectric resonators, thermostatic cycles, pre-concentrators, photo-acoustic spectroscopy, infrared (Frodl and Tille 2006), and sensor arrays. These methods are not yet foolproof and viable, and also do not necessarily contribute to both sensitivity and selectivity concurrently, but show reasonable improvements for future iterations of gas sensors.

1.11 Water Quality Sensors

Water quality sensors have similar drawbacks and limitations to gas sensors, with additional portability and placement issues to sense and monitor unhealthy water circumstances. Water quality sensors measure parameters such as pH levels, DO content, oxidation-reduction potential, conductivity (salinity) turbidity, temperature, and dissolved ions (such as Na^+, Ca^+, F^-, Cl, Br^-, I^-, CU_2^+, K^+, MG_2^+, NO_3^-). Applications for such sensors include portable water monitoring for common chemical parameters and pH, nitrates and DO, chemical leakage detection in rivers, pollution levels in oceans, corrosion and lime-scale deposit prevention, aquaculture monitoring and hydroponics. Many water-quality sensing systems require some sort of wireless transmission between the sensing module and the application software and database equipment. Technologies such as Zigbee® and the more recent Wifi protocols can provide such services, with battery-operated devices requiring little energy consumption to operate over extended periods of time, especially since monitoring is generally conducted in real-time over a series of days or weeks. Once an unhealthy substance is reported by a sensor, the source and spread of the contamination is not immediately identified and corrective and preventative action cannot be taken immediately. This leaves the contaminated area exposed for a longer period, during which measures must be taken to inform the general public and organizations responsible for adherence to quality. Water quality sensors generally consist of various single- or multi-parameter sensors, but implementation of both these systems is costly and limited to relatively few strategic points where water spread and destinations are known.

A variety of improved designs for water quality sensors are being researched and will be commercially available in the near future. Advancements in light-emitting diode technology in the lower ultraviolet (UV) range are being used for fluorometers to detect wastewater and other contaminants from urban and agriculture landscapes. Wet chemical sensors are increasingly being used for long-term in situ monitoring of soluble reactive phosphorus and ammonium in freshwater bodies. Not only chemical constituents are monitored, but also water levels, as decreasing water levels are equally hazardous to sustaining the environment. Optical, infrared, and microelectromechanical systems (MEMS) sensors offer selective, sensitive, relatively low-cost alternatives (Bhardwaj et al. 2015) to water level sensing and can provide long-term and accurate data for water bodies that require a steady and constant level. Contrary to gas sensors, water quality sensors are receiving somewhat less attention in design improvements of the sensing elements, as these measurements are less complex to achieve compared to air quality. Various research and development programs are focusing on the data capture and transmission between the sensor and the digital storage application, leading to enhanced techniques to monitor and log recorded information, with fast response to contamination detection a priority in these circumstances. Application-based research has led to smart sensors to monitor water quality in oceans (Adamo et al. 2015), solar powered devices (Kulkarni and Turkane 2013), and portable sensors with disposable electrodes (Grossi et al. 2013) to name a few. The important aspect in these examples is the connectivity, an integral part in the larger scope of applications, such as the Internet-of-Things.

1.12 The Internet-of-Things

Cisco® estimated the following statistics for mobile devices and networks in 2014; Global mobile traffic reached 2.5 exabytes (1 billion gigabytes) per month at the end of 2014. This is 30 times more than the monthly mobile network traffic in 2000. Mobile video traffic exceeded 50 % of total mobile traffic for the first time in 2012. Almost half a billion (497 million) mobile devices and connections were added to the global network in 2014. Smart devices accounted for 26 % of all mobile device connections in 2014. Mobile network (cellular) connection speeds grew 20 % in 2014, with a global downstream average of 1683 kilobits per second (kbps). There were 189 million laptops on the mobile network in 2014, with each laptop averaging 2.6 GB of data communication per month. By 2019, it is estimated that global mobile data traffic will exceed 24 exabytes per month and the number of mobile-connected devices will exceed the world's population.

Evidently, there are many *things* capable of digital data processing and advanced functions available globally. Each one of these *things* could potentially contribute to expanding the global availability of information regarding, for the sake of this book, the quality of our environment. Since it has been established that planning a smarter city could result in a city being branded smart, the fact that so many capable devices have already infiltrated these infrastructures highlights the potential of using the devices to achieve such a goal.

The Internet-of-Things (IoT) is a broad term that groups any device that is capable of manipulating data and exchanging such information with other devices. By this theory, billions of such devices are already used every day throughout the world. The applications of these devices are vast and no attempt is made to list them here. One inherent problem of having such a vast number of devices that gather and analyze data is the actual amount of data available but not necessarily being used or understood. Technology is evolving much quicker than our ability to harness its power to its full potential. There are almost 6.8 billion cellular telephone subscriptions online in 2015, and the global population is estimated to be exceeding 7.2 billion by this time. This means that essentially each person owns a device capable of digital processing (of course this is not true, since the statistics are not limited to one device per person) but it gives a general idea of the reach of such devices. It is somewhat over-zealous to expect all of these devices to work in harmony to provide sensible and accurate information on issues deemed important to the health of humans (and all other organisms for that matter), but it might not be as far-fetched as it seems. Accessing only 1 % of these devices to perform such meaningful tasks would already be a major step forward in establishing a global network of sensors, dedicated to reporting parameters such as current air and water quality of our immediate environment. Some possible applications and methods to take advantage of the availability of these devices are explored in this book, which aims to provide some guidelines on capturing their potential as outlined.

Projects such as the Air Quality Egg (AQE) is an open-source hardware, IoT platform developed through a crowd-sourced funding scheme to enable individuals (as opposed to government-run initiatives) to monitor airborne pollutants (NO_2 and CO). The data are uploaded to the internet and displayed on a map of the world, freely available for anyone to access. Each device reports its location and air quality at frequent intervals and has a built-in warning system if pollutant levels are above a safe threshold.

Accessibility of general packet radio service for mobile sensor arrays to monitor air pollution, as described by Al-Ali et al. (2010), uses city buses to collect information about CO, NO_2, and SO_2 from various mobile sensors, transmits the data to a central server and makes these data available over the internet. Such research projects aim to provide services to improve data availability on air pollutants in urban areas.

1.13 Wireless Sensing Networks

A wireless sensor network (WSN) consists of spatially distributed autonomous sensors to monitor physical environmental conditions and passes its data through (wireless) mesh network to a main location. The general idea of a WSN is therefore a series of nodes, each node consisting of a variety of sensors, depending on the application and required information gathering. The vision of these networks is based on the strength-in-numbers principle, since a multitude of sensors (depending on the cost of each sensor) work together to provide accurate information on each location in a system. Community sensing mechanisms based on incentives could potentially provide a good starting point to encourage citizens to participate in information gathering (Faltings et al. 2014). Typically, several main subsystems make up a node, consisting of a transceiver and antenna combination, electronic circuitry that interfaces with the sensors and energy source, generally a power source in the form of a battery to provide mobile implementation, or an embedded form of energy-harvesting (such as solar power) to provide long-term and sustainable operation of these components.

Various categories of application of such networks exist and can be categorized into process management for area monitoring, such as military intrusion detection and health care monitoring (Zhang et al. 2014), environmental/earth sensing for air pollution monitoring, forest fire detection, landslide detection, water quality monitoring, natural disaster prevention and chemical agent detection, and industrial monitoring, which includes machine health monitoring, data logging, wastewater monitoring and structural health monitoring. WSN are not limited to these applications and categories and can be implemented generally to monitor any parameter autonomously that could provide preventative and corrective maintenance faster because of its continuous monitoring and some sort of informing system.

There are several challenges when dealing with WSN systems and these are not easily overcome if not completely researched and understood when building such a network of devices. First and foremost is energy efficiency. These devices are generally placed in strategic locations that are less accessible. The devices can only operate if enough energy is supplied by a source, whether battery-powered, solar-powered or any other energy-harvesting technique. Ensuring that these devices operate consistently during times that energy can be harvested, and from reserves stored during this time, is essential for autonomous monitoring. In addition, the responsiveness of these devices may be critical in some applications (such as military intrusion detection), and choosing a sensor and components that make up a node able to provide this functionality is essential. The robustness of the devices could lead to lower maintenance costs and less frequent interaction with faulty devices. Another characteristic that benefits the longevity of these devices in a node is its ability to self-configure and adapt. As environmental conditions change and measured constituents vary, having the ability to adapt these sensors without replacing them could also lead to lower overall cost. Furthermore, scalability, heterogeneity, systematic design, privacy, and security are considerations during design and development of a WSN. WSN routing can be made robust and efficient by incorporating different types of local state information, such as the link quality, link distance, residual energy, and position information. Given unreliable and time-varying wireless links, it is important to provide robust delivery of information. Several techniques can be used to provide robust links, including the use of appropriate link quality metrics, multipath routing and wireless diversity-based routing techniques.

The size of each sensor varies greatly, depending on the application. MEMS advancements have enabled technology miniaturization to benefit WSN density of sensor placement, energy efficiency (since many sensors operate using passive structures) and a decrease in cost. MEMS devices in their most basic form (miniaturized mechanical and electro-mechanical elements) require three steps during manufacturing: the deposition of metal (or other) layers, patterning by photolithography, and etching to produce the required shapes. MEMS sensors are micro- and nanoscale devices that perform tasks in a similar way to their larger counterparts, with the benefit of a small footprint and increased resolution of sensing changes in the environment. The potential advantages of MEMS start to become apparent when these miniature sensors and/or actuators are merged onto a common silicon substrate together with CMOS integrated circuits. Components such as coplanar waveguides are commonly used to conduct electronic (generally microwave) signals on a dielectric substrate. This integration of technologies allows for true system-on-chip products. MEMS sensors and actuators already exist in commercially available products such as microphones, inkjet printers, accelerometers, gyroscopes, micro-fluidics and micro-scale energy-harvesting, with many more applications available. Nanotechnology offers the ability to manipulate matter at the atomic or molecular level to make something useful at the nano-dimensional scale. The one main criterion of MEMS is that there are at least some elements that have some sort of mechanical functionality whether or not these elements can move. In the case of micro-sensors, the device typically converts a measured mechanical signal into an electrical signal. These sensors can include accelerometers, gyroscopes, digital compasses, inertial modules, pressure sensors, humidity sensors, micro-phonic sensors, temperature sensors, touch sensors (similar to pressure sensors but with capacitive sensing), UV index sensors, air and water sensors and a multitude of applied technologies of these variants. The performance of MEMS devices is not only superior to that of their macro-scale counterparts, but their method of production is similar to fabrication techniques used in the integrated circuit (IC) industry and translates into low per-device production costs, as well as other benefits.

Smartphones are powerful and widely used devices with built-in wireless communication protocols that can be used in WSN infrastructure. These devices are especially useful when rapid response time is a requirement or when the monitored conditions change dynamically. A large set of integrated embedded sensors in smartphones collectively enables various low-cost monitoring solutions when in situ measurements are required. More recently, commercially available wearable smart devices have provided unobtrusive monitoring of health-related parameters such as fitness levels, heart rate and even air quality. Future improvements to this technology, especially considering the energy requirements of these devices, could open up an entirely new paradigm shift of wearable sensors. Smart-phones and smart wearable devices require no additional hardware modifications and applications can be downloaded via the internet, in many cases free of charge. Challenges in distributing the hardware and software, security and privacy concerns, and the availability of high-quality and trusted software applications still exist, and overcoming these limitations is of concern, but they do not limit advancement in these fields, as challenges can generally be overcome with innovative solutions.

1.14 Energy-Harvesting

Capturing and storing energy from natural sources for low-power devices provide much needed requirements for autonomous device operation. Sources such as solar power, thermal gradient energies, wind energy, salinity gradients and other forms of kinetic energy such as water flow can be used to generate and harvest energy. Despite the efficiency and amplitude being relatively low, it can be used for many applications, specifically for WSN modules. The conversion of natural and ambient energy to electrical energy requires some kind of mechanical device, such as MEMS or semiconductor devices that convert photons to electrical charge to generate power, quantified in watts [W]. For low-power devices, different energy sources can be used, ranging from kinetic energy for watches, photovoltaic solar radiation conversion, thermoelectric generators, micro wind turbines and piezoelectric crystals generating electrical voltage when deformed (Kim et al. 2014), to certain antennas that can collect energy from radio waves (Elanzeery et al. 2012).

Two types of wireless energy transfer and harvesting exist, near-field and far-field systems. Near-field systems are suitable for short-range energy-harvesting or power transfer, as it uses electric or magnetic induction to transfer energy with efficiency of up to 80 %. Far-field systems are used in longer range applications, using antennas to collect stray radio-frequency (RF) electromagnetic waves and electronic circuits such as rectifiers and charge-pumps to convert the RF wave to a usable direct-current (DC) signal. RF is the range of electromagnetic frequencies above the audio range and below infrared light, therefore from 10 kHz to 300 GHz. Although not as efficient as near-field harvesting, far-field applications have the advantage of longer range, necessary for certain applications. The power density of these sources specifies the amount of power that can be generated as a ratio of its area or volume, generally specified in watts per m^2 [W m^{-2}] or watts per m^3 [W m^{-3}]. Its output is measured in volts [V] and its ability to supply a current [I] to compute the total [W] of the source that is available to the connected device. Some of these sources can provide continuous output if the stimulus is available, for instance thermal energy, whereas piezoelectric energy sources are activity-dependent. Conversion efficiency varies greatly between these sources, where a high current generated relates to better efficiency of the source.

Solar power is most commonly used, with an established technology and active research. It features high power density during daytime of approximately 100 mW cm^{-2} and reported efficiencies of up to 40 %. Solar sources can also operate as hybrid models with other sources. Thermal energy sources harness the temperature difference in thermoelectric devices, using thermoelectric effects such as the Seebeck and Thomson effect for conversion. The energy densities of these sources are low compared to solar-powered devices, approximately 20–60 μW cm^{-2} when using the human body or ambient room temperature variations, but increases dramatically as the temperature of the source increases. The sources do, however, require relatively large areas to harvest practically useable amounts of power. Piezoelectric devices use mechanical deformation or pressure to generate power, requiring mechanical motion such as wind, generally intermittent. Power densities are generally around 250 μW cm^{-3}, but can increase if the motion of deformation is strong. Ambient RF energy sources have a much lower power density, in the vicinity of 1 μW cm^{-2}, and can be increased using high-gain antennas. The advantage of this source stems from the fact that RF sources are becoming almost a constant feature in many areas where television and radio signals are broadcast, and powering very low-power devices, in WSNs for example, where nodes are placed in difficult to-reach places, is possible with little probability of intermittent sources.

RF energy-harvesting has many applications, whereas RFID technology is mainly used to harness energy. RFID is an important facilitator of IoT. RFID uses radio waves to identify items that can be connected to the IoT and also provide a means of tracking items in real-time providing location and status information (Bari et al. 2013). RFID tags contain at least two main subsystems: an IC for storing and processing information, modulating and/or demodulating the incident RF signals, and receiving an incident DC signal from the reader. It should also contain an antenna for receiving and transmitting the signals. The tag information is stored on non-volatile memory. If no data processing is required on the device, the IC can be omitted from the structure, leaving only the antenna and memory on the device. Both of these structures can be created using silicon on insulator or MEMS technology. Harvested DC power from RF sources inherently undergoes propagation loss (L) and the total available power, P_{total}, can be approximated by adding and subtracting components, such that

$$P_{total}[dB] = P_T[dB] - L[dB] + G_R[dB] + \eta_{RF}[dB] \tag{1.2}$$

where P_T is the transmitted power at the source, G_T is the gain of the antenna at the source, transmitting the original RF signature, G_R is the gain of the receiving antenna (at the RFID), and η_{RF} is the efficiency of the electronic conversion circuit. All values are expressed in decibels [dB]. All components are also spectrally dependent, meaning that the absolute value differs at the varied frequency.

This book aims to provide a detailed background on design techniques, considerations and mathematical derivations required for effective energy-harvesting for low-power devices, such as sensors, placed in WSNs to monitor specifically air quality and water quality parameters. To enable designers to implement these equations and derivations, open-source software that require no initial cost are recommended to test these concepts.

1.15 Geographic Information Systems and Remote Sensing

A geographic information system (GIS) allows organizations in various industries to visualize, analyze and interpret data and understand relationships, patterns and trends using a multitude of visual aids. GIS systems use spatial and temporal location information as its key index variable for all additional information. An advantage of this characteristic is that GIS can relate seemingly unrelated information by appending its key index variable to the information. GIS's dependency on accurate source data can also be considered a disadvantage since high-resolution digital imagery and capable computing requirements are not always available for all regions globally (especially in developing countries) and influence the credibility of the final data. Other potential integration problems include; missing positional information, inconsistent classifications and methodologies, use of different spatial units, use of different levels of aggregation or resolution, the presence of spatial data gaps and variations in time references for the data (Mennecke and West 2001). A GIS can be used to optimize maintenance schedules, fleet movements and infrastructure planning and facilitates cost-savings through its efficiency and improved decision making. GIS-based maps and visualizations improves communication between groups involved in strategic planning. GIS systems provide a localized database of records about past trends, the current status and predicted change in geography. GIS information include data representation of objects such as roads, land usage, geographic elevation, vegetation, water bodies, geographic borders, natural occurrences such as rainfall and cloud distribution, anthropogenic objects such as housing and industrialized areas and various other mapping data.

Closely related to GIS is remote sensing. Remote sensing is a term that describes the science of obtaining information (commonly for GIS systems), generally using sensors mounted on high altitude objects such as aircrafts, balloons or satellites. The sensors measure variations in amplitude or phase of energy reflected from the earth at wavelengths dependent on the application. Active sensors induce energy directed towards the earth, varying the wavelength (for example using lasers or radar) for each application, whereas passive sensors use the reflected radiation from the sun onto objects on the earth to gather information.

Combining GIS and remote sensing to gather information about specific areas on the earth help to not only monitor established infrastructure on earth, but also in predicting the effects of new developments. Applications also include coastal monitoring and variations, oceanic circulation and currents, hazard assessments such as forest fires or flooding and natural resource management such as wildlife habitat protection. There are many additional applications for GIS and remote sensing in resource management, regional planning and economic development and with advances in technology, the amount of applications is growing.

There is however a strong relationship between the potential use of technology (such as GIS) to support governmental decision making, such as expanding of industrial areas, and the needs of its potential users, available funding and backing and availability of technical skills for implementation, management and maintenance. In developed countries these parameters are somewhat easier to address, but with an evident lack in supporting infrastructure and external factors such as corruption many times associated with developing countries, it becomes increasingly difficult to implement technologically advanced systems. This book aims to provide relevant information about GIS and remote sensing that could assist developing countries to form a basis for further research into the subject matter and apply some of the consideration especially targeted at developing countries.

1.16 Ubiquitous Computing

Ubiquitous computing describes a growing trend towards embedding microprocessors into everyday objects, effectively rendering them smart and ideally invisible to people because of their small size and unobtrusive placement. The phrase "ubiquitous computing" was invented by Mark Weiser, a chief technologist of the Xerox Palo Alto Research Center, around 1988 and has since then become anonymous with smart cities aiming to gather information regarding their environment through small, pervasive and non-obtrusive devices. Ubiquitous computing can be used for countless applications, and are ideal for monitoring air pollution and water pollution in urbanized (or rural) areas. The rapid growth rates of cities in developing countries are causing pollution and waste production to spiral out of control, and methods to monitor these causes and effects are crucial to ensure sustainable futures. Ubiquitous computing relies on devices with relatively low energy consumption (compared to for example domestic appliances and other household products) and are ideally powered by renewable energy-harvesting sources. WSNs and other smart-device topologies can be achieved by spatially distributing ubiquitous devices which monitor the environment and use ambient energy as their primary power source.

Ubiquitous computing has evolved to account for various applications ranging from critical environmental analysis such as earthquake monitoring to social awareness to analyze a user's location and provide recommendations for eating, relaxing, shopping or any other form of recreational activities.

Ubiquitous computing requires multiple computer science and engineering disciplines during its design and implementation and is an integral player in developing technological advancements and human-capital in the science and engineering fields. These systems are a result of close interaction between hardware developers, systems design and engineering, systems modeling and user interface designs using relevant software resources (where Android® and Apple® application availability is fast becoming the norm for any new developments).

This book presents ubiquitous computing in lieu of its multiple facets of inter-disciplinary approaches and possible distribution to social, economic and environmental applications.

1.17 Software Resources

This book uses the aid of general-purpose high-level programming language, Python®, and an open-source mobile operating system (OS) based on the Linux® kernel, Android®, to display design equation results and implement mobile applications based on the information provided. Design equations for air quality and water quality, electronic circuit design, electro-optical design, radiation transfer equations and other relevant data are manipulated and represented using these resources.

Python® was conceived in the late 1980s, developed in 1989, and introduced in 1991. It allows programmers to express concepts in code that is optimized for mathematical representation, which would take additional lines of code if implemented in, for example, the C® programming language. Python® supports multiple programming paradigms, including object-oriented, imperative and functional styles. It features a dynamic type system and automatic memory management and has a large and comprehensive standard library. Additional libraries are available and developed as community-based projects, and these libraries can be used extensively for specific application purposes and improved upon by developers and enthusiasts. This book takes advantage of the vast library and resources of the Python® infrastructure, its free-of-charge model, and its processing capabilities to represent equations and theorems highlighted in separate chapters graphically.

Android® is a mobile OS developed by Google®. Its initial release occurred on September 23rd, 2008. Android's® source code is released by Google under open-source license and is popular with technology companies requiring low-cost and customizable operating systems for high-tech devices. Android's® open nature has encouraged a large community of developers and enthusiasts to use the open-source code as a foundation for community-drive projects. Its default user interface is based on direct manipulation, using touch inputs that loosely correspond to real-world actions, such as swiping, tapping and pinching to interact with on-screen objects. Android® can easily be used to create rich and dynamic interfaces and software applications. Transfer protocols are built into most devices capable of running the Android® OS and this can be used to interact with remote sensing equipment, using standard protocols such as Bluetooth® or Wifi. These advantages, and a large network of compatible devices, allow potential distribution of applications developed to measure and monitor air quality and water quality indices, and this book provides general guidelines for building such an application.

1.18 Conclusion

This chapter provides a brief background on megacities, the recent increase in urbanization and the drive towards smart cities during expansion of city infrastructure. Developing countries, such as South Africa, are also planning entire new cities, based on the smart city model, to enter the global race for smart city status. As a result, controlling and limiting pollution, especially air and water pollution, constitute an increasing problem if proper planning is not adhered to during developmental phases, but also in established cities quickly reaching megacity status.

Common air and water pollutants are described in this chapter, as are the sources from where these pollutants stem. Knowing and understanding these sources can provide essential information to reduce and eliminate events causing diseases and pollution-related illnesses. Detailed analysis of these pollutants and sources will be provided in this book, supported by design equations and graphical representation as informative aids.

Awareness of pollution is important, but equally important is the ability to know where and if pollution does occur. Sensing and monitoring, and ultimately reporting pollutants are not an easy task, often requiring large and costly technologies to be implemented, generally through government intervention and limited to a few spatially distributed nodes. Miniaturization of sensing elements through advancements in technology such as MEMS has enabled designers and enthusiasts to make use of equipment, at relatively low cost, to develop applications for WSNs that are comprised of much larger distribution networks. The IoT, with special focus on smartphones and smart devices, has in addition enabled a large distribution network of software resources, coupled with an increasing number of sensors and actuators in these devices, to monitor parameters in the environment that could potentially provide information about immediate circumstances that affect individuals' daily routine and health.

Combining this information, together with guidelines on techniques to implement designs in software and hardware, forms the baseline of this book, which aims to provide designers and enthusiasts with the knowledge and tools they require to make a difference in their community, and possibly globally, considering the importance of lowering pollution and avoiding polluted areas. This should be of concern to everyone, if it is assumed that people are properly informed.

References

Adamo, F., Attivissimo, F., Carducci, C., Lanzolla, A. (2015). A Smart Sensor Network for Sea Water Quality Monitoring. *IEEE Sensors Journal*, 15(5), 2514–2522.

Al-Ali, A. R., Zualkernan, I., Aloul, F. (2010). A Mobile GPRS-Sensors Array for Air Pollution Monitoring. *IEEE Sensors Journal*, 10(10), 1666–1671.

Baglio, S. (2003). Bio-Geochemically Inspired Capacitive Sensors for Heavy Metals Pollution Monitoring. *IEEE Transactions on Instrumentation and Measurement*, 52(5), 1474–1481.

Bari, N., Mani, G., Berkovich, S. (2013). Internet of Things as a Methodological Concept. *2013 Fourth International Conference on Computing for Geospatial Research and Application*, San Jose, 22–24 July 2013, 48–55.

Bartram, J., Ballance, R. (1996). Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes. Published on behalf of the United Nations Environment Program (UNEP) and the World Health Organization (WHO).

Bhardwaj, J., Gupta, K. K., Gupta, R. (2015). A Review of Emerging Trends on Water Quality Measurement Sensors. *2015 International Conference on Technologies for Sustainable Development (ICTSD-2015)*, Mumbai, 4–6 February 2015, 1–6.

Chand, R., Cunningham, P. R. (1970). Characterization of Organic Solvents for Electrochemical Air Pollution Sensors. *IEEE Transactions on Geoscience Electronics*, 8(3), 158–161.

Davy, P. K., Gunchin, G., Markwitz, A., Trompetter, W. J., Barry, B. J., Shagijamba, D., Lodoysamba, S. (2011). Air Particulate Matter Pollution in Ulaanbaatar, Mongolia: Determination of Composition, Source Contributions and Source Locations. *Atmospheric Pollution Research*, 2, 126–137.

Elanzeery, H., Elbagouri, M., Guindi, R. (2012). Novel Radio Frequency Energy Harvesting Model. *In 2012 IEEE International Power Engineering and Optimization Conference (PEOCO2012)*, Melaka, 6–7 June 2012.

Faltings, B., Jingshi, J., Jurca, R. (2014). Incentive Mechanisms for Community Sensing. *IEEE Transactions on Computers*, 63(1), 115–128.

Frodl, R., Tille, T. (2006). A High-Precision NDIR CO2 Gas Sensor for Automotive Applications. *IEEE Sensors Journal*, 6(6), 1697–1705.

Gadkari, A. B., Shinde, T. J., Vasambekar, P. N. (2011). Ferrite Gas Sensors. *IEEE Sensors Journal*, 11(4), 849–861.

Gardner, J. W., Guha, P. K., Udrea, F., Covington, J. A. (2010). CMOS Interfacing for Integrated Gas Sensors: A Review. *IEEE Sensors Journal*, 10(12), 1833–1848.

Gerlach, T. (2011). Volcanic Versus Anthropogenic Carbon Dioxide. *Transactions of the American Geophysical Union*, 92(24), 201–202.

Grossi, M., Lazzarini, R., Lanzoni, M., Pompei, A., Matteuzzi, D., Riccò, B. (2013). A Portable Sensor with Disposable Electrodes for Water Bacterial Quality Assessment. *IEEE Sensors Journal*, 13(5), 1775–1782.

Karami, J., Alimohammadi, A., Modabberi, S. (2012). Analysis of the Spatio-Temporal Patterns of Water Pollution and Source Contribution Using the MODIS Sensor Products and Multivariate Statistical Techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4), 1243–1255.

Kim, S., Vyas, R., Bito, J., Niotaki, K., Collado, A., Georgiadis, A., Tentzeris, M. M. (2014). Ambient RF Energy-Harvesting Technologies for Self-Sustainable Standalone Wireless Sensor Platforms. *Proceedings of the IEEE*, 102(11), 1649–1666.

KPMG Africa Blog (2013). Retrieved May 26 from http://www.blog.kpmgafrica.com.

KPMG International. Infrastructure 100: World Cities Edition. (2012). Retrieved 26 May 2015 from http://www.kpmg.com.

Kulkarni, A. M., Turkane, S. M. (2013). Solar Powered Water Quality Monitoring System using Wireless Sensor Network. *2013 International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing*, Kerala, 22–23 March 2013, 281–285.

Liu, X., Cheng, S., Liu, H., Hu, S., Zhang, D., Ning, H. (2012). A Survey on Gas Sensing Technology. *Sensors*, 12, 9635–9665.

Mennecke, B. E., West, L. A. (2001). Geographic Information Systems in Developing Countries: Issues in Data Collection, Implementation, and Management. *Journal of Global Information Management*. 9(4), 45–55.

Mudgal, V., Madaan, N., Mudgal, A., Singh, R. B., Mishra, S. (2010). Effect of Toxic Metals on Human Health. *The Open Nutraceuticals Journal*, 3, 94–99.

Mackenzie, F. T., Mackenzie, J. A. (1995). Our Changing Planet. *Prentice Hall,* Upper Saddle River, NJ, 288–307.

Rajasegarar, S., Havens, T. C., Karunasekera, S., Leckie, C., Bezdek, J. C., Jamriska, M., Gunatilaka, A., Skvortsov, A., Palaniswami, M. (2014). High-Resolution Monitoring of Atmospheric Pollutants Using a System of Low-Cost Sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7), 3823–3832.

Hogan, C. M., McGinley, M. (2014). The Encyclopedia of Earth (EoE): Water Pollution. Retrieved May 26, 2015 from http://www.eoearth.org.

Unnasch, S., Chan, M., Waterland, L., Pont, J. Hooks, M., Rutherford, D., Stoner, S., Olson, T., Addy, M. (2007). Full Fuel Cycle Assessment Tank to Wheels Emissions and Energy Consumption. *California Energy Commission – Consultant Report (Draft)*, CEC-600-2007-003-D.

United Nations: World Urbanization Report (UNWUR) 2014. (2014). Retrieved May 31, 2015 from http://www.esa.un.org.

United Nations Environment Programme (UNEP): Urbanization. Retrieved May 26, 2015 from http://www.unep.org.

WHO (2013). Retrieved April 13, 2015 from http://www.who.int.

Zhang, Y., Sun, L., Song, H., Cao, X. (2014). Ubiquitous WSN for Healthcare: Recent Advances in Future Prospects. *IEEE Internet of Things Journal*, 1(4), 311–318.

Chapter 2
Population Growth in Developing Countries and Smart City Fundamentals. The Internet-of-Things and Wireless Sensor Networks

2.1 Introduction

This chapter explores population growth in cities and provides estimates on the growth of urbanization worldwide. Equations for growth estimates are given, with the carrying capacity of cities included where data are available. Two models are compared: the Malthusian model, which does not account for carrying capacity, and the Verhulst model, which takes into account the limits of growth of cities. Developed and developing countries' statistics are given based on historical data and future estimates of growth of populations. Smart city market shares and potential for investments, especially in developing countries, are highlighted and conclusions drawn. Sensor applications in smart city infrastructures are given with examples for each application. Expansion from Chap. 1 on the IoT and its industrial applications is provided in this chapter, and the growth of this technology is statistically analyzed. An in-depth look into WSNs describes their implementation possibilities, topologies and optimization.

2.2 Population Growth in Cities

The population is the number of people in an area based on specific categories such as ethnicity, age, income, gender and socio-economic status. The population is constantly changing due to births and deaths and the migration of families to find better sources of income. A population is measured by counting the actual number of people in a given area and measuring birth-to-death ratios. Centripetal and centrifugal forces predict how successful the country's economy is going to be; many people migrate in or out of the country based on the availability of jobs. Before a population of a city can exceed ten million, giving it megacity status, many external factors have to contribute to the rate of achieving this milestone.

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_2

Population growth is proportional to the existing population, immigration rates, age distribution and available resources such as housing, medical facilities and natural resources. As technology and medicine advance, people are starting to live longer because of better medical care and technology created to assist in difficult daily tasks. Because of rapid growth, natural resources are becoming scarce and lack of land causes overcrowding in many areas of the world. These factors affect the ratio of births and deaths in a region. A simplified relationship between the rate of growth of a population, r, also called the Malthusian parameter, the initial population, P_0, (the known population at time t_0) and the population at time t exists as an exponential function known as the Malthusian growth model. To determine the future population, $P(t)$ is represented by

$$P(t) = P_0 e^{rt} \tag{2.1}$$

where the population growth rate can also be determined through known population values expressed as a fraction of the initial population and some future value, such that

$$r = \frac{P(t_2) - P(t_1)}{P(t_2)(t_2 - t_1)} \tag{2.2}$$

where $P(t_1)$ and $P(t_2)$ are the two known population values at different times, generally expressed in years (t_1 and t_2). This derivation is, however, only valid for small time intervals and does not take into account large variations over longer terms. Equations (2.1) and (2.2) are derived from the logarithmic relationship

$$\log_{10} \frac{P(t)}{P_0} = rt \tag{2.3}$$

that relates the future population and initial population to a value comprising of rate of population growth and time. Determining the population growth rate by dividing both sides of (2.3) by the time interval results in a more accurate rate compared to (2.2). A positive growth rate indicates an increase in the population, whereas a negative value for r would suggest a decrease in population. Future values for r are generally estimated based on historical data, current trends and contributing factors such as disease outbreaks or natural disasters.

2.3 Carrying Capacity

As populations grow, the rate of growth is affected as individuals within the population interfere with one another by competing for critical resources such as food, water and space. Thomas Malthus was a philosopher who expected populations to grow in time as long as there were food and shelter. However, because of rapid growth of the population, diseases and scarcity of resources were bound to be influencing factors. Verhulst derived a logistic equation to describe this self-limiting growth of (any) biological population. The equation is given as a differential equation,

$$\frac{dP(t)}{dt} = rP(t)\left(1 - \frac{P(t)}{K}\right) \tag{2.4}$$

where K is known as the carrying capacity. K is the maximum population size that the environment can sustain indefinitely given the food, water and space resources in the area and not accounting for expansion of these. The value of K therefore determines when the function changes from population increase towards maturity, when $P(t)$ ceases to increase. This is mathematically represented by

$$\lim_{t \to \infty} P(t) = K \tag{2.5}$$

and solving the differential equation given in (2.4), the future population can be determined by

$$P(t) = \frac{KP_0 e^{rt}}{K + P_0(e^{rt} - 1)} \tag{2.6}$$

where P_0 represents the initial population, which is a known value. Figure 2.1 represents the population growth of Nairobi, Kenya, for the ten consecutive years starting from 2014, with a growth rate of 4.38 % and a constant carrying capacity of $K = 8$ million.

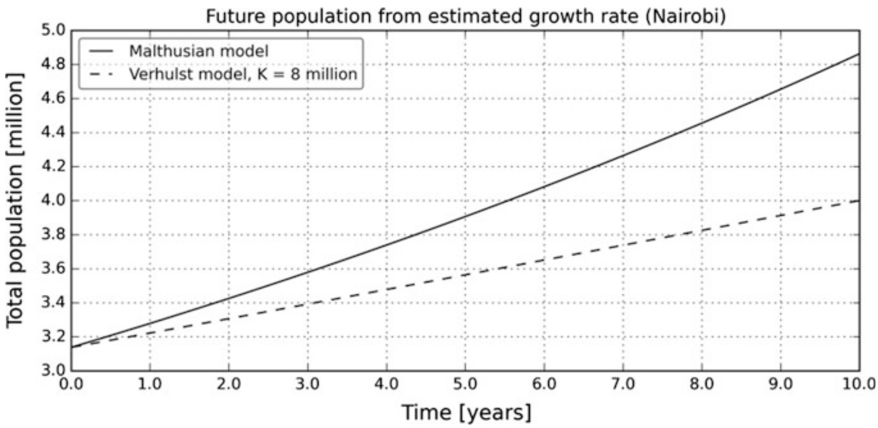


Fig. 2.1 Future population models for 2014 to 2024 for Nairobi, Kenya, Malthusian model versus Verhulst model ($K = 8$ million)

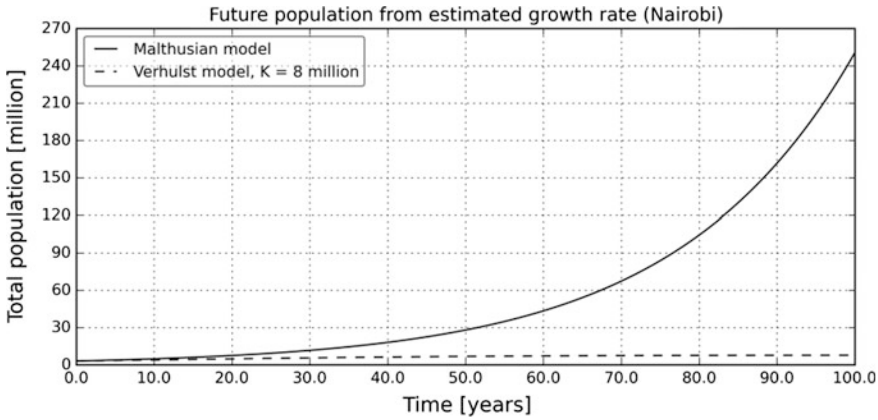


Fig. 2.2 Future population models for 2014 to 2114 for Nairobi, Kenya, Malthusian model versus Verhulst model (K = 8 million)

As is evident from Fig. 2.1, the Verhulst self-limiting property of the population growth as in (2.6) results in slower growth relative to growth where no limits are set on the maximum population in (2.1). The Verhulst model would reach its asymptote at eight million people over a longer time span, and the Malthusian model would keep increasing indefinitely and exponentially. Plotting the same data over a 100 years results in the growth depicted in Fig. 2.2.

The self-limiting feature of the Verhulst model limits the population to eight million people using the growth rate of 4.38 %, whereas no limitation on space and resources would result in a population of more than 240 million in 2114 because of urbanization.

To determine the actual growth rate over a period while accounting for carrying capacity, the growth rate can be determined by rewriting (2.6) as follows:

$$r = \frac{\log_{10}\left(\frac{P_0 P(t) - P(t)K}{P_0 P(t) - P_0 K}\right)}{t}. \tag{2.7}$$

This result is useful when obtaining statistics on estimated versus actual growth rate to determine the requirements for expanding an area to accommodate a larger population.

2.4 Variable Carrying Capacity

The carrying capacity K is also time-varying; it changes with the eironmental condition and human intervention, such as expansion resulting in a varied carrying capacity $K(t)$ and it follows that

$$\frac{dP(t)}{dt} = rP(t)(1 - \frac{P(t)}{K(t)}) \tag{2.8}$$

and this equation is particularly focused on periodical changes in the carrying capacity, typically in 1 year cycles. This is represented by

$$K(t+T) = K(t) \tag{2.9}$$

where it can be shown that in such circumstances, independently from the initial value (assuming $K(0) > 0$), $P(t)$ tends to a unique periodic solution with period T. Periodical changes are generally tied to weather variations, therefore suggesting a yearly cycle. In addition, the carrying capacity can be considered as a function of the population at an earlier time, leading to a logistic delay equation. This function becomes complex and for the purpose of this book, the general differential equation will be used in its simplified, but relatively accurate, approach.

2.5 Developing Country Growth Statistics

Table 2.1 represents population statistics for seven developing megacities in Africa for 1950, 2000, and 2014, and the estimated growth rate for these periods determined by (2.3). The population figures also account for cities with fewer than 10 million inhabitants, not branded megacities in the classical sense, but for developing countries these figures show the rapid expansion of their populations into urban areas. The growth rate does not account for self-limiting population statistics because of the limited availability of these data.

From Table 2.1, it can be seen that in the 1950s, large cities housed generally fewer than a million citizens, Cairo being the exception with almost 2.5 million inhabitants already in 1950. The growth of these cities between 1950 and 2000 is largely attributed to urbanization. Growth rates during this time were large,

Table 2.1 African megacity population statistics for 1950, 2000, and 2014 with calculated growth rates for these periods

City	1950 (million)	2000 (million)	2014 (million)	Growth rate (%)	Growth rate (%)
				1950–2000	2000–2014
Johannesburg	0.899	2.732	4.434	2.22	3.45
Lagos	0.325	7.280	11.867	6.21	3.49
Kinshasa	0.201	5.414	9.5	6.58	4.01
Cairo	2.493	10.169	12	2.81	1.18
Nairobi	0.137	1.698	3.138	5.03	4.38
Accra	0.177	1.673	2.291	4.49	2.24
Khartoum	0.183	3.505	5.274	5.90	2.91

averaging around 4.75 % in this period, with Lagos and Kinshasa both showing growth rates above 6 %. In the 2000 to 2014 period, the combined average of growth rates of all the listed cities decreased to 3.1 %, although not because of less demand for urban lifestyles, but as a result of the lower growth of each city's self-limiting factors and carrying capacity. More recent growth rates are also increasing with expansion of city real estate and improvements to infrastructure, but at a slower pace compared to natural population increases.

2.6 Developed Country Growth Statistics

Table 2.2 lists the total population for seven developed cities, Tokyo, Moscow, New York, Beijing, Paris, London and Singapore, for 1950, 2000, and 2014. The growth rates for these cities are also calculated using (2.3) and can be compared to growth rates for African (developing) megacities in Table 2.1.

From Table 2.2, it can be observed that the growth rates for cities in developed nations are lower compared to developing cities listed in Table 2.1. The total populations of these cities are however larger, as these cities reached megacity status earlier. Tokyo, for example, has a population density of 8,817 people per km^2, the world's highest metropolitan population density. The population density of Nairobi in comparison is 4,850 people per km^2. The average growth rate for the period between 1950 and 2000 is $\sim$1.62 % and for the period between 2000 and 2014 it is 1.15 %. Asian countries such as China and Singapore contribute to increasing the average growth rates because of their relatively recent developed nation status. China and India are the countries with the largest populations in the world. Because of vast amounts of land and lack of contraception, the population is growing at a rapid rate in these countries. Developed countries such as the United States outsource to China and India because the labor cost is cheap. Labor laws in

Table 2.2 Developed world megacity population statistics for 1950, 2000, and 2014 with calculated growth rates for these periods

City	1950 (million)	2000 (million)	2014 (million)	Growth rate (%) 1950–2000	Growth rate (%) 2000–2014
Tokyo[1]	11.274	34.449	36.932	2.23	0.49
Moscow	5.356	10.004	11.471	1.24	0.97
New York	12.338	17.845	20.104	0.73	0.85
Beijing	1.671	10.162	14.999	3.61	2.78
Paris	6.283	9.739	10.516	0.87	0.54
London	8.360	8.224	8.923	-0.03	0.58
Singapore	1.016	3.919	5.086	2.69	1.86

[1]Urban population of the Greater Tokyo area consisting of the Kantō region as well as the prefecture of Yamanashi

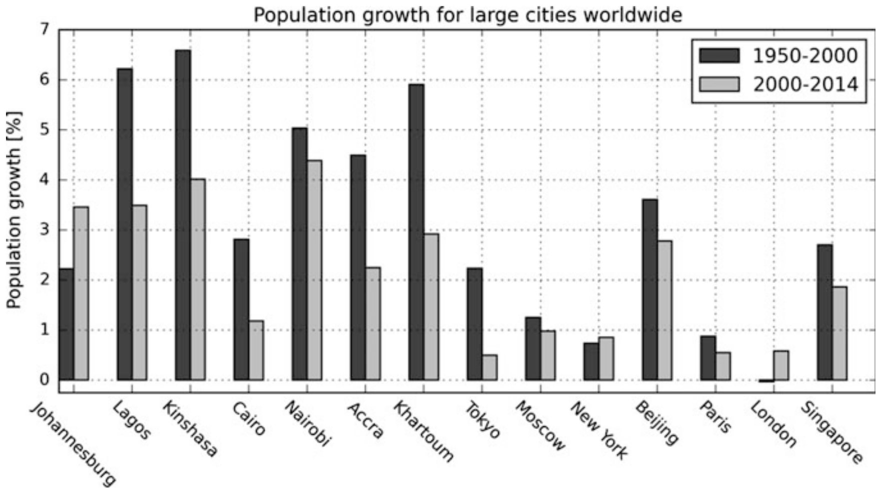


Fig. 2.3 Population growth rates in percentage for fourteen cities worldwide for the periods 1950–2000 and 2000–2014

these countries are not regulated, allowing these countries to take shortcuts, resulting in more products to export. The negative growth rate in London between 1950 and 2000 of 0.03 % is due to small cities emerging outside London after World War 2 ended in 1945 and a large proportion of its population moved to settlements outside the city. The city population has been increasing steadily since its decline in 1945. The growth rate for seven developing and seven developed nations are shown in Fig. 2.3.

From Fig. 2.3 it can be seen that the general trend in population growth shows a decrease in growth rate for the period 2000–2014, except for Johannesburg, New York and London. Developing cities, Johannesburg, Lagos, Kinshasa, Cairo, Nairobi, Accra and Khartoum, show larger growth rates for both periods compared to developed cities, Tokyo, Moscow, New York, Beijing, Paris, London and Singapore. This results from recent interest in these countries, as investors are managing to take advantage of Africa's rich natural resources to stimulate job creation in these countries. More opportunities in these countries could have a positive effect towards urbanization if monitored, managed and abundant educational and training facilities are available in these countries.

2.7 Smart Cities

Smarter cities of all sizes should take advantage of technology advancements and innovative solutions to transform their internal systems, operational structure and service delivery. International Business Machines® (IBM®) published smart city

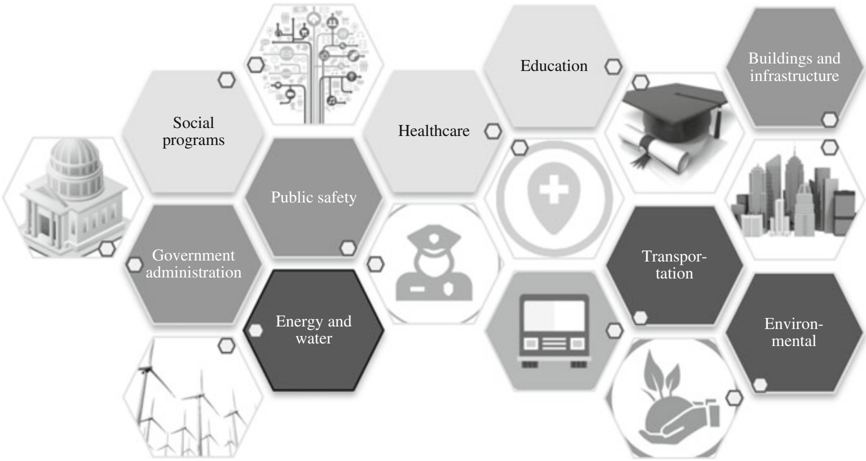


Fig. 2.4 Infographic summarizing smart city evolvement through improvements in planning and management, infrastructure and people development

transformation techniques and summarized the process as improvements on three fronts, namely planning and management, infrastructure and the people in a city. Planning and management consist of public safety and emergency management procedures, law enforcement advancements, smarter buildings, better city planning and operations, and important considerations during government and other agency administration. Infrastructure development includes better water services and quality, transportation solutions and energy management. Regarding the people aspect in a city, IBM® proposes interactive social programs, smarter care programs for citizens and focus on education, especially during early development of the youth in a city. Figure 2.4 captures this information in a summarized infographic adapted from the information available from IBM®.

In Fig. 2.4 the shaded shapes indicate the respective categories of each entry, consisting of the three main solutions for building a smarter city. The planning and management component is shaded medium (public safety, government administration, and buildings and infrastructure), infrastructure planning is shaded dark (energy and water, and transportation), and people development is shaded light (social programs, healthcare, and education).

2.8 Smart City Investments

According to the White Paper published by IDC (2014), a smart city focuses on five key areas for improving urban life:

- Fostering economic development by attracting new businesses, growing existing ones and creating new jobs;
- Improving sustainability through reductions in energy use and greenhouse gas emissions, among other efforts;
- Engaging with citizens, businesses and community groups as active partners in city operations and decision-making;
- Developing a partnership ecosystem with like-minded vendors, universities and colleges, utilities, and other organizations to produce new products and services for constituents;
- Taking an innovative approach to improving the quality of life for residents, and solving longstanding urban challenges.

The information provided in Fig. 2.5 summarizes several highlights of smart city development and the large potential market (and its growth) in future infrastructure

US $506.78 billion (2012)

- The global smart city market value. Estimated to grow 14% per year to US $1265.85 billion in 2019.

US $7 billion smart city in South Africa (2015)

- Cost of a Chinese-financed smart city in eastern Johannesburg in South Africa, with planning for 35 000 houses, an educational center, hospitals and a sport stadium.

US $35 billion estimated for Songdo (2000)

- The projected cost of completion of the smart city of Songdo in South Korea in 2015 was US $35 billion. Climate, energy consumption, water consumption and leisure activities are monitored in real-time.

US $640 million power plant replacement in Denmark (2011)

- The cost of project Amagerforbraending, replacing a 40-year-old adjacent power plant initiative in Denmark.

US $108 billion investment in smart city technology

- Research forecasts of investment costs in smart city technology development and implementation from 2010 to 2020.

2.5 billion urbanized population (2009)

- The majority of the world population growth will occur in urban areas of developing countries with a population increase estimated from 2.5 billion in 2009 to 5.2 billion in 2050.

US $34 billion annually (2014)

- The smarter cities information technology market opportunities are expected to be US $34 billion anually and Cisco estimates 30% improved efficiency in 20 years.

US $1 trillion global savings (2013)

- The estimated global savings that could be attained by optimizing city infrastructure. In addition, US $57 trillion in infrastructure investments will be required between 2013 and 2030.

Fig. 2.5 Summary of smart city investment potential, research forecasts and estimated efficiency improvements due to smart city evolvement strategies. Up to 80 % of future economic growth in developing nations will occur in cities. The environmental impact of smart cities is also significant since 60–80 % of total energy consumption globally occurs in cities and 60 % of water allocated for domestic use is routed to cities. The World Bank estimates that US $14 billion worth of potable water is lost annually owing to leaks, theft and unbilled usage.

Figure 2.5 highlights the potential of investments in smart cities, not only to increase revenue for participating industries, but also to increase energy efficiency, decrease potable water losses and improve air quality and the distribution of information about air quality to the citizens. Transport, one of the main human-caused sources of air pollution, can be controlled better in a smart city environment. Humans all over the world travel up to 30 billion miles (48.2 billion km) and this number is expected to increase threefold by 2050. Analyzing transportation effects on the environment becomes a requirement when taking these numbers into consideration.

2.9 Smart City Initiatives

The main concern in smart city development is addressing the question of how to make a smart city a sustainable city. Stakeholders focus on different (but related) issues when addressing this question. The IBM® Smarter Cities initiative focuses on information and communications technology (ICT), data analytics, cloud computing and intelligent platforms. Cisco's Smart + Connected Communities focus on ICT, networking equipment and the IoT. The Smart City Planning Corporation, Inc. focuses on infrastructure planning, energy and software, whereas Siemens Sustainable Cities is aiming to contribute to energy, transportation, water, waste, and healthcare. Microsoft's CityNext initiative plans advancements in ICT, software, cloud data and big data. These are only a few of the larger stakeholders in the smart city initiatives and it is of importance for each of these players to interact during planning and development to contribute collectively to achieving their goals.

2.10 Applications and Typical Environments

Table 2.3 lists applications of sensors in categories of smart cities, smart environment, smart water, smart metering, security and emergencies, retail, logistics, industrial control, smart agriculture, smart animal farming, home automation and smart health monitoring.

In smart city implementation initiatives, there are significant changes and transformations in work processes and cultures in organizations and departments. Smart city development occurs when multiple initiatives coordinate to (IDC 2014):

Table 2.3 Possible applications for sensors used in smart city environments

Application	Description
Smart cities	
Smart parking	Monitoring of parking spaces available in a city
Structural health	Monitoring of vibrations and material conditions in buildings, bridges, and historical monuments
Noise urban maps	Sound monitoring in bar areas and centric zones
Smartphone detection	Detecting smartphone devices that are WiFi or Bluetooth enabled
Electromagnetic field levels	Measurement of the energy radiated by cell stations and WiFi routers
Traffic congestion	Monitoring of vehicles and pedestrian levels to optimize driving and walking routes
Smart lighting	Intelligent weather and adaptive lighting in streetlights to conserve energy
Waste management	Detection of rubbish levels to optimize trash collection routes
Smart roads	Intelligent highways with warning messages and diversions in reaction to unexpected events
Smart environment	
Forest fire detection	Monitoring of combustion gases and preemptive fire conditions to define alert zones
Air pollution	Controlling air pollutants from factories, cars, and generated by farming
Snow level monitoring	Snow level measurement for real-time information on quality of ski tracks
Landslide and avalanche prevention	Monitoring of soil moisture, vibrations and earth density to detect dangerous land patterns
Earthquake early detection	Distributed control in specific places of tremors
Smart water	
Potable water monitoring	Monitoring the quality of tap water in cities
Chemical leakage detection in rivers	Detecting leakages and wastes of factories in rivers and natural water bodies
Swimming pool remote measurement	Controlling swimming pool conditions such as temperature and pH remotely
Pollution levels in oceans	Controlling leakages and waste into the ocean
Water leakages	Detection of liquid presence outside tanks and pressure variations along pipes
Rover floods	Monitoring of water level variations in rivers, dams and reservoirs
Smart metering	
Smart grid	Energy consumption monitoring and management
Tank level	Monitoring of water, oil and gas levels in storage tanks and cisterns
Photovoltaic installations	Monitoring and optimization of performance in solar energy plants
Water flow	Measurement of water pressure in water transportation systems
Silos stock calculations	Measurement of emptiness level and weight of goods in silos
Security and emergencies	
Perimeter access control	

(continued)

Table 2.3 (continued)

Application	Description
	Access control to restricted areas and detection of people in non-authorized areas
Liquid presence	Liquid detection in data centers, warehouses, and sensitive building grounds to prevent corrosion
Radiation levels	Distributed measurement of radiation levels in nuclear power stations and leakage alerts
Explosive and hazardous gases	Detection of gas levels and leakages in industrial environments, chemical factories and mines
Retail	
Supply chain control	Monitoring of storage conditions along the supply chain and product tracking for traceability
NFC[1] payment	Payment processing based in location or activity for public transport, theme parks, etc.
Intelligent shopping applications	Information on customer habits, preferences, expiring dates, etc.
Smart product management	Controlling rotation of products in shelves and warehouses to automate restocking processes
Logistics	
Quality of shipment conditions	Monitoring of vibrations, strokes, container openings or cold chain maintenance
Item location	Searching for individual items in warehouses or harbors
Storage incompatibility detection	Warning emission on containers storing inflammable goods close to explosive material
Fleet tracking	Controlling routes followed for delicate goods like medical drugs, jewelry or dangerous merchandise
Industrial control	
M2M[3] applications	Machine auto diagnosis and assets control
Indoor air quality	Monitoring of toxic gas and oxygen levels inside chemical plants
Temperature monitoring	Controlling temperature inside industrial and medical refrigerators with sensitive merchandise
Ozone presence	Monitoring ozone levels during the meat-drying process in food factories
Indoor location	Indoor asset location using active and passive tags
Vehicle auto diagnostics	Information collection on vehicle parameters to alert drivers of possible dangers
Smart agriculture	
Wine quality enhancing	Monitoring soil moisture and trunk diameter in vineyards to control sugar content in grapes
Green houses	Controlling micro-climate conditions to maximize production of fruit and vegetables and their quality
Golf courses	Selective irrigation in dry zones to reduce water resources required
Meteorological station network	Studying weather conditions in fields to forecast ice formation, rain, drought, snow or wind changes

(continued)

Table 2.3 (continued)

Application	Description
Compost	Controlling humidity and temperature levels in alfalfa, hay and straw to prevent fungus contamination
Smart animal farming	
Hydroponics	Controlling the exact conditions of plants grown in water to get the highest crops
Offspring care	Controlling growing conditions of the offspring on animal farms to ensure their survival and health
Animal tracking	Location and identification of animals grazing in open pastures or location in big stables
Toxic gas levels	Studying ventilation and air quality on farms and detection of harmful gases from excrement
Home automation	
Energy and water use	Energy and water supply consumption monitoring to obtain advice on how to save cost and resources
Remote control appliances	Switching on and off appliances remotely to avoid accidents and save energy
Intrusion detection	Detection of window and door openings and violations to prevent intruders entering
Art and goods preservation	Monitoring of conditions inside museums and art warehouses
Health	
Fall detection	Assistance for elderly or disabled people living independently
Medical refrigerators	Controlling conditions inside freezers storing vaccines, medicine and organic elements
Care of sports people	Vital signs monitoring in high-performance centers and fields
Patient surveillance	Monitoring of conditions of patients in hospitals and old age homes
Ultraviolet (UV) radiation	Measurement of UV sun rays to warn people not to be exposed at certain hours

Descriptions can vary for each application and additional applications exist
[1]Near field communication to establish radio communication in close proximity
[2]Machine-to-machine technology allowing systems to communicate with devices of the same type

- Leverage technology investments across an entire department, organization, city, and city ecosystem (with utilities, schools and universities, even other cities);
- Use common platforms to decrease service time/maintenance costs;
- Share data across work groups and systems;
- Tie IT investments to 'smart' missions.

Initiatives that involve the communities are also aiding shareholders through participation of the local community in highlighting the issues that are being experienced and offering potential solutions and ideas for resolving these issues, especially in youth unemployment, ageing, obesity, civic engagement, public health, social capital, economic development, transportation, energy and the immediate environment.

Smart city strategy initiatives require a number of collaborators to provide input to achieve a common goal. Local city authorities must be included during funding considerations. Cities should also individually commit themselves to including projects in the larger scope of the project. Regional authorities, which are responsible for structural funds in some countries, can be approached to mobilize the funds for smart city development. Furthermore, funding organizations must be included when granting funds, and private and public utilities could have incentive-based strategies to use investment opportunities. Research organizations offer valuable insights into smart city development and similar incentive strategies to ensure that collaboration can be implemented. National ministries, citizens and non-governmental organizations should also be included during discussions, development and possible integration.

2.11 Internet-of-Things

The internet has made the availability of information accessible to almost half of the people on earth. In 2015 it has been estimated that there are three billion unique internet users worldwide. Open access to the internet has revolutionized the way individuals communicate and collaborate, entrepreneurs and corporations conduct business, and governments and citizens interact. In parallel, the internet established a revolutionary open mode for its own development and governance encompassing all stakeholders. Fundamentally, the internet is a network of networks, globally, and its protocols are designed to allow networks to interoperate. Initially the information on the internet was governed and managed by academic, government and research communities. As it evolved and was commercialized, vendors and operators joined the open protocol development process and contributed to the immense growth and innovation through the internet (Jin et al. 2014). The spirit of collaboration that lies at the foundation of the internet has extended from standards to a multi-stakeholder governance model for shared resources (Keoh et al. 2014). The internet created a mergence between traditional media such as broadcasting and telecommunications, allowing civil society, businesses and governments to take advantage of this model to gather and share information. These traditional sources are not the only channels that have been changed through the internet. It can be used to deliver services and goods, provide financial services and serve as a channel for entertainers to reach a large audience, eliminating, in some sense, the need for expensive advertising. The internet effectively enables endless possibility for all users to share their interests and services much more easily compared to traditional methods.

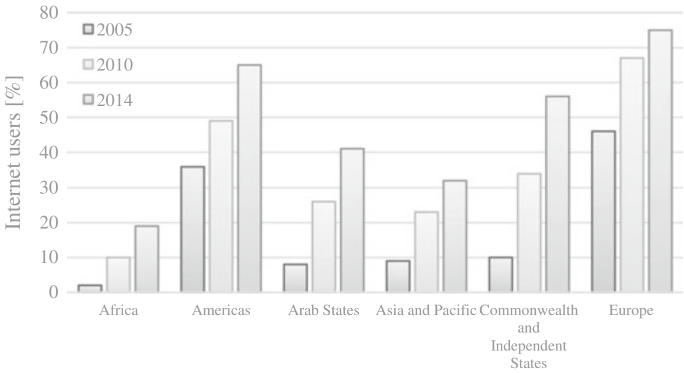


Fig. 2.6 Internet users' growth between 2005 and 2014 for regions globally

2.12 Internet Access in Developing and Developed Nations

The internet, both in terms of infrastructure and content, has grown rapidly since its inception. Figure 2.6 depicts the growth in internet users since 2005 up to 2014 for the regions encompassing Africa, the Americas, the Arab States, Asia and Pacific, Commonwealth and Independent States, and Europe.

The worldwide number of internet users in 2005 was estimated at ~1 billion people; at the time 51 % of the developed world had internet access. Only 8 % of the developing world had access to the internet, and 84 % of people globally were not using the internet. By 2013 the statistics had changed in favor of developing nations; 31 % of developing countries had access to the internet, 77 % of developed nations and 39 % (2.8 billion) of the global population. This still meant that a large percentage (61 %) of the global population in 2013 was not using the internet and it was considered a luxury in those areas (Evans 2011). Large growth in the number of internet users is visible globally estimated at ~3 billion by May 2015 (Internet Society Global Internet Report 2014), with Africa still trailing behind in total internet users, with only 2 % of the population having internet access in 2005, up to ~19 % in 2014. It is believed that growth in use will increase to 2020 and that the percentage of internet users in Africa will reach around 50 % by that time.

2.13 Addressable Networked Devices

Each computer, mobile phone or any other device that can access a network or the internet has identification, an internet protocol (IP) address. Internet host numbers are growing significantly, increasing from ~1.3 million in January 1993 to over 1 billion in January 2014. The growth rate of internet hosts is over 37 % annually and has been constant over 21 years, indicating that future connectivity could grow at similar rates. IP addresses are limited; IPv4 is a protocol described in 1981 for use on packet-switched networks (the transfer of data between devices). This protocol uses a 32-bit addressing scheme, resulting in ~4.3 billion addresses (2^{32}). This limit was already reached in 2011, though delayed in some way by classless inter-domain routing and network address translation, but its successor, IPv6 is designed to extend the number of addresses for networked devices. IPv6 uses a 128-bit addressing scheme, resulting in approximately 2^{128} or 3.4×10^{38} unique addresses, and has been in commercial deployment since 2006. As approximately half a billion mobile devices and connections had been added to the global network annually (since 2014), the addressing scheme needed to be changed and in practical terms IPv6 should be able to cope with the increasing number of addressable devices.

2.14 IoT Market Share

Many things are capable of digital data processing and advanced functionality. This term, things, has given rise to the collective term, the IoT. For a device to be classified in the so-called IoT, it must have (at least) three basic properties: a constant data connection, low power usage and the ability to communicate over short distances. Gartner® has estimated that IoT product and service suppliers will generate incremental revenue exceeding US $300 billion by 2020. The industry market share was estimated at around US $100 billion in 2010 and growth between 2010 and 2015 was ~5.3 % (from (2)) but higher growth of 11.6 % in the industry is anticipated between 2015 and 2020 as the technologies mature. Figure 2.7 depicts the estimated market value of the IoT industry from 2010 to 2020.

Between 2013 and 2022, US $14.4 trillion of net profit will be shared by enterprises worldwide, with industrial internet having the potential to add US $10–15 trillion to the global gross domestic product value. Figure 2.8 depicts the growth of internet-connected devices (IoT devices) compared to the global population from 2003 to 2020.

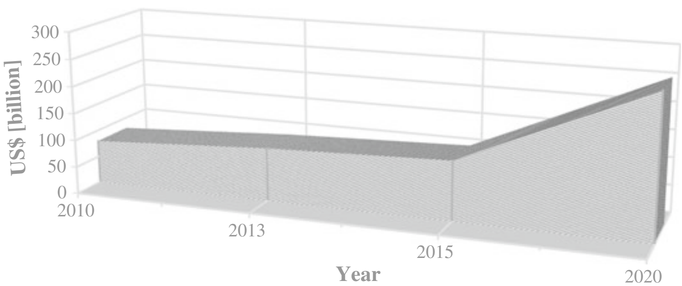


Fig. 2.7 Estimated market value of IoT industry from 2010 to 2020

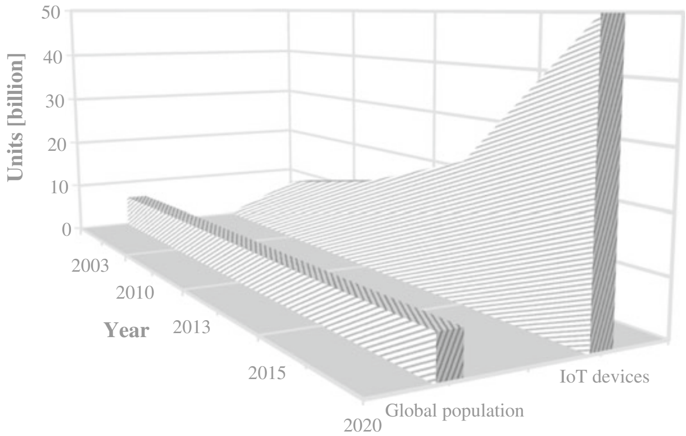


Fig. 2.8 Global population versus number of IoT devices from 2003 to 2020

Figure 2.8 shows that in 2003 there were more people on earth (6.3 billion) than IoT devices (500 million). According to Cisco®, between 2008 and 2009 the number of IoT devices surpassed the global population and by 2010 there were 1.84 IoT devices for each human. By 2020, the number of worldwide internet users will grow to 5 billion people, the global population is estimated to be 7.6 billion, and an estimated total number of 50 billion things will be connected to the internet (Internet Society 2014).

These values are mentioned to highlight the analytic investment opportunity into the IoT for future endeavors. Water and air pollution monitoring should take advantage of the growth of this industry. Google® is reportedly working on an Android®-based operating system to run specifically on the emerging class of low-power devices (IoT). This operating system requires few resources, as little as 32 megabytes (MB) of random access memory (RAM), and marks a significant departure for Google® from its classical Android® operating system, demanding at least 512 MB of RAM. Qualcomm® stated in May 2015 that it had made US $1 billion in revenue in the 2014 fiscal year on chips used in a variety of city infrastructure projects, home appliances, cars and wearables. Qualcomm® said that 120 million smart home devices were shipped with Qualcomm® chips in the previous financial year and 20 million cars equipped with its chips. Moreover, silicon chips were used in 20 types of wearable devices. Although most of its revenue came from supplying chips for smartphone devices, the company estimated that 10 % of its revenue would come from non-smartphone chips during 2015. Unlike the smartphone business, the IoT market is not likely to be dominated by a handful of massive players such as Apple® or Samsung®; instead, it is expected to consist of several smaller players competing across a variety of segments. However, problems and challenges arise when having such large-scale integration of devices and some of these obstacles are highlighted. Firstly, security and data privacy is of major concern, especially for organizations that deal with sensitive data, such as governments and military institutions. Overall, security is said to be only as effective as its weakest point. It is therefore the responsibility of each developer to ensure that his or her device conforms to strict security standards. Customer demand is another challenging obstacle to overcome, in which semiconductor companies will be tasked with supplying integrated circuits to a variety of devices connected to the IoT. In addition, the lack of standards is particularly troublesome among emerging protocols demanding low power consumption, such as long-term evolution, IEEE[1] 802.11ah, Sigfox and OnRamp. Market fragmentation as a result of the diversity of the IoT applications could mean that no single chip is likely to be suitable for all applications. Such fragmentation represents a concern, since it limits economies of scale and raises production costs. Finally, technology shortcomings could also play an important role in slower uptake of devices and services. Technology in its current form might seem to fulfil all the required characteristics; however, many improvements, especially in low power consumption and energy harvesting devices, must be made before the commercial market will adapt.

2.15 IoT and Smart City Word Usage

The IoT and smart cities are synonymous with each other and show similar trends in its usage since the invention of the word internet-of-things. Figure 2.9 compares the historical word usage for internet-of-things and smart city, both case-insensitive and accounting for variations of the words in literature (IoT represented as Internet of Things, internet-of-things or internet of things).

The word internet-of-things was only invented around 2000, and has since seen a noticeable increase in its usage as seen in Fig. 2.9. Smart city is, however, not a new word, and has existed for more than 200 years; however, its meaning and descriptive qualities have changed historically. An increase in its usage is evident during the 1940s. During this time, World War 2 was the main driving force towards new technologies and smarter cities. Its use declined again after the war ended and has showed an increase in usage with gaining popularity of the internet, following a similar trend compared to the words internet-of-things.

[1]Institute of Electrical and Electronics Engineers.

Fig. 2.9 Comparison of historical data of the words internet-of-things and smart city

2.16 Industrial IoT

As the IoT focuses on commercial and consumer electronics, the industrial sector is also harnessing the IoT to create sustainable business models to grow and automate core processes. The industrial internet-of-things (IIoT) is another fast-growing network of increasingly intelligent connected devices, machines and physical objects. Businesses have made progress in applying IIoT to reduce operational expenses, boost productivity or improve worker safety. For example, drones are being used to monitor remote pipelines and intelligent drilling equipment can improve productivity in mines. The IIoT is similar to the IoT, the primary difference being in its business model to improve productivity, much like the early internet applications. This transformation in business has dramatic implications for the workforce. The IIoT can digitize some jobs that have resisted automation, its ethical approach arguably depending on the application. These automation processes also create jobs, as they replace some jobs. Routine and sometimes mundane work can be replaced by more engaging work, as technology allows workers to apply new skills to tasks that require human input. An infographic highlighting applications of IIoT is depicted in Fig. 2.10.

From Fig. 2.10, examples of uses for IIoT can be recognized, including weather-based agriculture control, automated transportation, adaptive water consumption, pharmaceutical product-tracking and regulated construction. The meaning of industrial internet, henceforth called IIoT, is similar to that of cyber-physical systems (CPS) and its design can follow the “5C Architecture”—the five levels of designing a CPS. The five levels are a smart connection, data-to-information conversion, cyber, cognition and configuration. Smart connection is a necessity for establishing a connection between the cyber space and the physical space by acquisition of data from industrial equipment. These data can be collected from various sources, similar to the classical sense of the IoT, by sensors, controllers, inspection, maintenance logs, and for example simple alarm systems. Data-to-information conversion requires processing large amounts of gathered data

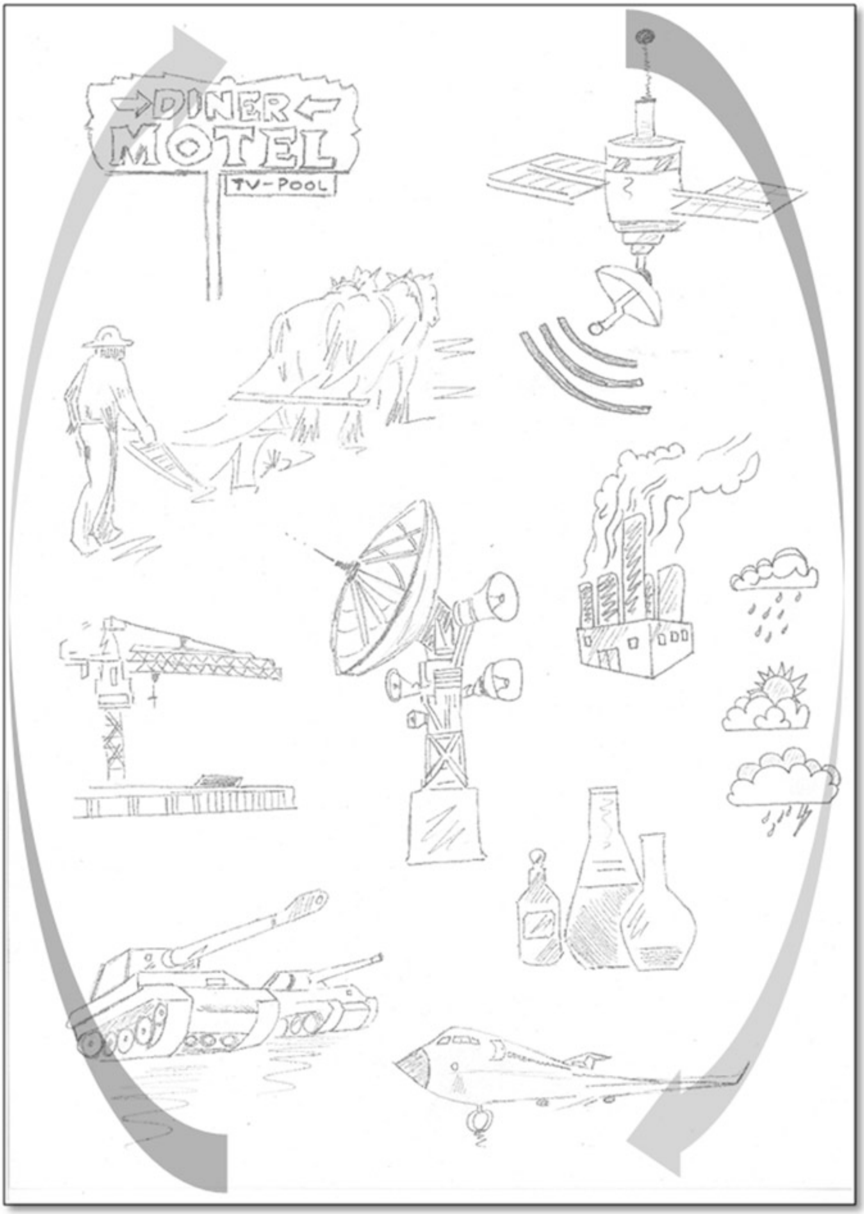


Fig. 2.10 Business model development for sustainable process management through harnessing the IIoT

and generating useful and significant information. This is achieved through the development of machine learning, statistical analysis and data mining techniques. The term cyber refers to the fundamental difference between CPS or industrial internet and conventional data-driven modeling frameworks. The cyber level serves as a central information hub and information is parsed to this entity. Customized analytics are performed to extract knowledge and information from the data. This extraction equips the machine with the ability of self-learning, teaching itself from past information and events. Cognition follows to generate an understanding of the monitored machines or assets to visualize the acquired information for users. Finally, configuration levels realize feedback from the cyber space towards the physical space. Corrective and preventative decisions can be taken to improve the efficiency of actions taken. An example of this process is the Google® driverless car, which takes environmental data from roof-mounted remote sensing equipment to measure distances through reflected light and machine vision to identify road geometry and obstacles to control the vehicle's throttle, brakes and steering mechanism.

2.17 SCADA

A system concept that encompasses a centralized monitoring system to control industrial plants (and can be expanded to other applications) is the supervisory control and data acquisition (SCADA) system. SCADA consists of remote terminal units (RTUs) and programmable logic controllers (PLCs) with host control restricted to basic overriding or supervisory level intervention by operators (Otani and Kobayashi 2013). SCADA systems generally consist of RTUs, PLCs, telemetry systems, data acquisition servers, a human-machine interface, time-stamped data, a supervisory computerized system sending commands to the SCADA controller, defined communication protocols and analytical instrumentation. SCADA systems have evolved through approximately four generations, the first (early 1970s) being a monolithic solution where computing was done by minicomputers but common network services did not exist. These systems were independently operated and RTUs were accessed during alarm events to troubleshoot the system. Second generation SCADA systems were networked through a local area network (LAN); however, the network protocols were not standardized and mainly system operators could determine how secure and efficient an installation was. The third generation SCADA systems could be spread across multiple network connections and be separated geographically. Large-scale systems could be implemented and relatively easily managed by supervisors, since network protocol standardization was established during this time. SCADA has since the third generation adopted cloud computing and the IoT to reduce cost significantly and increase the scale of integration and the complexity of maintenance. Modern SCADA systems allow real-time data from the plant floor to be accessed from anywhere in the world. This access to real-time information allows governments, businesses and individuals to make data-driven decisions about how to improve processes (Sahin and Isler 2013). The introduction of modern information technology standards and practices such as structured query language and web-based applications into SCADA software has greatly improved the efficiency, security, productivity and reliability of SCADA systems. SCADA is at the core of many modern industries, including the energy, food and beverage, manufacturing, oil and gas, power, recycling, transportation, and water and waste industries. Recent progress (Kulkarni et al. 2015) in terms of SCADA systems has been made towards distribution systems capable of solving problems caused by the connection of numerous distributed data generators.

2.18 Wireless Sensor Networks

WSNs offer great potential to increase system efficiency, automation and process control. WSNs for industrial application such as mobile sensor networks, remote surgery and general healthcare (Zhang et al. 2014), as well as industrial automation, use spatially distributed sensors, actuators and controllers connected by wireless communication systems. These sensors are generally referred to as smart sensors since there are additional elements that contribute to internal data processing, albeit a small portion of overall requirements to limit energy requirements (Sharma et al. 2013). A representation of the components required in smart sensors and their function for the user is depicted in Fig. 2.11.

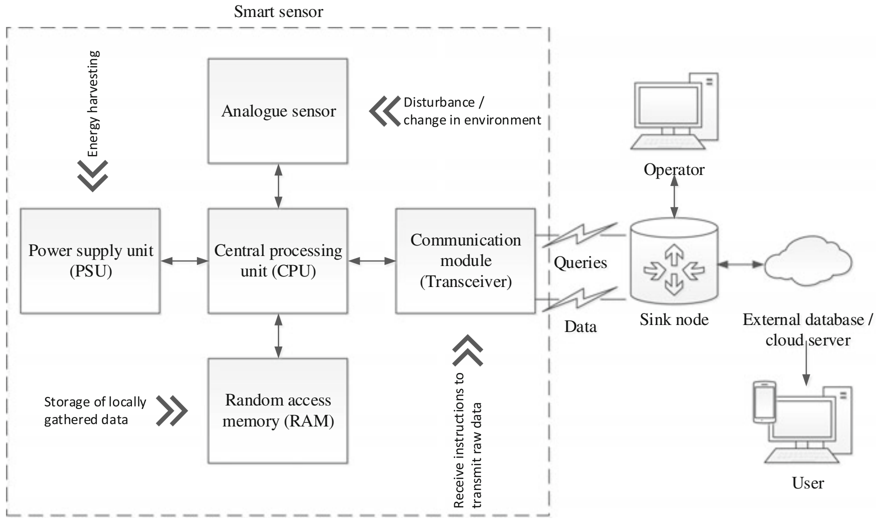


Fig. 2.11 WSN smart sensor components and their role in communicating with the operator and end-user

From Fig. 2.11 it is evident that a smart sensor consists of (at least) a central processing unit capable of analyzing and sorting incoming raw data and memory to store the calculations before sending them via a transceiver module once queried. Each of these components requires energy to operate and this is provided by the power supply, generally batteries or solar cells. The data can be requested by an operator locally and supplied to the end-user, commonly in the form of tables or graphs. Wireless networks have the advantage over wired networks because of simplified deployment, low installation and maintenance cost, increased mobility, and no cabling requirement and are appealing technology for a smart infrastructure. The manufacturing techniques and industrial applications for WSN are not yet widespread globally and mainstream adaption is slow. Control systems based on WSNs must support correct decision-making at crucial times, despite network failures and traffic congestion. Furthermore, an energy-efficient network is critical in view of the limited battery power available for small devices. Research into power-efficient devices with failsafe operation is an active field and improvements are implemented in new infrastructures. Future improvements to already installed systems are also not always viable, and scalability and adaptation during infrastructure design are sometimes overlooked. Communication protocols suffer from non-standardized approaches as active research leads to frequent changes in protocols, making it troublesome to follow a one-size-fits-all approach. Currently, the IEEE 802.15.4 standard is used for many low-data-rate, wireless, ubiquitous personal area networks (Park 2011). The standard specifies the physical layer and media access control protocols, maintained by the IEEE 802.15 working group (IEEE 2011), as defined in 2003. Figure 2.12 represents the layer structure for implementing the IEEE 802.15.4 wireless protocol where the physical (PHY) layer and medium access control (MAC) layer are defined and additional layer definitions depend on the implemented standard. The Open Systems Interconnection (OSI) model of network operation and the defined IEEE 802.15.4 layers (PHY and MAC) are depicted in Fig. 2.12.

According to Fig. 2.12 the connectivity between the physical medium and the PHY layer towards the MAC layer is defined in the IEEE 802.15.4 standard. Additional layers include a service-specific convergence sub-layer (SSCS), logical link control (LLC) layer, and security implementations within the network (NWK) layer. Various additional layers can be implemented to define a protocol properly, and the structure generally ends with the application (APL) layer presented to the user. IEEE 802.15.4 forms the basis of ZigBee (Han and Lim 2010), RF4CE, SyncroRF, ISA100.11a, HART, WirelessHART, and MiWi specifications, each of these further extending the standard by developing the upper layers of the protocol not defined in the IEEE 802.15.4 standard (Torabi et al. 2015). The IEEE supports many working groups to develop and maintain wireless and wired communications standards. For example, 802.3 is wired Ethernet and 802.11 is for wireless LANs, also known as Wi-Fi. The 802.15 group of standards specifies a variety of wireless personal area networks for different applications (Sallabi et al. 2014). For instance, 802.15.1 is Bluetooth, 802.15.3 is a high-data-rate category for ultra-wideband technologies, and 802.15.6 is for body area networks. There are several others. The

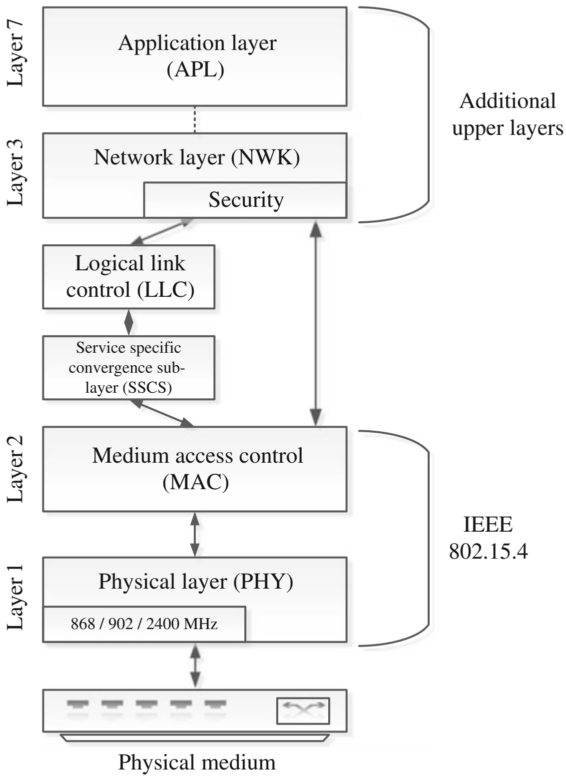


Fig. 2.12 OSI model of network operation within the IEEE 802.15.4 WSN communication structure

802.15.4 category is probably the largest standard for low-data-rate short-distance wireless networks. It has many subcategories. The 802.15.4 category was developed for low-data-rate monitor and control applications and extended-life low-power-consumption uses. The basic standard with the most recent updates and enhancements is 802.15.4a/b, with 802.15.4c for China, 802.15.4d for Japan, 802.15.4e for industrial applications, 802.15.4f for active (battery-powered) radio-frequency identification uses, and 802.15.4 g for smart utility networks for monitoring the Smart Grid. All these special versions use the same base radio technology and protocol as defined in 802.15.4a/b. The IEEE 802.15.4 features are compared in Table 2.4 to other existing wireless standards in use, however it is the only standard that encourages low power usage at lower data rates.

Table 2.4 highlights various standards and their main features, where the IEEE 802.15.4 protocol is focused on low power and data rate to ensure long operation times for low-power devices in WSNs. The basic framework conceives a 10 m communication range with transfer rates up to 250 kbit/s with range increasing to above 100 m at lower data rates (20 kbps) (Khanafer et al. 2014). Lower rates can be implemented, resulting in lower power consumption, depending on the

Table 2.4 Comparison of features of IEEE 802.15.4, GSM/GPRS/CDMA, IEEE 802.11, and Bluetooth 4.0 technologies

	IEEE 802.15.4	GSM/GPRS/CDMA	IEEE 802.11	Bluetooth 4.0
Typical application	Monitoring and telemetry	Voice and data	High-speed internet	Device connectivity
Battery life	>1 year	<5 days	>1 week	<1 week
Throughput	250 kbps	<2 Mbps	<54 Mbps	<24 Mbps
Typical range	+100 m	Several km	50–100 m	<100 m
Advantages	Low power and cost	Existing infrastructure	Speed and ubiquity	Device compatibility

Table 2.5 Industrial Scientific and Medical (ISM) frequency assignment and channel information of IEEE 802.15.4 for regions globally

	Europe	Americas	Worldwide
Frequency assignment	868–868.6 MHz	902–928 MHz	2.4–2.4835 GHz
Number of channels	1	10	16
Channel bandwidth	600 kHz	2 MHz	5 MHz
Data rate	20 kbps	40 kbps	250 kbps
Modulation scheme	BPSK	BPSK	QPSK

application. A summary of the IEEE 802.15.4 frequency assignment and channel information is given in Table 2.5.

Data rates can be lowered in the IEEE 802.15.4 standard if the application permits, where lower data rates offer a larger communication range. The modulation scheme for the lower data rates (20 and 40 kbps) is binary phase shift keying (BPSK) and quadrature phase shift keying (QPSK) for the higher data rate option (250 kbps). BPSK is commonly used as the preamble or beacon frame for channels where low-data-rate modulation is sufficient. QPSK has similar characteristics compared to BPSK, but allows transmission of two bits on a single carrier and therefore allows higher data rates. The probability of a bit error P_b, or bit error rate (BER) (Martalo et al. 2013) for both modulation schemes is given by

$$P_b = Q\sqrt{\frac{2E_b}{N_0}} \tag{2.10}$$

where E_b is the energy per bit and N_o is the noise power spectral density. Q is the complementary error function given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{\left(t^2/2\right)} dt. \tag{2.11}$$

Although the BER for QPSK is similar to BPSK, for this scheme to achieve the same BER as BPSK, it uses twice the power, as two bits are transmitted simultaneously, hence the increase in bandwidth and data rate.

Europe and the Americas operate these devices at frequencies below 1 GHz, whereas the rest of the world operate these devices within the 2.4 GHz ISM band. Important network quality parameters for WSNs, implemented through networked control systems, include system bandwidth, sampling and delay times, and packet dropouts (reliability). Each of these parameters is application-specific; system bandwidth might be a lower priority during water-level monitoring compared to satellite telemetry and system reliability is generally less stringent in commercial applications than in military applications.

2.19 Layered WSN Nodes

WSNs are typically designed with a large number of redundancies to achieve fault tolerance and to maintain the desired network lifetime and coverage (Buratti et al. 2009). It is possible to determine the optimal number of sensor nodes required in a WSN with a layered structure. The minimum sensor counts in a layered WSN to maintain network lifetime and coverage are given by Parikh et al. (2009)

$$N_n = \left[cd^2\left(\sum n + 1\right)\tau N_{active} T_{total}\right] / (E_{node} T) \tag{2.12}$$

where N_n is the sensor count in an n layer network, c is a proportionality constant depending on the electromagnetic conditions in the immediate atmosphere, d is the distance between two adjacent layers in meters, τ is the transmitting time for a node in seconds, N_{active} is the number of active nodes required in layer n, T_{total} is the required network lifetime, E_{node} is the energy available at each node and T is the time period between two transmissions for the same node.

2.20 WSN Topologies

The development and deployment of WSNs entail taking traditional dynamic network topologies and creating wireless spatially distributed networks. Common WSN topologies include bus, tree, star, ring, mesh, circular and grid topologies. The most commonly used topologies are the star, tree and mesh topologies. Each network is divided into three main components, namely a human interface device responsible for interfacing data to the operator, a routing device (sink node) that serves as a data-hopping device (Zhao et al. 2012) and sensor nodes that gather information such as temperature, pressure, wind direction and speed,

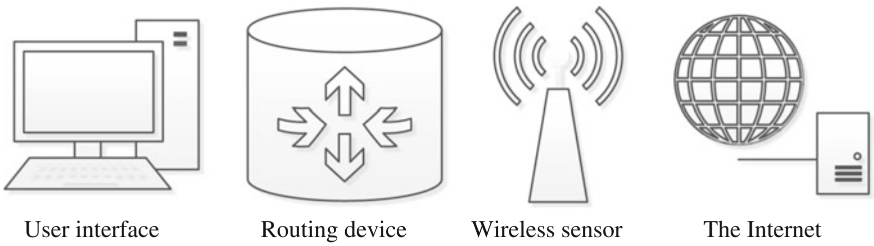


Fig. 2.13 WSN topology legend description

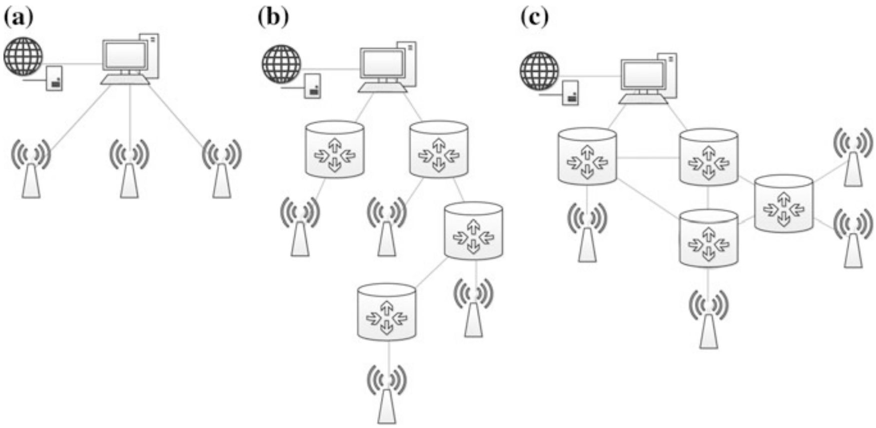


Fig. 2.14 Representation of **a** star, **b** tree, and **c** mesh topologies for WSN infrastructures describing the dependencies of the user interface, routing devices and wireless sensors

water levels and a multitude of additional parameters, depending on the application and sensor type. Finally, the information is transferred to an internet-based service, or a cloud-connected device, for further processing. Figure 2.13 depicts these components used to describe a WSN.

The three commonly used topologies, star, tree and mesh, are represented in Fig. 2.14, with reference to the legend description in Fig. 2.13.

As seen in Fig. 2.14, the tree topology uses a central hub called a root node as the main communication router. In this hierarchy, the central hub is one level below the root node. This lower level effectively forms a star network topology. The tree network topology can therefore be considered a hybrid of the star and the peer-to-peer (bus) topology. However, if the main communication path is broken or interrupted, the active route to the user is also broken. Tree topology networks are connected to a centralized communication hub, or sink, and the sensor nodes cannot communicate directly with one another. All communication must be routed through the centralized hub. Each node is considered a client while the central hub is the server or sink node. The main disadvantage of this topology is the single-path communication structure. In the mesh topology, the message can take any route or path from the sensor to the user interface, in various directions. Full mesh and partial mesh topologies exist if each node is connected to every other node, or if the sensor nodes only connect to each other indirectly.

2.21 WSN Modelling

WSN nodes change their states frequently and are viewed as transitions, usually between active and passive (idle) monitoring (Gračanin et al. 2004). The probabilities associated with state changes are called transition probabilities. The entire process is characterized by a state space, or a transition matrix describing the probabilities of specific transitions, and initial states. System changes occur randomly and it is generally not possible to predict with 100 % certainty the state of a node in the future; however, it is statistically possible to predict the future states based on the current state. These characteristics of a WSN enable modelling the topology as a Markov chain event. A Markov chain is a sequence of random variables, X_1, X_2, X_3, …, X_n, where the present, future and past states are independent. From this definition, the variables can be defined as

$$\Pr(X_{n+1} = x | X_n = x_n) \tag{2.13}$$

and the possible values of X_n form a countable set in the state space, S. The state space is labeled in terms of each sensor node and a transition matrix is created showing the probabilities of each node connecting to the routing device. In the transition matrix, P, the distribution over states can be written as a stochastic row vector, x, with the relation

$$x^{(n+1)} = x^{(n)}P \tag{2.14}$$

where n defines a single point in time. The reliability of a WSN depends on the communication link and the reliability of the WSN nodes. This dependency can be modelled by a reliability block diagram (RBD). An RBD enables the user to represent and evaluate the reliability of a closed system as a set of building blocks, each block representing the reliability of an element within the system, such as the reliability of the routing device, communication link, or the power supply. RBD systems have input and output points, representing the source and destination (target) of a link. These blocks can be arranged as a series connection, parallel connection, or a combination of both.

2.22 WSN Reliability

The essential characteristic that makes RBD the preferred one is its binary operation where each block has two possible states, a failed state and a working state. All failures are independent of one another. The reliability (Dâmaso et al. 2014) of each block is represented in the RBD, such that block B_i has a reliability $R_i(t)$ associated to its working state at time t. Its failure state is henceforth represented by $(1 - R_i(t))$. Series, parallel, and combined RBDs are shown in Fig. 2.15.

If a system is functioning correctly, each component of the RBD operates in a working state, and the system is represented by a series combination. If a series combination is represented by n independent components where

$$p_i(t) = P\{x_i(t) = 1\} \tag{2.15}$$

are the functioning probabilities of blocks b_i, the probability of the entire system to be operational is therefore

$$P\{\phi(\bar{x}(t)) = 1\} = \prod_{i=1}^{n} p_i(t) = 1 \tag{2.16}$$

and the system reliability $R_s(t)$ is

$$R_S(t) = P\left\{\phi(\bar{x}(t)) = 1 = \prod_{i=1}^{n} R_i(t)\right\} \tag{2.17}$$

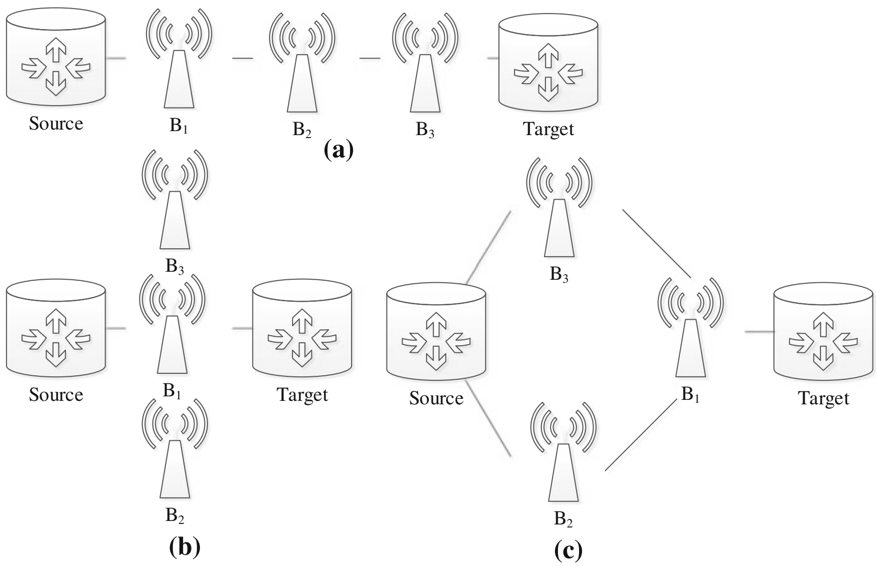


Fig. 2.15 Reliability block diagrams for **a** series, **b** parallel, and **c** combined topologies

where $R_i(t)$ is the reliability of block b_i. From this, a system with n components in series combination has reliability at time t equal to the product of the reliability of the individual blocks. For example, if the system contains three blocks of reliability estimated at 0.9, 0.8, and 0.7 (90, 80, and 70 %), the system reliability is the product of the three, 0.9 × 0.8 × 0.7 = 0.504 (50 %). In a parallel system, with n components (blocks) making up the entire topology, the reliability of the system is equal to the complement of products of unreliability of all the blocks at time t such that

$$R(t) = 1 - \prod_{i=1}^{n}(1 - R_i(t)) \tag{2.18}$$

and the reliability of the system differs from a series combination. The above system with three blocks (Fig. 2.15b), reliabilities of 0.9, 0.8, and 0.7, would therefore equate to a total reliability of 1 − (0.1 × 0.2 × 0.3) = 0.994 (99.4 %). System reliability is therefore improved in parallel combination, and combining blocks in series and parallel combinations can improve range and reliability for optimal operation. In addition, a minimal path is a set of components organized in series that guarantees system operation and a minimal cut is a set of components that implies failure. The minimal path in Fig. 2.15 a) is ($\{B_1B_2B_3\}$) and the minimal cuts are ($\{B_1\}$, $\{B_2\}$, and $\{B_3\}$). To simplify a system can be a complex operation, and Boolean algebra can be applied to create a less complex equivalent system. The sum of disjoint products (SDP) method is such a method, based on Boolean algebra, which evaluates the probability of system operation by the union of the minimal paths or system failure by the union of the minimal cuts. If two events A and B have components in common, the following equation is used to evaluate the probability of union of the events ($A \cup B$) such that

$$R(A \cup B) = R(A) + R(\bar{A}B). \tag{2.19}$$

A system with n events (A_1, A_2, …, A_n) has probability of union equal to

$$R(A_1 \cup A_2 \cup \ldots \cup A_n) = R(A_1) + R(\overline{A_1}A_2 + \cdots + R(\overline{A_1A_2}\ldots\overline{A_{(n-1)}}A_n))$$

so if two or more events have no components in common, the probability of at least one of the events occurring is the sum of the probabilities of the individual events. The reliability of communication links and WSN nodes can be modeled using RBD blocks as depicted in Fig. 2.16.

From Fig. 2.16, the reliability of each node is composed by the series reliability of its components, namely the application software, operating system, hardware components such as sensors, its power supply (generally a battery), and the radio transmitter and receiver. Each of these blocks has an associated reliability defined by the user and/or manufacturer of the component. This value could also be obtained through simulation and measurements. The path model is defined by accounting for all basic blocks of the communication link and WSN nodes. Each

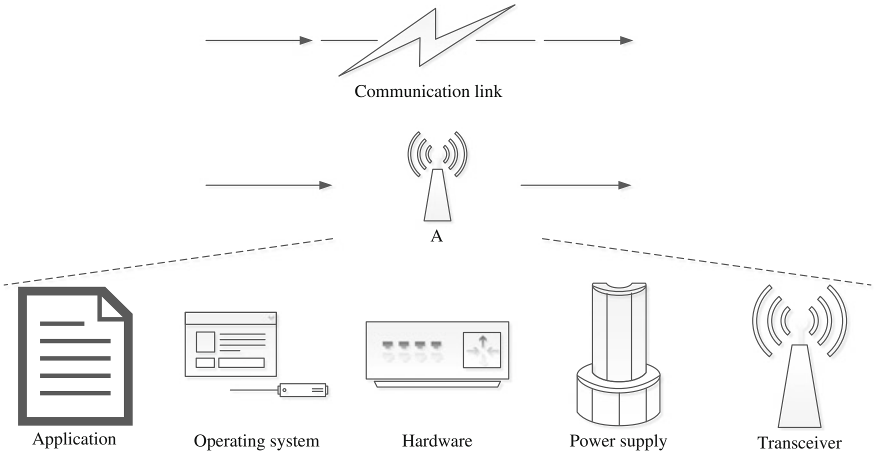


Fig. 2.16 Reliability building blocks consisting of communication link and WSN nodes—inclusive of individual reliabilities of each node

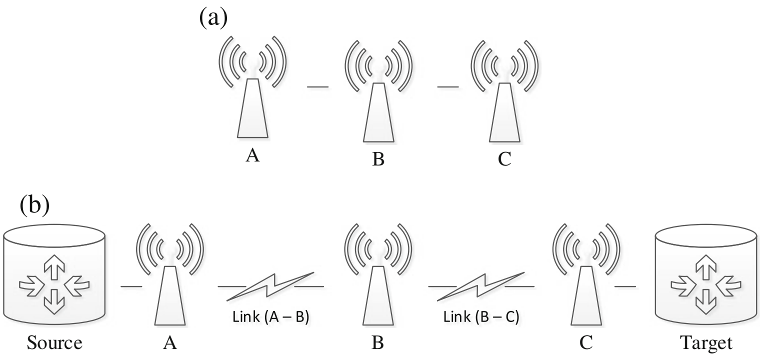


Fig. 2.17 WSN node **a** path model and **b** RBD model

path includes at least two WSN nodes, of which one node is the source and the other the target or destination. For a system consisting of more than two nodes, these nodes are placed between the source and destination. To illustrate this configuration, Fig. 2.17 depicts a path with sensor nodes *A*, *B*, and *C*. Sensor node *A* is the source, *C* is the target and sensor node *B* routes the packet between the source and the target. The corresponding RBD model is also given in Fig. 2.17b.

As seen in Fig. 2.17b, the RBD model is a success-oriented model, therefore all system components *Node A*, *Link (A – B)*, *Node B*, *Link (B – C)*, and *Node C* are arranged in series combination, meaning that if one of these components fails, the entire system fails at time *t*. This topology yields one minimal path, namely {*Node A*, *Link (A – B)*, *Node B*, *Link (B – C)*, *Node C*}. To improve the system reliability,

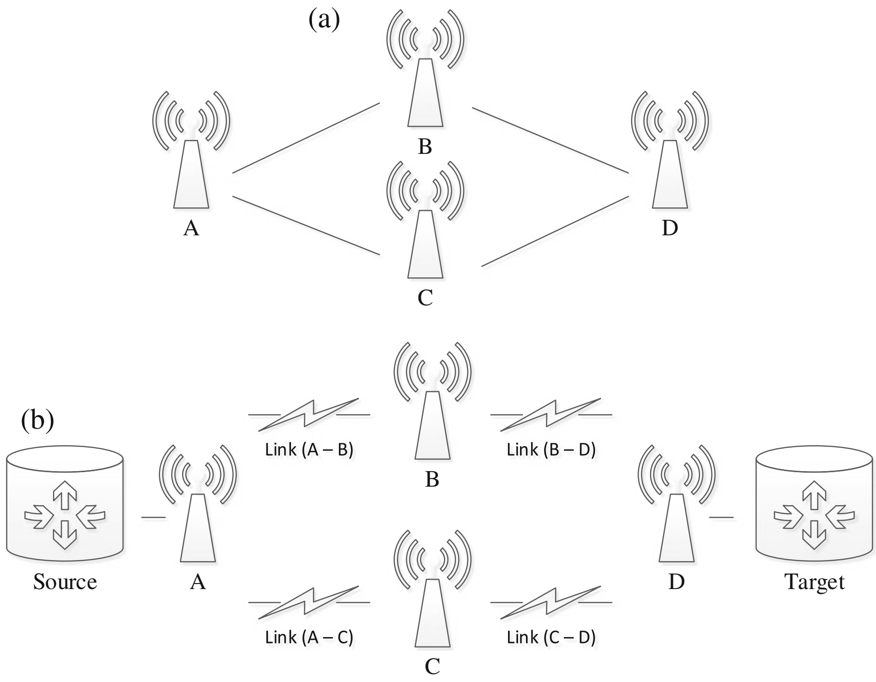


Fig. 2.18 WSN node **a** multipath model and **b** RBD model

multiple path communication can be implemented so that a sensor node can send information through various routes. Such a configuration is shown in Fig. 2.18.

As depicted in Fig. 2.18a, the multipath model shows that two or more nodes forward the same information packet from the same transmitter. To build the RBD model, the basic building blocks must first be established, namely *Node A*, *Node B*, *Node C*, *Node D*, *Link (A – B)*, *Link (A- C)*, *Link (C – D)*, and *Link (B – D)*. There are thus two minimal paths: {*Node A*, *Link (A – C)*, *Node C*, *Link (C – D)*, *Node D*} and {*Node A*, *Link (A – B)*, *Node B*, *Link (B – D)*, *Node D*}. Multipath fading contributes heavily to the unreliability of wireless links, causing fairly large deviations from link quality predictions based on path loss models. Its impact on WSNs is considerable. Although analytical models provide a probabilistic description, multipath fading is a deterministic phenomenon. Moreover, in the case of static nodes, fading is time-invariant. Since data transmission from the target area towards the sink node is the main task of WSNs, the method used to forward data packets between each pair of source-sink nodes is an important issue that must be addressed in developing these networks. Because of the limited capacity of a multi-hop path and the high dynamics of wireless links, a single-path routing approach is unable to provide efficient high data rate transmission in WSNs. Nowadays, the multipath routing approach is broadly utilized as one of the possible solutions to cope with this limitation.

In a wireless channel the delay spread Δt represents the time it takes the radio signal to cover the path difference Δl and the phase change relative to Δl is

$$\frac{\Delta l}{\lambda} = \Delta t \times f \tag{2.20}$$

where λ is the wavelength of operation, related to the frequency f by $\lambda = 2\pi f$. If the phase change is 2π it results in $\Delta l/\lambda = 1$; the coherence bandwidth W_c can be defined by

$$W_c \approx \frac{1}{2\pi \Delta t} \tag{2.21}$$

and the path difference Δl depends on the topology of the environment where the nodes are deployed.

2.23 Conclusion

Developing and developed countries have different trends in populatn growth historically and in terms of future estimations. Developed countries (more specifically megacities in these countries) are showing slower growth compared to developing cities, as can be expected, since carrying capacity compels slower growth, as city expansion must first be addressed. Although the growth of these cities is slower, they do tend to exhibit higher growth in smart city development. Upgrading of existing infrastructure towards smarter information gathering and sharing is taking place in these cities, whereas developing cities are addressing smart initiatives together with coping with large portions of the population moving to the cities and attempting to provide the basic necessities to the population, such as waste management and power. The IoT is a term that groups together smart devices capable of gathering information, using sensors and sending data to remote locations. These devices, or things, form the baseline of implementing WSNs across cities and provide the people living in the cities with valuable information about their health and resources. This chapter explored these terms and provided information and a starting point in designing networks of networks used for smart cities.

References

Al-Ali, A. R., Zualkernan, I., Aloul, F. (2010). A Mobile GPRS-Sensors Array for Air Pollution Monitoring. *IEEE Sensors Journal*, 10(10), 1666–1671.

Buratti, C., Conti, A., Dardari, D., Verdone, R. (2009). An Overview on Wireless Sensor Networks Technology and Evolution. *Sensors*, 9, 6869–6896.

Dâmaso, A., Rosa, N., Maciel, P. (2014). Reliability of Wireless Sensor Networks. *Sensors*, 14, 15760–15785.

Evans, D. (2011). Cisco, The Internet of Things: How the Next Evolution of the Internet is Changing Everything. Retrieved 14 May 2015 from http://www.cisco.com.

Gračanin, D., Eltoweissy, M., Olariu, S., Wadaa, A. (2004). On Modelling Wireless Sensor Networks. *Proc. of the 18th International Parallel and Distributed Processing Symposium (IPDPS'04)*, Santa Fe.

Han, D., Lim, J. (2010). Smart Home Energy Management System using IEEE 802.15.4 and ZigBee. *IEEE Transactions on Consumer Electronics*, 56(3), 1403–1410.

IEEE Standard for Local and Metropolitan Area Networks: Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs). (2011). Retrieved 19 May 2015 from http://www.ieeexplore.ieee.org.

Internet Society Global Internet Report 2014: Open and Sustainable Access for All. (2014). Retrieved 3 May, 2015 from http://www.internetsociety.org.

Internet Society: IPv6. (2014). Retrieved 3 May 2015 from http://www.internetsociety.org.

Jin, J., Gubbi, J., Marusic, S., Palaniswami, M. (2014). An Information Framework for Creating a Smart City Through Internet of Things. *IEEE Internet of Things Journal*, 1(2), 112–121.

Keoh, S. L., Kumar, S. S., Tschofenig, H. (2014). Securing the Internet of Things: A Standardization Perspective. *IEEE Internet of Things Journal*, 1(3), 265–275.

Khanafer, M., Guennou, M., Mouftah, H. T. (2014). A Survey of Beacon-Enabled IEEE 802.15.4 MAC Protocols in Wireless Sensor Networks. *IEEE Communications Surveys and Tutorials*, 16(2), 856–875.

Kulkarni, A., Payne, J., Mistretta, P. (2015). Integrating SCADA, Load Shedding, and High-Speed Controls on an Ethernet Network at a North American Refinery. *IEEE Transactions on Industry Applications*, 51(2), 1360–1368.

Martalò, M., Buratti, C., Ferrari, G., Verdone, R. (2013). Clustered IEEE 802.15.4 Sensor Networks with Data Aggregation: Energy Consumption and Probability of Error. *IEEE Wireless Communications Letters*, 2(1), 70–73.

Otani, T., Kobayashi, H. (2013). A SCADA System Using Mobile Agents for a Next-Generation Distributed System. *IEEE Transactions on Power Delivery*, 28(1), 47–57.

P. Park. (2011). Modelling, Analysis, and Design of Wireless Sensor Network Protocols. Doctoral Thesis, KTH Electrical Engineering, Stockholm.

Parikh, A., Joshi, H., Xing, L., Vokkarane, V. M., Liu, H. (2009). Optimizing Sensor Count in Layered Wireless Sensor Networks. *International Journal of Performability Engineering*, 5(3), 296–298.

Sahin, S., Isler, Y. (2013). Microcontroller-Based Robotics and SCADA Experiments. *IEEE Transactions on Education*, 56(4), 424–429.

Sallabi, F. M., Gaouda, A. H., Salama, M. M. A. (2014). Evaluation of ZigBee Wireless Sensor Networks under High Power Disturbances. *IEEE Transactions on Power Delivery*, 29(1), 13–20.

Sharma, D., Verma, S., Sharma, L. (2013). Network Topologies in Wireless Sensor Networks: A Review. *The International Journal of Electronics and Communication Technology (IJECT)*, ISSN: 2230–7190 (Online) 2230-9543 (Print) 4(3), 93.97.

Torabi, N., Rostamzadeh, K., Leung, V. C. M. (2015). IEEE 802.15.4 Beaconing Strategy and the Coexistence Problem in ISM Band. *IEEE Transactions on Smart Grid*, 6(3), 1463–1472.

Zhang, Y., Sun, L., Song, H., Cao, X. (2014). Ubiquitous WSN for Healthcare: Recent Advances in Future Prospects. *IEEE Internet of Things Journal*, 1(4), 311–318.

Zhao, W., Tang, Z., Yang, Y., Yu, J., Wang, L. (2012). Routing in Wireless Sensor Network based on Field Theory. *Journal of Computational Information Systems*, 8(5), 1–8.

Chapter 3
Modelling of Air and Water Pollution Sources

3.1 Introduction

Air pollution is the contamination of the indoor or outdoor environment by any chemical, physical, or biological agent that modifies the natural characteristics of the atmosphere. Household combustion devices, motor vehicles, industrial facilities, and forest fires are common sources of air pollution. Pollutants of major public health concern include particulate matter, carbon monoxide, ozone, nitrogen dioxide and sulfur dioxide. Outdoor and indoor air pollution cause respiratory and other diseases, which can be fatal. Almost 3 billion people, in low- and middle-income countries mostly, still rely on solid fuels (wood, animal dung, charcoal, crop wastes, and coal) burned in inefficient and highly polluting stoves for cooking and heating. In 2012 alone, 4.3 million children and adults died prematurely from illnesses caused by such household air pollution, according to estimates by the WHO. Together with widespread use of kerosene stoves, heaters and lamps, these practices also result in many serious injuries and deaths from scalds, burns, and poisoning. In order to conduct accurate environmental impact, risk analysis, and emergency planning due to air pollution, models are used to predict the effect and behavior of pollution sources. Sources and effects of indoor and outdoor (ambient) air pollution are briefly described in the following sections.

3.2 Indoor Air Pollution

Kitchen stove emissions are the key driver of indoor air quality deterioration. At any given ventilation rate and room size, stoves that emit quantities of pollutants into the indoor environment are likely to increase the average indoor air pollution (IAP). The frequency and duration of stove use change over time due to factors such as the seasonal demands for heating or cooking, the availability and cost of

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_3

fuel, and the condition of the stove. Indoor air quality (IAQ) is also impacted by general outdoor air pollution with factors such as re-infiltration of stove emissions when chimneys are used, nearby burning of trash or crop residues, cooking outdoors, and pollution from nearby traffic, industry, or other households. The International Agency for Research on Cancer has concluded that indoor emissions from household combustion of coal are carcinogenic. In parts of the world where coal is used extensively as a household fuel and the evidence base is strongest, it contains toxic elements, such as fluorine, arsenic, lead, selenium, and mercury. These elements are not destroyed by combustion and technical constraints make coal difficult to burn cleanly in households. South Asia for example has nearly 1.5 billion inhabitants, who account for approximately a quarter of the world's population. Since nearly 70 % of the population of this region lives in rural areas (WHO 2005a) and approximately 74 % relies on solid fuels for household energy requirements (Rehfuess et al. 2006), the region accounts for a major fraction of global exposure to indoor air pollution from smoke that is attributable to combustion of solid fuels. Recent estimates of disease burdens calculated by the WHO indicate that nearly 4 % of the disease burden in the region may be attributable to consequent exposures, with women and children under the age of 5 years being affected the most (WHO 2005b). Nearly all countries in the region are classified as belonging to medium or low human development categories (UNDP 2001) and the profile of several determinants of indoor air pollution that results from cooking and heating is similar within countries of the region. A wide variety of fuels is used in households in developing countries for cooking and heating. Solid fuels refer to both biomass fuels and coal. The most common fuel used for cooking and heating is wood, followed by other solid biomass fuels such as charcoal, dung, agricultural residues, and sometimes even leaves and grass. These fuels are often collected from the local environment in rural areas or are purchased through markets in urban areas. Fuels differ in their energy densities and efficiency. An example of a modern fuel is liquefied petroleum gas (LPG), which has the highest energy content per kilogram of fuel at approximately 45 MJ kg^{-1}. In contrast, crop residues and dung have energy densities of about 14 MJ kg^{-1} of fuel. The efficiency of a fuel is measured by the amount of energy used for cooking compared with that which escapes from the stove without actually heating the food. The efficiency of cooking with LPG is estimated to be approximately 60 % compared with only 12 % for agricultural residues burnt in traditional stoves. All fuels are burned in various types of devices to provide the heat necessary for cooking. The device can be relatively efficient or inefficient and be associated with high or low levels of pollution. Figure 3.1 presents the energy content in MJ kg^{-1} for various types of fuel and other energy sources.

From Fig. 3.1 it is seen that LPG (60 % propane and 40 % butane), diesel fuel, kerosene (jet fuel), 87 octane gasoline, and methane gas have relatively high energy content (approximately 40 MJ kg^{-1} and above), coal, natural gas, wood, and butanol are below 40 MJ kg^{-1} but still viable, whereas battery cells are generally considered to have low energy content per weight. Conversion efficiencies for kerosene stoves range from 35 % for wick stoves to 55 % for pressure stoves; those

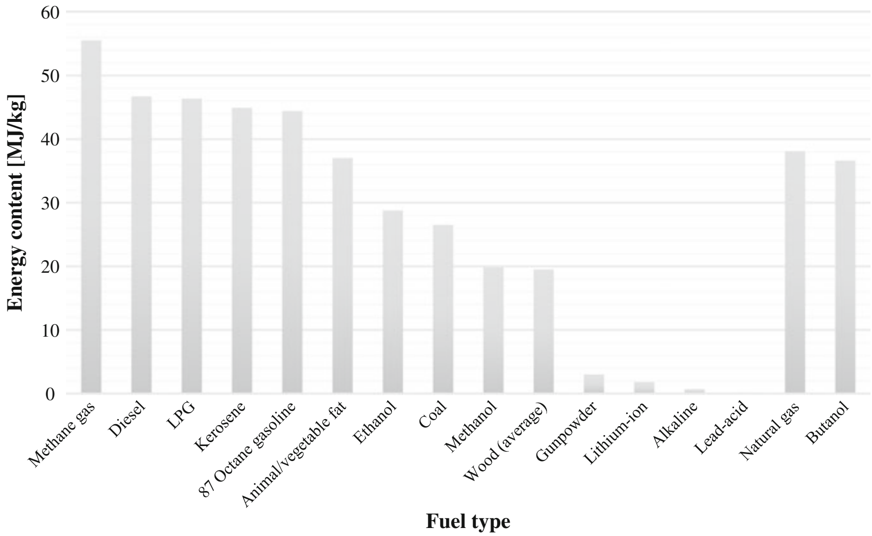


Fig. 3.1 Energy content for various energy sources used in areas worldwide for domestic purposes

for fuelwood stoves range from 15 % for traditional stoves to 25 % for improved stoves. Improved stoves have the potential to reduce indoor air pollution levels, to burn wood or other biomass more efficiently, and sometimes to reduce average cooking times. Table 3.1 presents data adapted from (WHO 2006) regarding the household use of main cooking fuels in selected developing countries.

Table 3.1 represents the household use of main cooking fuels in selected developing countries for rural, urban, and combined national levels. The percentage use of solid fuels and modern fuels for rural areas is graphically displayed in Fig. 3.2.

From Fig. 3.2 it can be noted that a large percentage of developing countries use mainly solid fuels for cooking in households and the general trend shows that fuel types are rarely combined and households tend to only use one of the fuel types (solid fuel usage generally above 85 % in the rural areas). Several countries have adapted using modern fuels as its availability is higher and cost is relatively low, these include predominantly Latin American countries such as Brazil with only 38.3 % of the rural areas using solid fuel, Colombia, Costa Rica, Uruguay, and to some extent Paraguay and El Salvador.

To model IAQ, several types of models are implemented that combine mass rate of pollutant emissions, room geometry, and ventilation and extraction parameters. These models range from simple constructs to complex computer-based simulations.

The simplest construct is the single-zone model which assumes that the air in a zone (such as a room) is perfectly mixed such that any pollutant emitted into the room is uniformly mixed throughout the entire space. The room receives fresh air at a given rate through natural infiltration or by mechanical means and this supply is

Table 3.1 Household use of main cooking fuels in selected developing countries, national household surveys 1996–2003 (WHO 2006)

Country	% Solid fuels			% Modern fuels		
	Rural	Urban	National	Rural	Urban	National
African countries						
Benin	98.7	87.5	94.6	1.3	12.5	5.4
Burundi	99.9	98.1	99.8	0.2	1.9	0.2
Cameroon	98.2	62.2	82.8	1.8	37.8	17.3
Eritrea	97.4	30.4	79.7	2.6	69.6	20.3
Ethiopia	99.9	72.9	95.4	0.1	27.1	4.6
Ghana	99.4	88	95.8	0.6	12	4.2
Kenya	94.7	33.8	81.8	5.1	66.1	18.1
Madagascar	98.8	96.2	98.2	1.1	3.7	1.7
Malawi	99.6	83	97.4	0.4	17	2.6
Mali	99.8	98.4	97.9	0.2	1.6	0.4
Niger	98.4	94.8	97.8	1.6	5.2	2.2
Nigeria (eight states)	94.2	57.4	85.7	5.9	42.6	14
Rwanda	99.9	98.1	99.8	0.1	1.9	0.2
Uganda	98.7	85	96.8	1.3	15	3.2
Zambia	98.1	62.4	85.9	1.9	37.6	14.1
Zimbabwe	93.6	4.7	59.7	6.4	95.3	40.3
Latin American countries						
Bolivia	80.4	7.1	34.4	19.6	92.9	65.6
Brazil	38.3	2.7	9.3	61.7	97.3	90.7
Colombia	48.2	3.4	19.5	51.8	96.6	80.5
Costa Rica	23.9	3.6	11.8	76.1	96.4	88.2
El Salvador	71.7	17.6	37.9	28.3	82.4	62.1
Paraguay	71.3	22	43.3	28.7	78	56.7
Uruguay	1.8	0.4	1.1	98.2	99.6	98.9
Haiti	99.6	91	96.4	0.4	9	3.6
Nicaragua	93.3	46.1	64.4	6.8	53.9	35.6
Asian countries						
India	90.2	29.2	73.7	8.5	66.3	24.3
Nepal	95.6	39.9	89.7	4.4	60.1	10.3
Pakistan	95	28	76	5	72	24
Cambodia	98.7	82	96.3	1.3	18	3.7
Indonesia	83.2	20.4	72.2	16.8	79.6	27.8
Papua New Guinea	98.2	34.4	89.6	1.7	65.5	10.3
Republic of Yemen	53.1	3	41.6	46.9	97	58.4

matched by an outflow, again naturally or mechanically, at the same rate. Non-ventilation pollutant loss mechanisms such as particle deposition onto room surfaces may be included. A constant pollution emission rate can be considered to

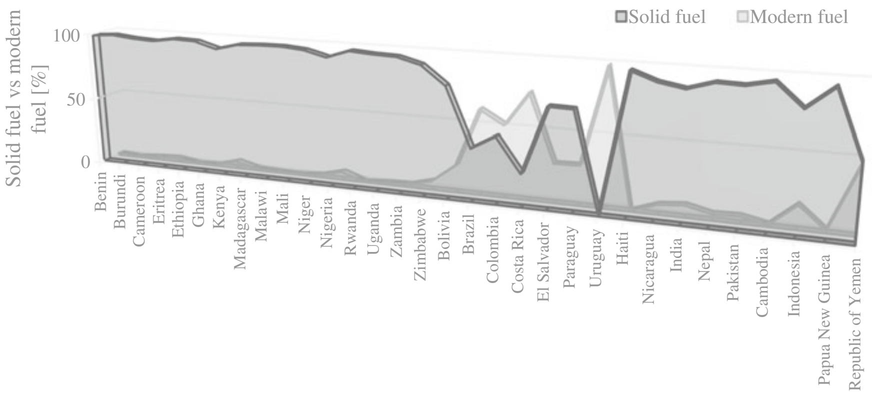


Fig. 3.2 Graphical representation of percentage uses of solid fuels and modern fuels for rural areas in developing countries worldwide, adapted from Table 3.1

decrease the complexity, but varied emission is also sometimes used. The duration of emission rate can be set to reflect the time the source emits into the zone. Fractional terms can be applied to the model to reflect the effect of an exhaust chimney which removes emitted pollutants before it mixes with the room air. Based on these parameters the concentrations in a room can be estimated over a given time.

A more complicated but still tractable construct is a three-zone model, originally formulated to model exposure due to welding. Since perfect mixing is practically not realistic the three-zone model addresses variation in concentration across locations in a room. The model is, as the name suggests, partitioned into three zones. The first zone in a cooking stove scenario is the thermal plume rising upwards to the ceiling. The second zone accounts for the warm air within a given distance to the ceiling, and the third zone is the rest of the room where occupancy (exposure) can occur. It is assumed that the air on each zone is perfectly mixed but that there is a limited air flow between the zones. The thermal plume from the stove creates a circulating airflow pattern in the kitchen, as the warm ceiling-level air cools, it falls into the zone of occupancy and is drawn back into the thermal plume. Similar to the single-zone model, the three-zone model can account for a deposition or loss parameters in the different zones, as well as fractional terms for the venting of emissions out of the ceiling and kitchen zones, and the duration of the emissions rate can be set to reflect the time the source emits into the zone.

Physics-based computer-intensive modelling is achieved through computational fluid dynamics (CFD) models that consider the forces by which air and pollutants are transported within a room. The entire space is divided into thousands or millions (depending on the computational power available) of smaller volumes of air by a mesh of intersecting lines. The points of intersection are termed nodes. The results of CFD models are strongly dependent on the resolution of the nodes. A system of equations account for momentum, thermal energy, and conservation of mass and these equations are solved within a computer program. An application of CFD models is the prediction of three-dimensional velocity fields that describe how air pollutants move between these positions by defining nodes in a room to visualize pollutant dispersion from a biomass cooking stove for example. Emission sources can be modelled in conjunction with the velocity field, allowing pollutant concentrations at these positions to be estimated. An experimental procedure was developed in (Kurabuchi et al. 2013) and the following CFD capture efficiency calculation is presented:

$$\mu = \frac{C_e - C_{bg}}{M} \times Q \times 100 \tag{3.1}$$

where μ is the capture efficiency (given as a percentage), C_e is the exhaust duct concentration, C_{bg} is the background concentration, M is the generated contaminant quantity given in $m^3\ h^{-1}$, and Q is the exhaust flow rate in $m^3\ h^{-1}$. The capture efficiency measurement dealt with the combustion waste gases from above the stove and the cooking effluence from a water surface. Indoor air pollution can to some extend be controlled by educating residents, improving stoves and fuels, and improving extraction mechanisms. Outdoor air pollution is more difficult to control and exposure is not necessarily noticed and its sources might not be known by those affected.

3.3 Outdoor Air Pollution

The WHO reports that in 2012 around 7 million people died (one in eight global deaths) as a result of air pollution exposure (indoor and outdoor air pollution). This finding more than doubles previous estimates and confirms that air pollution is the world's single largest environmental health risk. Globally, 3.7 million deaths were attributable to ambient air pollution (AAP) in 2012. About 88 % of these deaths occur in low- and middle-income (LMI) countries, which represent 82 % of the world population. The Western Pacific and South East Asian regions have the most related deaths with 1.67 million and 936 000 deaths, respectively. About 236 000 deaths occur in the Eastern Mediterranean region, 200 000 in Europe, 176 000 in Africa, and 58 000 in the Americas. The remaining deaths occur in high-income countries of Europe (280 000), Americas (94 000), Western Pacific (67 000), and Eastern Mediterranean (14 000). This data is represented graphically in Fig. 3.3.

The large increase in deaths compared with a previous estimate of 1.3 million deaths from AAP in 2008 is due to additional evidence that has become available on the relationship between exposure and health outcomes and the use of integrated exposure-response functions, an increase in non-communicable diseases, the inclusion of the rural population, whereas the previous estimate only covered the urban population, and the use of a lower counterfactual, i.e. the baseline exposure against which the effect of air pollution is measured. Fine PM is associated with a

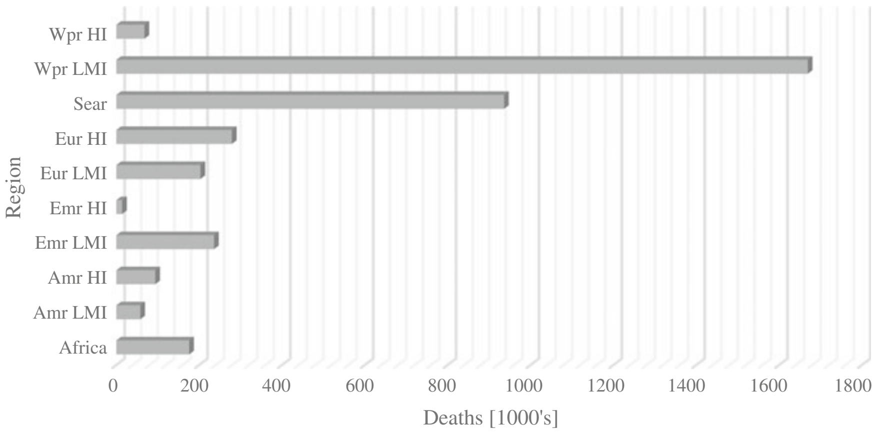


Fig. 3.3 Graphical representation of total number of deaths per region attributable to AAP in 2012—*Amr* America, *Afr* Africa, *Emr* Eastern Mediterranean, *Sear* South-East Asia, *Wpr* Western Pacific, *LMI* low- and middle income, *HI* high-income

broad spectrum of acute and chronic illness, such as lung cancer, chronic obstructive pulmonary disease (COPD), and cardiovascular diseases. Worldwide, PM pollution is estimated to cause about 16 % of lung cancer deaths, 11 % of COPD deaths, and more than 20 % of ischemic heart disease and stroke. PM pollution, a large contributor to AAP, is an environmental health problem that affects people worldwide, but low- and middle-income countries disproportionately experience this effect. The following section describes air pollution modelling with focus on outdoor air pollution since it generally affects large proportions of areas and is difficult to control on individual levels (the general population could be exposed without their knowledge or consent).

3.4 Ambient Air Pollution Modelling

Regional scale air pollution models are generally used in highly polluted cities to restrict industry and vehicle emissions. Recently, models are implemented using software algorithms as opposed to hand calculations from earlier years. These algorithms are implemented to predict the concentration of pollutants downwind from an outdoor source based on acquired knowledge about the emission characteristics such as stack exit velocity, plume temperature, and stack diameter. The immediate terrain surface roughness and topography impact the distribution of the plume and the atmospheric conditions such as wind speed and direction, temperature, and humidity must also be accounted for during calculations. All these variables combined can form complex and time consuming, sometimes inaccurate equations to predict pollution distribution. Several assumptions on the type of source, the terrain, and the environment can be made to simplify calculations with reasonably accurate results. Examples of such assumptions include assuming that the pollutant concentrations do not affect the flow field, supposing a passive dispersion model. Molecular and along-wind diffusion can also be neglected since these effects are generally considered small. In atmospheric assumptions, it is assumed that the wind velocities and concentrations can be decomposed into mean and fluctuating components with the average value of the fluctuating component equal to zero. The mean values are generally based on time averages ranging from 10 to 60 min. Turbulent fluxes are assumed to be linearly related to the gradients of the mean concentration, and the mean lateral and vertical wind velocities set to zero, and the terrain in many cases assumed to be flat.

Air pollution modelling dates back to the 1920s on the battlefield, where military scientists aimed to estimate the dispersion of toxic chemical agents released in various conditions to minimize the effect on their soldiers. The first regulatory air pollution models were only developed in the United States of America during the 1950s and 1960s. Resulting from this and further research in the 1970s, the Gaussian dispersion modelling of plumes was developed and is still relevant for simplified modelling of air pollution. This model captures the essentials of dispersion physics and provides reasonable estimates for downwind dispersion modelling. Today, many regulatory models are adaptations of the original Gaussian dispersion models.

There are two approaches to modelling pollutant dispersion. The first being a simplified and easy to use gross screening model that accounts only for a single elevated stack source and worst-case predictions of meteorological conditions. This method can be applied to determine an estimated figure of pollutant concentration at a distance from a source, where the next approach would be to increase the accuracy and trustworthiness of the results through more complex models. These models require extensive data sets for meteorology and source emissions and include multiple point, area, and/or volume sources, complex terrain, layered atmospheric conditions, and flow around buildings and large structures. These models are therefore computationally intense and require large amounts of computing power. Additionally, there are specialized models used to predict dispersion of hazardous gases for military and chemical warfare, or by chemical industries to model the behavior of accidental chemical releases into the atmosphere that may cause devastating effects to the population. An example of such a disaster is the industrial accident that occurred on July 10th 1976 north of Milan, Italy. A chemical steam release reaction occurred during the removal of ethylene glycol from the reaction mixture to produce 1,2,4,5-tetrachlorobenzene and resulted in an exothermic side reaction that affected 37,000 citizens, killed 3,300 animals and by 1978 it was believed that over 80,000 animals contracted various diseases from the hazardous gases and were subsequently slaughtered. Industrial safety regulations were passed in the European Community in 1982 called the Seveso Directive, and this also sparked interest (and need) to model polluted air dispersion to quantify the reach of such an event.

3.5 Pollutant Concentration Conversion

Air pollution from a point, area, or volume source is distributed depending on the wind speed and direction, the concentration and molecular mass of the pollutant, the height of the source above ground, the surface roughness of the environment, temperature of the pollutant at the source, and the temperature and humidity of the environment. Each of these factors can also be assumed constant, which is generally acceptable in non-extreme cases, or functionally dependent on each other, leading to complex and time consuming calculations. Complexity and processing requirements increase if real-time weather conditions are accounted for, additionally increasing if it is considered that the pollutant itself changes the immediate environment, for instance ambient temperature. To estimate the effect of the pollutant to the surrounding environment, these factors must be accounted for, either as constant effects or as varying quantities.

In certain circumstances it might be required to convert between units of pollutant concentration. Two commonly used units are parts per million per volume (ppmv), and mg m^{-3} (or µg m^{-3} for convenience). The conversion of air pollutant concentrations depends on the temperature at which the conversion is required, generally at 20–25 °C (293–298 K) and at an ambient sea level atmospheric pressure of 1 atm (101.325 kPa), such that

$$\text{ppmv} = \frac{\text{mg}}{\text{m}^3} \times \frac{0.08205 \times T}{M} \tag{3.2}$$

where mg m^{-3} is milligrams of pollutant per cubic meter of air at sea level atmospheric pressure, T is the ambient temperature in Kelvin, ppmv is the air pollution concentration in parts per million by volume, and M is the molecular mass of the air pollutant in g mol^{-1}. To compute for mg m^{-3} from ppmv, it follows that

$$\frac{\text{mg}}{\text{m}^3} = \text{ppmv} \times \frac{12.188 \times M}{T}. \tag{3.3}$$

To perform these conversions, the material's molecular weight is required, and Table 3.2 lists 12 common toxic gases and the molecular weight of each.

To compare the effect of the molecular weight of the chemical compound, the concentrations in ppm for ammonia, sulfur dioxide, nitrogen dioxide, and ozone for 1 mg m^{-3} concentrations are represented in Fig. 3.4.

From Fig. 3.4 it can be seen that the larger the molecular weight of the chemical compound, the smaller the concentration at a target distance from the source. In this example, ammonia (17.031 g mol^{-1}) has a larger concentration evident in the atmosphere at 5 km range from the source compared to sulfur dioxide (64.066 g mol^{-1}). Heavier compounds tend to diffuse into the atmosphere and require larger wind speeds to reach similar contamination at target distance compared to lighter compounds. Altitude variations also affect concentration at the target distance, as described in the following section.

Table 3.2 Chemical compound, formula, and molecular weight of common toxic gases

Compound	Chemical formula	Molecular weight (g mol^{-1})
Fluorine	F_2	18.998
Carbon monoxide	CO	28.01
Phosphine	PH_3	33.997
Hydrogen sulfide	H_2S	34.08
Chlorine	Cl_2	35.453
Nitrogen dioxide	NO_2	46.005
Ozone	O_3	48
Cyanogen	C_2N_2	52.031
Sulfur dioxide	SO_2	64.066
Arsine	AsH_3	77.95
Bromine	Br_2	79.904
Silicon tetrafluoride	SiF_4	104.079
Sulfur tetrafluoride	SF_4	108.07
Boron trichloride	BCl_3	117.17
Silicon tetrachloride	$SiCl_4$	169.9

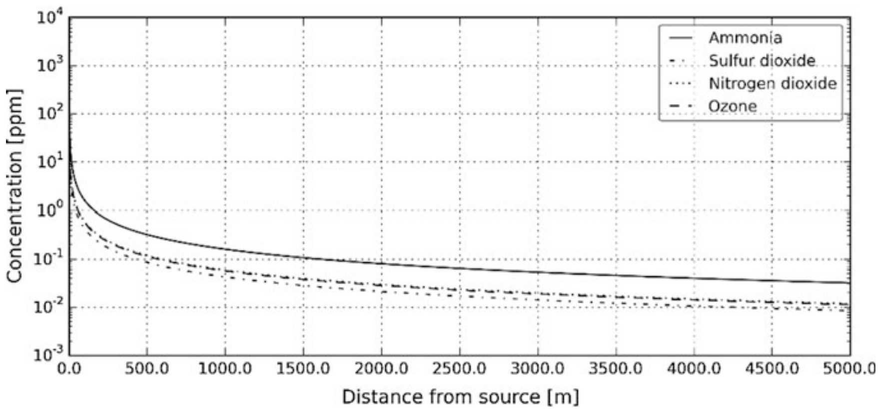


Fig. 3.4 Mean concentration downwind from a point source for various chemical compounds and presented in ppm

3.6 Concentration in Altitude Variations

Air pollutant concentrations expressed as mass per unit volume of atmospheric air at sea level will decrease with increasing altitude, as a result from a decrease in atmospheric pressure. This variation is used to determine the air pollutant concentration at different altitudes at a specified temperature, C_z, from the relationship

$$C_z = C \times \left(\frac{288 - 6.5z}{288}\right)^{5.2558} \tag{3.4}$$

where z is the altitude where the concentration is to be determined in m, and C is the air pollution concentration in mass per unit volume at sea level atmospheric pressure specified at temperature T in Kelvin. A graphical representation of (3.4) showing the exponential decline of air pollution as a function of altitude is given in Fig. 3.5.

Since the pollutant concentration varies with temperature, and temperature of the atmosphere varies with height, determining the variation in temperature due to height leads to determining the pollutant concentration variation. To determine the rate at which the atmospheric temperature varies with increasing altitude, temperature lapse rate is commonly used. This equation is valid for altitudes below 20 km, therefore within the earth's stratosphere. Temperature decreases with altitude starting at sea level, but variations in this trend begin above 11 km, where the temperature stabilizes through a large vertical distance through the rest of the troposphere. In the stratosphere, starting above about 20 km, the temperature increases with height, due to heating within the ozone layer caused by capture of significant ultraviolet radiation from the sun by the dioxygen and ozone gas in this region. A lapse (decline) rate γ as expected below 11 km above sea level is in general defined by

$$\gamma = -\frac{dT}{dz} \tag{3.5}$$

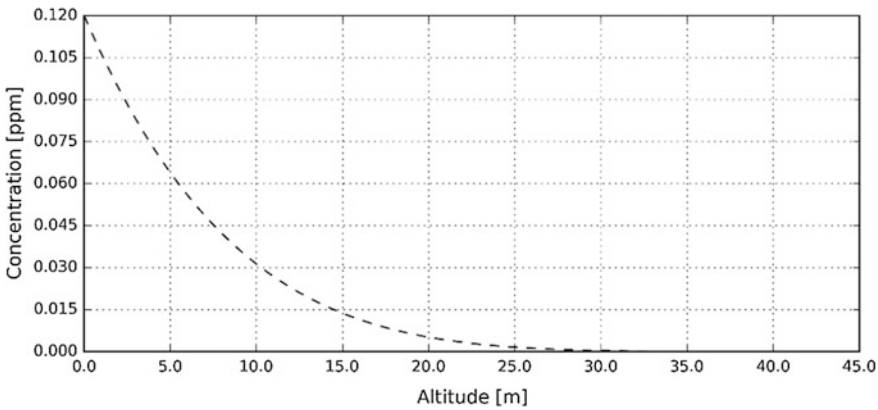


Fig. 3.5 Air pollutant concentration as a function of altitude

where γ is a function of changing temperature T in Kelvin and height above sea level z in m. Two types of lapse rate occur, an environmental lapse rate (ELR) for a stationary atmosphere at a given time and location and an adiabatic lapse rate assuming no exchange of heat with its surroundings. The ELR is defined at an international standard atmosphere by the International Civil Aviation Organization at −6.49 K(°C) per 1000 m from sea level up to 11 km above sea level. From 11 km upwards, ELR is defined constant at −56.5 °C (216.65 K).

The term adiabatic means that no heat transfer occurs into or out of a parcel of air, a fair assumption since air has low thermal conductivity and the bodies of air involved in these calculations are large and transfer of heat by conduction can be neglected. Adiabatic lapse rates are also categorized as dry (unsaturated) or wet (saturated) adiabatic lapse rates and are described below.

3.7 Unsaturated Adiabatic Lapse Rate

Unsaturated adiabatic lapse rate is the rate of temperature decrease as a function of altitude for a parcel of dry or unsaturated air rising under adiabatic[1] conditions. Unsaturated air has less than 100 % relative humidity and therefore its actual temperature is higher than its dew point. From the laws of thermodynamics, the equation for a dry adiabatic lapse rate Γ_d is given by

$$\Gamma_d = \frac{g}{c_p} \tag{3.6}$$

where g is the gravitational constant on the earth of 9.8 m s^{-2} and c_p is the specific heat at constant pressure in J kg^{-1} K^{-1}. c_p depends on the temperature at which the calculations are required and the pressure at this point. As an example, defining the specific heat for water in its liquid phase is written as c_p = 4185.5 J kg^{-1} K^{-1} at 288 K and 101.325 kPa atmospheric pressure. The equation for c_p is given by

$$c_p = \frac{C_p}{m} \tag{3.7}$$

where C_p is heat capacity of the body at constant pressure and m is the mass of the body in kg. The internal energy for unsaturated air changes at a specified rate of −9.8 K(°C) km^{-1}. If the air is saturated with water vapor at its dew point, the lapse rate varies strongly with temperature and is described below.

[1] An adiabatic process is one that occurs without transfer of heat or matter between a system and its surroundings.

3.8 Saturated Adiabatic Lapse Rate

To determine the saturated adiabatic lapse rate, it is first required to define w_s as the saturation mixing ratio of water vapor and dry air where

$$w_s = \varepsilon \frac{e_s}{p - e_s} \quad \text{following that} \quad e_s \approx \frac{w_s p}{\varepsilon} \tag{3.8}$$

where e_s is the saturation vapor pressure in Pa, p is the total pressure in the system in Pa, and ε is a dimensionless ratio of the specific gas constant of dry air to the specific gas constant for water vapor and is approximated as $\varepsilon = 0.622$. There is a temperature dependence of the latent heat of water vapor L_v and the saturation vapor pressure e_s, and this dependency ($e_s(T)$) is given by August-Roche-Magnus in their formula, also known as the Magnus-Tetens approximation,

$$e_s(T) = 6.1094\exp\left(\frac{17.625T}{T + 243.04}\right). \tag{3.9}$$

From the first law of thermodynamics,

$$c_p dT - \alpha dp = dq = -L_v dw_s \tag{3.10}$$

where L_v is the specific latent heat of evaporation of water at 250.1 kJ kg^{-1}, the heat exchange between the air parcel and its environment is zero, but heat is released within the air parcel by the phase change from water vapor to liquid water/ice. As vapor is converted to liquid water, $w = w_s$ decreases and $dq = - L_v dw_s$ where $dw_s < 0$. Neglecting the weak dependence of w_s on pressure, thus using $\alpha dp = -gdz$, it can be shown that (3.10) becomes

$$-L_v dw_s = c_p dT + g dz \tag{3.11}$$

$$\frac{dw_s}{dz} = -\frac{c_p}{L_v}\left[\frac{dT}{dz} + \frac{g}{c_p}\right] \tag{3.12}$$

$$\frac{dw_s}{dz} = -\frac{c_p}{L_v}[\Gamma_d - \Gamma_s] \tag{3.13}$$

which is a constant value. To derive the saturated adiabatic lapse rate by noting

$$\Gamma_s = -\frac{dT}{dz} = \frac{g}{c_p} + \frac{L_v}{c_p}\frac{dw_s}{dz}, \tag{3.14}$$

it can be approximated by applying the general Clausius-Clapeyron relationship,

$$\ln \frac{P_1}{P_2} = -\frac{L}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \tag{3.15}$$

where P_1 and P_2 indicate the difference in pressure, T_1 and T_2 indicate a difference in temperature, L is the specific latent heat, and R is the specific gas constant. Applying this function to meteorology, atmospheric saturation water vapor under typical atmospheric conditions can be modelled as a rate of change, such that

$$\frac{de_s}{dT} = \frac{L_v e_s T}{R_v T^2} \tag{3.16}$$

where R_v is the specific gas constant of water vapor. Rewriting (3.14) as

$$\frac{dw_s}{dz} = \left(\frac{\partial w_s}{\partial p}\right)_T \frac{dp}{dz} + \left(\frac{\partial w_s}{\partial T}\right)_p \frac{dT}{dz} \tag{3.17}$$

and since $dp/dz = -pg$ and $\Gamma_d = -dT/dz$, (3.17) can be written such that

$$\frac{dw_s}{dz} = -\left(\frac{\partial w_s}{\partial p}\right)_T pg - \left(\frac{\partial w_s}{\partial T}\right)_p \Gamma_s \tag{3.18}$$

and solving for (3.18) and (3.13), the saturated adiabatic lapse rate is given by

$$\Gamma_s = \Gamma_d \frac{1 + \frac{L_v w_s}{R_d T}}{1 + \frac{\varepsilon L_v^2 w_s}{c_p R_v T^2}} \tag{3.19}$$

where R_d is the specific gas constant of dry air in J kg^{-1} K^{-1}. From (3.19) it is therefore evident that the saturated adiabatic lapse rate is dependent on the dry/unsaturated lapse rate given by (3.6). The level at which an air mass attains saturation by adiabatic ascent is called the lifting condensation level (LCL). If vertical ascent continues above the LCL, condensation will occur and the latent heat of phase change will be released. This latent heating will cause the temperature to decrease more slowly with pressure above the LCL than below. This means that the lapse rate above the LCL will be smaller than the lapse rate below the LCL. To describe the behavior of the parcel a pseudo-adiabatic process or irreversible saturated adiabatic process is used where all the condensed water or sublimated ice falls out of the air parcel as soon as it is produced. A pollution dispersion screening model for first order effect approximation is described in the following section.

3.9 Pollution Dispersion—Screening Model

Applying a simplified screening analysis or model to air pollution dispersion before implementing complex and extensive models is useful in determining an estimate of the overall effects and possible range of contamination. One such model to estimate the worst case mean concentration (C_{wc}) downwind of a point source is a gross screening model suggested by (Hanna et al. 1996), and this model is given by

$$C_{wc} = \frac{10^9 Q}{U H_{wc} W_{wc}} \tag{3.20}$$

where Q is the source strength or emission rate of gas or particulate in kg s^{-1}, C_{wc} is the worst case concentration in μg m^{-3}, U is the worst case wind speed at height z = 10 m (approximately 1 m s^{-1}), W_{wc} is the worst case cloud width in m (generally assumed to be $W_{wc} = 0.1x$ where x is the distance to the source, x = 5000 m for this example and therefore W_{wc} = 500 m), and H_{wc} is the worst case cloud depth (approximately H_{wc} = 50 m). Again, some assumptions must be made when implementing this model, such as assuming that the mean concentration is inversely proportional to the wind speed, directly proportional to the source release rate, and inversely proportional to the plume cross-sectional area. The following figure depicts a graphical representation of this formula for varying target distance from the source, from zero to 5000 m (5 km). The equation is implemented for a mean mass of 1 kg gas over a 30-minute period and assuming a wind speed of 1 m s^{-1}.

From Fig. 3.6, it is noticeable that the effect (concentration) of the toxic gas decreases logarithmically with distance from the source. Initial concentration of the pollutant is approximately 20^4 μg m^{-3} (160 mg m^{-3}) and at 5 km from the source, it decreases to approximately 22.2 μg m^{-3} (0.022 mg m^{-3}). This can give analysts a

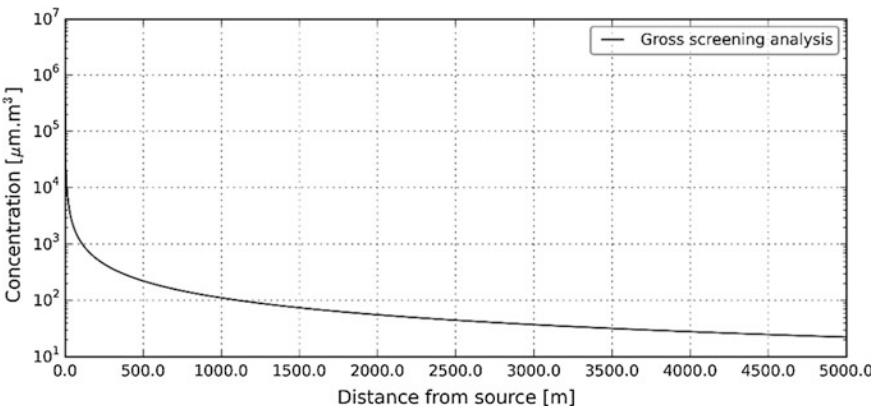


Fig. 3.6 Worst case mean concentration downwind from a point source, simplified model as presented in (3.20)

good first-order estimation on expected contamination in the proximity of the source, assuming the largest effects being in the down-wind velocity. Generally, the concentration in μg m^{-3} is converted to parts per million (ppm), using (3.2) and (3.3). Therefore, assuming in the above example that the contaminant is ammonia, with molecular weight of 17.031 g mol^{-1}, the concentration in ppm is 0.032 ppm at a distance of 5 km, and 230 ppm at the source. If taking into account plume characteristics and stack height, the Gaussian plume model can be implemented, as described in the following section.

3.10 Pollution Dispersion—Gaussian Plume

For increased accuracy a more complex dispersion model can be used by combining the diffusion equation and the Gaussian plume model and performing a mass balance on a small control volume that describes a continuous cloud of material dispersing in a turbulent flow. This model can be represented as follows:

$$\frac{dC}{dt} + U\frac{dC}{dx} = \frac{d}{dy}\left(K_y\frac{dC}{dy}\right) + \frac{d}{dz}\left(K_z\frac{dC}{dz}\right) + S \tag{3.21}$$

where x is the along-wind coordinate measured from the source, y is the cross-wind coordinate direction, z is the vertical coordinate measured from the ground as reference, C is the mean concentration of diffusing substance at a point *(x,y,z)* in kg m^{-3}, K_y and K_z are the eddy diffusivities in the direction of the x- and y-Cartesian axes in m^2 s^{-1}, U is the mean wind velocity along the x-axis in m s^{-1}, and S is the source or sink term which represents the net production or destruction of the pollutant due to its source or removal mechanism in kg m^{-2} s^{-1}. A graphical representation of such a system is given in Fig. 3.7.

The analytical solutions to this equation were already studied in the 1920's where the eddy diffusivities K_y and K_z were used to relate turbulent fluxes of material against the mean gradients of the concentration (these values are generally unknown in practical situations and need to be estimated). By term-wise interpreting (3.21), the following assumptions or observations are made to simplify the characterization procedure. Inspecting each of the terms in (3.21), it can be identified that

$$\frac{dC}{dt} + U\frac{dC}{dx} \tag{3.22}$$

represents the time rate of change *dt* and advection of the cloud by the mean downwind component *dx*. The turbulent diffusion of the material relative to the center of the pollution cloud is given by the term involving the eddy diffusivities

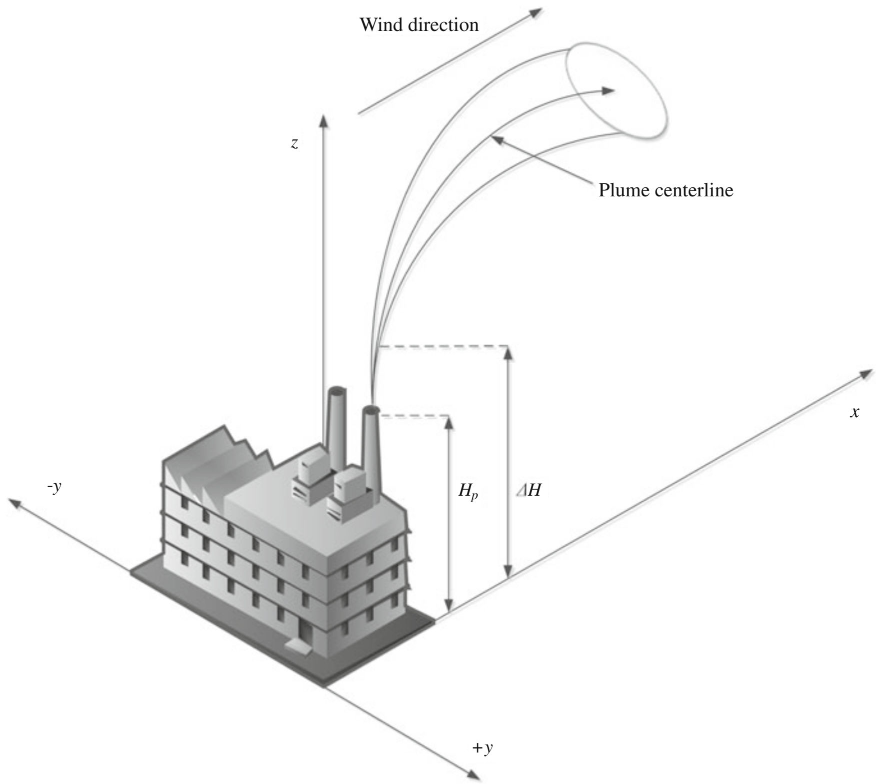


Fig. 3.7 Gaussian distribution model graphical representation showing distributed parameters

$$\frac{d}{dy}\left(K_y\frac{dC}{dy}\right) \text{ and } \frac{d}{dz}\left(K_z\frac{dC}{dz}\right) \tag{3.23}$$

whereas the final term in (3.21), S, represents the source term of the net production or destruction of the pollutant. For a continuous point source released at the origin in a homogenous turbulent flow, the analytical solution to the diffusion Eq. (3.21) is therefore

$$C(x,y,z) = \frac{Q}{4\pi x\sqrt{K_yK_z}}\exp\left(\frac{-y^2}{4K_y\left(\frac{x}{U}\right)}\right)\exp\left(\frac{-z^2}{4K_z\left(\frac{x}{U}\right)}\right). \tag{3.24}$$

The following Gaussian parameters are defined:

$$\sigma_y = \sqrt{2K_y\frac{x}{U}} \tag{3.25}$$

and

$$\sigma_z = \sqrt{2K_Z \frac{x}{U}}. \tag{3.26}$$

For an elevated plume defined by the stack height of the source H_p and the plume rise ΔH, resulting from initial buoyancy and momentum of the plume release from the source, if by setting $z = H_p$ and $x = U_p$ then (3.24) is written as

$$C(x,y,z) = \frac{Q}{2\pi U_p \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left[\exp\left(-\frac{\left(z-H_p\right)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{\left(z+H_p\right)^2}{2\sigma_z^2}\right)\right]. \tag{3.27}$$

The second z-exponential term in (3.27) is added to account for the fact that the pollutant gas cannot diffuse downward through the ground at $z = 0$. This image term can be visualized as an equivalent source located at $z = -H_p$ below the ground where an increase in altitude is defined by a positive value of z. Assuming that it is required to determine the pollutant concentration at ground level (a valid assumption since the concentration is generally measured where humans are affected) the value of z is set equal to zero. From this, the Gaussian plume equation becomes

$$C(x,y,z=0) = \frac{Q}{2\pi U_p \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \exp\left(-\frac{H_p^2}{2\sigma_z^2}\right) \tag{3.28}$$

Using (3.28), and again using ammonia as example, the Gaussian distribution from a stacked point source for wind speeds 0.5, 1, and 2 m s^{-1} is depicted in Fig. 3.8. In this example, the source production rate is 0.5 kg m^{-2} s^{-1}, the eddy

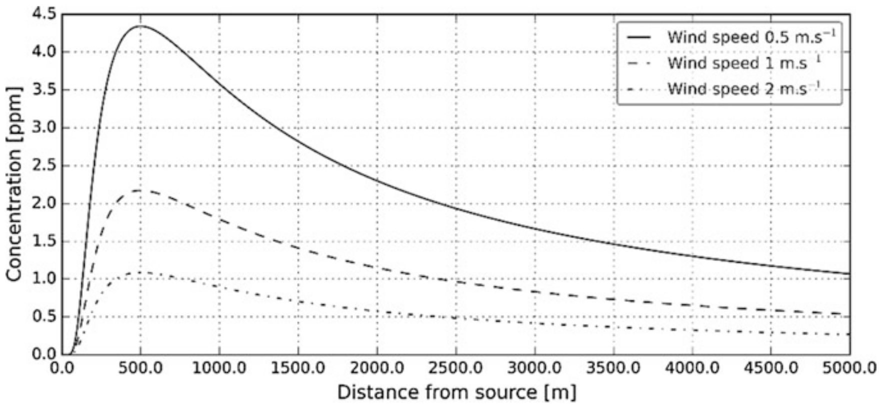


Fig. 3.8 Gaussian distribution model for varied wind speeds up to 5 km from the source using (3.28)

diffusivities K_y and K_z are 0.3 and 0.5 $m^2\ s^{-1}$ respectively, stack height is 10 m and the reference point for measurements is on the ground at $z = 0$ m. The results are converted to ppm using (3.2).

In Fig. 3.8, the chemical compound is not varied as in Fig. 3.4 since similar results are expected based on the molecular weight of the compound. An important consideration to note here is the effect of the wind speed in the direction of the target and how it influences the concentration downwind. Additionally, it is noted that there is zero concentration of the chemical compound at the source, and this concentration increases up to a maximum at approximately 500 m from the source in this example. This is due to the initial velocity and buoyancy of the emitted plume from the source, and the measurement point in the *z*-axis of the Cartesian plane. Looking at a constant height near the source, the plume is ejected above this height and starts contaminating this region downwind as the plume distributes outward. To estimate vertical distribution, a length scale is introduced and described in the following section.

3.11 Vertical Distribution

To determine the vertical distribution due to a ground-level source, H_p is set to zero, simplifying (3.27) to

$$C(x, y, z) = \frac{Q}{2\pi U_p \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \exp\left(-\frac{z^2}{2\sigma_z^2}\right) \quad (3.29)$$

but this equation is generally considered to be inaccurate since for a ground-level source the vertical profile varies by $\exp(-z^{1/2})$ as opposed to the Gaussian profile $\exp(-z^2)$ due to large variations of the diffusivity K_z near the ground. To allow for the vertical variation of K_z and the vertical variation of the velocity profile, a non-Gaussian model is written as

$$C(x, y, z) = \frac{Q}{\sqrt{2\pi} U_p \sigma_y} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) f(z) \quad (3.30)$$

where *f(z)* is a normalized function which describes the vertical distribution of material in the plume. The model must satisfy $\iint CUdydz = Q$ to ensure conservation of mass, where the integration is taken over the *y* and *z* planes perpendicular to the plume axis. Inclusion of a logarithmic wind profile due to surface roughness can additionally be accounted for to better describe wind profiles. This follows that wind speed U is not a constant parameter but varies by height z, such that

$$U(z) = u_z = U_{10}\left(\frac{z}{10}\right)^p \tag{3.31}$$

where U_{10} is the measured wind speed at 10 m above the reference point and p is a power law coefficient which increases with surface roughness. Its value ranges from approximately 0.07 for flat surfaces such as lakes or smooth seas, to about 0.4 for large forest of urban areas.

Atmospheric convection also modifies the shape of the mean wind speed profile and such a profile for use in stable environments modifies the general logarithmic wind profile relation such that

$$U(z) = u_z = \frac{u_*}{k}\left[\ln\frac{z}{z_0} + 4.5\frac{z}{L_{MO}}\right] \tag{3.32}$$

where k is the von Karman constant (k = 0.4), z_0 is an approximate roughness height, u_* is the frictional or shear velocity, and L_{MO} is the Monin-Obukhov length scale. To determine u_* and L_{MO} could result in complex calculations and a brief guideline to this process is described in this section. Shear velocity is the velocity profile near the boundary of flow and is given by

$$u_* = \frac{ku_{ref}}{\ln(z_{ref}/z_0) - \Psi_m(z_{ref}/L_{MO}) + \Psi_m(z_0/L_{MO})} \tag{3.33}$$

where u_{ref} is the wind speed at reference height z_{ref}. The stability terms Ψ_m are equated by

$$\Psi_m(z_{ref}/L_{MO}) = 2\ln\left(\frac{1+\mu}{2}\right) + \ln\left(\frac{1+\mu^2}{2}\right) - 2\tan^{-1}\mu + \pi/2 \tag{3.34}$$

and

$$\Psi_m(z_0/L_{MO}) = 2\ln\left(\frac{1+\mu_0}{2}\right) + \ln\left(\frac{1+\mu_0^2}{2}\right) - 2\tan^{-1}\mu_0 + \pi/2 \tag{3.35}$$

for which the values for μ and μ_0 are given by

$$\mu = \left(1 - \frac{16z_{ref}}{L_{MO}}\right)^{1/4} \tag{3.36}$$

and

$$\mu_0 = \left(1 - \frac{16z_0}{L_{MO}}\right)^{1/4} \tag{3.37}$$

respectively. The initial step in the iteration to solve u_* from (3.33) is to assume a neutral limit $\Psi_m = 0$ and setting $\mu = \mu_{ref}$. From this point,

$$L_{MO} = -\frac{\rho c_p T_{ref} u_*^3}{kgH} \tag{3.38}$$

where H is the surface heat flux in W m^{-2}, g is the acceleration of gravity of 9.8 m s^{-2}, c_p is the specific heat of air at constant pressure, ρ is the density of air, and T_{ref} is the ambient temperature representative of the surface layer. The values of L_{MO} and u_* are iteratively calculated, initially $L_{MO} = \infty$, until the value of L_{MO} converges and changes by less than 1 % assuming an initial neutral conditions. To determine the total plume rise from initial conditions of momentum and buoyancy, the following section is presented.

3.12 Momentum and Buoyancy

For most dispersion models the total plume rise ΔH must be calculated depending on the momentum flux and buoyancy flux as emitted from the source stack. These initial exit conditions determine the real range of the plume downwind. The equations for momentum flux (F_m) and buoyancy flux (F_b) are given by

$$F_m = w_s^2 R_s^2 \frac{T_a}{T_s} \tag{3.39}$$

and

$$F_b = g w_s^2 R_s^2 \frac{(T_s - T_a)}{T_s} \tag{3.40}$$

where w_s is the stack exit velocity (not to be confused with the mean wind velocity U), R_s is the stack radius, T_a and T_s are the ambient and source temperatures respectively, and g is the gravitational constant of 9.8 m s^{-2}. It can be seen that momentum and buoyancy flux increase exponentially proportional to the stack diameter and exit velocity. Both fluxes have units m^4 s^{-2} and require several assumptions to be made in order to find a closed-form solution for (3.39) and (3.40). These include assuming constant wind speed above the stack, a fully bent over plume (not rising after a finite distance), the plume being driven downwind by the mean wind direction and velocity, and that the plume and ambient density are equal for momentum flux calculations but not for buoyancy flux calculations. Accepting these assumptions to hold true, the resulting plume rise trajectory in a neutrally stable atmosphere can be written in the following format:

$$\Delta H = \left(\Delta H_m^3 + \Delta H_b^3 + H_0^3\right)^{1/3} - \Delta H_0 \tag{3.41}$$

where ΔH_m is the rise component due to its initial momentum, given by

$$\Delta H_m = \left(\frac{3}{\beta^2}\right)^{1/3} \frac{F_m^{1/3}}{U_p^{2/3}} x^{1/3}, \tag{3.42}$$

ΔH_b is the rise component due to its initial buoyancy, given by

$$\Delta H_b = \left(\frac{3}{2\beta^2}\right)^{1/3} \frac{F_b^{1/3}}{U_p} x^{2/3}, \tag{3.43}$$

and ΔH_0 is a constant that accounts for the initial size of the plume, calculated by

$$\Delta H_0 = \frac{R_s}{\beta}\left(\frac{T_a}{T_s}\frac{w_s}{U_p}\right)^{1/2}. \tag{3.44}$$

The coefficient β is a measure of the rate at which ambient air is entrained by the plume and a value of $\beta = 0.6$ can be used, as recommended by (Briggs 1984) and (Davidson 1989). Figure 3.9 represents the contribution of each term (ΔH_m, ΔH_b, and ΔH_0) towards the total plume rise trajectory. The assumed values for this example are $w_s = 6$ m s^{-1}, $R_s = 15$ m, $T_a = 300$ K, $T_s = 373$ K, $g = 9.8$ m s^{-2}, $\beta = 0.6$, $U_p = 3$ m s^{-1} (in the x-direction). The rise trajectory is given for the first 100 m from the source exit.

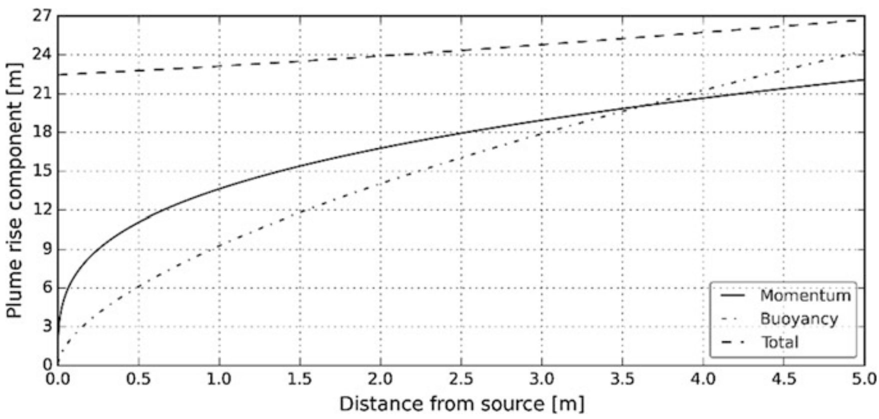


Fig. 3.9 Total plume rise trajectory and separate contributions as given in (3.41)

From Fig. 3.9 it is seen that the momentum and buoyancy contributions to the total plume rise trajectory are relatively large where initially momentum contributes the largest component, and after a certain distance from the source, the buoyancy component dominates momentum. The constant component of ΔH_0 contributes about a third of the overall plume rise. To account for wind speed variations, the following section is presented.

3.13 Wind Speed Variations

Derived in Chen et al. (1998), the wind u_z speed at height z above the ground can be determined by the following logarithmic law, similar to (3.32) but with a generalized L_{MO} term:

$$U(z) = u_z = \frac{u_*}{k}\left[\ln\left(\frac{z}{z_0}\right) + \psi(z, z_0, L_{MO})\right] \tag{3.45}$$

where generally this equation is considered under neutral stability and therefore z/L_{MO} is equal to zero and ψ can be discarded to decrease the complexity of (3.32). Assuming $\psi = 0$, (3.32) and therefore (3.45), such that in its general form,

$$U(z) = u_z = \frac{u_*}{k}\ln\left(\frac{z}{z_0}\right). \tag{3.46}$$

Collecting terms with $\psi = 0$, (3.46) can be written as

$$u_z = b_1 + b_2 \ln z \tag{3.47}$$

where

$$b_1 = -\frac{1}{k}u_* \ln z_0 \tag{3.48}$$

and

$$b_2 = \frac{u_*}{k} \tag{3.49}$$

For wind speed at reference height 10 m, (3.45) becomes

$$u_{10} = \frac{u_*}{k}\ln\left(\frac{10}{z_0}\right) \tag{3.50}$$

and calculating u_z/u_{10} from (3.45) and (3.50) as

$$\frac{u_z}{u_{10}} = \frac{u_*/k \ln(z/z_0)}{u_*/k \ln(10/z_0)} = \frac{\ln z - \ln z_0}{\ln 10 - \ln z_0} = 1 + \frac{\ln(z/10)}{\ln(10/z_0)} \tag{3.51}$$

which eliminates the factor k, and therefore;

$$u_z = u_{10}\left(1 + b \ln \frac{z}{10}\right) \tag{3.52}$$

where

$$b = \frac{1}{\ln 10 - \ln z_0} \tag{3.53}$$

it follows that

$$b_1 = u_{10} \quad \text{and} \quad b_2 = u_{10}^b. \tag{3.54}$$

This logarithmic law stems from the turbulent boundary layer theory that was developed based on practical experiments. It is however cumbersome since parameters u_* and z_0 cannot be directly measured and must be adjusted based on first approximations. For this reason, the power law for wind velocity modelling is used. The power law has the following form:

$$u_z = b_1 z^{b_2}. \tag{3.55}$$

The term b_2 is the power law exponent that depends on surface roughness and atmospheric stability. The relationship between wind speed at 10 m and wind speed at any height from 0 to 10 m is therefore given by

$$u_z = u_{10}\left(\frac{z}{10}\right)^{b_2} \tag{3.56}$$

which is similar to (3.31) where $p = b_2$ which depends on surface roughness and atmospheric stability and the value of this parameter is given from various sources in Table 3.3 (Chen et al. 1998).

Figure 3.10 presents (3.56) with varied values for b_2, where $b_2 = 0.05$, 0.142, and 0.8.

From Fig. 3.10 it can be seen that the wind speed increases to approximately 2.8 m s^{-1} for all three cases, but the value of b_2 determines the rate at which the maximum wind speed is reached, where a higher value for b_2 results in maximum wind speed reached at lower distances from the ground.

Table 3.3 Exponent b_2 for power law wind speed profile

Power law exponent b_2	References
1/7 (0.142)	Sutton (1953)
0.05–0.6	Irwin (1979)
0.0–1.0	Wark and Warner (1976)
1/7–0.40 (0.142–0.40)	Simiu and Scanlan (1978)
1/10–1/3 (0.1–0.33)	ANSI (1982)
0.07–0.55	Turner (1994)
0.080–0.0624	Touma (1997)
0.1–0.8	Strom (1976)

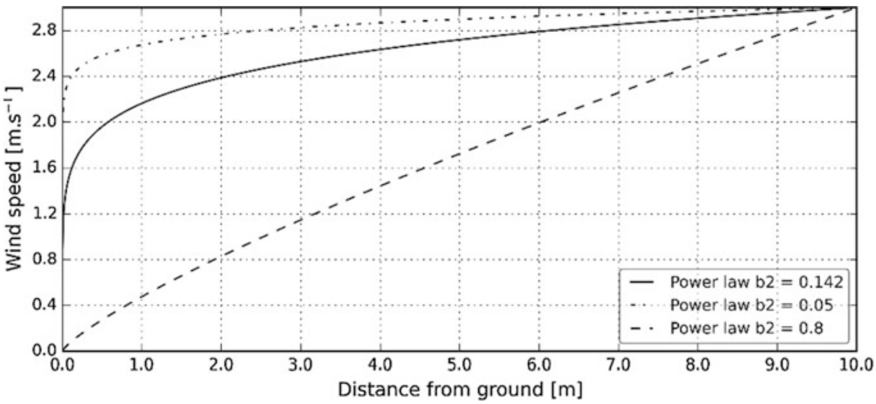


Fig. 3.10 Wind power law profiles for varied exponent b_2 (0.05, 0.142. and 0.8)

3.14 Water Pollution Modelling

Water quality models can be applied to many different water systems such as rivers, lakes, reservoirs, oceans, streams, coastal waters, and estuaries. Models are used to describe the water quality estimations and typically require hydrological and constituent inputs such as water flow and pollutant volume and concentration. Models include terms for dispersion and/or advection transport depending on the hydrological and hydrodynamic characteristics of the water body. Also included are terms for the biological, chemical, and physical reactions among constituents. Advection transport occur due to a liquid in bulk motion, found predominantly in rivers and strong streams. Dispersion transport describes how pollutants disperse into water bodies (similar to air pollution modelling) such as estuaries (river mouths) and is subject to tidal action. Estimating pollution distribution in for example lakes requires a combination of advection and dispersion effects and external factors such as wind speed and direction which increase the complexity of these models. Describing the transport patterns in lakes is generally more difficult compared to rivers and streams, whereas oceans and coastal waters can be described relatively accurately by using large scale flow patterns and tides as the key transport mechanisms. Water quality distribution models can in some way be considered more difficult to implement than air pollution models, accuracy and efficiency of these models vary significantly with confidence of input parameters. Simple models can give a good overview of expected disruption, whereas more complex models generally require accurate inputs to yield useful results.

3.15 Types of Water Pollution

Various types of water pollution from different sources can be distinguished from each other and reacting to each type of pollution is dependent on identifying the type accurately.

Certain wastewater, fertilizers, and sewage contain high levels of nutrients. If these nutrients end up in water bodies, it encourages the growth of algae and weeds. Large amounts of algae can deplete the water body of its oxygen content resulting in natural water organisms dying from oxygen starvation. Nutrient polluted water is generally considered undrinkable and can cause unwanted build-up or clogging of water filters. Damage to drinking water supplies costs municipalities excessive amounts of money to repair and raises the living cost of citizens in these areas. Excessive algae growth damages drinking water supplies, degrades the recreational and aesthetic values of lakes, streams and rivers, and harms aquatic life. Toxic algae blooms, called cyanobacteria, occur in nutrient abundant lakes in warm climates and can be fatal to livestock and pets. In humans, skin contact with cyanobacteria can cause itchy eyes and throat, skin rashes, and hives. Swimmers and other recreationists that ingest these toxins risk a variety of unpleasant to severe health problems, including stomach cramps, vomiting, diarrhea, fever, headache, severe muscle or joint pain, seizures, or convulsions. Children are more vulnerable and may experience more severe reactions such as liver and central nervous system damage. Major sources of nutrient pollution include agricultural runoff and municipal wastewater discharge.

Surface water pollution is associated to pollution of natural water bodies such as rivers, lakes, lagoons, and oceans. Pollution occurs when hazardous materials such as chemicals come into contact with the water surface, combining with the natural source and polluting the larger body. The concentration of the hazardous material is of concern in this type of pollution, and the spread of the contamination depends on the introduced concentration and the flow of the water.

Oxygen depletion, as mentioned in nutrients pollution, starves micro-organisms from oxygen resulting in the decay of these organisms. If too much biodegradable matter is introduced into the water body, it encourages the growth of micro-organisms which use up more oxygen. If the oxygen is depleted, aerobic organisms die and anaerobic organisms grow which can produce harmful toxins in the water body such as ammonia and sulfides.

Pesticides and chemicals used to control disease in farmlands can be washed away by strong rain and eventually diffuse into the soil, and if these chemicals reach the natural underground water, they will cause ground water pollution. Wells and boreholes that are used to extract underground water may then produce toxic and polluted water, and if reused, may spread the toxins further, causing large-scale pollution.

Microbiological organisms like viruses, bacteria, and protozoa are natural entities found in natural water sources. If humans drink from these natural sources without sterilizing the water, these organisms may cause serious illness to humans.

Some substances, chemicals, and particles cannot dissolve in water and are called particulate matter. These materials can eventually settle on the water body floor and can cause damage to aquatic life, either by mistaking it as a food source, choking, or attaching and limiting movement of the aquatic organism that can cause death. Metals can exist in both dissolved and particulate forms. The partition coefficient is defined as the ratio of metal concentration in particulates to the dissolved concentration. Exposure to mercury for example most often happens when humans consume fish contaminated with the element. The major sources of mercury pollution in lakes include coal-fired utilities, taconite processing, and the unintentional release of mercury used in product manufacturing. Examples of the last are municipal trash incineration, fluorescent light bulb breakage, scrapped cars, and thermometers and other gauges. Mercury released into the air or ground can make its way into lakes, where it can potentially be converted to methyl mercury, the form that is readily taken up by fish. The methyl mercury accumulates in fish tissue, and is most concentrated in predator fish.

Industries and farmers commonly use chemicals during manufacturing or other practices, and these chemicals may end up in water bodies. This leads to chemical water pollution and is common with point-source pollution. These chemicals include pesticides that control weeds, insects, and pests. Metals and solvents from industries can also end up in water bodies, which are poisonous to aquatic life and could slow its development, making it infertile or killing it. Estimates of the bioavailability and toxicity of chemical pollutants are needed to make predictions about their ecological impact.

Oil spillage is generally construed as a localized event, but the effect can spread over large distances causing death to aquatic life or contaminating the feathers of birds, causing them to lose their ability to fly and escape from natural dangers, or lose their ability to forage. Oil has a range of physical and chemical properties that can vary considerably between different types of oil. Properties such as its viscosity, volatility, and density determine its rate of spread, evaporation, and dispersion into the water body. Oil is subject to advection and diffusion as it is less dense than water, much of the oil travels as surface slick affected by wind, waves, and surface current.

3.16 Water Quality Treatment

Water quality and treatment standards usually come in one of three forms:

Technological standards which specify that the treatment must include certain technologies or processes. Frequently this standard is expressed in the form that secondary treatment or biological treatment is required without identifying either the water quality of the effluent to be produced, or the exact details of the technical process to be used.

Effluent standards are specifications of the physical, biological, and chemical quality of the effluent to be produced by the treatment. These usually include allowable concentrations of biochemical oxygen demand, chemical oxygen demand, suspended solids, nitrogen, and phosphorus.

Uniform effluent standards are the most common, in which target concentrations of pollutants in the effluent are standardized across the entire country or state. This approach has the advantage of simplicity and ease of implementation. Unfortunately, they are usually inefficient, leading to excessive treatment in some cases, and insufficient treatment in others.

Ambient or stream quality standards, by contrast, seek to fix standards for the quality of the water body receiving the waste and allocate higher scores to drinking water. If, however natural processes of dilution and biodegradation can improve the quality sufficiently before there is significant use of the resource, then the stream standards can be lower. Once the desired stream quality is known, environmental scientists can reverse-engineer to determine the maximum concentrations of each pollutant allowed in each wastewater discharge; these concentrations become the standards for that discharge. Under this arrangement, different discharges in different environmental contexts must meet different effluent standards. Ambient standards have the advantage that they ensure resources are efficiently allocated to address local environmental conditions. The major disadvantages are that the approach requires considerable planning and skilled environmental scientists.

3.17 Water Pollution Statistics

Figure 3.11 summarizes some statistics on global freshwater related issues including sanitation, adapted from (Unwater 2010).

Freshwater ecosystems sustain a disproportionately large number of identified species, including a quarter of known vertebrates. Such systems provide more than US $75 billion in goods and ecosystem services for people, but are increasingly threatened by a host of water quality problems (Vié et al. 2009). The greatest single service freshwater ecosystems provide—marshes in particular—is water purification and the assimilation of waste and is valued at US $400 billion worldwide (Costanza et al. 1997). With the Millennium Development Goals (MDG), the international community committed to halving the proportion of people without

Water quality

- One in nine people worldwide does not have access to improved sources of drinking water
- One in three people worldwide does not have access to improved sanitation
- The major sources of water pollution are from human settlements and industrial and agricultural activities
- 80% of sewage in developing countries is discharged, untreated, directly into water bodies
- Industry dumps an estimated 300-500 megatonnes of polluted waste globally each year
- Nitrate from agriculture is the most common chemical contaminant in the world's groundwater aquifiers
- Approximately 3.5 million people die each year due to inadequate water supply, sanitation, and hygiene
- The biodiversity of freshwater ecosystems has been degraded more than any other ecosystem

Water and biodiversity

- Freshwater holds over 10% of all life on the planet and 35% of all vertebrates
- Between 1970 and 2000 the populations of freshwater species declined by 55%
- Marine and terrestrial species populations have declined 32% in the same time

Water and disasters

- During the period 2000 to 2006, a total of 2 163 water-related disasters were reported globally in the Emergency Disasters Database
- These disasters killed 290 000 people globally and affected more than 1.5 billion people
- US $422 billion worth of damages was inflicted as a result of water disasters worldwide
- Floods, droughts, and windsorms account for 88.5% of the thousand most disastrous events
- More than 83% of all water-related disasters occured in Asia
- Since 1900 more than 11 million people have died and 2 billion affected as a consequence of drought

Water scarcity

- Water use has been growing at more than twice the rate of the global population increase in the last century
- An increase in water withdrawals by 2025 of 50% in developing countries and 18% in developed countries is estimated
- By 2025, 1 800 million people will be living in countries or regions with absolute water scarcity
- By the same time, two thirds of the world population could be under stressed water scarcity conditions

Sanitation

- 1.8 billion people gained access to improved sanitation facilities between 1990 and 2010
- 15% of the world population still practice open defecation
- Access to sanitation, good practice of hygiene, and safe water supply could save 1.5 million children a year
- Every 20 seconds a child dies as a result of poor sanitation

Water and urbanization

- More than half of the world's population lives in cities
- Every second the urban population grows by two people
- 96% of urban populations and 81% of rural populations have improved access to drinking water in 2010
- 79% of urban populations and 47% of rural populations have access to improved sanitation facilities in 2010
- 827 million people live in slums, often lacking adequate water and sanitation services
- 227 million people have moved out of slum conditions between 2000 and 2010 mainly as a result of slum upgrading
- 888 million people is expected to live in slum conditions by 2020 if current trends continue

Fig. 3.11 Statistics on freshwater-related issues including sanitation as reported by (UNwater 2010) access to safe water and sanitation by 2015. Meeting this goal means some 322 million working days per year gained, at a value of nearly US $750 million (SIWI 2005), and an annual health sector cost saving of US $7 billion. Overall, the total economic benefits of meeting the MDG target have been estimated at US $84 billion (SIWI 2005). Poor countries with access to clean water and sanitation services experienced faster economic growth than those without: a study found the annual economic growth rate of 3.7 % among poor countries with better access to improved water and sanitation services, while similarly poor countries without access had annual growth of just 0.1 % (Sachs 2001).

In a comparison of domestic, industrial, and agricultural sources of pollution from the coastal zone of Mediterranean countries, agriculture was the leading source of phosphorus compounds and sediment (UNEP 1996). Nutrient enrichment, most often associated with nitrogen and phosphorus from agricultural runoff, can deplete oxygen levels and eliminate species with higher oxygen requirements, affecting the structure and diversity of ecosystems. Nitrate is the most common chemical contaminant in the world's groundwater aquifers (Spalding and Exner 1993). Mean nitrate levels have risen by an estimated 36 % in global waterways since 1990 with the most dramatic increases seen in the Eastern Mediterranean and Africa, where nitrate contamination has more than doubled (GEMS 2004). According to various surveys in India and Africa, 20–50 % of wells contain nitrate levels greater than 50 mg L^{-1} and in some cases as high as several hundred milligrams per liter.

Point-of-use drinking water treatment through chlorine and safe storage of water could result in 122.2 million avoided disability adjusted life years, a measure of morbidity, at a total cost of US $11.4 billion. (UN WWAP 2003). Nearly 70 million people living in Bangladesh are exposed to groundwater contaminated with arsenic beyond WHO recommended limits of 10 ug L^{-1} (UN WWAP 2009). The naturally occurring arsenic pollution in groundwater now affects nearly 140 million people in 70 countries on all continents (UN WWAP 2009). Even drinking water quality in developed countries is not assured. In France, drinking water testing uncovered that 3 million people were drinking water where the quality did not meet WHO standards, and 97 % of groundwater samples did not meet standards for nitrate in the same study. To combat water quality degradation, equations based on the mass balance principle have been developed and are described in the following sections.

3.18 Mass Balance Equations

Mass balance is considered as the key principle in water quality models. This principle divides the water body into segments used as computational cells and the collective cells are referred to as the grid. Each cell must have a mass balance for each water constituent over time and be divided into Δx, Δy, or Δz directions (or a combination of these) of any rectangular or cubic shape. Most water quality simulation models simulate quality over a consecutive series of discrete time periods, Δt.

The time is divided into discrete intervals t and the flows are assumed constant within these time periods. For each cell and each time period the mass should balance for the segment. Components of the mass balance for each segment include changes by transport mechanism (Tr) into and out of the segment, variations by physical or chemical processes (P) in each segment, and changes of sources to or from the segment (S). The mass balance equation can be presented in the following form:

$$M_i^{t+\Delta t} = M_i^t + \Delta t\left[\left(\frac{\Delta M_i}{\Delta t}\right)_{Tr} + \left(\frac{\Delta M_i}{\Delta t}\right)_P + \left(\frac{\Delta M_i}{\Delta t}\right)_S\right] \tag{3.57}$$

where the mass in computational cell i at the beginning of a time step t is given by M_i^t, and the mass at the end of the time step is given by $M_i^{t+\Delta t}$. Changes in the computational cell i are given by the three division terms with relevant subscripts (transport, physical/chemical, or source). Changes in transport include both advection and dispersion transport variations. Changes by physical or chemical processes include re-aeration and settling, adsorption, transformation, denitrification, and production and predation on phytoplankton. Variations in sources include addition of mass by waste loads and the extraction of mass by intakes. Advection and dispersion transport mechanisms are described in the following section.

3.19 Advection-Dispersion Transport

To model contaminant/pollution transport in porous media such as ground, the advection-dispersion transport model in one direction (x) is represented in its basic form as a partial derivative, from a schematic representation given in Fig. 3.12.

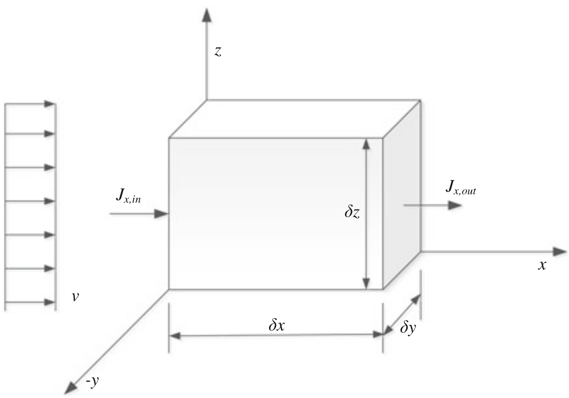


Fig. 3.12 Schematic representation for water pollution dispersion model

Figure 3.12 presents the flux in and out ($J_{x,in}$ and $J_{x,out}$) in the x-direction on the Cartesian plane as a result from moving substance through a defined small area or volume with velocity v. The one-directional partial derivative is given by

$$D\frac{\partial^2 c}{\partial x^2} - v\frac{\partial c}{\partial x} = R\frac{\partial c}{\partial t} \tag{3.58}$$

where D is the dispersion coefficient of both molecular and mechanical dispersion ($D = D_{mol} + D_{mech}$), v is the average linear velocity of the flowing liquid in the advection term, R is the retardation term representing the net result of the process transport through the porous medium, c is the concentration of the substance, and t represents time. Diffusion describes the spread of particles through random motion from regions of higher concentration to regions of lower concentration. The molecular diffusion coefficient D_{mol} in $m^2\ s^{-1}$ is solved by Fick's laws of diffusion, given by

$$J = -D_{mol}\frac{\partial c}{\partial x} \tag{3.59}$$

where J is the diffusion flux of material flowing through an area during a specified time interval with units mol $m^{-2}\ s^{-1}$. The diffusion coefficient depends on the material and the direct environment and can be measured for accuracy. The mechanical dispersion generally dominates the system and is defined by

$$D_{mech} = \alpha v \tag{3.60}$$

where α is the dynamic dispersion coefficient. The advection term v is defined by

$$v = -\frac{K}{n_e}\frac{dh}{dx} \tag{3.61}$$

where n_e is the effective porosity of the medium, K is the hydraulic conductivity, and dh/dx is the hydraulic gradient. Setting the advection to zero to determine the dispersion concentration profile, therefore

$$D\frac{\partial^2 c}{\partial x^2} = R\frac{\partial c}{\partial t} \tag{3.62}$$

it can be shown that through stating initial and boundary conditions the concentration profile as a function of time and space (in the x direction) is given by

$$c(x,t) = \frac{M}{\sqrt{4\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right) \tag{3.63}$$

which is useful for determining the concentration spread over time, and extraction of information regarding the width of the dispersion can be found. Comparing

(3.63) with (3.24) it is seen that a similar analysis as for the Gaussian plume model is undertaken. Taking the first integral of (3.63), results in

$$\int_{-\infty}^{\infty} c dx = M = \text{constant} \tag{3.64}$$

and gives the total amount of the substance M, as expected. Performing a second integral by inserting the coordinate x to inject the notion of distance,

$$\int_{-\infty}^{\infty} x c dx = 0 \tag{3.65}$$

results in zero due to symmetry, but gives information about the mean position, x, of the solution. Taking the third integral

$$\int_{-\infty}^{\infty} (x - \bar{x})^2 c dx \tag{3.66}$$

results in a non-zero, positive, function of time representing the squared distance to the mean position, and therefore the average distance to the center of the solution defining the width profile of the solution dispersion. The time evolution of this width should provide information regarding the rate of spreading. By defining a normalized quantity,

$$\sigma^2 = \frac{1}{M} \int_{-\infty}^{\infty} (x - \bar{x})^2 c dx \tag{3.67}$$

where M is defined by (3.64) and

$$\bar{x} = \frac{1}{M} \int_{-\infty}^{\infty} x c dx, \tag{3.68}$$

then the term σ has a dimension of length proportional to the width of the substance spread. Solving for

$$\sigma^2 = \frac{1}{M} \int_{-\infty}^{\infty} x^2 \frac{M}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right) dx \tag{3.69}$$

it is found that

$$\sigma = \sqrt{2Dt} \tag{3.70}$$

showing that the increase in width is proportional to the square root of diffusivity and time. Furthermore, by normalizing the distance x by σ and the concentration c by its maximum value c_{max}, where

$$c_{max} = \frac{M}{\sqrt{4\pi Dt}}, \tag{3.71}$$

(3.69) becomes

$$\frac{c}{c_{max}} = \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right] \tag{3.72}$$

which graphically represents the universal Gaussian curve. An important property of this curve is that 95 % of the concentration under the curve lies in the interval -1.96 < (x/σ) < + 1.96, which leaves only 2.5 % of the area under each tail. Therefore, by assuming 1.96 as approximately 2, it is concluded that the interval -2σ < x < + 2σ contains 95 % of the pollutant. Defining the width of the substance herewith as $(2 \times 2\sigma = 4\sigma)$, it follows that $4\sigma = 4\sqrt{2Dt} = 5.66\sqrt{Dt}$. Pollutant substances can be hazardous in very small concentrations, and thus the 2.5 % in each tail should not be discarded and precaution from a conservative approach recommends calculating the width of the polluted substance dispersion at 6σ or more (0.13 % in each tail).

The following section briefly describes advection transport of a solute as the bulk movement of porous groundwater, specifically between underground sources such as wells.

3.20 Advection Transport

Advection transport is the transport of a solute by the bulk movement of groundwater. The velocity of the bulk movement of groundwater is an average linear velocity, generally referred in context to advection velocity. In studying the motion of a solid particle through a fluid, it is found that the force of friction opposing the motion is proportional to the velocity of the particle. Similarly, in flow through a porous medium, it can be assumed that the frictional forces opposing the flow are proportional to the fluid velocity. The proportional relationship between the instantaneous discharge rate through a porous medium Q, the viscosity of the fluid μ, and the pressure drop $(h_b - h_a)$ over a certain distance Δl is described in equation form by Darcy's law, given as

$$Q = \frac{-\kappa A}{\mu}\frac{(h_b - h_a)}{\Delta l} \tag{3.73}$$

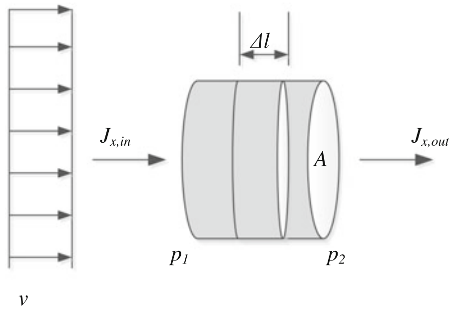


Fig. 3.13 Schematic representation for water pollution advection model

where κ is the hydraulic conductivity, and A is the cross-sectional area of flow. The graphical representation of Darcy's law in (3.73) for a cylindrical channel is given in Fig. 3.13.

As an example of applying Fig. 3.13 to water pollution advection, consider an unconfined aquifer with two wells, A and B, located 2 km from each other separated by ground. If the water takes 1.5 years to move from well A too well B and assuming a hydraulic conductivity of 100 m d^{-1}, and a drop in height of 15 m, the discharge rate if 3.5×10^5 m^3 water passes through the cross-section of the aquifer in 21 days is found by

$$Q = \frac{V}{t} = \frac{8.42 \times 10^5}{14} = 16.67 \times 10^4 \text{ m}^3.\text{d}^{-1} \tag{3.74}$$

The Darcy velocity of the water can be found by

$$v_D = \frac{\kappa \Delta H}{\Delta l} = \frac{(135)(33)}{3940} = 0.75 \;\; \text{m.d}^{-1} \tag{3.75}$$

and determining the actual velocity of the water is found by

$$v = \frac{\Delta l}{t} = \frac{(3940)}{(547.5)} = 3.65 \, \text{m.d}^{-1}. \tag{3.76}$$

In meteorology and physical oceanography, advection often refers to the horizontal transport of some property of the atmosphere or ocean, such as heat, humidity or salinity, and convection generally refers to vertical transport (vertical advection). Advection is important for the formation of orographic clouds (terrain-forced convection) and the precipitation of water from clouds, as part of the hydrological cycle.

3.21 Conclusion

This chapter discusses the types of indoor and outdoor air pollution and water pollution. It also provides relevant statistics on the types of pollution and models used to estimate its effects. A guideline to convert between concentrations based on molecular weight is given, used for air and water pollution. Focus is placed on outdoor/ambient air pollution, with equations provided for saturated and unsaturated adiabatic lapse rate, pollution dispersion, vertical dispersion, and the effects of momentum and buoyancy of pollution plume (smoke) from a stack source. Additionally, wind speed variations can also be accounted for and a brief outline is provided. Water pollution, water quality and treatment guidelines, and water pollution statistics are also provided in this chapter, where mass balance equations are provided to model advection and dispersion transport of water pollution through porous media. The information provided in this chapter can be applied to determine the effect of air and water pollution in large areas, and provides information on how to qualify and quantify its effects and how to treat and to prevent such outbreaks.

References

ANSI. (1982). A58.1, Minimum Design Loads for Buildings and Other Structures. New York: American National Standard Institute.

Balakrishnan, K., Mehta, S., Ghosh, S., Johnson, M., Brauer, M., Zhang, J., Naeher, L., Smith, K. R. (2013). WHO Indoor Air Quality Guidelines: Household Fuel Combustion – Review 5: Population Levels of Household Air Pollution and Exposures. Retrieved May 9, 2015 from http://www.who.int

Bartram, J., Ballance, R. (1996). Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes. Published on behalf of the United Nations Environment Program and the World Health Organization.

Chen Y. C., Bundy, D., Hoff, S. (1998). Modeling the Variation of Wind Speed with Height for Agricultural Source Pollution Control. *ASHRAE Transactions: Symposia*, 26(2), 1685–1691.

Cimorelli, A. J., Perry, S. G., Venkatram, A., Weil, J. C., Paine, R. J., Lee, R. F., Peters, W. D., Brode, R. W., Paumier, J. O. (2004). AERMOD: Description of Model Formulation. *United States Environmental Protection Agency*, 1–91.

Daly, A., Zannetti, P. (2007). Air Pollution Modeling – An Overview. *Chapter* 2 *of Ambient Air Pollution, The Arab School for Science and Technology (ASST)*, 1–28.

Emeis, S., Turk, M. (2007). Comparison of Logarithmic Wind Profiles and Power Law Wind Profiles and their Applicability for Offshore Wind Profiles. *Wind Energy: Proceedings of the Euromech Colloquium*, 61–64, ISBM 978-3-540-33865-9.

Engelbrecht, J. P., Swanepoel, L., Zunckel, M., Chow, J. C., Watson, J. G., Egami, R. T. (2000). Modelling PM10 Aerosol Data from the Qalabotjha Low-Smoke Fuels Macro-Scale Experiment in South Africa. *Ecological Modelling*, 127, 235–244.

Heldstab, J., de Haan van der Weg, P., Künzle, T., Keller, M., Zbinden, R. (2003). Environmental Documentation No. 169 – Air – Modelling of PM10 and PM2.5 Ambient Concentrations in Switzerland 2000 and 2010. *Published by the Swiss Agency for the Environment, Forests and Landscape SAEFL*, Berne, 1–118.

Hussain, S. A., Atshan, W. G., Najam, Z. M. (2005). Mathematical Model for the Concentration of Pollution and Dissolved Oxygen in the Diwaniya River (Iraq). *American Journal of Scientific Research*, 78, 33–37.

Irwin, J.H. (1979). A theoretical variation of the wind profile power-law exponent as a function of surface roughness and stability. *Atmospheric Environment*, 13, 191–194.

James, I. D. (2002). Modeling Pollution Dispersion, the Ecosystem and Water Quality in Coastal Waters: A Review. *Environmental Modelling & Software*, 17, 363–385.

Johnson, M., Edwards, R., Morawska, L., Smith, K. R., Nicas, M. (2011). WHO Indoor Air Quality Guidelines: Household Fuel Combustion – Review 3: Model for Linking Household Energy Use with Indoor Air Quality. Retrieved May 9, 2015 from http://www.who.int

Kiesewetter, G., Borken-Kleefeld, J., Schöpp, W., Heyes, C., Thunis, P., Bessagnet, B., Terrenoire, E., Fagerli, H., Nyiri, A., Amann, M. (2015). Modelling Street Level PM_{10} Concentrations Across Europe: Source Apportionment and Possible Futures. *Atmospheric Chemistry and Physics*, 15, 1539–1553.

Kurabuchi, T., Sosa, H., Lee, S. (2013). Feasability Study of CFD Prediction for Evaluation of Commercial Kitchen Environments. *13th Conference of International Building Performance Simulation Association*, Chambéry, 26–28 August 2013, 776-783.

Loucks. D. P., van Beek, E. (2005). Water Resources Systems Planning and Management: An Introduction to Methods, Models, and Applications. *Studies and Reports in Hydrology: UNESCO Publishing*, ISBN 92-3-103998-9.

Macdonald, R. (2003). Theory and Objectives of Air Dispersion Modelling. *Modelling Air Emissions for Compliance*, 1–27.

Masuric, G. (2013). A Study on the Mathematical Modeling of Water Quality in 'River-Type' Aquatic Systems. *WSEAS Transactions on Fluid Mechanics*, 2(8), 80–89.

Simiu, E., Scanlan, R. H. (1978). Wind Effects on Structures. New York: John Wiley & Sons, Inc.

Strom, G. H. (1976). Transport and Diffusion of Stack Effluents. In: Air pollution. Edited by A. C. Stem. New York: Academic Press.

Sutton, O. G. (1953). Micrometeorology: A Study of Physical Processes in the Lowest Layers of the Earth's Atmosphere. New York: McGraw-Hill.

Touma, J.S. 1977. Dependence of the wind profile power law on stability for various locations. Journal of the Air Pollution Control Association, 27 (9): 863–866.

Turner, D. B. (1994). Workbook of Atmospheric Dispersion Estimates. Ann Arbor: Lewis Publishers.

Wark, K., Warner, C. F. (1976). Air Pollution, its Origin and Control. New York: Thomas Y. Crowell Company, Inc.

Whitehead, P. G., Lack, T. (1982). Dispersion and Self-Purification of Pollutants in Surface Water Systems: A Contribution to the International Hydrological Programme. *UNESCO Publishing*, ISBN 92-3-102049-8.

WHO. (2005a). Indoor Air Quality Guidelines: Household Fuel Combustion. Retrieved May 18, 2015 from http://www.who.int

WHO. (2005b). Burden of Disease from Ambient Air Pollution for 2012. Retrieved May 18, 2015 from http://www.who.int

WHO. (2014). WHO's Ambient Air Pollution Database – Update 2014. Retrieved May 18, 2015 from http://www.who.int

Chapter 4
Harvesting Energy from Ambient Sources: Wind Energy, Hydropower, Radiation and Mechanical Deformation

4.1 Introduction

To generate energy, fossil fuels are predominantly used, formed by natural processes such as anaerobic decomposition of organisms containing large amounts of carbon. These energy-generating compositions include coal, petroleum and natural gas, or derivatives such as kerosene and propane. The demand for energy is always increasing, especially with the rise in urbanization and the industrial revolution. Although fossil fuels are the major source of energy, overconsumption still leads to air pollution and health risks to humans and other mammals. Since fossil fuels are derived from prehistoric fossils, they are non-renewable; once used these sources cannot be re-introduced. Fossil fuels have affected industry and evolvement immensely, which cannot be overlooked, but they have several disadvantages that should not be ignored.

One major disadvantage is the formation of pollution, mainly air pollution, from the burning of fossil fuels. Global warming, a controversial term but a factual and measurable quantity, is a result of the formation of greenhouse gases, as carbon monoxide is generated in the burning process. Global warming refers to a global rise in temperature that leads to the melting of Arctic polar ice caps, leading to flooding of low-lying areas from a rise in sea levels. In addition, this effect can eliminate entire ecosystems and lead to the extinction of species such as coral reefs, which are highly sensitive to small changes in temperature. According to the National Geographic Society, average global temperatures have risen by 0.8 °C since 1880; most of this increase has occurred in recent decades, according to NASA's Goddard Institute for Space Studies. According to these climate studies, the rate of temperature increase is also rising and the last two decades have been the hottest in the last 400 years, possibly the last several millennia. According to a study by the Intergovernmental Panel on Climate Change (IPCC), conducted by 2500 scientists in more than 130 countries, humans have caused a large percentage of global warming, or more accurately, the rate at which these changes take place.

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_4

There is also the possibility that natural cycles in the earth's orbit alter the planet's exposure to sunlight, which could result in the current trends, for example complicated solar activity or just a natural cycle spanning thousands of years. Whichever it may be, it cannot simply be ignored that humans are damaging the earth by burning fossil fuels and alternative approaches can only benefit the planet and its inhabitants. A more tangible effect of air pollution is acid rain when sulfur dioxide gas mixes with rainwater. Acid rain degrades brick and marble and causes damage to monuments, other buildings and farmland crops.

To produce large amounts of constant coal power, power plants require vast amounts of regular, uninterrupted coal supplies. Coal is delivered to the plants by trains or trucks that also burn fossil fuels and pollute the air. The amounts of reserves required at these plants are constantly growing as the demand for electricity increases and some nations are unable to cope with this increase in demand. The cost-effectiveness of supplying electricity with coal can in extreme cases become negative. The mining of coal is yet another activity that causes physical harm to humans, destroys large pieces of natural land and creates ecological imbalances.

Middle-Eastern countries have the largest supply of oil and natural gas. Other countries are dependent on commodities to supply them with constant and sufficient fuel. Over 40 % of the world's oil production is dependent on the Organization of Petroleum Exporting Countries (OPEC), a group of thirteen countries including Iran, Iraq, Kuwait, Qatar, Saudi Arabia and the United Arab Emirates. OPEC constantly monitors the volumes of oil consumed and adjusts its own production to maintain its desired barrel price. This results in large fluctuations of the oil price worldwide, as many external factors are at play that influence the price. As a result, the increase in the price of oil and natural gas is having a negative impact on some economies that cannot afford these increases. If an alternative could be implemented, each country could generate its own electricity only dependent on local environmental conditions, making countries less reliant on worldwide events that do not necessarily concern their economy or political bias.

Technology is progressing, albeit slower than ideally envisioned, to harness large-scale ambient energy such as solar power, wind turbines, tidal generators and even compost. These sources are renewable, and if implemented efficiently, could in practical terms sustain energy requirements indefinitely. Research into evolving these technologies to be as efficient as fossil fuels is being conducted worldwide and major strides have been made in recent years towards implementation.

There is, however, another source of energy, one that might not be as obvious as fossil fuels or renewable large-scale energies. This source focuses more on small quantities of energy that is lost in daily activities such as walking, running, cycling, or any movement or heat generated during normal human activity. Energy harvesting is a process of capturing this lost energy, although only in very small quantities that can be used to power small devices such as sensors. Wireless sensor networks can, for example, be powered by small quantities of energy, scavenged using energy-harvesting techniques. The devices are effectively maintenance-free, requiring no replacement of power sources as do battery-powered devices, they are environmentally friendly since they do not contain disposable chemicals and metals, and could be used in locations where batteries might not be practical, such as in underwater or solar applications. The most promising energy-harvesting technologies are capturing vibration, movement and sound with piezoelectric materials, capturing heat from lost energy during fossil fuel burning, using thermoelectric and pyroelectric materials, stray electromagnetic energy from transmitted radio waves, and electrostatic/capacitive energy harvesting (Vullers et al. 2009). Capturing heat from the sun's radiation can also be considered a form of energy harvesting, since most of this energy is lost daily across the surface of the earth.

All these ambient sources can potentially replace, or at least contribute to, energy production using fossil fuels. A summary of the available ambient sources is given in the following sections. To fully describe the process of converting radiation and mechanical deformation to electrical energy, photodiodes and piezoelectric materials are discussed in this chapter as well.

4.2 Ambient Sources

Ambient energy is resident energy that is often agile and transient in nature, generally considered of no ordinary value or usefulness. The methodology that is most effective in harnessing the benefits of ambient energy is the leveraging of the properties of matter and the physics of nature in such a manner as to effectively do work without artificial power or mechanical means.

Ambient energy sources, including wind, geothermal heat, flowing water and radiation from the sun, have been used successfully for power production since ancient times. The Great Pyramid in Giza is one structure supposedly built to honor the Lord as described in the Book of Isaiah in the Christian Bible. An empty stone chest was found in the center of the pyramid, which archeologists called the King's Chamber, believing that the place was a tomb like the others. However, the chamber was discovered to be empty when the pyramid was explored, with no sarcophagus inside the stone chest. The stone chest itself is different from other sarcophagus containers in the other tomb emplacements, as is its location in the pyramid. The Great Pyramid with its distinct King's Chamber may have been designed for another function. The Great Pyramid has an energy-conductive structure because its walls were built with quartz-infused stones; if a super-conductive device or highly advanced battery should be installed there, the pyramid could self-generate enough wireless electricity for all of Egypt. Some claim that the radioactive nature of the Ark of the Covenant as described in the Bible reveal or hint that it was the same device set in the King's Chamber, given its specific dimensions matching the stone chest in the "reactor room". Although these data have not been proven and sound more like fiction, the idea of generating power from ambient sources even in those times is intriguing.

In most locations on earth, there is abundant ambient electromagnetic energy (EM) from cellular, television and radio transmitters, as well as satellite and other wireless communication systems. Alternative sources include physical motion, visible light, heat from the sun and the earth's interior, wind, ocean waves, ocean tidal currents, river currents and sound waves. All are established means of producing electricity in fixed-site facilities. In contrast, relatively little use has been made of these resources in portable power devices. Other forms of energy, such as broadcast radio and television, microwave communications and EM fields of electric power lines, add to the list of potentially useful energy sources. In addition, human presence or activities in the vicinity of applications represent another underutilized potential for generating electric power. Human activities inevitably produce waste energy that can go to productive use if harnessed and stored. For example, when people walk down the street, they move their arms and legs, generating kinetic energy to propel themselves, and also thermal energy through friction with clothes, the surrounding air and the pavement beneath their feet. An ambient energy-scavenging device can convert most or all of this wasted energy into electrical energy that can be used for example to charge a cell phone battery. The US Military has been testing prototype "Kinetic Boots" which can generate up to 1.5 W of usable power per boot per step. These energy harvesting boots are still in its prototype phase (as of 2015) and have received some criticism in terms of its actual practicality and claims that the shoes themselves cause fatigue as they are heavier compared to normal military boots. Table 4.1 presents several means of energy production, storage, and conservation that can be found in nature and utilized; the efficiencies of these technologies might be low and the techniques not perfected, but the potential still exists.

Table 4.1 Examples of ambient energy production, storage and conservation from natural sources

Atomic energy	Battery cell	Bio-alcohol	Biodiesel
Biomass	Bio-nano generator	Bitumen	Coal
Cold fusion	Combustion	Compound turbine	Compressed air
Solar power	Deep lake water cooling	Diesel	Electrical grid
Energy tower	External combustion	Fischer-Tropsch	Fossil fuel
Francis turbine	Fuel cell	Fusion power	Gas turbine
Geothermal pump	Wind power	Hydroelectricity	Hydrogen
Implosion	Kaplan turbine	Light crude	Liquid nitrogen
RF power	Magneto hydrodynamic	Methane clathrate	Methanol
Natural gas	Oil well	Osmotic power	Ocean thermal
Oxidation	Peat	Petroleum	Photovoltaics
Piezoelectricity	Pneumatics	Propellant	Pyrolysis
Solar satellite	Team turbine	Stirling engine	Vegetable oil
Tesla turbine	Tidal power	Uranium	Vacuum energy
Water turbine	Wind energy	Wood fuel	Wood gas

Since IoT devices generally require little energy to operate, the means to generate this energy does exist in several forms. Lampi suggests that an IoT node requires approximately 124 μW to perform basic tasks. Considering this value, some alternatives to powering these devices are solar cells, magnetic coupling with power lines, harvesting radio signals, magnetic capture of vibration, piezo capture of vibration and strain, thermoelectric capture through the Peltier effect or air movement. The following sections are dedicated to exploring energy harvesting through wind power, hydro-electricity, detection of radiation through photo-detectors and piezoelectric materials converting mechanical energy to electrical energy, with a brief introduction to MEMS devices used in, for example, piezoelectric devices.

4.3 Wind Energy

Wind is caused by uneven heating of the atmosphere by the sun, the roughness of the earth's surface, and the rotation of the earth. The flow pattern of wind is changed by the topography of the land, water, and plants on the surface of the earth. To harness its energy, wind turbines use the flow of air to generate electrical energy through conversion of the available kinetic energy. Wind energy is one of the cleanest and most reliable ways to generate electrical energy, as it does not produce toxic emissions that contribute to global warming. Wind energy is also widely available, but depending on the geometric location on earth, wind speeds vary significantly and this randomness is a drawback of wind energy. It some parts of the world wind is likely to be an everyday occurrence, but in some parts wind rarely reaches speeds capable of generating significant energy. A significant advantage of wind power is that the power output of turbines increases with the cube of wind speed, meaning that if wind speed doubles, power output increases theoretically eight times.

Most wind turbines have rotating blades that turn in the presence of wind. The kinetic energy of the rotating blades is transferred to a shaft that provides power to an electric generator that produces electricity. The generator produces electricity using a magnetic rotor inside the shaft that spins inside loops of copper wire wound around an iron core. This rotor creates electromagnetic induction, which creates electrical energy. The current is regulated and after the conversion process it can be used or stored in battery banks.

Wind energy is the kinetic energy of moving air. The kinetic energy E of a packet of air with mass m and with velocity v is given by

$$E = \frac{1}{2}mv^2. \tag{4.1}$$

To find the mass of the packet of air passing through an area A perpendicular to its velocity such as through a rotor, the volume after a certain time t has passed is multiplied by the air density ρ and the wind energy is found by

$$E = \frac{1}{2}\rho A v^3 t. \tag{4.2}$$

By differentiating (4.2) with respect to time, the rate of increase of energy is determined, leading to the total wind power given by

$$P = \frac{dE}{dt} = \frac{1}{2}\rho A v^3 \tag{4.3}$$

where P is the total wind power in watts. From (4.3) the cubed relationship between wind power and air velocity is shown in the v^3 term. Wind speed behind a turbine is lower than the ambient air speed in front of the turbine. Since the mass flow must be continuous, the area behind the wind turbine is therefore larger than the area in front of the turbine, as shown in Fig. 4.1.

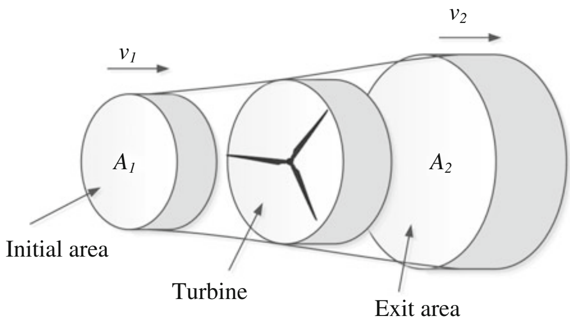
Taking into consideration the relationship between A_1 and A_2 as shown in Fig. 4.1, the equations below are derived. The effective power in a turbine system is the difference between the wind power in front of and behind the turbine, such that

$$P_{eff} = P_{A1} - P_{A2} \tag{4.4}$$

where P_{A1} and P_{A2} are the power of the wind in front of and behind the turbine respectively. Using (4.3), (4.4) can be rewritten in terms of the variation in volume, the air density, time, and wind speed, such that

$$P_{eff} = \frac{\Delta V \rho}{2t}\left(v_1^2 - v_2^2\right) \tag{4.5}$$

Fig. 4.1 Wind speed velocity and area relationship in front of and behind a wind turbine



which can be further simplified to

$$P_{eff} = \frac{\rho A}{4}(v_1 + v_2)\left(v_1^2 - v_2^2\right). \tag{4.6}$$

From (4.6) it can be seen that if the difference between the speed in front of and behind the turbine were to be zero, there would be no effective power. If this difference is too large, the air flow through the rotor is slowed down too much. An ideal power coefficient, $c_{p,}$ can be obtained, where

$$c_p = \frac{P_{eff}}{P_{wind}} \tag{4.7}$$

and P_{ref} is the reference wind power. The power coefficient is determined by Betz's law, where

$$c_p = \frac{P_{eff}}{P_{ref}} = \frac{(v_1 + v_2)\left(v_1^2 - v_2^2\right)}{2v_1^3} \tag{4.8}$$

is used to derive the power coefficient. If a constant density fluid and the conservation of mass is implied, then

$$A_1 v_1 = A_2 v_2 = \frac{A(v_1 + v_2)}{2} \tag{4.9}$$

and if the ratio v_2/v_1 is replaced by the variable x; (4.8) can be written as

$$c_p = \frac{(1 + x)(1 - x^2)}{2}. \tag{4.10}$$

To find the maximum value of v_2/v_1 and hence x, (4.10) is differentiated and set to zero. Solving for x results in a value of $x = 1/3$ and substituting the relationship $v_2 = v_1/3$ into (4.8), the ideal power coefficient is found to be

$$c_p = \frac{P_{eff}}{P_{ref}} = \frac{16}{27} \approx 0.59. \tag{4.11}$$

Modern large wind turbines achieve peak values for c_p in the range of 0.45–0.50, about 75–85 % of the theoretical maximum with a larger drive to 98 % efficiency in recent developments. In high wind speed where the turbine is operating at its rated power, the turbine rotates (pitches) its blades to lower c_p to protect itself from damage. In basic terms, 59 % efficiency is the theoretical maximum a conventional wind turbine can achieve in extracting power from moving air. Losses in the power generator equipment, inter-connections to the grid and energy transfer to the end-user decrease the maximum useable power.

Global wind power generation capacity has been increasing steadily in recent years and many African countries have been following the trends towards sustainable and renewable wind energy. Table 4.2 highlights countries worldwide and the wind power capacity of these countries in MW, for data generated in 2014 as reported by the Global Wind Energy Council (GWEC).

According to Table 4.2, the world total installed wind power capacity is 369 597 MW, to which Africa only contributes 2 535 MW (0.69 %), pointing to much room for expansion, especially since Africa is the second largest continent on earth after Asia, which contributes 141 964 (38 %) of the total capacity. The results are displayed in Fig. 4.2, depicting each region's overall contribution.

Figure 4.2 is given as a waterfall distribution; therefore, each region's contribution adds up to the total global capacity of wind power. It can be seen from Fig. 4.2 that Africa and the Middle East, Latin America and the Caribbean, and the Pacific Region (although the smallest in land area) make the lowest contributions worldwide. Developments in Africa could increase its contribution significantly, since it has the natural resources (land area and wind distribution) available. Average wind speeds in Africa are given in Fig. 4.3; the data were made available by the International Renewable Energy Agency (IRENA) Global Atlas.

From Fig. 4.3, it can be seen that the wind speeds in Africa, measured at a height of 80 m, reach 4.8–7.2 m s^{-1} across a large percentage of the continent. Lower wind speeds below 3.6 m s^{-1} are recorded in Central Africa, around Equatorial Guinea and the Democratic Republic of Congo. According to the African Development Bank Group (2013), based on the individual capacity of wind projects, Africa's installed wind energy capacity increased 12-fold between 1995 and 2010, with most of the growth taking place from early 2000. Notably, between 2000 and 2010, Africa's installed wind capacity grew at a rate of 41 %, much faster than the average global growth rate of 27 %. Africa's great performance is a reflection of the embryonic nature of the African market, which was characterized by limited initial capacity. Africa's growth rate between 2000 and 2010 is very similar to the one reported for other regions during the early stages of their market development (35 % over the period 1995–2000). It is worth noting that the bulk of Africa's completed projects are located either in coastal areas or on islands. The exception is Kenya, which has its largest potential inland around the Lake Turkana highlands, while having significant coastal potential as well. Both its pilot facility and planned projects are located inland. The wind farms in Algeria, Nigeria and Tanzania are also located inland. In order of installed capacity, NREA (Egypt) is the largest owner/operator of wind energy capacity on the continent then followed by other North African public utilities, ONE in Morocco and STEG in Tunisia. The profile of private players is heterogeneous in nature and includes both established international operators and recently established African-based firms, some of which are joint ventures between African and international firms. Table 4.3 summarizes the key operators and developers of wind energy projects in Africa, including the top ten prospective entrants. French, Spanish and Dutch investments have already been

Table 4.2 Global Wind Energy Council global wind statistics for 2014

Region	Country	Installed wind power [MW]	Country	Installed wind power [MW]
Africa and Middle-East	Morocco	787	Egypt	610
	South Africa	570	Tunisia	245
	Cape Verde	24	Ethiopia	171
	Other	129	**Total**	**2 535**
Asia	People's Republic of China	114 609	Thailand	223
	India	22 465	Pakistan	256
	Japan	2 789	Philippines	216
	Taiwan	633	Other	167
	South Korea	609	**Total**	**141 964**
Europe	Germany	39 165	Turkey	3 763
	Spain	22 987	Romania	2 954
	United Kingdom	12 440	Netherlands	2 805
	France	9 285	Ireland	2 272
	Italy	8 663	Austria	2 095
	Sweden	5 425	Greece	1 980
	Portugal	4 914	Rest of Europe	6 543
	Denmark	4 883	**Total**	**134 007**
	Poland	3 834	–	–
Latin America and Caribbean	Brazil	5 939	Honduras	152
	Chile	836	Peru	148
	Uruguay	464	Caribbean	250
	Argentina	271	Other	83
	Costa Rica	198	**Total**	**6 526**
	Nicaragua	186	–	–
North America	United States of America	65 879	–	–
	Canada	9 694	–	–
	Mexico	2 551	–	–
	Total	**78 124**	–	–
Pacific Region	Australia	3 806	–	–
	New Zealand	623	–	–
	Pacific Islands	12	–	–
	Total	**4 441**	–	–
World total		369 597		

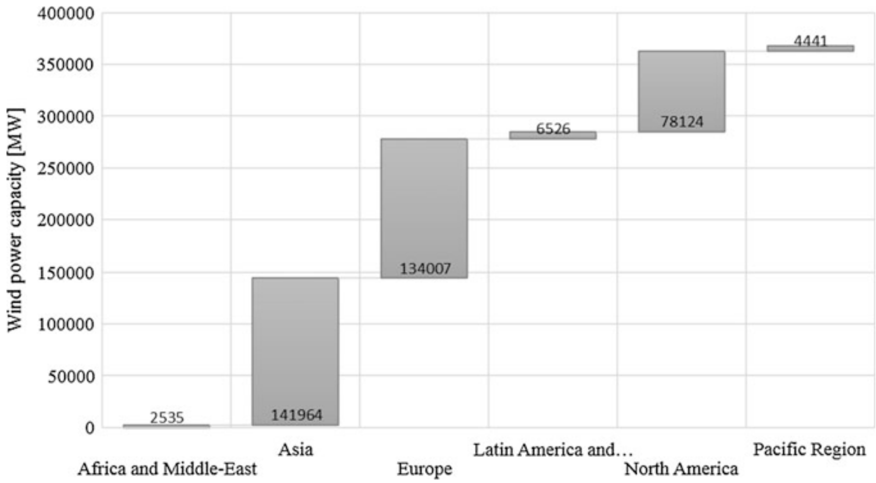


Fig. 4.2 Regional contribution to global wind power capacity in MW

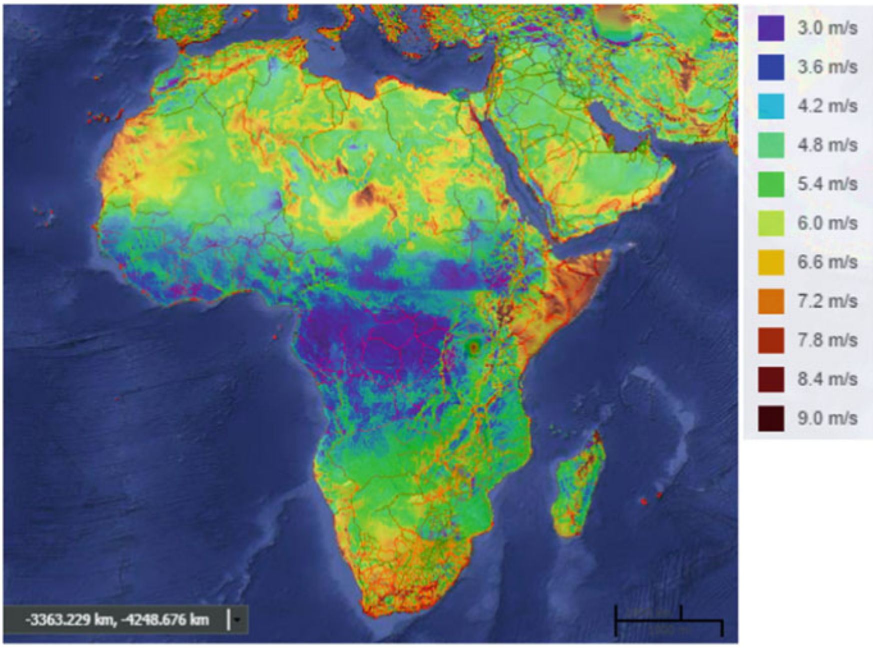


Fig. 4.3 Africa 3TIER's Global Wind Dataset 5 km onshore wind speed at 80 m height as generated by IRENA Global Atlas

Table 4.3 Key operators in the wind energy market in Africa (African Development Bank Group 2013)

Investor	Ownership	Domiciliation	Country	Capacity (MW)
NREA	Public	Egypt	Egypt	550.2
ONE	Public	Morocco	Morocco	304.3
STEG	Public	Tunisia	Tunisia	53.6
Chaâbi Group	Private	Morocco	Morocco	70
Compagnie Eolienne du Détroit (CED)	Private	France	Morocco	50
La cimenterie de Tetouan	Private	Morocco	Morocco	32.2
Cabeolica	PPP	Cape Verde	Cape Verde	28
Nareva Holding	Private	Morocco	Morocco	70

made on the continent. With the exception of prospective IPP investments by the world's third largest wind energy developer, Iberdrola, foreign direct investments in Africa's wind energy market is driven by relatively small and emerging players.

With regard to African firms, South Africa will soon have the fastest growing wind energy industry, as indicated by the number of prospective private operators in the market. Over 500 MW is expected to be installed by South African private operators, though all operations will be concentrated in South Africa (Mukasa et al. 2013). Emerging south-south cooperation is also noteworthy: some well-established African firms are already seeking investment opportunities elsewhere on the continent, for example Egypt's El Sewedy, which has plans to invest in the Ghanaian market (African Development Bank Group 2013).

Another market that offers substantial gains in the energy market is hydropower, generating electricity from flowing water.

4.4 Hydropower

Hydropower is electricity generated using the energy of moving water (Castaldi et al. 2003). Rain or melted snow, usually originating in hills and mountains, create streams and rivers that run towards the oceans. The energy of this moving water can be substantial and can be converted to a useable resource. This energy has been exploited for centuries. Farmers since the ancient Greeks have used water wheels to grind wheat into flour. In the late 19th century, hydropower became a source for generating electricity. The first hydroelectric power plant was built at Niagara Falls in 1879. In 1881, street lamps in the city of Niagara Falls were powered by hydropower. In 1882 the world's first hydroelectric power plant began operating in the United States of America, in Appleton, Wisconsin.

A typical hydro plant is a system with three parts: an electric plant where the electricity is produced, a dam that can be opened or closed to control water flow, and a reservoir where water can be stored. The water behind the dam flows through an intake and pushes against blades in a turbine, causing them to turn. The turbine spins a generator to produce electricity. The amount of electricity that can be generated depends on how far the water drops and how much water moves through the system.

Hydroelectric power provides almost one-fifth of the world's electricity. China, Canada, Brazil, the United States and Russia were the five largest producers of hydropower in 2014. The world's largest hydro plant is at Three Gorges on China's Yangtze River with an instantaneous generating capacity of 22,500 MW). The dam is 2.3 km wide and 185 meters high and took 17 years to finish the construction. The reservoir for this facility started filling in 2003 and reached full capacity by 2009. By August 2011 the plant had generated 500 TWh of electricity and energy production during 2014 was just below 100 TWh.

There are four broad hydropower typologies: Run-of-river hydropower entails a facility that channels flowing water from a river through a canal or penstock to spin a turbine. Typically, a run-of-river project will have little or no storage facility. Run-of-river generation provides a continuous supply of electricity (base load), with some flexibility of operation for daily fluctuations in demand through water flow that is regulated by the facility.

Storage hydropower typically involves a large system that uses a dam to store water in a reservoir. Electricity is produced by releasing water from the reservoir through a turbine, which activates a generator. Storage hydropower provides base load as well as the ability to be shut down and started up at short notice according to the demands of the system (peak load). It can offer enough storage capacity to operate independently of the hydrological inflow for many weeks or even months.

Pumped-storage hydropower provides peak-load supply, harnessing water that is cycled between a lower and upper reservoir by pumps that use surplus energy from the system at times of low demand. When electricity demand is high, water is released to the lower reservoir through turbines to produce electricity.

Offshore hydropower is a less established but growing group of technologies that use tidal currents or the power of waves to generate electricity from seawater.

These technologies often overlap. For example, storage projects often involve an element of pumping to supplement the water that flows into the reservoir naturally, and run-of-river projects may provide some storage capability.

Damming rivers may destroy or disrupt wildlife and other natural resources. Some fish, such as salmon, may be prevented from swimming upstream to spawn. Technologies such as fish ladders help salmon go up over dams and enter upstream spawning areas, but the presence of hydroelectric dams changes their migration patterns and hurts fish populations. Hydropower plants can also cause low dissolved oxygen levels in the water, which are harmful to river habitats.

Facilities range in size from large power plants that supply many consumers with electricity to small and micro plants that individuals operate for their own energy needs or to sell power to utilities. Although definitions vary, the United States Department of Energy (DOE) defines large hydropower as facilities that have a capacity of more than 30 mW. DOE defines small hydropower as facilities that have a capacity of 100 kW–30 mW and micro hydropower plants have a capacity of up to 100 kW. A small or micro-hydroelectric power system can produce enough electricity for a home, farm, ranch or village. The power in watts available from falling water as used to drive hydropower plants (P_h) can be calculated from the flow rate and density of water, the height of fall and the local acceleration due to gravity and is given by the following equation:

$$P_h = \eta Q \rho g \Delta h \tag{4.12}$$

where η is the efficiency of the turbine, Q is the water flow in $m^3.s^{-1}$, ρ is the density of the water in kg m^{-3}, g is the gravitational acceleration, and Δh is the difference in height between the inlet and outlet in meter (called the head). The efficiency of the turbine is a function of its design and technology. Design improvements have made large strides towards improving the efficiency of power conversion to 95 %. Two main categories of turbines are generally used, impulse turbines and reaction turbines. Choosing a turbine type depends on the head, water flow, water volumes, required efficiency and cost. The impulse turbine generally uses the velocity of the water to move the runner and discharges to atmospheric pressure. The water stream hits each bucket on the runner. There is no suction on the down side of the turbine and the water flows out the bottom of the turbine housing after hitting the runner. An impulse turbine is generally suitable for high-head, low-flow applications. Examples of impulse turbines include Pelton wheels (with Turgo wheels a variation on the Pelton wheel) and cross-flow turbines. A reaction turbine develops power from the combined action of pressure and moving water. The runner is placed directly in the water stream flowing over the blades rather than striking each individually. Reaction turbines (Liangliang et al. 2012) are generally used for sites with lower heads and higher flows compared with impulse turbines. Reaction turbines include propeller turbines (bulb, straflo, tube and Kaplan), as well as Francis turbines and kinetic turbines. All these turbines operate on the fundamental Euler pump and turbine equations, derived from the conservation of angular momentum, where Euler's pump equation (ΔW_c) is given by

$$\Delta W_c = U_2 c_{\theta 2} - U_1 c_{\theta 1} > 0 \tag{4.13}$$

and Euler's turbine equation (ΔW_t) by

$$\Delta W_t = U_1 c_{\theta 1} - U_2 c_{\theta 2} > 0 \tag{4.14}$$

where c is the tangential component perpendicular to the flow area, and U is the rotational speed of the blade, given by

$$U = \Omega r \tag{4.15}$$

where Ω is the angular velocity and r is the radius of the blade.

According to the African Development Bank Group (2013), the choice of wind turbine manufacturers for Africa is dominated by global leaders in the industry (Table 4.4). Data show that the world leaders, Gamesa and Vestas (fourth and

Table 4.4 Wind turbine manufacturers distributing in Africa (African Development Bank Group 2013)

Manufacturer	Domiciliation	Country	Installed capacity (MW)	Pipeline capacity (MW)
Gamesa	Spain	Egypt, Morocco, Tunisia	630.8	590
Vestas	Denmark	Egypt, Kenya, Morocco, South Africa, Cape Verde, Algeria	136.3	55
Alstom Ecotècnia	Spain	Morocco	100	–
Nordex	Germany	Egypt	63	–
Fuhrländer	Germany	South Africa	5.5	–
Wind World	Denmark	Namibia	0.2	

largest wind turbine manufacturers respectively on the world market), dominate in terms of installed capacity and geographic reach.

The manufacturing landscape will change as more global players increase their presence in Africa. In South Africa such developments have been encouraged through the Renewable Energy Independent Producers Procurement Programme (REIPPP) that is targeting installation of 1 850 MW of on-shore wind power with at least 35 % local content requirements. Both South Africa and Egypt owe their success in the manufacturing industry to the existence of adequate regulatory and policy frameworks, well-established research and development institutions and the relatively low cost of doing business in the two countries. However, more needs to be done to boost the industry. Policy makers not only need to focus on ways in which international technology providers could be attracted to the continent, but also on taking advantage of regional and intra-continental economies of scale and encourage skills and technology transfer through partnerships between local firms and established global manufacturers (IRENA 2013).

A photodiode is a semiconductor device that converts light into current. On a small scale (compared to wind and hydro-power), this is also regarded as a form of harvesting energy from renewable ambient resources (radiation).

4.5 Photodiodes

Photodiodes operate by absorption of photons or charged particles and generate a flow of current in an external circuit that is proportional to the incident power on the device. Photodiodes can be used to detect the presence of small quantities of ambient light and their accuracy can be calibrated from intensities below 1 pW cm^{-2} to above 100 mW cm^{-2}. These devices have a diverse range of applications, including spectroscopy, photography, analytical instrumentation, optical positioning sensors, beam alignment, surface characterization, laser range

Table 4.5 Constant values, descriptions and units used in this chapter

Symbol	Description	Value	Units
k	Boltzmann's constant	1.38065×10^{-23}	Joule/Kelvin (J K^{-1})
c	Velocity of light in free space	2.99792×10^{8}	m/s
h	Planck's constant	6.62606×10^{-34}	Joule-second (J-s)
Z_0	Impedance of free space	376.73	Ω
ε_0	Permittivity of free space	8.85418×10^{-12}	Farad/metre (F m^{-1})
m_e	Electron mass	9.10938×10^{-31}	kilogram (kg)
q	Electron charge	1.60217×10^{-19}	Coulomb (C)
μ_0	Permeability of free space	1.25663×10^{-6}	Henry/metre (H m^{-1})
E	1 Electronvolt (eV)	$1.60217646 \times 10^{-19}$	Joules (J)

finders, optical communications and medical imaging instruments. In this book, the terms photodiode and photo-detector are used interchangeably.

For convenience, Table 4.5 lists some of the constant parameters used in this discussion and can be used as reference throughout the chapter.

Planar photodiodes are in essence p-n junction diodes, formed by either diffusing a p-type impurity (such as zinc) into an n-type substrate, or an n-type impurity (such as phosphorus) into a p-type substrate. The diffused area defines the photodiode active area, where light photons are captured and the change in electric field is converted to electrical energy. To form an ohmic contact, additional impurity diffusion into the bottom of the substrate is performed. The impurity is an n-type for a p-type active area and vice versa. The contact pads, usually gold/titanium, are deposited on the front active area and on the entire backside. The active area is then passivated with an antireflection coating to reduce the reflection of the light for a specific wavelength. The thickness and material (generally SiO_2 or SiN_x) determine the optical properties of the coating. The non-active area on top is covered with a layer of metal to avoid stray light entering the active area. By controlling the thickness of the bulk substrate, the speed and responsivity of the photodiode can be controlled.

The material bandgap is the gap between the valence band and the conduction band. At absolute zero temperature the valence band is completely filled and the conduction band is vacant. As the temperature increases, the electrons become excited and escalate from the valence band to the conduction band by thermal energy. The electrons can also be escalated to the conduction band by particles or photons with energies greater than the bandgap. The resulting electrons in the conduction band are free to conduct current. The relationship between the band gap energy (E_g) and the photodiode cut-off wavelength (λ_{co}) plays an important role in choosing the correct material and/or process for the required application. The bandgap energy is heavily dependent on the material and the doping profile of the material, and is a compound structure (such as a Group III-V binary compound semiconductor). The equation that relates the bandgap and the cut-off wavelength is given by

$$\lambda_{co} = \frac{hc}{E_g(eV)} \tag{4.16}$$

where h is Planck's constant in J s^{-1}, and c is the speed of light in free space in m s^{-1}. Therefore, at a specified wavelength, the energy bandgap for a material is a constant value dependent on the speed of light. Table 4.6 gives common materials (with respect to photo-detectors) and their respective bandgaps at 300 K. The cut-off frequency is also specified in the table, with a reference to the band of operation (infrared (IR) or visible).

Table 4.6 gives an indication of which materials are used to detect light photons at various wavelengths. Shorter wavelengths (<900 nm) can be detected by InP or GaAs, wavelengths within the near IR range are detected by InGaAs, Ge, or Si, and medium IR signals (>3000 nm) can be detected by InSb or InAs.

Resulting from the gradient in concentration, the diffusion of electrons from the n-type region to the p-type region and the diffusion of holes from the p-type region to the n-type region develop a built-in voltage across the junction. This voltage is determined by

$$\varphi_i = \frac{kT}{q} \ln \frac{N_a N_d}{n_i^2} \tag{4.17}$$

where k is Boltzmann's constant in J K^{-1}, T is temperature in Kelvin, q is the electron charge in Coulomb, N_a and N_d are the electron and hole concentrations in cm^{-3} respectively, and n_i is the intrinsic impedance of the material in Ω. At 300 K, the value for kT/q is 25 mV. The inter-diffusion of electrons and holes between the n- and p-regions across the junction results in a region with no free carriers. This is the depletion region. The built-in voltage across the depletion region results in an

Table 4.6 Energy bandgap and cut-off wavelengths of commonly used materials in photodetectors

Material	Symbol	E_g (eV)	λ_{co} (nm)	Band
Indium gallium arsenide	$In_xGa_{1-x}As$	0.73 to 0.47	1 700 to 2 600	Near IR
Indium gallium arsenide	$In_{0.53}Ga_{0.47}As$	0.75	1 655	Near IR
Indium phosphide	InP	1.35	919	Visible
Germanium	Ge	0.66	1 800	Near IR
Gallium arsenide	GaAs	1.42	875	Visible
Indium arsenide	InAs	0.36	3 400	Near IR
Indium antimonide	InSb	0.17	5 700	Medium IR
Silicon	Si	1.12	1 130	Visible

electric field with maximum at the junction and no field outside the depletion region. The width of the depletion x_d region can be determined by

$$x_d = \sqrt{\frac{2\varepsilon_s(N_a + N_d)}{q(N_a N_d)}(\varphi_i - V_a)} \tag{4.18}$$

where ε_s is the permittivity of the material, and V_a is the applied bias voltage to the device. Any applied reverse bias adds to the built-in voltage and results in a wider depletion region. The electron-hole pairs generated by light are swept away by drift in the depletion region and are collected by diffusion from the undepleted region. The current generated is proportional to the incident light or radiation power. The light is absorbed exponentially with distance and is proportional to the absorption coefficient.

The electrical characteristics of a photodiode can be described by representing the diode as a noise equivalent circuit, as depicted in Fig. 4.4.

In Fig. 4.4, the current generated as a result of incident light photons on the active area of the device is given by I_{light}. The diode (device) in its ideal form is represented by D_1, where the current flowing through the device under normal operation is I_{diode}. The capacitance as a result of the open area (window) on the top layer (window layer) towards the absorption layer is given by C_j. The resistance R_{SH} is the internal shunt resistance, where voltage (heat) is generated by the leakage/dark current I_{dark}. The series resistance R_S appears between the device and the load represented by the output current I_{out} and the load resistance R_{load} respectively. The output voltage v_{out} is measured at the designated ports (*Out* + and -). The output voltage must be processed by a high-precision amplifier.

The shunt resistance R_{SH} is the slope of the current-voltage (I-V) curve of the photodiode with 0 V external bias applied. An ideal photodiode should have infinite R_{SH}, but practical values are commonly in the order of MΩ. The shunt resistance is mainly used to determine the noise current of the photodiode in photovoltaic mode (zero bias) and a high resistance is desired.

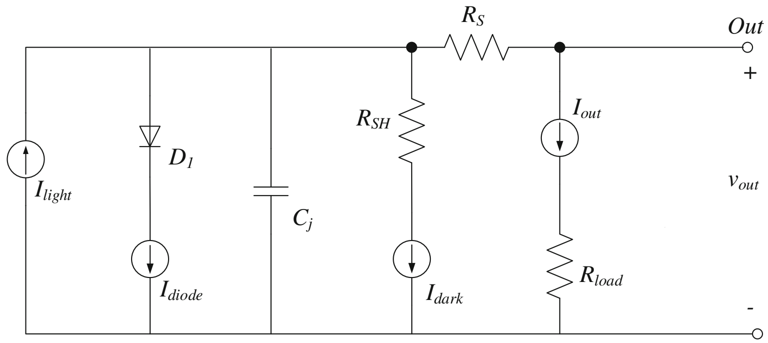


Fig. 4.4 Equivalent circuit and noise sources of a photodiode

The series resistance R_S arises from the resistance of the ohmic contacts and the resistance of the undepleted region, and is given by

$$R_s = \frac{(W_s - x_d)\rho}{A} + R_c \tag{4.19}$$

where W_s is the thickness of the substrate, x_d is the width of the depletion region, A is the diffused area of the junction, ρ is the resistivity of the substrate and R_c is the contact resistance. The series resistance is used to determine the linearity of the photodiode in photovoltaic mode and a 0 Ω resistance is desired, although practical values for series resistance are usually somewhere below 100 Ω.

The junction capacitance C_j is a capacitance value modelled owing to the boundaries of the depletion region acting as the parallel plates of a capacitor. The junction capacitance is proportional to the diffused area and inversely proportional to the width of the depletion region. Higher resistivity substrates have lower junction capacitance, and the value is furthermore dependent on the reverse bias, given by

$$C_j = \frac{\varepsilon_s \varepsilon_0 A}{x_d} \tag{4.20}$$

where ε_0 is given in Table 4.5, and ε_s is the permittivity of the substrate. The value of C_j determines the 3 dB bandwidth of the photodiode, since

$$f_{3dB} = \frac{1}{2\pi R_L C_j} \tag{4.21}$$

where R_L is the load resistance connected at the output. From f_{3dB}, the rise time t_r (and fall time t_f), which is defined as the time for the signal to rise from 10 to 90 % (or from 90 to 10 % for fall time) of its final value is represented by

$$t_r \cong \frac{0.35}{f_{3dB}} \tag{4.22}$$

from which the charge collection time of the carriers in the depleted region, in the undepleted region, and the RC time constant of the photodiode are defined.

The I-V characteristics of a photodiode with no incident light are similar to a rectifying diode. If the photodiode is forward-biased there is an exponential increase in the current. If a reverse-bias is applied, a small reverse saturation current appears across the anode and cathode terminals, related to the dark current by the following:

$$I_D = I_{SAT}\left(e^{\frac{qV_A}{kT}} - 1\right) \tag{4.23}$$

where I_D is the photodiode dark current and I_{SAT} is the reverse saturation current. Illuminating the photodiode with optical radiation shifts the I-V curve by the amount of the photocurrent I_P, and the total current I_{TOTAL} is given by

$$I_{TOTAL} = I_{SAT}\left(e^{\frac{qV_A}{kT}} - 1\right) - I_P. \quad (4.24)$$

As the applied reverse bias increases, there is a sharp increase in the photodiode current. The applied reverse bias at this point is referred to as the breakdown voltage. This is the maximum applied reverse bias and the photodiode should be operated below this point.

Finally, the noise characteristics of the photodiode are an important design parameter and are dependent on the shunt resistance, photocurrent and dark current of the device. A figure of merit of photodiodes is its noise equivalent power (NEP). The NEP is the optical signal power required to create an electronic signal such that the signal-to-noise ratio is unity. Lower values of NEP imply better overall performance of the device, and the units are in W. The NEP is defined by

$$\text{NEP} = \frac{I_t}{R_\lambda} \quad (4.25)$$

where I_t is the total noise current in the photodetector and R_λ is the wavelength-dependent responsivity of the photodetector in A/W. The total noise current is given by

$$I_t = \sqrt{I_s^2 + I_J^2} \quad (4.26)$$

where I_s is the shot noise and I_J is the Johnson (thermal) noise, given by

$$I_s = \sqrt{2q(I_P + I_D)}\Delta f \quad (4.27)$$

and

$$I_J = \sqrt{\frac{4kT\Delta f}{R_{SH}}} \quad (4.28)$$

respectively. In (4.27) and (4.28), Δf is the noise measurement bandwidth. Shot noise is the dominant noise in photoconductive (biased) mode and Johnson noise is dominant in photovoltaic (unbiased) mode. Important to note is the contributions of the dark current and photocurrent in biased mode, and of the shunt resistance in unbiased mode. The responsivity of the photodiode is a measure of the sensitivity to light and is defined as the ratio of photocurrent (I_P) to the incident light optical power P_{inc}, such that

$$R_\lambda = \frac{I_P}{P_{inc}} \tag{4.29}$$

and the responsivity is used to determine the photodiode quantum efficiency (QE) through the relationship

$$\mathrm{QE} = R_\lambda \frac{hc}{\lambda q} \tag{4.30}$$

where the values for h, c, and q are given in Table 4.5. The QE is the fraction of the incident photons that contribute to the photocurrent.

The optical power incident on the photodetector is dependent on the radiation intensity and wavelength of its source and could be in the form of various sources, such as the sun or heat radiated from, for example, a plume of smoke. To detect this radiation, the intensity profile of the heat source can be modelled as a Planck radiator (a blackbody), and a discussion on this is presented below.

4.6 Radiation Sources

Any object with a temperature above absolute zero (0 K) radiates and absorbs thermal energy. All matter consists of continuously randomly moving particles, constantly colliding with one another and the walls of the matter in which they are contained (cavity). This random movement (Brownian motion) is kinetic energy and its quantity is dependent on the temperature of the matter. These charged particles are the atoms that make up the matter and the kinetic energy as a function of temperature is observed as electromagnetic radiation. A blackbody (Lynch 1977) at thermal equilibrium emits electromagnetic radiation, referred to as blackbody radiation. The spectrum of this radiation is a function of temperature and is described by Planck's law of radiation. To derive Planck's law (Ragheb and Hamid 1987), it is assumed that the radiation is isotropic (uniform in all orientations), spatially homogeneous (uniform composition of space), unpolarized, and incoherent, enclosed in a cavity to retain equilibrium between the electromagnetic waves and the energy in the cavity wall. Accounting for these conditions and integrating overall wavelengths of light, it is possible to determine the thermal radiation as the total radiant exitance from the blackbody at temperature T in Kelvin through

$$M_e(T) = \sigma_e T^4 \tag{4.31}$$

where σ_e is the Stefan-Boltzmann constant of $\sigma_e = 5.67 \times 10^{-8}$ in units W $(\mathrm{m}^2\ \mathrm{K}^4)^{-1}$ and assuming that the environment is at 0 K with no incident flux. Exitance is the

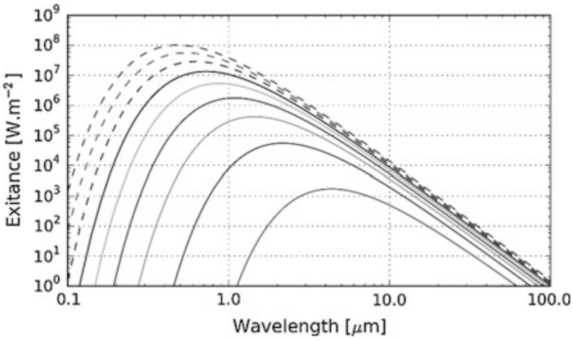


Fig. 4.5 Spectral radiant exitance from various sources, dependent on the temperature of the radiative source

areal density flux on the source surface area, with flux flowing outward from the surface with no regard to angular density. The exitance leaving the surface can be due to reflected light, transmitted light, emitted light or any combination thereof. Figure 4.5 presents the spectral radiant exitance from sources ranging from 300 K (room temperature) to 6000 K (estimated temperature of the sun as a blackbody) for wavelengths from 0 to 100 μm.

Figure 4.5 presents the radiative exitance from blackbody sources based on its temperature, where the lowest curve (bottom) is at 0 K and the highest curve (top) is at 6000 K. However, it is required to model the thermal radiator as a function of how it approximates a blackbody, as no real sources represent a perfect blackbody because of atmospheric losses. To quantify this ratio, emissivity is used and defined as

$$\varepsilon_\lambda = \frac{L_{object,\lambda}}{L_{bb,\lambda}} \tag{4.32}$$

which relates the radiance of the object ($L_{object,\lambda}$) to the radiance of a blackbody ($L_{bb,\lambda}$). Emissivity is therefore the material's effectiveness in emitting thermal radiation. The ratio falls between 0 and 1, where 1 represents a perfectly radiating source and 0 a non-radiative source. An example of a material with low emissivity is aluminum foil (0.03), whereas anodized aluminum has high (0.9) emissivity. A material with high emissivity (approaching unity and therefore a perfect blackbody) has low reflectance, whereas a material with low emissivity has high reflectance. The atmosphere can be modelled as the channel in which the thermal radiation from an object travels and inherently imposes losses on the signal. These losses are wavelength-dependent and are caused by physical attributes of matter in the atmosphere, such as aerosol and scattering, oxygen, carbon dioxide, ozone, water vapor, smoke, fog, clouds and additional particles in the environment. Combining all of these effects (assuming these are some of the largest contributors of atmospheric emissivity) the transmittance (inversion of emissivity) can be calculated for a given path and environment (atmospheric pressure relating also to altitude, temperature, and density) and represented as transmittance windows.

Mathematically it is convenient to represent the spectral transmittance $\tau(R)$ for a homogeneous medium by

$$\tau(R) = e^{-\gamma R} \tag{4.33}$$

where γ is the attenuation coefficient in units m^{-1} and R is the path length in m. In addition, atmospheric path radiance must also be considered in conjunction with atmospheric transmittance, accounting for the zenith angle and observer's point of reference, i.e. looking from a warm environment at a colder environment, and similarly for the opposite case. Atmospheric transmittance is generally divided into atmospheric windows grouped by their spectral quantities. These windows are divided into visible, near IR, short-wave IR (SWIR), medium-wave IR (MWIR), and long-wave IR (LWIR) regions, as summarized in Table 4.7.

According to Table 4.7, the visible region falls in the 0.5–0.75 μm region. This region is defined by human vision, the wavelengths that the human eye can see; it is largely unaffected by molecular absorption and has approximately 50 % transmittance (a relatively good percentage considering the absorption in the other regions). The near IR region consists of several narrow absorption spectral bands within the window of 0.75–1.25 μm. MWIR regions between 3 and 5 μm are commonly used to observe hot targets such as aircraft signatures. Sensors used in this region are rarely used for low ambient temperature targets, since they are not extremely sensitive to these temperatures. LWIR sensors are commonly used to observe lower temperature targets, such as at 300 K and above, since the spectral region between 8 and 12 μm is not largely affected by reflected sunlight from low reflectivity surfaces.

A blackbody at room temperature appears black, as most of the energy it radiates is IR and cannot be perceived by the human eye. Because the human eye cannot perceive color at very low light intensities, a blackbody, viewed in the dark at the lowest just slightly visible spectral radiance, subjectively appears grey, even though its objective physical spectrum peaks in the IR range. When it becomes a little hotter, it appears dull red. As its temperature increases further, it eventually becomes blindingly brilliant blue-white. Although planets and stars are neither in thermal equilibrium with their surroundings nor perfect blackbodies, blackbody radiation is used as a first approximation for the energy they emit. A blackbody radiates energy at all frequencies, but its intensity rapidly tends to zero at high frequencies (short wavelengths). For example, a blackbody at room temperature

Table 4.7 Atmospheric transmittance windows spectral regions

Atmospheric window	Spectral region (μm)
Visible	0.5–0.75
Near IR	0.75–1.25
SWIR	1.5–2.5
MWIR	3–5
LWIR	8–12

(300 K) with one square meter of surface area will emit a photon in the visible range (390–750 nm) at an average rate of one photon every 41 s, meaning that for most practical purposes, such a blackbody does not emit in the visible range. As the temperature decreases, the peak of the blackbody radiation curve moves to lower intensities and longer wavelengths. Exitance (and henceforth radiance) of a blackbody can be equated by Planck's law, given as

$$M_{ev}(T) = \frac{2\pi h v^3}{c^2\left(e^{\frac{hv}{kT}} - 1\right)} \tag{4.34}$$

where h is Planck's constant, v is the speed of light, k is Boltzmann's constant, T is the temperature in Kelvin and the unit for exitance is W.(m^2 Hz)$^{-1}$. The spectral radiance is given by $L_\lambda = M_{ev}(T)/\pi$ and completely defines the Planck radiator. The radiation at the destination (photodetector/sensor) is dependent on the solid angle between the device and the source. For example, consider a 1 mm^2 silicon sensor, viewing the sun at zero azimuth, and with a 1 130 nm cut-off frequency (see Table 4.5). The spectral response of the photodetector at wavelengths between 0 and 1.25 µm is as shown in Fig. 4.6.

According to Fig. 4.6, the spectral response of the photodetector is seen for wavelengths between 0 and 1.25 µm. The photodetector converts photons into electronic charges and the output signal is proportional to the number of incident photons, described by its responsivity. Photodetector responsivity depends on the incoming flux, the spectral QE of the detector, the gain of the detector material and the constant of electronic charge. For this specific photo-detector, a peak responsivity of 0.6 is achieved at approximately 900 nm. Since the photodetector responsivity is dependent on the incident flux, it is dependent on the source that generates the flux. Any source with temperature above 0 K contributes to radiant flux; for this example, the sun is considered the dominant source. The radiant exitance from the sun, estimated at 6000 K as a function of wavelength between 0 and 1.25 µm, is shown in Fig. 4.7.

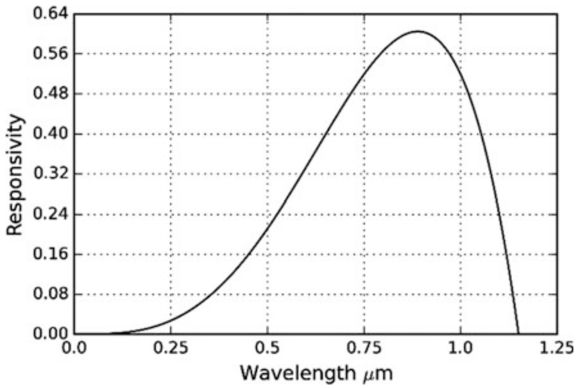


Fig. 4.6 Spectral response of silicon photodetector

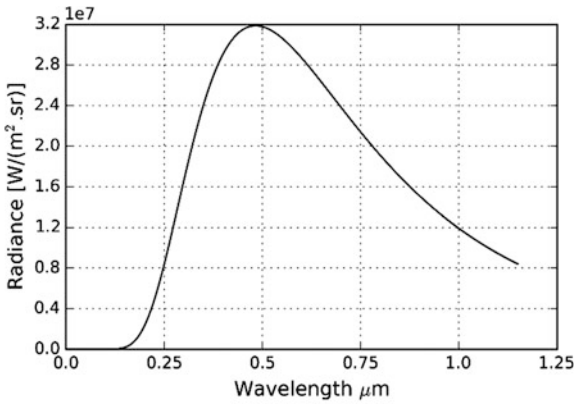


Fig. 4.7 Radiance of the sun as a blackbody in spectral band of the silicon photodetector

In Fig. 4.7, the exitance from the sun is seen to peak (in the specified band) at approximately 500 nm (0.5 μm). The amount of radiance energy at this point is around 32 MW m^{-2}. An interesting point in this regard is that the human eye has maximum spectral sensitivity at 507 nm under night conditions, and 555 nm during daylight, falling perfectly within the peak wavelengths that the sun offers (luck, or evolution?). The total amount of exitance from the sun fortunately does not reach the earth and the radiation available on earth depends on the solid angle formed between the sun and the earth. This solid angle φ is a function of the distance between the sun and earth (R_{sun}) and the area of the sun as seen from the side (A_{sun}). The solid angle can be determined by the following equation, since $R >> A$,

$$\varphi = \frac{A_{sun}}{\pi R_{sun}^2} \tag{4.35}$$

and is found to be $\varphi = 2.18 \times 10^{-5}$ sr sr^{-1}. Because of the solid angle and the areal density of flux on the receiving surface area, the irradiance in W m^{-2}, is given in Fig. 4.8.

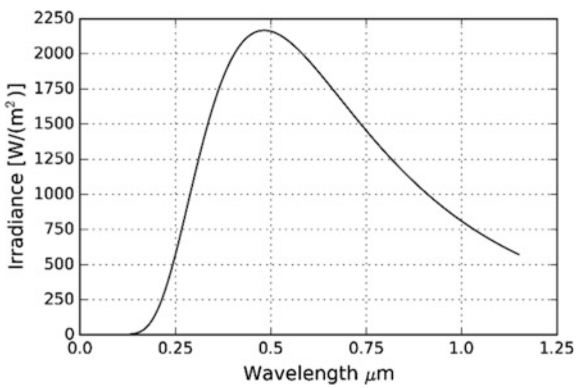


Fig. 4.8 Irradiance of the sun as a blackbody in spectral band of the silicon photodetector

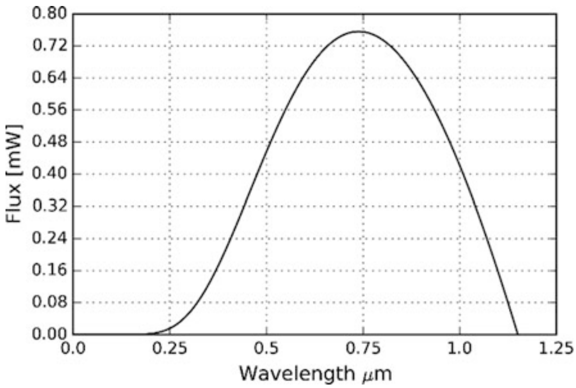


Fig. 4.9 Flux from sun on silicon photodetector with 1 mm^2 optical area

According to Fig. 4.8 the irradiance on the earth due to the sun as the 6000 K source is found to peak at approximately 2200 W m^{-2}. This is the available flux that humans can convert to electrical energy. The conversion efficiency of semiconductor devices, however, determines the maximum power output that can be harvested from this energy. If the silicon photodetector, with responsivity shown in Fig. 4.4, and with 1 mm^2 area, is used to convert the irradiance to electrical energy, the total available flux is given in Fig. 4.9.

Figure 4.9, representing the flux on the silicon photodetector, shows that the converted energy peaks at approximately 0.75 mW (750 μW) for the defined optical area. Cascading large amounts of these photodetectors will improve the power output, and enhancements in technologies aim to improve the conversion efficiency further.

Increased system integration has allowed solar energy harvesters, in the form of passive photodiodes, to be implemented on the same silicon die as active circuitry, which can be powered by the harvested energy. These integrated photodiodes can be modeled after a passive pixel architecture, which can form the basis for CMOS imagers. For many wireless systems, photovoltaics is a viable source for energy harvesting. Integrating the solar energy harvesting on the same die as other parts of the system enables reduced system cost and size. Piezoelectric energy harvesters offer some of the highest efficiency and power output by size and cost. However, there are also challenges of reliability and broadband performance that need to be addressed.

4.7 Piezoelectric Energy

Piezoelectric materials convert mechanical energy from vibrations, force, strain and pressure into electrical signals (Erturk and Inman 2011; Sarker et al. 2012). The piezoelectric effect is reversible, in other words, it can be applied from two perspectives. The direct effect ensues when a charge separation occurs in response to mechanical stress and the transducer is used as a sensor, and the converse effect is the occurrence of stress and strain when an electric field is applied to surface electrodes and the device is used as an actuator. The general forms of the constitutive equations for a piezoelectric material are given by

$$\delta = \frac{\sigma}{Y} + dE \tag{4.36}$$

and

$$D = d\sigma + \varepsilon E \tag{4.37}$$

where δ is mechanical strain, σ is mechanical stress in N m^{-2}, Y is Young's modulus of elasticity, d is the piezoelectric strain coefficient in m V^{-1}, E is the electric field in V m^{-1}, D is the electrical displacement in C m^{-2}, and ε is the dielectric constant (permittivity) of the piezoelectric material in F m^{-1}. The piezoelectric strain constant d is defined as the ratio of developed free strain to the applied electric field. The subscript d_{ij} implies that the electric field is applied or charge is collected in the i direction for a displacement or force in the j direction. To define the basis of piezoelectric materials operation, consider Fig. 4.10.

To describe the electromechanical equations for a linear piezoelectric material fully, consider the following forms of (4.36) and (4.37).

$$\delta_i = S_{ij}^E \sigma_j + d_{mi} E_m \tag{4.38}$$

and

$$D_m = d_{mi} \sigma_i + \varepsilon_{ik}^{\sigma} E_k \tag{4.39}$$

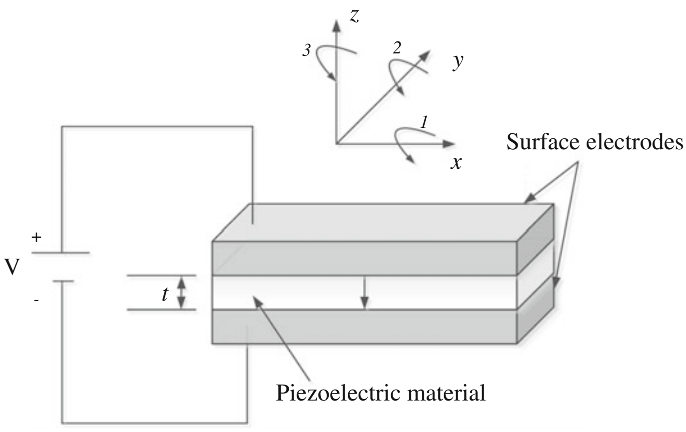


Fig. 4.10 Schematic diagram of a linear piezoelectric transducer

where S_{ij} is the elastic compliance constant and the indices $i, j = 1, 2, \ldots, 6$ and m, $k = 1, 2, 3$ refer to the different directions in the material coordinate system, as depicted in Fig. 4.10. The converse piezoelectric effect is described by (4.38) and the direct effect by (4.39). For applications that involve sensing, rewriting (4.38) and (4.39) in the following forms is convenient.

$$\delta_i = S_{ij}^D \sigma_j + g_{mi} D_m \tag{4.40}$$

and

$$E_i = g_{mi} \sigma_i + \beta_{ik}^{\sigma} D_k \tag{4.41}$$

where g represents the matrix of piezoelectric constants in $m^2\ C^{-1}$ and β represents the inverse of the permittivity component in $m\ F^{-1}$. These equations again represent the converse (4.40) and direct (4.41) piezoelectric effects. The converse effect is often used to determine the piezoelectric coefficients. In matrix form, Eqs. (4.38)–(4.39) can be written as

$$\begin{bmatrix} \delta_1 \\ \vdots \\ \delta_6 \end{bmatrix} = \begin{bmatrix} S_{11} & \cdots & S_{16} \\ \vdots & \ddots & \vdots \\ S_{61} & \cdots & S_{66} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_6 \end{bmatrix} + \begin{bmatrix} d_{11} & \cdots & d_{31} \\ \vdots & \ddots & \vdots \\ d_{16} & \cdots & d_{36} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{4.42}$$

and

$$\begin{bmatrix} D_1 \\ \vdots \\ D_3 \end{bmatrix} = \begin{bmatrix} d_{11} & \cdots & d_{16} \\ \vdots & \ddots & \vdots \\ d_{61} & \cdots & d_{66} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_6 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11}^{\sigma} & \cdots & \varepsilon_{13}^{\sigma} \\ \vdots & \ddots & \vdots \\ \varepsilon_{31}^{\sigma} & \cdots & \varepsilon_{33}^{\sigma} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{4.43}$$

where the notations for shear stress σ_4, σ_5, and σ_6 can be written as τ_{23}, τ_{31}, and τ_{12} and for shear strain, δ_4, δ_5, and δ_6 as γ_{23}, γ_{31}, and γ_{21} respectively. If it is assumed that the device is poled along axis 3 (see Fig. 4.10), and viewing the piezoelectric material as a transversely isotropic material, a fair assumption for piezoelectric ceramics, then some of these parameters in (4.42) and (4.43) equate to zero and are related to others. The non-zero compliance coefficients are $S_{11} = S_{22}$, $S_{13} = S_{31} = S_{23} = S_{32}$, $S_{12} = S_{21}$, $S_{44} = S_{55}$, and $S_{66} = 2(S_{11} - S_{12})$. The non-zero piezoelectric strain constants are $d_{31} = d_{32}$ and $d_{15} = d_{24}$. Also, the non-zero dielectric coefficients are $\varepsilon_{11}^{\sigma} = \varepsilon_{22}^{\sigma} = \varepsilon_{33}^{\sigma}$. Taking these simplifications into account, (4.42) and (4.43) reduce to

$$\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & 0 & 0 & 0 \\ S_{12} & S_{11} & S_{13} & 0 & 0 & 0 \\ S_{13} & S_{13} & S_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(S_{11} - S_{12}) \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \tau_{23} \\ \tau_{31} \\ \tau_{12} \end{bmatrix} + \begin{bmatrix} 0 & 0 & d_{31} \\ 0 & 0 & d_{31} \\ 0 & 0 & d_{33} \\ 0 & d_{15} & 0 \\ d_{15} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{4.44}$$

and

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11}^{\sigma} & 0 & 0 \\ 0 & \varepsilon_{11}^{\sigma} & 0 \\ 0 & 0 & \varepsilon_{33}^{\sigma} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{4.45}$$

The actuation matrix in (4.45) is valid for piezo-ceramic materials and should be modified for polyvinylidene fluoride (PVDF) materials to reflect the fact that in PVDF films the induced strain is non-isotopic on the surface of the film. An electric field applied in the direction of the polarization vector will result in different strains in one and two directions. The actuation matrix for PVDF is given by

$$\begin{bmatrix} 0 & 0 & d_{31} \\ 0 & 0 & d_{32} \\ 0 & 0 & d_{33} \\ 0 & d_{25} & 0 \\ d_{15} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{4.46}$$

The piezoelectric coefficient d_{ij} is the ratio of the strain in the j direction to the electric field applied along the i direction when all external stresses are held constant. If a voltage V is applied to a piezoelectric transducer polarized in direction 3, it follows that the generated electric field is given by

$$E_3 = \frac{V}{t} \tag{4.47}$$

where t is the thickness of the piezoelectric material seen in Fig. 4.10. This field strains the transducer such that the deformation of the solid material is given by

$$\delta_1 = \frac{\Delta l}{l} \tag{4.48}$$

where l is the initial length of the piezoelectric material and the change in length Δl is given by

$$\Delta l = \frac{d_{31} V l}{t}. \tag{4.49}$$

The piezoelectric constant d_{31} is generally a negative value, since a positive electric field generates a positive strain in direction 3. The coefficient d_{ij} can also be seen as the ratio of short circuit charge per unit area flowing between connected electrodes perpendicular to the j direction to the stress applied in the i direction. If a force F is applied to the transducer in the 3 directions it generates stress by

$$\sigma_3 = \frac{F}{lw} \tag{4.50}$$

where w is the width of the piezoelectric material and the resultant electric charge is given by

$$q = d_{33} F, \tag{4.51}$$

flowing through the short-circuit. If stress is applied equally in the 1, 2, and 3 directions and the electrodes are perpendicular to axis 3, the resulting short-circuit charge per unit area, divided by the applied stress, is denoted as d_p.

The piezoelectric constant g_{ij} signifies the electric field developed along the i direction when the material is stressed along the j direction. The applied force F therefore results in a voltage given by

$$V = \frac{g_{31} F}{w}. \tag{4.52}$$

In addition, this coefficient can be interpreted as the ratio of strain developed along the j direction to the charge deposited on the electrodes perpendicular to the i direction, per unit area. If an electric charge of Q is deposited on the surface electrodes, the length of the piezoelectric element will change by

$$\Delta l = \frac{g_{31} Q}{w} \tag{4.53}$$

and the piezoelectric transducer will be subject to the applied charge.

The elastic compliance constant S_{ij} is the ratio of the strain in the i direction to the stress in the j direction given that there is no change of stress along the other two directions. Direct strains and stresses are denoted by indices 1–3. Shear strains and stresses are denoted by indices 4–6. This means that S_{12} signifies the direct strain in direction 1 when the device is stressed along direction 2 and stresses along directions 1 and 3 are unchanged. The superscript E is used to state that the elastic compliance and is measured with the electrodes short-circuited, and the superscript

D denotes that the measurements are taken when the electrodes are in open-circuit operation. Mechanical stress results in an electrical response that can increase the resultant strain.

The piezoelectric coupling coefficient k_{ij} represents the ability of a piezo-ceramic material to transform electrical energy to mechanical energy as well as the other way around. This transformation of energy between mechanical and electrical domains is employed in both sensors and actuators made from piezoelectric materials. The ij index indicates that the stress or strain is in the j direction and the electrodes are perpendicular to the i direction. Therefore, if a piezo-ceramic is mechanically strained in direction 1 as a result of electrical energy input in direction 3 while the device is under no external stress, then the ratio of stored mechanical energy to the applied electrical energy is denoted as k_{31}^2. To measure k_{ij}, one method involves applying a force to the piezoelectric element while leaving its terminals in open-circuit operation. The piezoelectric device will deflect and this deflection Δz can be measured. The mechanical work done (W_M) can be determined by the applied force through the following relationship.

$$W_M = \frac{F\Delta z}{2} \tag{4.54}$$

and because of the piezoelectric effect, electric charge will accumulate on the transducer electrodes with electrical energy W_E equal to

$$W_E = \frac{Q^2}{2C_p} \tag{4.55}$$

where C_p represents the piezoelectric capacitor that stores this energy. According to (4.19) and (4.55) the coefficient k_{33} is determined by

$$k_{33} = \sqrt{\frac{W_E}{W_M}} = \frac{Q}{\sqrt{F\Delta z C_p}}. \tag{4.56}$$

In order to harvest electrical energy from a piezoelectric device, pressure can be applied to the device and in turn electrical energy can be generated. The above discussion leads to using piezoelectric sensors to generate usable voltages, generally sufficient to power low-power nodes such as WSNs and RFID tags. Piezoelectric sensors offer superior signal-to-noise ratio and better high-frequency noise rejection compared to strain gauges. These sensors are also compact and require relatively low-complexity support circuitry. If a piezo-ceramic sensor is subjected to a stress and assuming the applied electric field is zero, the resulting displacement vector is given as

$$\begin{Bmatrix} D_1 \\ D_2 \\ D_3 \end{Bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} \tag{4.57}$$

and the generated charge can be determined from

$$q = \iint [D_1 \quad D_2 \quad D_3] \begin{bmatrix} dA_1 \\ dA_2 \\ dA_3 \end{bmatrix} \tag{4.58}$$

where dA_1, dA_2, and dA_3 are the differential electrode areas in the 2–3, 1–3, and 1–2 planes respectively. The generated voltage V_p can be related to the charge by

$$V_p = \frac{q}{C_p} \tag{4.59}$$

and if the voltage V_p is measured, then the strain can be determined by solving the integral in (4.58). If the stress field is only applied along direction 1, then the capacitance can be determined by

$$C_p = \frac{lw\varepsilon_{33}^{\sigma}}{t} \tag{4.60}$$

and the sensor voltage generated from the applied pressure is given by

$$V_s = \frac{d_{31} Y_p w}{C_p} \int_0^l \delta_1 dx \tag{4.61}$$

where Y_p is the Young's modulus of the sensor and δ_1 is averaged over the length of the sensor through the integral. The strain can be calculated by

$$\delta_1 = \frac{C_p V_s}{(1 - v) d_{31} E_p lw} \tag{4.62}$$

where v is the Poisson ratio, accounting for whether the sensor is strained along more than one axis, else $v = 0$.

The market value of piezoelectric sensors is expanding rapidly, especially in the sectors pertaining and managing pavements, roads, and railroads, consumer electronics, industrial switches, remote controls, pushbutton industrial sensors, electronic locks/access control devices, toys and gadgets, military, aerospace and vehicle sensors as well as healthcare. Many miniaturized components, such as RFID tags and piezoelectric components (Wen et al. 2010), are manufactured using MEMS technology.

4.8 Microelectromechanical Structures

RFID tags, for example, contain at least two main subsystems. The first is an integrated circuit for storing and processing information, modulating and/or demodulating the incident RF signals, and receiving an incident DC signal from the reader. It should also contain an antenna, for transmitting and/or receiving the signals. The tag information is stored in non-volatile memory. If no data processing is required on the device, the IC can be omitted from the structure, leaving only the antenna, memory and read-out circuit on the device. Both these structures can be created using silicon-on-insulator or MEMS technology.

MEMS in its most basic form (miniaturized mechanical and electro-mechanical elements) require three steps during manufacturing: the deposition of metal (or other) layers, patterning by photolithography, and etching to produce the required shapes. Silicon-on-insulator works on the same principle, but the differences in process steps are not discussed in this work, as this falls outside the scope of the proposed topic.

The potential advantages of MEMS start to become apparent when these miniature sensors and/or actuators are merged onto a common silicon substrate together with CMOS ICs. Components such as coplanar waveguides are commonly used to conduct electronic (generally microwave) signals on a dielectric substrate. This integration of technologies allows for true system-on-chip products. MEMS sensors and actuators already exist in commercially available products such as microphones, inkjet printers, accelerometers, gyroscopes, micro-fluidics and micro-scale energy harvesting, with many more applications available. For example, MEMS micro-cantilevers can interact selectively with a particular immunogenic based on the antibody layer coating and sense its presence and content in a specimen. These chemical sensors are used in biomedical applications, but MEMS cantilever applications can be extended to RF filters and resonators, as well as stress sensors through monitoring their deflection under certain conditions. The micro-cantilever deflection (δ) can be determined by Stoney's equation, related to the amount of applied stress (σ) through the following equation;

$$\delta = \frac{3\sigma(1 - v)}{E}\frac{L^2}{t} \tag{4.63}$$

where v is Poisson's ratio, E is Young's modulus, L is the beam length, and t is the cantilever thickness.

Antennas, more specifically micro-antennas capable of RF energy harvesting, can be manufactured in various MEMS technologies. Two distinct categories of MEMS antennas exist, namely generic RF MEMS antennas and RF MEMS antenna circuits. RF MEMS antennas are physical radiating structures whose dimensions are proportional to the frequency of operation. Inherently, this can become problematic, since at very high frequencies, such as millimeter-wave (mm-wave) (30–300 GHz), the device size is comparable to its supporting structures and even interconnects can induce electromagnetic coupling. At lower frequencies, the devices can become relatively large and might influence the viability of using such devices for insect tagging and/or tracking. This results from the fact that the length (L) of the antenna depends on the wavelength (λ), which is given by

$$\lambda = \frac{n \times c}{f} \tag{4.64}$$

where n is the intrinsic impedance of free-space (377 Ω), c is the speed of light in a vacuum ($c = 300 \times 10^6$ m s^{-1} in free space) and f is the frequency of operation.

RF MEMS antenna circuits are resonant circuits containing non-radiating RF MEMS components to represent the architecture of a radiating structure, with similar transfer characteristics across its frequency band.

4.9 Conclusion

This chapter describes alternative renewable energy sources such as wind energy, hydropower, photodiodes, radiation sources, piezoelectric energy and microelectromechanical structures. These energy sources could serve as viable alternatives to pollution-generating fossil fuels used as the primary source of energy in most countries. Many countries are striving to replace fossil fuel energy plants with more sustainable and cleaner sources; where developed countries are leading the way towards these implementations. Developing countries are also adopting these energy sources as alternatives, but at a slower pace compared to developed countries; mainly due to financial constraints and lack of governmental backing. Design considerations and fundamental equations for each of these renewable energy sources are presented in this chapter to identify critical design parameters when considering such alternatives.

References

Castaldi, D., Chastain, E., Windram, M., Ziatyk, L. (2003). A Study of Hydroelectric Power: From a Global Perspective to a Local Application. *2003 Center for Advanced Undergraduate Studies and Experience*. The Pennsylvania State University.

Erturk, A., Inman, D. J. (2011). Piezoelectric Energy Harvesting. ISBN 9780470682548.

Li, P. Wen, Y. Young, P., Li, X., Jia, C. (2010). A Magnetoelectric Energy Harvester and Management Circuit for Wireless Sensor Networks. *Sensors and Actuators A: Physical Review*, 157, 100–106.

Liangliang, Q., Qijuan, C., Junyi, L. (2012). Study on the Development of Smart Hydropower Plants. *IEEE Conference Publications*, 330–333.

Lynch, A. C. (1977). Early History of Planck's Radiation Law. *IET Journals & Magazines*, 23(3), 244.

Mukasa, A., D., Mutambatsere, E., Arvanitis, Y., Triki, T. (2013). Development of Wind Energy in Africa. African Development Bank Group.

Ragheb, H., Hamid, M. (1987). An Approximation of Planck's Formula for the Inverse Black Body Radiation Problem. *IEEE Transactions on Antennas and Propagation*, 35(6), 739–742.

Sarker, M. R., Ali, S. H., Othman, M. Islam, S. (2012). Designing a Battery-Less Piezoelectric based Energy Harvesting Interface Circuit with 300 mV Startup Voltage. *3rd ISESCO International Workshop and Conference on Nanotechnology (IWCN2012) – Journal of Physics*, Conference Series 431, 1–8.

Vullers, R. J. M., van Schaijk, R., Doms, I., Van Hoof, C., Mertens, R. (2009). Micropower Energy Harvesting. *Solid-State Electronics*, 53, 684–693.

Chapter 5
The Sub-systems of an Energy Harvesting Device: Focus on RFID Fundamentals

5.1 Introduction

A good example of energy-harvesting circuits is RFID devices. These devices convert EM waves to information, using the signals as a power source. Energy harvesting for wireless sensors and RFID smart tags is a critical design feature on the path to the grand vision of ubiquitous computing. An important subset of energy-harvesting solutions is the use of the RF reader signal to deliver energy and power up smart RFID tag circuitry. Figure 5.1 depicts a general energy-harvesting block diagram with proposed sub-systems required to harness natural energy and convert it to usable power for low-powered devices.

In Fig. 5.1, the commonly found sub-systems of an energy-harvesting system are depicted. Such a system firstly requires a renewable ambient energy source such as wind, solar power, vibrations, movement, electromagnetic waves or heat. This source is captured and converted to electrical energy by a device such as a piezoelectric, thermoelectric, pyroelectric, capacitive or solar device. The converted energy is generally in AC and is converted to DC using a rectifier circuit. Low-power semiconductor devices are commonly used to perform this task. The electrical voltage is maintained at a constant level even if ambient source input changes by using a regulator circuit to stabilize the DC voltage. The regulated voltage is then amplified to a usable level, again using semiconductor devices and low-power integrated circuits. This voltage or current is either used by the load, usually a sensor, wireless sensor network (Safarian et al. 2014), or even a biomedical device (Ma et al. 2011), or if the energy is not immediately required it can be stored in a storage device such as a capacitor of battery cell. If the load device is activated, the captured data must be processed. This can be done by using ultra-low power microcontroller devices, designed for functionality and power efficiency, and transmitted to a central station using the transceiver module. These devices are optional in such a system, and all contribute to energy consumption.

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_5

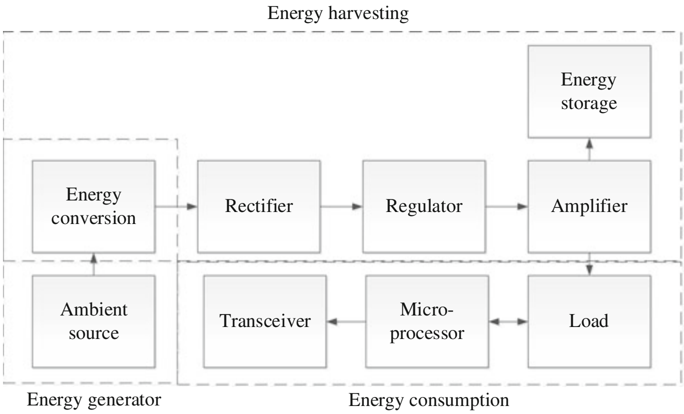


Fig. 5.1 General block diagram of energy-harvesting sub-systems

The essence of an energy-harvesting circuit, however, lies in converting ambient sources to usable electrical energy.

The sub-systems that make up an energy-harvesting system are discussed in the following sections. Each sub-system is described in its general form, with additional information from the current body of knowledge supplied about performance enhancements and circuit improvements as the focus. Ongoing research into each of these sub-systems aims to provide more efficient circuits with lower power requirements, a smaller footprint on integrated circuit level and higher efficiency levels using either improved techniques in available technology, or new technology implementations to test and verify its effectiveness. The first sub-system to be discussed is the energy conversion itself.

Before discussing the sub-systems that make up an energy-harvesting system, the RFID technique is briefly discussed.

5.2 Radio-Frequency Identification

Several techniques are used to actively track or passively tag objects, which is beneficial in expanding smart cities. Depending on size, cost, and power limitations, these include high-frequency radio and radar implementations for tracking, and RFID, optical, or combined techniques.

Passive tags do not require their own power source and obtain energy from a tag reader. This is generally the easier option, but it limits the reading distance of the object. An active tag has its own power source built into the tag itself. This increases the complexity, the cost, and the physical dimensions and weight of the device. The advantage of such a system is the increased reading distance from the source, allowing for tracking when the object is outside the allocated range.

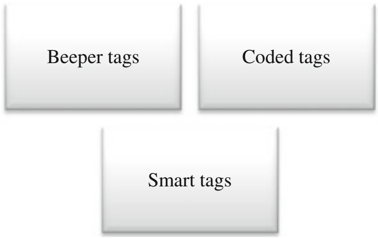


Fig. 5.2 Active radio tag categories

The reading distance depends on the frequency of operation, therefore the size of the antenna, and the power transmitted by the reader. The device on the object must also be small enough not to interfere with its natural activities, especially for example if used on flying insects.

From an active tag perspective, radio tags are divided into three general categories as shown in Fig. 5.2.

The active radio tags in Fig. 5.2 include beeper tags that only inform about the detection of the object, coded tags may include additional information such as the current temperature or humidity from sensors where each tag could share the frequency of operation and smart tags include a microprocessor and memory and could implement bidirectional communication and a global positioning system. Smart tags are generally larger in size and not suitable for insects. Coded tags include scanning radar and harmonic radar tags that can provide location and distance from the transmitter. Radar tags can also estimate the velocity of the object (from the Doppler effect) and estimate the size of the object. Radar tag applications are advantageous for large-scale monitoring of for example swarms of locust, but the complexity of receiver-side electronic signal demodulation can become cumbersome.

RFID tags are less complex and therefore allow for smaller footprint devices. A passive RFID requires a reader to transmit energy to the device. The tag does not emit its own RF carrier but modulates the existing carrier, generally using amplitude shift keying. This energy is relayed back to the reader through inductive or capacitive coupling that picks up a sudden change in amplitude through envelope detection. Each tag would contain an identification code (binary barcode) unique to the object to which it is attached. The received code can be up to 60 dB lower than the transmitted signal, therefore sensitivity and good noise suppression are required at the receiver. The communication of such a device works on the same principle as any generic transponder system, where the Friis free-space equation can be used to determine the distance and power requirements of the system;

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{5.1}$$

where $P_r(d)$ is the power received by the receiving antenna (in watt) at distance d (in m), P_t is the amount of power transmitted (in watt), G_t is the gain of the transmitter and G_r is the gain of the receiver (dimensionless), λ is the wavelength of the transmitted signal (in meters), and L is the system loss factor ($L \geq 1$). A value of $L = 1$ would suggest that there are no losses in the system due to transmission line attenuation, filter losses, or arbitrary antenna losses. It is important to note that from (5.1) the received power is attenuated proportionally to the square of the distance (d) between the transmitter and the receiver. Since this would be a passive implementation, G_r would represent a loss (as opposed to a potential gain) within the system as well.

Miniaturization of RFID tags is an active field of research and Hitachi®, for example, has created a 50 × 50 μm tag, called the mu-chip®. This device has been created in silicon-on-insulator and can store a 38-digit number using 128-bit read-only-memory. Currently, the greatest challenge is attaching the antennas, limiting the read-range to a few millimeters.

Since active RFID devices on small objects such as insects are limited by the requirement to supply enough power to the devices to transmit signals over a practical distance, passive RFID devices can be manipulated more easily by increasing the energy transmitted from the central transceiver. For bees, for example, the reader can be positioned at a specially crafted entrance to the hive, where it senses the RFID transponder when the insect passes through the entrance compartment of the hive. The immediate environment could also contain additional readers, with output energy of the RF signal proportional to the required distance of detection. Sensing readers are rarely used as singular devices and an array of devices would be beneficial in tracking objects better in the immediate environment.

Analysis of generated data would require several layers of data acquisition. The layers are depicted in Fig. 5.3.

As shown in Fig. 5.3, a data record layer consists of the physical RFID tags mounted on the objects. The data collection layer would generally consist of multiple RFID readers (transponders) transmitting at various levels, dependent on the number of readers and the distances between them. Data aggregation would require networked devices to capture the data to a data sharing/backup layer where a network database would store and analyze the acquired data.

Different countries allocate different bands of the radio spectrum to RFID, so no single technology optimally satisfies all the requirements of existing and potential markets. The industry has worked to standardize three main RF bands—low frequency (LF), high frequency (HF), and ultra-high frequency (UHF)—with recent developments in the super-high frequency (SHF) band, although this is used less often since low-bandwidth, power-efficient frequencies are preferred. Most countries have assigned the 125 or 134 kHz areas of the spectrum to low-frequency RFID systems, and 13.56 MHz is generally used around the world for high-frequency RFID systems. Typical RFID energy-harvesting frequency allocations are summarized in Table 5.1.

UHF RFID systems (Jinpeng et al. 2013) have only been around since the mid-1990s and countries have not agreed on a single area of the UHF spectrum for RFID. Accordingly, different countries have different bandwidth and power

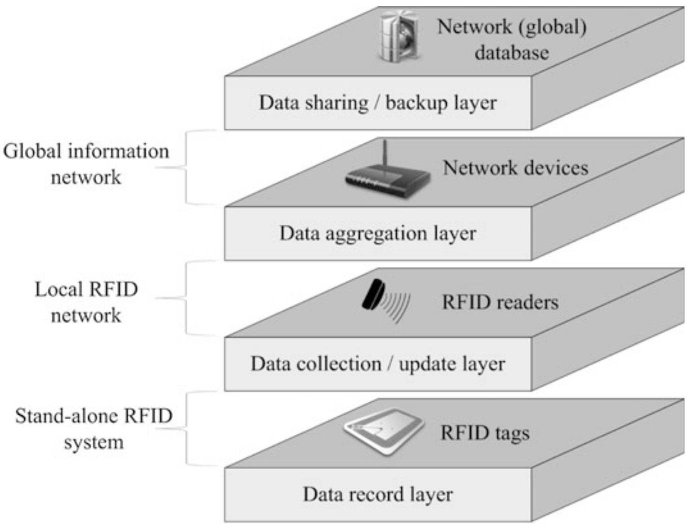


Fig. 5.3 Data acquisition layers for a RFID system

Table 5.1 RFID frequency bands and spectrum allocations worldwide

RFID frequency band	Frequency band description	Typical range	Applications
125–134.2 kHz and 140–148.5 kHz	Low frequency	Up to 50 cm	Animal ID, car key locks
6.765–6.795 MHz	Medium frequency	Up to 1 m	Smart cards, clothing ID
13.553–13.567 MHz	High frequency (often called 13.56 MHz)	Up to 1 m	
26.957–27.283 MHz	High frequency	Up to 1 m	Special applications
433 MHz	Ultra-high frequency	Up to 10 m	Remote car locks (Europe)
858–930 MHz	Ultra-high frequency	1–10 m	Supply chain tracking
2.400–2.483 GHz	Super high frequency	>3 m	Highway toll collection, vehicle fleet ID, long-range active tags
2.446–2.454 GHz	Super high frequency	>3 m	
5.725–5.875 GHz	Super high frequency	>3 m	

restrictions for UHF RFID systems. Across the European Union, UHF RFID ranges from 865 to 868 MHz, with RFID readers able to transmit at maximum power (2 W effective radiated power (ERP)) at the center of that bandwidth (865.6–867.6 MHz). In North America, the UHF RFID frequency ranges from 902 to 928 MHz with readers able to transmit at maximum power (1 W ERP) for most of that bandwidth.

Most other countries have adopted either the European Union or North American standard, or they are using a subset of one of the two bandwidths. The only two African countries that have definite allocated frequencies for RFID applications are South Africa and Nigeria. Nigeria uses 865.6–867.6 MHz with 2 W ERP, regulated by the Nigerian Communications Commission (NCC), and South Africa the 865.6–867.6 MHz with 2 W ERP, and 915.4–919 and 919.2–921 MHz with 4 W EIRP, regulated by the Independent Communications Authority of South Africa.

5.3 Sub-system Definitions

The following sections describe the sub-systems in Fig. 5.1 with varying degrees of complexity, and aim to adapt the working principles. The first component that is discussed is the antenna, which is responsible for intercepting EM signals from the environment and feeding them to the rest of the device.

5.4 Antenna

The theoretical operation of an antenna is described here and is applicable to energy-harvesting devices. The directional characteristics, the gain of an antenna, and the resonant frequency are considered the focal design considerations. These parameters determine the types of antennas that can be used in the application and the physical size of the antenna. Gain and efficiency are performance characteristics of the antenna. Also considered here is the impedance of the antenna and the effective aperture that contributes to the amount of power delivered to the terminals of the antenna related to its effective area. Table 5.2 summarizes equations used in antenna characterization.

To determine the maximum directivity (D_{max}) in Table 5.2, the maximum radiation intensity (U_{max}) in W/unit solid angle, and the total radiated power (P_{rad}) in watts are required. In the equation for the isotropically radiated gain, P_{in} is the power (in watts) received by the antenna. The current through the radiating structures is given by I_0. In the equation for the electric field, η is the intrinsic impedance of free-space (377 Ω), and $P_r(d)$ is the received power specified in watts at distance d.

The directivity of an antenna is defined as the ratio of radiation intensity in a specified direction from the antenna to the average radiation intensity calculated in all directions around the same antenna, and is considered an important figure of merit in antenna design. This unit specifies that the maximum directivity is radiated in a specific direction as compared to the average directivity of an isotropic antenna, which radiates equally in all directions. Measurements of antenna electric and magnetic field propagation are done in specific field regions considered at a given distance from the antenna. These regions are usually divided into three categories, namely the reactive near-field, radiating near-field (or Fresnel zone), and the

Table 5.2 Antenna characterization parameters

Parameter	Equation
Maximum directivity	$D_{max} = D_0 = \frac{4\pi U_{max}}{P_{rad}}$
Inner boundary of far-field region (outer boundary is infinity)	$R = 2D^2/\lambda$
Isotropically radiated gain	$G = 4\pi \frac{U}{P_{in}}$
Antenna efficiency	$\eta = \frac{R_r}{R_r + R_L}$
Effective aperture	$A_{em} = \frac{\lambda^2}{4\pi} D_0$
Radiation resistance	$R_r = \frac{2P_{rad}}{\lvert I_0 \rvert^2}$
Electric field	$\lvert E \rvert = \sqrt{\frac{\eta \times P_r(d)}{A_{em}}}$

far-field (Fraunhofer) regions. The boundaries that separate these regions are not unique, but there are some techniques that identify these regions.

Firstly, the reactive near-field region lies immediately in the area surrounding the antenna where the reactive field dominates. Strong inductive and capacitive effects occur from the currents and charges in the radiating source, and this behaviour is very different from effects in other regions. Absorption of radiated power in this region can be fed back to the transmitter and may increase the impedance of the antenna that the transmitter sees, resulting in apparent mismatched conditions. The outer boundary of this region can be defined at a distance (R) from the radiating source, and is approximately $R < 0.62(D^3/\lambda)^{1/2}$ from the antenna surface, where D is the largest physical dimension of the antenna, and λ is the wavelength.

The radiating near-field region lies between the reactive near-field and the far-field region. In this region, radiation fields predominate and the angular field distribution is dependent on the distance from the antenna. It is stated that if the antenna has a maximum physical dimension that is not large compared to the wavelength, this region may not exist. For the case where the maximum dimension of the antenna is large compared to the wavelength, the inner boundary of this region lies at a distance of $R \geq 0.62(D^3/\lambda)^{1/2}$ and the outer boundary at a distance of $R < 2D^2/\lambda$.

The far-field region is where the angular field distribution is essentially independent of the distance from the antenna. The field components are therefore essentially transverse and the angular distribution is independent of the radial distance where the measurements are made. In the far-field region, the total energy per unit area at a distance R is approximately proportional to $1/R^2$. Subsequent to defining the field-regions, the gain and efficiency of the antenna are discussed.

The gain (G) of the transmitting and receiving antenna is the ratio of power received through free-space at distance d, $P_r(d)$, to the power received by the antenna transmission lines, P_{in}, such that the equation for the antenna gain is given as

$$G = 10 \log_{10} \frac{P_r(d)}{P_{in}}. \tag{5.2}$$

The radiation intensity corresponding to the isotropically radiated power is equal to the power (in watts) accepted by the antenna divided by 4π. The effective aperture of the antenna is used to determine the directivity and hence the radiation patterns and intensity of the source.

The input impedance of a radiating source is defined as the impedance presented by an antenna at its terminals or the ratio of the voltage to current at a pair of terminals. The impedance of the antenna (Z_A in Ω) with no load connected is defined as

$$Z_A = R_A + jX_A \tag{5.3}$$

where R_A is the real impedance in Ω at the antenna terminals, and X_A is the complex antenna reactance also measured in Ω. The real impedance consists of two components, given by

$$R_A = R_r + R_L \tag{5.4}$$

where R_r is the radiation resistance of the antenna and R_L is the load loss resistance (both measured in Ω). Losses at the input terminals of the antenna and within the structure are taken into account to determine the efficiency of the antenna. These losses can occur in response to reflections caused by mismatches between the transmission line and the antenna. Conduction and dielectric losses (I^2R) also influence the efficiency of the antenna. The efficiency of the antenna can be determined by the real impedance and the load loss resistance such that its relationship determines the antenna efficiency (refer to Table 5.2).

To describe the amount of power that an antenna can absorb if a wave impinges on it, the effective area, or effective aperture, of the antenna is defined. This area is defined as a ratio of the available power at the receiving terminals of the receiving antenna, to the power flux density of the incident wave. In general, the maximum effective aperture (A_{em}) of any antenna is directly related to its maximum directivity (refer to Table 5.2).

To determine and visually represent the electric field (E) as a result of the transmitted power, the equation is derived to relate the electric field to the distance travelled by the signal, and the effective aperture of the antenna.

These parameters can be used to define and characterize the type of antenna and its radiating properties fully. The signal received by the antenna is a variable AC signal and it has to be converted to DC. A rectifier circuit is used to convert the signal.

5.5 Rectifier Circuits

Rectifier circuits in electronic applications convert AC to DC and include a device that only allows one-way flow of electrons: the semiconductor diode. The simplest kind of rectifier circuit is the half-wave rectifier; however, full-wave rectifiers and

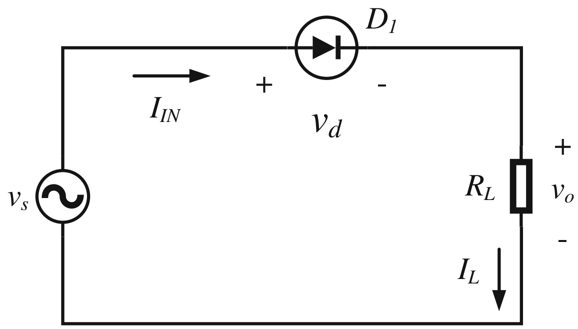


Fig. 5.4 Basic diode rectifier circuit

three-phase rectifiers are generally preferred (Mandal and Sarpeshkar 2007). Rectifier circuits may be single-phase or multi-phase (three being the most common number of phases). Most low-power rectifiers for domestic equipment are single-phase, but three-phase rectification is very important for industrial applications and for the transmission of energy as DC. From the early days of microelectronics, a simple p-n type junction diode was used to convert an alternating signal to a unidirectional signal. Based on the diode circuit configurations, several types of diode rectifiers exist. The basic half-wave diode rectifier circuit is shown in Fig. 5.4.

Consider the circuit in Fig. 5.4, the basic half-wave rectifier. The supply voltage to the circuit v_s is described by

$$v_s = V_m \sin \omega t \tag{5.5}$$

where V_m is the maximum DC amplitude, and ω is the operating frequency in rad. s^{-1}. The rectified voltage is obtained across the load resistor R_L. During the positive half-cycle of the input voltage, the diode is forward-biased and conducts current, therefore operating similar to a closed switch (Ashry et al. 2008). The source voltage v_s is available and reproduced at the load. During the negative half-cycle of the input voltage, the diode is reverse-biased and does not conduct, operating similarly to an open switch. The output voltage across the load is zero during this interval. The average voltage (DC rectified output voltage) across the load is given by

$$V_{ave} = \int_0^T v_o dt = \int_0^{T/2} V_m \sin(\omega t) dt + \int_{T/2}^{T} 0 dt \tag{5.6}$$

and from (5.6), the average output voltage is derived as

$$V_{ave} = \frac{2V_m}{\omega T}\left[\cos 0 - \cos\frac{\omega T}{2}\right] = \frac{2V_m}{2\pi}[\cos 0 - \cos \pi] = \frac{V_m}{\pi} \tag{5.7}$$

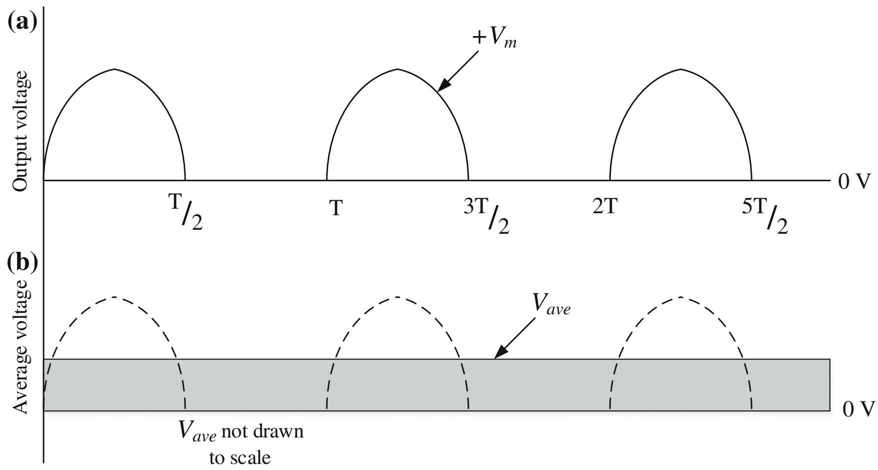


Fig. 5.5 **a** Output voltage across load (*top*) and **b** average output voltage plus ripple (*bottom*)

which is the maximum peak voltage of the input divided by π. The output voltage and average voltage are given in Fig. 5.5.

The output voltage can be seen as an average DC voltage plus a ripple component, which is evidently large in half-wave rectifiers, and should ideally be minimized. The peak current at the load can also be converted to average current I_L by

$$I_L = \frac{V_{ave}}{R_L} \tag{5.8}$$

and can in practice be measured by an ammeter. The maximum amount of reverse bias to which a diode will be exposed is called the peak-inverse-voltage (PIV), and for a half-wave rectifier this value is simply V_m. For the half-wave rectifier, during reverse bias, the load has zero voltage across its terminals and hence the PIV appears only across the diode. The PIV is an important characteristic of rectifiers, since it determines the maximum voltage that will appear across the diode, affecting the type and size of the diode. Various techniques exist to increase the average voltage produced at the load; the introduction of a capacitor in parallel with R_L is the most commonly used technique (Burasa et al. 2014). The capacitor smooths the output voltage as it discharges through R_L during the negative cycle of the supply voltage, effectively lowering the ripple voltage. If this technique is used, the average voltage at the load can be approximated as

$$V_{ave} \cong V_m \tag{5.9}$$

and in practice a very large capacitor at the output is used. A large variety of rectifier circuits is available, all with advantages and disadvantages in terms of

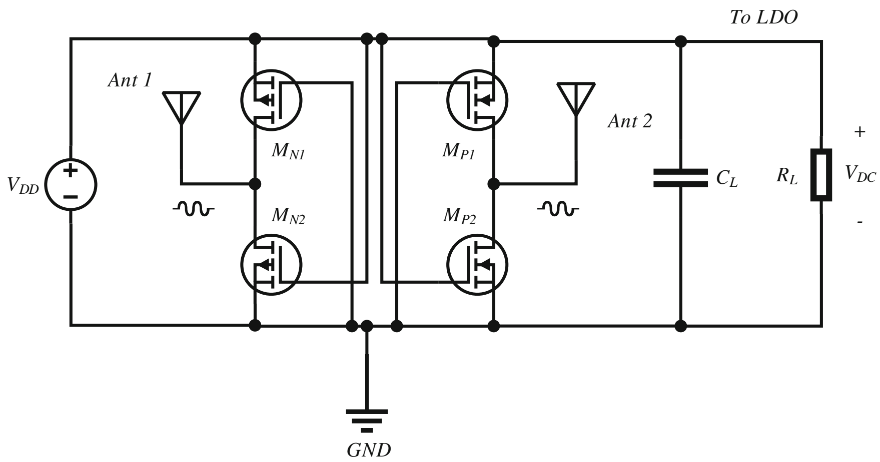


Fig. 5.6 Cross-coupled bridge configuration CMOS rectifier

linearity, power consumption, output voltage and ripple. However, the CMOS cross-coupled differential configuration is preferred for low-power, variable input voltage, RFID and the energy-harvesting circuit (Liu et al. 2011). The full bridge diode rectifier is popular because of its small size, light weight and relatively low PIV. Each unit stage has a cross-coupled differential CMOS configuration with a bridge structure. The operational principle of one stage is depicted in Fig. 5.6.

The four MOSFET diodes are arranged in series pairs with only two MOSFETs conducting current during each half cycle of the input RF signal from the antennas *Ant 1* and *Ant 2*: during the positive half cycle, M_{P1} and M_{N2} conduct in series while M_{N1} and M_{P2} are reverse-biased; during the negative half cycle, M_{N1} and M_{P2} conduct in series while M_{P1} and M_{N2} are reverse-biased.

The structure consists of a combination of two cross-connected gate structures that provide a complementary bridge rectifier. In the circuit, the PMOS transistor delivers the highest voltage to the load, whereas the NMOS transistor provides the lowest voltage. The transistor operates in the triode region, which behaves as a switch, thus having a smaller voltage drop compared to the MOS diode. The output voltage is given as

$$V_{DC} = 2V_{RF} - V_{switch} \tag{5.10}$$

where V_{RF} is the amplitude of the differential signals and V_{switch} is the loss due to switch resistance and reverse conduction. The maximum output voltage is limited to $2V_{RF}$. To increase the output voltage, N cells of the structure can be cascaded. The differentials signal of the first stage is directly connected to the RF source, whereas the preceding stages are capacitive-coupled to the RF source. This structure behaves as a charge pump voltage multiplier and the expected output voltage at the *N*th stage is

$$V_{DC} = N(2V_{RF} - V_{switch}). \tag{5.11}$$

However, in practice the output voltage is lower because the V_{switch} increases with the increase in the number of cells due to the increase in the body bias of the transistor. The loss is low in this circuit because there are no diode-wired MOSFETs and their associated threshold voltage problems. The circuit also has the advantage that the ripple frequency is doubled because there are effectively two voltage doublers, both supplying the output from out-of-phase clocks. The primary disadvantage of this circuit is that stray capacitances are much more significant than with the Dickson multiplier and account for the larger part of the losses in this circuit. To regulate this varying signal, a regulator circuit must be used. The low dropout (LDO) voltage regulator is typically used in RFID circuits.

5.6 Regulator Circuits (Low Dropout Regulator)

Dropout voltage is the input-to-output differential voltage at which a circuit ceases to regulate against further reductions in its input voltage, therefore as the amplitude of the input voltage approaches the output voltage. All linear voltage regulators reduce an input voltage to a constant output voltage across a load. These circuits are not capable of amplifying the voltage, like switch mode power supplies and charge pumps. Pulse-width modulated switching regulators typically operate between 50 kHz and 1 MHz and produce electromagnetic interference that can disrupt both analog and RF circuits. In contrast, the switching in LDOs occurs in the bandgap reference and the level is in the microvolts root-mean-square range over a defined bandwidth, a level that is considerably lower than a switching regulator. This is a major design advantage in noise-sensitive, variable input applications (Lee et al. 2012). LDO circuits also require fewer passive components in the circuit, leading to circuit miniaturization. In the past, many so-called LDO devices found their way onto the market, reducing the needed voltage difference between input and output from volts to millivolts, at the cost of regulation stability. The functionality of reducing the required voltage difference is especially advantageous in high-frequency RFID circuits, since the amplitude of the received signal varies greatly with time, environment, reading intervals and signal strength. Some low-supply current types of regulators use less than 1 µA of self-supply current. Because of this feature, LDO regulators can maintain supply current of the electronic devices as low as possible when these devices are in standby mode. Since these regulators can also provide the benefit of CMOS miniaturization technology, they offer great potential to mobile electronic devices that require a low profile and high precision. The key selection criteria for designing or choosing an LDO voltage regulator for a specific application are input voltage range, output voltage, fixed or

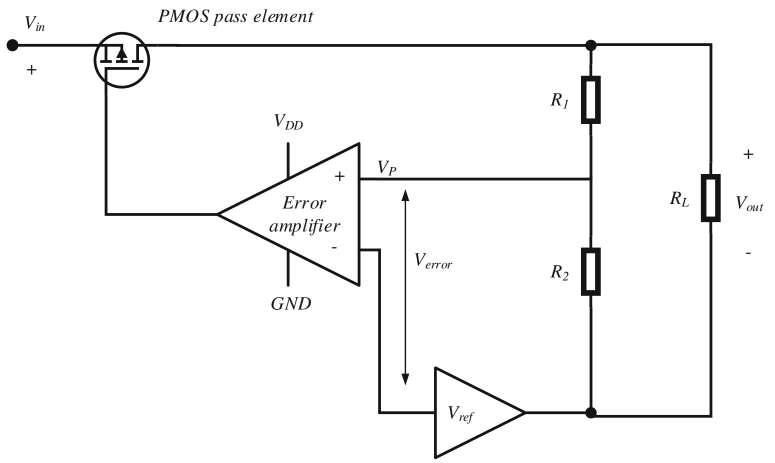


Fig. 5.7 Simplified schematic of LDO voltage regulator

adjustable output, output accuracy over line, load, and temperature, load current requirement, dropout voltage, power supply rejection ratio, output noise, quiescent current and shutdown current. Figure 5.7 depicts a basic NMOS LDO voltage regulator.

The general idea of linear voltage regulators (Fig. 5.7) is to adjust the resistance of a series pass element in such a way that the voltage across the load remains constant. The input voltage may be varying and the load itself may not be constant, as often devices show a large difference in supply current between standby and actual operation. In all these cases, the output voltage must remain constant. The output voltage of the LDO linear voltage regulator as shown in Fig. 5.7 can be determined by

$$V_{out} = V_{ref}\left(1 + \frac{R_1}{R_2}\right). \tag{5.12}$$

Generally, PMOS transistors are used in LDO circuits as opposed to NMOS or bipolar devices. The circuit could be operated in two regions, the linear region and saturation region. In the linear region, the series pass element acts like a series resistor and in the saturation region, the device becomes a voltage-controlled current source. LDO voltage regulators are usually operated in the saturation region to provide constant output functionality. LDO circuit topologies can be partitioned into four functional blocks: the reference circuit, the pass element, the sampling resistor and the error amplifier. The voltage reference is the starting point when designing regulators. This is usually a function of the bandgap type, since this kind of reference can also work at low supply voltages and provides enough accuracy and thermal stability to meet the less-stringent performance requirements of regulators. Bandgaps typically have an initial error of 0.5–1.0 % and a temperature

Table 5.3 LDO linear voltage regulator topology features

Parameter	NPN-Darlington	NPN	PNP	NMOS	PMOS
$I_{out,max}$	High	High	High	Medium	Medium
I_q	Medium	Medium	Large	Low	Low
$V_{dropout}$	$V_{sat} + 2V_{be}$	$V_{sat} + V_{be}$	$V_{ce(sat)}$	$V_{sat} + V_{gs}$	$V_{ds(sat)}$
Speed	Fast	Fast	Slow	Medium	Medium

coefficient of 25–50 ppm/°C. The function of the error amplifier is to sample a scaled-down value of the output and compare it to the reference voltage to adjust the output voltage by using the series pass element to the value required to drive the error signal as close as possible to zero. The pass element structure can consist of NPN, PNP, NMOS, PMOS, or NPN-Darlington configurations (Texas Instruments 1999). The bipolar devices can deliver the highest output currents where the MOS devices minimize quiescent current flow but offer limited drive performance with strong dependence on transistor aspect ratio and voltage-gate drive. Table 5.3 provides a summarized comparison of LDO linear voltage regulator topologies.

The topologies are depicted in Fig. 5.8.

As shown in Fig. 5.8 (a), the control circuit should be operated at a voltage higher than 0.6 V at the output pin, as a result of the voltage drop across the base-emitter of the transistor. Since the control circuit is operated by the input power source, a dropout voltage of 0.6 V is required. In (b) and (c) the transistor will turn on when the input voltage is lower than the base/gate voltage. For both these circuits, there is no limitation on the input power source voltage relative to the output voltage and the dropout voltage is small, since the circuit operates only if a base/gate voltage is applied. In (d) the control circuit must be 1.2 V higher, since there are two base-emitter voltage drops across the two transistors. These circuits are able to output large currents, since the base current of the load transistor is amplified by the pre-driver transistor.

The PNP bipolar transistor was traditionally used in LDO linear voltage regulators and the development thereof, primarily since it allowed a very low dropout voltage. These devices did, however, suffer from high quiescent current, high power requirements and low efficiency. NMOS devices followed, their main advantage being their low on-resistance, but the gate drive requirements limited their adoption. For low-power circuits, especially looking at RFID circuits and low-power nodes within the IoT framework, NMOS LDO regulators have been highly developed and their performance levels exceed those of most bipolar-based LDO circuits.

The efficiency of a LDO linear voltage regulator is limited by its quiescent current and the input versus output voltage, such that

$$Eff(\%) = \frac{V_{out} I_{out}}{\left(I_{out} + I_{quiescent}\right) V_{in}} \times 100 \quad (5.13)$$

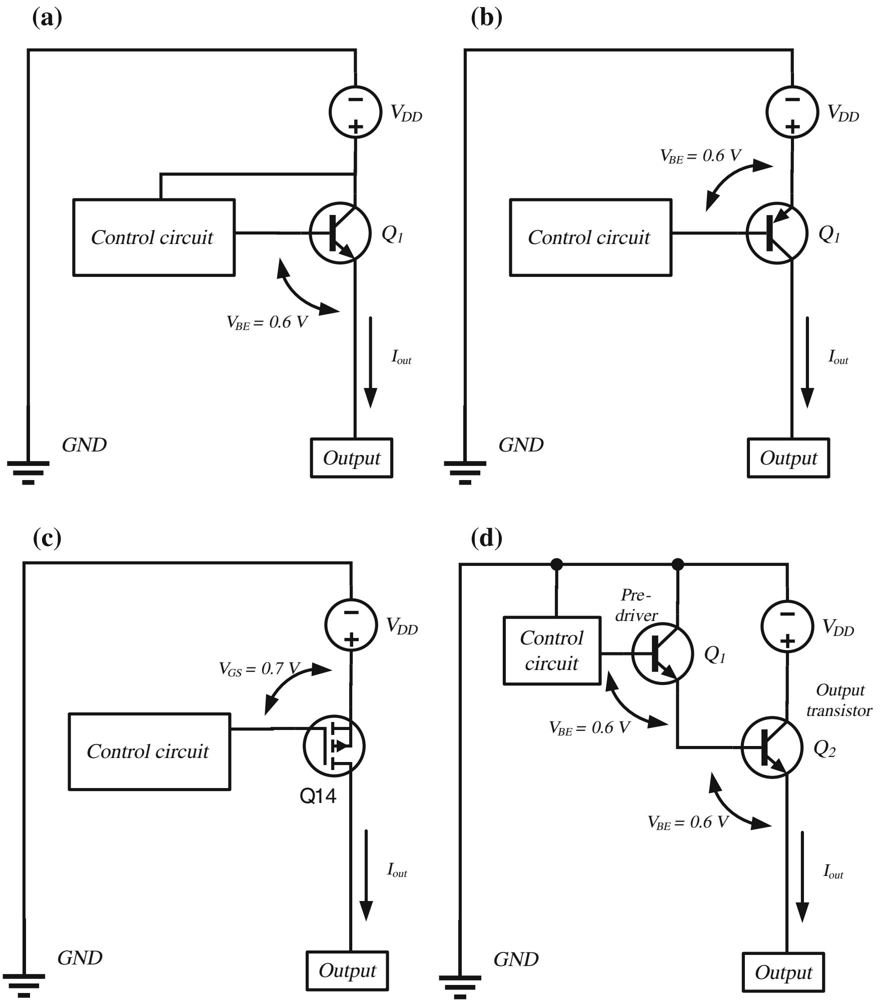


Fig. 5.8 LDO linear voltage regulator output driver topologies, **a** NPN emitter-follower output, **b** PNP transistor output, **c** PMOS transistor output, **d** NPN Darlington output

where V_{out} and I_{out} represent the output voltage and current across the load respectively, V_{in} is the input voltage to the LDO circuit from the rectified voltage obtained by the RFID device, and $I_{quiescient}$ is the current used in the standby state. Quiescent current consists of bias current such as bandgap reference, sampling resistor, and error amplifier currents and the gate drive current of the series pass element, which do not contribute to output power. For bipolar transistors, the quiescent current increases proportionally with the output current, because the series pass element is a current-driven device. In addition, in the dropout region the quiescent current can increase owing to the additional parasitic current path between the emitter and the base of the bipolar transistor, which is caused by a lower base voltage than that of the output voltage. For MOS transistors, the quiescent current has a near constant value with respect to the load current, since the device is voltage-driven. The only aspects that contribute to the quiescent current for MOS transistors are the biasing currents of the bandgap, sampling resistor and error amplifier. In applications where power consumption is critical, or where small bias current is needed in comparison with the output current, an LDO voltage regulator using MOS transistors is essential. The stability of LDO regulators remains one of the main drawbacks in their application. Techniques exist where using an output capacitor with as low equivalent series resistance (ESR) as possible improves the stability. With multilayer ceramic capacitors having high capacity and very low ESR, this technique is somewhat nullified, since the stability of an LDO requires a minimum ESR (Lee 1999). The transient response of an LDO regulator shows the maximum allowable output voltage variation for a load current step change and is a function of the output capacitance C_{out}, the bypass capacitor C_b, the maximum load current $I_{out,max}$, and the output capacitor ESR. The maximum transient voltage variation is given by

$$\Delta V_{max} = \frac{I_{out,max}}{C_{out} + C_b} \Delta t_1 + \Delta V_{ESR} \tag{5.14}$$

where Δt_1 corresponds to the closed-loop bandwidth and ΔV_{ESR} is the voltage variation resulting from the presence of the ESR of the output capacitor. Once the input signal is rectified and regulated, it can be amplified to a level as required by the load device.

5.7 Amplifier Circuits

To describe power amplification, first consider the general transmitter and receiver functional diagrams in a generic application. In the vast majority of applications, transmitters and receivers (transceivers) are a variation of superheterodyne radios, as depicted in Fig. 5.9.

The basic concept of operation of the components in Fig. 5.9 is described as follows: For the receiver in Fig. 5.8 (a), the signal that is received from the antenna is amplified in the RF stage. The output of the RF stage is used as one input of the mixer stage. A local oscillator (LO) is applied to the second input of the mixer. These signals combine and produce an output, referred to as the intermediate frequency (IF). The gain and bandwidth of the output are typical trade-offs in amplifier design and the product of these two properties are referred to as the gain-bandwidth product (GBP). The signal is then modulated, where the modulation bandwidth is typically much smaller compared to the RF carrier frequency. The signal is again fed to a mixer stage, where the signal is converted to baseband and demodulated to extract information from it. Modulation and demodulations schemes vary greatly

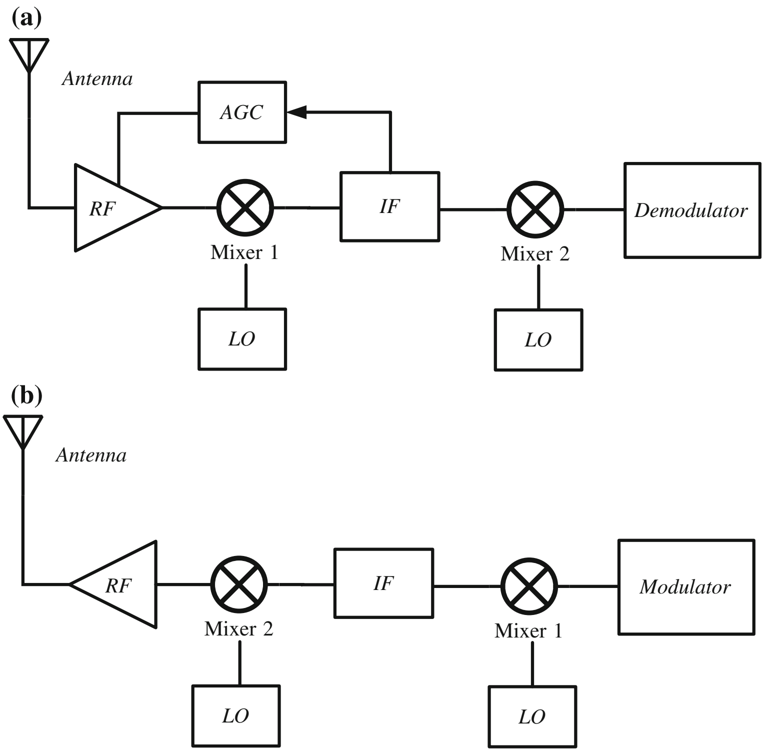


Fig. 5.9 Generalized superheterodyne **a** receiver and **b** transmitter

with application, where schemes such as amplitude modulation, frequency modulation and phase modulation such as quadrature amplitude modulation could be implemented. The transmitting functional block in Fig. 5.9 (b) operates similarly to the receiver; however, the operation is effectively a reverse of the receiver operation and the mixers convert the frequencies up towards the carrier instead of down to baseband. Analogue-to-digital conversion circuits have evolved enough to enable the modulation and demodulation of the signal in the digital domain, effectively removing the IF sampling stage and simplifying the mixer stages.

Power amplification (Berglund et al. 2006) is required on the transmission side of RF circuits to drive the antennas with enough power to relay the signal towards the receiver, where power and range are inversely proportional. Another trade-off in the amplification stage is the efficiency of the conversion and linearity of the signal. The limitation of supply voltage in low-power devices such as RFID applications constrains the design alternatives as well. A general representation of a power amplifier (PA) is depicted in Fig. 5.10.

From Fig. 5.10, the circuit voltage supply is V_{DD}, resulting in a DC current flow of I_{DC} through the main path of the circuit. To ensure a constant current through the driver, a RF choke (*RFC*) is placed between the driver and V_{DD}. An output filter is

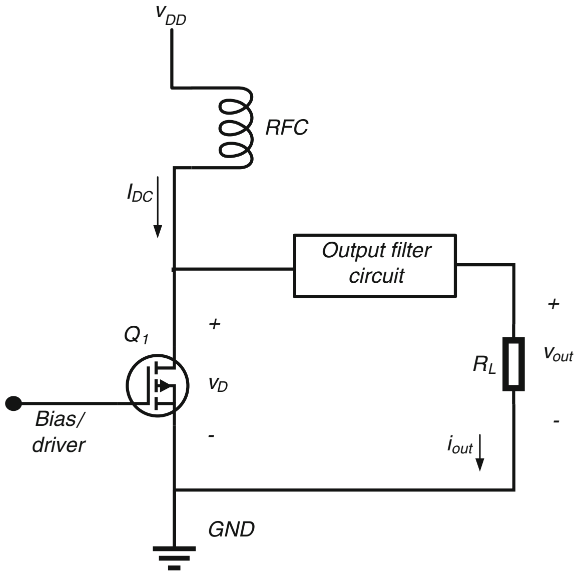


Fig. 5.10 Generalized model of a power amplifier

used for harmonic tuning, wave shaping or impedance matching. The driver circuit is typically a power transistor, and MOS, HBT, or BJT transistors can be used in this stage, each with its own advantages and disadvantages in terms of quiescent current, power output, linearity and output impedance. The power available at the load (R_L) is generally given by

$$P_{out} = \frac{V_{DD}^2}{2R_L} \tag{5.15}$$

and is proportional to the input supply (generally limited in low-power devices), and inversely proportional to the load impedance (ideally minimized through matching networks). The overall power consumption of the amplifier circuit should again be minimized in battery-operated, low-power, wireless devices such as WSNs and RFIDs, and the DC power consumption is given by

$$P_{DC} = V_{DD} I_{DC} \tag{5.16}$$

where I_{DC} is the DC component of the current waveform through the driver transistor. To quantify the performance of the amplifier, the efficiency η is equated as a ratio of the power delivered to the load, and the supplied input power, given as

$$\eta = \frac{P_{out}}{P_{DC}} \tag{5.17}$$

where higher efficiency is generally obtained by maximizing the output power and limiting the consumed power in the circuit through various techniques, such as choosing the type of amplifier. Another method of quantifying the performance of the amplifier, is through the power-added efficiency (PAE), rating the efficiency of a PA taking into account the gain of the amplifier, given by

$$\mathrm{PAE} = 100\left(\frac{P_{out} - P_{in}}{P_{DC}}\right) \tag{5.18}$$

where P_{out}/P_{in} is the gain (G) of the amplifier. Rewriting (5.18) as

$$\mathrm{PAE} = \eta\left(1 - \frac{1}{G}\right) \tag{5.19}$$

shows that the PAE increases with the gain of the amplifier, although the overall bandwidth of the device will be lower (GBP). For a very low gain amplifier, the PAE could become negative and is generally avoided in circuit design. Through conjugate matching of the input side of the PA, maximum power transfer to the circuit can be achieved; however, this technique does not apply to the output side of the PA. At the output side, power transfer can be maximized by using either the load-line approach or load-pull analysis to determine the optimal load, where the required impedance often differs from the standard 50 or 75 Ω load impedance.

A PA is classified based on its GBP, type of operation (linear or constant envelope), the nature of its voltage and current waveform, output power and its efficiency. Amplifier classes represent the amount of output signal, which varies within the amplifier circuit over one cycle of operation when excited by a sinusoidal input signal. The classification of amplifiers ranges from entirely linear operation (for use in high-fidelity signal amplification) with very low efficiency, to entirely non-linear (where a faithful signal reproduction is not critically important) operation, but with much higher efficiency, while others are a compromise between the two. Amplifier classes (Jiang et al. 2008) are mainly lumped into two basic groups. The first are the classically controlled conduction angle amplifiers forming the more common amplifier classes of A, B, AB and C, which are defined by the length of their conduction state over some portion of the output waveform, such that the output stage transistor operation lies somewhere between being fully on and fully off. The second set of amplifiers comprises the newer so-called switching amplifier classes of D, E, F, G, S, T, which use digital circuits and pulse width modulation to switch the signal constantly between fully on and fully off, driving the output hard into the transistor's saturation and cut-off regions. A summary of these classes is given in Table 5.4.

RFID and IoT applications require high efficiency power amplifiers to ensure the highest transfer of already limited input power. For this reason, recent RFID applications operating within the 13.56 MHz band (Wang et al. 2015) use Class-E amplifiers to realize high efficiency amplification of the information signal.

Table 5.4 Amplifier classes, conduction angle, and theoretical efficiencies

Amplifier class	Conduction angle	Efficiency
A	2π	50
B	π	78.5
AB	π < θ < 2π	50–78.5
C	θ < π	78.5–100
D	0	100
E	0	>90

Class-E amplifiers are classified as switching amplifiers and can exhibit efficiencies approaching 100 %. In Class-E, the transistor operates as an on/off switch (in saturation mode) and the load network shapes the voltage and current waveforms to prevent simultaneous high voltage and high current in the transistor; that minimizes power dissipation, especially during switching transitions. Figure 5.11 depicts a simplified schematic of a low-order single-ended Class-E PA.

As shown in Fig. 5.11, the output network of a Class-E power amplifier starts with a shunt capacitor C_1 that absorbs the output capacitance of the transistor. Current passes through the capacitor when the transistor channel is closed. If it is assumed that the *RFC* inductor L_1 has a high value, the output capacitance of the transistor is independent of the switching voltage, and the transistor is an ideal switch with zero resistance and zero switching time. The value of the optimum load resistance, which delivers the highest power output, can be calculated by

$$R_L = \frac{2}{\pi^2/4+1} \frac{(V_{CC} - V_{CEsat})^2}{P} \tag{5.20}$$

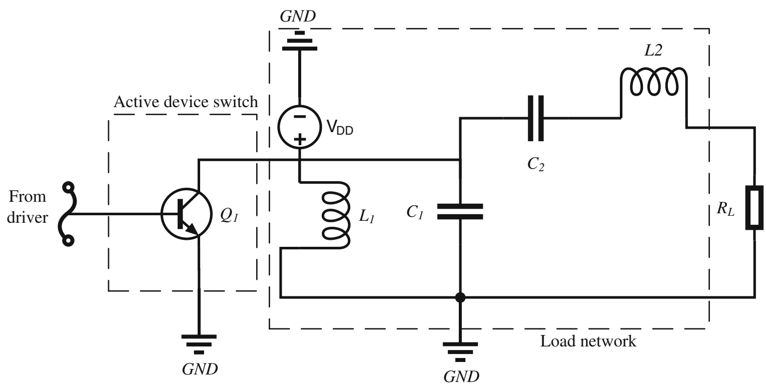


Fig. 5.11 Schematic of a low-order single-ended Class-E PA

where V_{CEsat} is the saturation voltage of the transistor, and P is the power in the circuit. To determine the value of L_2, the following equation can be used:

$$L_2 = \frac{Q_L R_L}{\omega} \tag{5.21}$$

where Q_L is the desired quality factor of the output resonant tank circuit, and ω is the operating frequency in rad.s^{-1} ($\omega = 2\pi f$). The shunt capacitance C_1 can be calculated using

$$C_1 = \frac{1}{\omega R_L(\pi^2/4+1)(\pi/2)} \tag{5.22}$$

and the resonant capacitance C_2 is found by

$$C_2 = \frac{1}{\omega^2 L_2}\left[1 + \frac{1.42}{Q_L - 2.08}\right]. \tag{5.23}$$

The capacitance C_1 includes the parasitic capacitance of the transistor Q_1 and therefore in some instances the parasitic capacitance is sufficient and C_1 can be omitted from the design. The efficiency of the Class-E amplifier is determined by

$$\eta = \frac{1 - \frac{(2\pi A)^2}{6} - \frac{V_{CEsat}}{V_{CC}}\left[1 + A - \frac{(2\pi A)^2}{6}\right]}{1 - \frac{(2\pi A)^2}{12}} \tag{5.24}$$

where

$$A = \left(1 + \frac{0.82}{Q_L}\right) f t_f \tag{5.25}$$

and t_f is the collector current fall time, ideally equal to zero. The Class-E amplifier can operate at arbitrarily low frequencies. Below about 3 MHz, one of the three switching-mode Class-D amplifier types might be preferred. Each can be as efficient as the Class-E, with about 1.6 times as much output power per transistor, but with the possible disadvantage that transistors must be used in pairs, therefore symmetrically matched on-chip, versus the single Class-E transistor. Class-E is preferable to Class-D at frequencies higher than 3 MHz because it is more efficient, the transistor input port is easier to drive, and Class-E has fewer detrimental effects from parasitic inductance in the output-port circuit. Low-order Class-E amplifiers are useful up to the frequency at which the achievable turn-off switching time is about 17 % of the RF period. In a Class-B amplifier, the turn-off transition time is 25 % of the period.

Silicon lateral double-diffused MOSFET (LDMOS) transistors have been the dominating technology for high-power RF amplifiers for a long time and could remain the leading technology in Class-AB applications at frequencies below 3 GHz. However, for higher frequencies and emerging switch-mode architectures, fundamental limitations, such as comparatively low f_t/f_{max} and high, lossy parasitic output capacitance, call for alternative technologies. A comparison of intrinsic material properties shows that high electron mobility transistors, which use AlGaN/GaN heterostructures (Chen et al. 2012), clearly stand out as the most promising of emerging technologies. The energy gap of GaN is three times that of silicon, resulting in reduced performance degradation at high temperatures. Similarly, breakdown at a sixfold electric field and two-and-a-half-fold carrier saturation velocity enables much greater power densities, resulting in the same output power capability at a much higher impedance level. To reap the full benefit, devices must be optimized for a particular application. Therefore, the performance of GaN transistors in future switch-mode architectures might depend on what manufacturers see as the main application for these devices. Compared with silicon LDMOS, GaN technology is still immature, hampered by basic manufacturability and reliability issues, and is far from being competitive in terms of cost. To ensure maximum power transfer to the load, impedance matching is required.

5.8 Matching Networks

To achieve maximum real-power to complex load impedances, the delivered source impedance should be equally matched to the complex conjugate of the load impedance (Freescale Semiconductors 2005). Two- and three-component matching networks offer easy and intuitive implementations for impedance matching. As circuit wavelengths are reducing and parasitic elements from lumped elements increasing, microstrip-line matching networks are increasing in popularity. Stub-line matching networks remove all lumped elements and cancel the reactive part of the presented impedance, depending on the length of the stub. In general RFID circuits, impedance matching is required between the PA and the antenna, as depicted in Fig. 5.12.

Matching networks are used not only to match complex source and load impedances; they also contribute to maximizing power-handling capabilities, linearizing the frequency response, and minimizing the noise added to the existing circuit. The impedance of the load (as in the case of transmitting antennas) can vary as a result of configuration, environmental characteristics and operating frequency. Ensuring optimal power flow under all of these varying conditions is crucial and is achieved with matching networks.

The simplicity of a two-component matching network allows for a unique solution to be found, but limits the capabilities of the network when designing not only for impedance matching (additional criteria could include bandwidth requirements). Forbidden regions exist for these networks and create conditions where matching to a 50 Ω source impedance is impossible. By including a third component, an additional degree of freedom is introduced to the circuit. Through

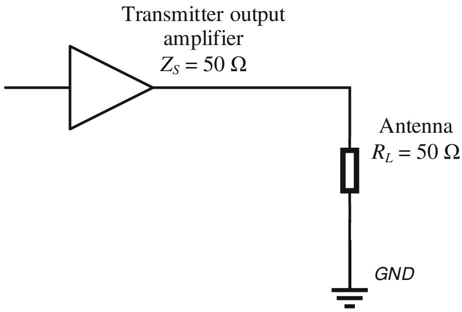


Fig. 5.12 Basic principle of impedance matching in RFID circuits

this, it is possible to also design for a loaded quality factor with narrower bandwidth requirements, parasitic effects, or harmonic rejection, which adds flexibility and additional design criteria. The number of possible solutions to match load and source impedances with these networks is not limited to one, making the networks difficult to tune over a range of frequencies. Traditional two- and three-component matching networks are given in Fig. 5.13.

Input matching not only achieves maximum power transfer, but also affects the noise figure (NF) of the amplifier. For a two-port amplifier, the minimum noise figure (NF_{min}) is achieved by minimizing the noise admittance towards its optimal value, $Y_{s,opt}$, and the noise resistance, R_n. For an arbitrary source admittance Y_s, the noise figure of the amplifier is given by

$$NF = NF_{min} + \frac{R_n}{G_s}\left|Y_s - Y_{s,opt}\right|^2 \tag{5.26}$$

where G_s is the real part of Y_s. From (5.26) it is evident that the NF can be minimized when $Y_s = Y_{s,opt}$ and R_n determines the sensitivity of the NF to deviations of Y_s from $Y_{s,opt}$. Generally, Y_s for noise matching differs from the optimum source admittance for a conjugate match.

An equivalent lumped circuit model of a RFID tag is given in Fig. 5.14.

From Fig. 5.14, the RFID on-chip complex impedance Z_c is given by

$$Z_c = R_c + jX_c \tag{5.27}$$

and the complex impedance of the antenna Z_a is given by

$$Z_a = R_a + jX_a \tag{5.28}$$

where R and X in (5.27) and (5.28) represent the real and imaginary terms of the impedances respectively. The voltage source represents an open-circuit RF voltage developed over the terminals of the receiving antenna, as would be the case during a RFID transponder event. Important considerations in low-power RFID-type

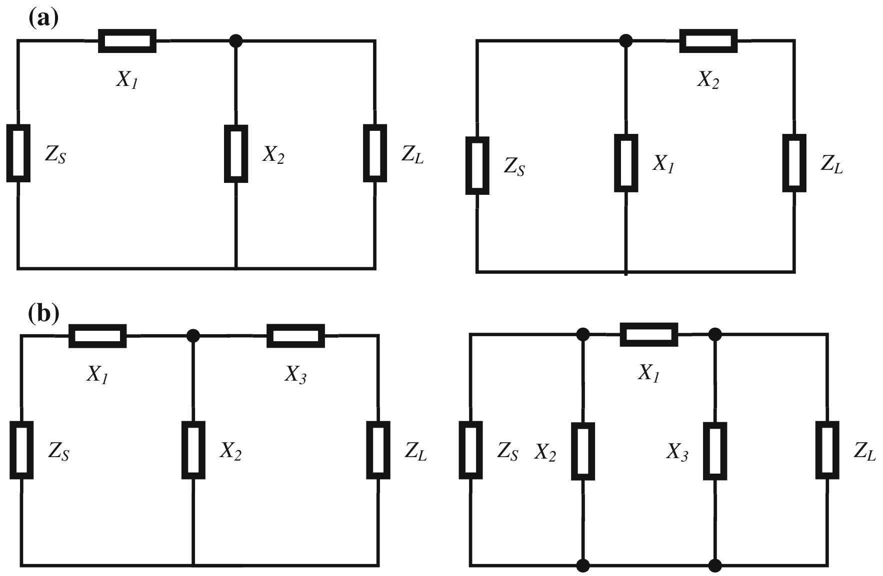
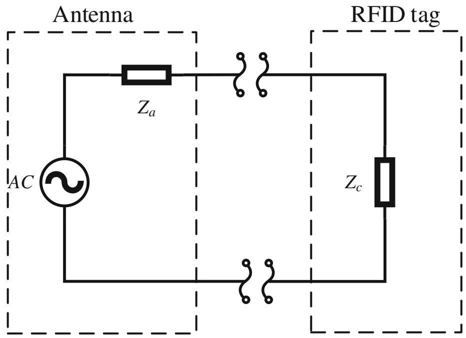


Fig. 5.13 **a** Two-component matching networks with passive component parallel to the load (*left*) and source (*right*). **b** Three-component matching network T-network topology (*left*) and π-network topology (*right*)

Fig. 5.14 Basic principle of impedance matching in RFID circuits



applications are that not only are Z_c and Z_a frequency-dependent, but Z_c could vary with the power absorbed by the on-chip circuitry under varying circumstances. The antenna is generally matched to the RFID tag at the minimum threshold power level required for the tag to respond. The amount of power that can be absorbed by the tag from the antenna, P_c, is given by

$$P_c = P_a \tau \tag{5.29}$$

where P_a is the power received by the antenna and hence the power that can be dissipated by the tag if matched to the antenna, therefore $Z_c = Z_a^*$, and τ is the power transmission coefficient given by

$$\tau = \frac{4R_c R_a}{|Z_c - Z_a|^2} \tag{5.30}$$

which directly characterizes the degree of impedance match between the tag and the antenna. The resonant factor of the RFID tag for the given frequency and absorbed power is given by

$$Q = \frac{X_c}{R_c}. \tag{5.31}$$

The read range of the transponder is the maximum distance at which the tag receives just enough power to turn on and scatter back. An additional limitation is the maximum distance at which the reader can detect the scattered signal. Typically, the reader's sensitivity is high enough so the read range is determined by the limitation on the tag, since the reader can be provided with constant and higher power supply. The power received by the RFID antenna can be calculated using the Friis free-space equation

$$P_a = P_t G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2 \tag{5.32}$$

where P_t is the power transmitted by the reader, G_t is the gain of the reader antenna, G_r is the gain of the receiver tag antenna, λ is the wavelength of operation, and d is the distance between the reader and the tag. If the minimum threshold power required to power the tag is given by P_{th}, the read range can be calculated by rearranging (5.26) such that

$$r = \frac{\lambda}{4\pi}\sqrt{\frac{P_t}{P_{th}} G_t G_r \tau} \tag{5.33}$$

where the peak of the range in the frequency domain is referred to as tag resonance, primarily determined by τ.

If component values become a limitation for on-chip matching networks, and parasitic effects increase the complexity of these designs to the extent that their use becomes impractical, microstrip-line matching can replace the use of lumped elements (assuming the wavelength is sufficiently small compared to circuit dimensions). The tuning capabilities of these types of networks (through capacitive tuning and component locations) ensure popularity of these circuits in prototyping. Single-stub matching networks eliminate the use of lumped elements, but placement at a variable distance from the load can be difficult to achieve in practice, confining its use to fixed networks, and matching is only done at one frequency.

Double-stub matching networks remove this variable distance problem. These networks are often replaced in implemented designs with more compact and broadband applications. In CMOS power amplifier design, short-circuited, LC, on-chip stub-matching networks are preferred because of their low-loss, good noise performance and attenuation of out-of-band components, and are generally used for input and output stage matching. The performance of matching networks is characterized by parameters such as the load reflection coefficient, nodal quality factor, bandwidth and the standing wave ratio.

5.9 Energy Storage

Two types of energy storage are used to apply the energy provided by harvesting devices in an efficient way and to be able to guarantee a minimum lifetime even if no energy is provided by the harvesting device. These are primary (single use) and secondary (multiple use) energy storage. Primary energy storage is not rechargeable, but guarantees a certain energy and thus lifetime, whereas secondary energy storage is able to be charged/discharged several times depending on the amount of energy provided by the energy-harvesting device.

More recent harvesting devices, operating in the UHF bands, are mostly based on primary storage devices, therefore their lifetime is limited but guaranteed for a certain period of time with regard to operating conditions such as operations per day, charging cycles and operating temperatures. If the lifetime is surpassed, the battery can be replaced or the RFID tag can be changed. Examples of primary storage devices are nuclear batteries, electrochemical cells and microcells. Nuclear batteries (even if the name sounds dangerous), are actually based on the piezo-generator principle and fall under the beta voltaic principle. These cells operate similarly to solar cells, the difference being the source of energy. Solar cells depend on the sun for radiation energy, whereas piezo-generator cells are coated with radioactive material fixed to a MEMS cantilever to excite particles and generate energy from mechanical movement. Electrochemical cells use chemical compounds such as zinc-carbon, alkaline-manganese dioxide, or lithium to generate energy. Each cell contains a positive and negative terminal, and an electrolyte to allow ions to move between the electrodes and generate a flow of electricity. Microcells operate in a similar way to electrochemical cells, the main difference being the materials used inside the cell and the size (and therefore total capacity) of these cells.

Secondary storage devices can be recharged, and depending on the application and available resources, the harvested energy used to recharge the cells can be taken from the environment, such as radiation or wind energy. The two types of secondary storage devices are electrochemical capacitors and electrochemical cells. Electrochemical capacitors are basically capacitors with very high energy density. A high energy density is achieved by applying two electrochemical layers and high conductivity materials. The main difference between electrochemical capacitors and rechargeable batteries is the practically infinite number of charge/discharge cycles of the capacitors without noticeable degradation of the device. Electrochemical cells operate similarly to the primary cells; however, different electrolyte materials are used, which enable these cells to be recharged. These electrolytes include lead-acid, nickel-cadmium, nickel metal hydride, lithium ion and lithium ion polymer variants.

Relatively new devices are solid state batteries, which are rechargeable energy storage devices manufactured on silicon wafers using semiconductor fabrication processes. They can be packaged as stand-alone components or co-packaged with other integrated circuits. A solid state battery has both solid electrodes and solid electrolyte, resulting in superior transfer of ions between electrodes. The high ionic conductivity minimizes the internal resistance of the battery and therefore enables the high energy density, while the high resistance reduces the self-discharge rate. Materials used for solid state batteries include Ag_4RbI_5, LiI/Al_2O_3 mixtures and $NaAl_{11}O_{17}$ compounds.

The capacity and discharge rate t of a battery are related, as shown by

$$t = H\left(\frac{C}{IH}\right)^k \tag{5.34}$$

where H is the time it takes for the battery to discharge, given the load conditions (usually practically determined) in hours, C is the rated capacity of the battery in Ampere-hours, I is the discharge current drawn by the connected load, and k is Peukert's constant, a dimensionless quantity. The value of Peukert's constant for batteries is seldom supplied by the manufacturer and can be determined by loading the battery and forcing a full discharge. The constant will, however, degrade with time, as will the battery performance. Some ranges for Peukert's constant are as follows:

- between 1.05 and 1.15 for absorbed glass mat batteries,
- between 1.1 and 1.25 for gel batteries, and
- between 1.2 and 1.6 for flooded batteries.

Battery packs combined with solar-powered energy harvesting have become more popular and affordable in recent years. Tesla® for example offers home-based battery packs that can store 10 kWh of power harvested from solar energy at US $3,500. Advancements in battery technology which include using blends of nickel, manganese and cobalt oxide in the cathode and stacks of graphite in the anode have contributed to the practicality and lifetime of these systems. This technology and its global adoption is relatively slow, especially in developing countries primarily due to the largest initial cost associated with installation and maintenance cost over its lifetime, but many households in developed countries are embracing renewable sources to lessen their dependence on power supplied by utilities.

5.10 Conclusion

There are various techniques and devices that can harvest (or scavenge) energy from the ambient environment and apart from installation cost and maintenance cost (which is unfortunately still relatively high) this energy can be considered free. The advantages of renewable energy harvesting are different for developed and developing countries, with some overlapping in the interest of such devices. In developed countries, people opt for energy harvesting devices which generates relatively small amounts of energy for commercial products such as wearables and smart devices. Military applications such as energy-harvesting boots are being researched by large military corporations such as Lockheed Martin® to supply soldiers with alternative energy if required in remote areas. Industry and smart cities are also looking at energy harvesting to distribute smart systems realized with technologies such as WSNs and RFID throughout large areas in urban and rural environments. Energy-harvesting provides a means to power small devices or nodes which form part of these larger, information gathering and distributing, systems. In developing countries, the adoption of energy-harvesting as a sustainable technique to power electronic devices are slow compared to developed countries and their application often differ due to different requirements and needs in these communities. These communities are very reliant on power utilities, which are often not reliable and a move towards sustainable and renewable energy through energy harvesting is encouraged. In many developing countries, especially in Africa due to the large open areas and low population density on average (resulting from the sheer size of the continent), investors are flocking to take advantage of the benefits that the ambient environment provides.

This chapter discusses the fundamental properties of energy-harvesting with focus on RFID systems. Sub-systems of RFID systems include various frequently-used electronic circuits which require efficient designs to limit losses in these systems since the amount of energy harvested from the ambient environment is generally small. Sub-system considerations of RFID devices include effective antenna design and placement, rectifiers to convert AC signals to DC, regulators to stabilize the output voltage during varying input conditions, amplifiers to increase the usable signal, matching networks to match the output of these circuits with subsequent devices and applications as well as devices and circuitry to store the gathered energy such as batteries.

References

Ashry, A., Sharaf, K., Ibrahim, M. (2008). A Simple and Accurate Model for RFID Rectifier. *IEEE Systems Journal*. 2(4), 520-524, December 2008.

Berglund, B., Johansson, J., Lejon, T. (2006). High Efficiency Power Amplifiers. *Ericsson Review*, 3, 92-96.

Burasa, P., Constantin, N. G., Wu, K. (2014). High-Efficiency Wideband Rectifier for Single-Chip Batteryless Active Millimeter-Wave Identification (MMID) Tag in 65-nm Bulk CMOS

Technology. *IEEE Transactions on Microwave Theory and Techniques*. 62(4), 1005-1011, April 2014.

Chen, W., Chinga, R. A., Yoshida, S., Lin, J., Chen, C., Lo, W. (2012). A 25.6 W 13.56 MHz Wireless Power Transfer System with a 94 % Efficiency GaN Class-E Power Amplifier. *2012 IEEE MTT-S International Microwave Symposium Digest*. Montreal, 17-22 June 2012, 1-3.

Freescale Semiconductor. Impedance Matching Networks Applied to RF Power Transistors. *Application Note*. AN721, 2005.

Jiang, F., Guo, D., Cheng, L. L. (2008). Analysis and Design of Power Generators on Passive RFID Transponders. *Progress in Electromagnetics Research Symposium*. Hangzhou, 24-28 March 2008, 1357-1362.

Jinpeng, S., Bo, W., Shan, L., Xin'an, W., Zhengkun, R., Shoucheng, L. (2013). A Passive UHF RFID tag with a Dynamic-Vth-Cancellation Rectifier. *Journal of Semiconductors*. 34(9), 095005-1-095005-5, September 2013.

Lee, B. S. (1999). Understanding the Terms and Definitions of LDO Voltage Regulators. *Application Report – Texas Instruments*. SLVA079, October 1999.

Lee, M. C., Hu, C. C., Lin, Z. W. (2012). Implementation of Low Dropout Regulator with Low Bandgap Reference Voltage Circuit for RFID Tag Applications. *2012 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference*. New Taipei City, 40-43, 23-27 July 2012.

Liu, D. S., Li, F. B., Zou, X. C., Liu, Y., Hui, X. M., Tao, X. F. (2011). New Analysis and Design of a RF Rectifier for RFID and Implantable Devices. *Sensors*. 11, 6494-6508.

Ma, Q., Haider, M. R., Massoud, Y. (2011). A Low-Loss Rectifier Unit for Inductive-Powering of Biomedical Implants. *2011 IEEE/IFIP 19th International Conference on VLSI and System-on-Chip*. 86-89.

Mandal, S., Sarpeshkar, R. (2007). Low-Power CMOS Rectifier Design for RFID Applications. *IEEE Transactions on Circuits and Systems – I: Regular Papers*. 54(6), 1177-1188, June 2007.

Safarian, Z., Hashemi, H. (2014). Wirelessly Powered Passive Systems with Dynamic Energy Storage Mechanism. *IEEE Transactions on Microwave Theory and Techniques* 62(4). 1012-1021, April 2004.

Texas Instruments. (1999). Fundamental Theory of PMOS Low-Dropout Voltage Regulators. *Application Report*. SLVA068, April 1999.

Wang, D., Hu, J., Tan, H. Z. (2015). A Highly Stable and Reliable 13.56-MHz RFID Tag IC for Contactless Payment. *IEEE Transactions on Industrial Electronics*. 62(1), 545-554, January 2015.

Chapter 6
Geographic Information Systems and Remote Sensing

6.1 Introduction

A contributing factor to create a smart and sustainable city is the ability to map its assets, changes in environmental conditions, and location information with a geographic information system. A GIS enables and facilitates the electronic management and visualization of spatial data. Its purpose is to gather, model, store, manipulate, analyze, recovery, and present geometric data to enable users to easily understand and analyze geographic patterns. A GIS can use any information that includes location data and the type of data is only limited by our imagination and technology restrictions. Ideally, a city would require a single, central GIS system able to share data between departments, for example between forest fire monitoring and local fire stations. To further encourage city-wide cooperation, the cost of GIS systems could be shared between outside organizations such as utilities and phone service providers. Many GIS installations implement forms of open-encoding and interface standards to allow systems to communicate to other GIS installations and to a variety of mobile devices, emergency response systems, smart grids, sensor webs, or smart vehicles, only to name a few. GIS technology specifically allows for different data types to be overlaid on a single map and this is where these systems are considered extremely powerful and useful. Examples of practical GIS contributions include mapping crime data and sharing it with safety authorities such as the local police, monitoring pipes and pumps to actively manage the water infrastructure, or environmental impact assessments during infrastructure expansion. A GIS could potentially present a map of sites that generate pollution overlaid on a map that presents natural bodies that are susceptible to pollution, highlighting at-risk areas. Further examples of applications are explored in this chapter, supported by an overview of how these systems use geospatial data to present its users with visual maps. This chapter explores GIS, its subsystems, and its applications, and highlights various GIS projects already in place in Africa to emphasize the practicality and cost-effectiveness of such a system even in developing countries.

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_6

6.2 Spatial Technologies

Administrations of large cities, mega cities in particular, are often confronted with a key logistic and more serious problems such as informal settlements, traffic management, natural hazards, unclear responsibilities and mandates, uncoordinated planning, drinking and waste water management, provision of electrical power, visual, air, and water pollution (Kelly 2008). To manage such problems adequately, urban governance needs comprehensive, reliable and easy accessible spatial data in an established spatial data infrastructure (SDI). Local government can benefit from reliable and current spatial information during its own planning and management. Spatial data can contribute to the analysis of situations and risks by increasing efficiency and effectiveness and reducing uncertainty. The value of the information and the effectiveness of the decision-making and planning processes are related to the quality of the information and the manner in which it is presented. Key components here are data access, management, integration, analysis, standards and communication. An information management system should comprise all of these aspects, and operate as a closed loop continuous cycle. The key principles as a general framework for an information management system are depicted in Fig. 6.1.

As seen in Fig. 6.1, the general framework for an information management system consists of key concepts that operate within a cycle, and as one key component fails, all others may suffer its consequences. Starting arbitrarily in Fig. 6.1 from the top, an important aspect to consider is the hardware that is employed city-wide to gather information about the environments. The integrity of the hardware, as well as its installation and management, also determines the integrity of the

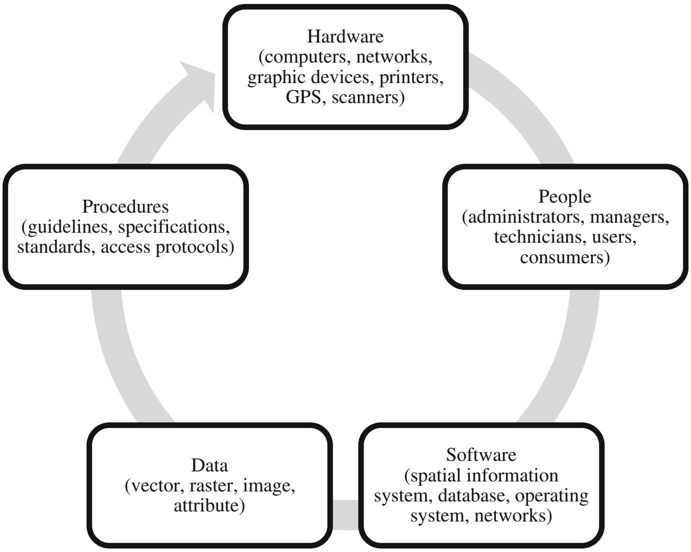


Fig. 6.1 General framework for an information management system

data. Hardware investments are sometimes costly and limited by technology trends, and could in some cases be somewhat under-estimated. Investing in state-of-the-art technology usually leads to better performance and longer lifetime in the long run. Procuring the hardware would not have much effect if there were no skilled persons to install, manage, and maintain the components. The general framework in Fig. 6.1 therefore suggests that high-quality investments in the workforce will be tasked to operate the system, which is true for any system, and in particular spatial information management systems through GIS. Training of personnel, quality contact-time in the field, and a high level of technical skills are required to ensure a successful endeavor. Spatial information gathering and GIS are equally reliant on software as it is on hardware. Quality instruments may collect valuable data, but if its interpretation and presentation is inadequate, the data is not of much value. Spatial information systems need to have enough resources to store data, generally large amounts of data are recorded in relatively short intervals, and storage capacity must be maintained. Operating systems should be trustworthy with regard to up-time and network communications equally so. Many GIS implementations are based on open-source standards to enable and encourage deeper control and also collaboration between institutes. If the data is presented, either in vector, raster, or image formats, proper procedures, specifications, standards, and access protocols must be well-documented and firmly in place to ensure data reproducibility. It is frequently required to expand on systems, or upgrade current technologies, and backward compatibility should always be maintained. Documented and approved procedures must be in place at all times to reduce the time to setup these systems, as GIS are generally large networks of devices monitoring large areas to increase its usefulness to the local population. Such an information management system forms the baseline for a well thought out and trustworthy GIS. GIS basically refers to any system capable of gathering and displaying applicable data, as briefly outlined below.

6.3 Geographic Information System

GIS can be traced back to as early as 1854 when cholera hit the city of London, England. British physician John Snow began mapping outbreak locations, roads, property boundaries, and water lines during this period of the outbreak. He noticed that cholera cases were commonly found along the water line. This could be considered the first time that data was gathered and analyzed, and visually displayed on a geographical map to realize a trend of a present health risk, in some ways, a classical GIS. GIS has evolved as a computer-based (to handle vast amounts of data without requiring a person to perform these sometimes tedious calculations) tool that analyzes, stores, manipulates and visualizes geographic information on a map. GIS requires skilled individuals from various disciplines to effectively display data in an organized and easy to use way, these disciplines include cartographers, database managers, programmers, remote sensing analysts, spatial analysts, and land surveyors, some already outlined and discussed in Fig. 6.1. GIS is used in many professions, including agriculture, archaeology, architecture, business, education, engineering, environmental studies, consumer science, forestry, recreation, media, law, medicine, military sciences, public administration, public policy, real estate, social work, transportation, and water services. These systems are not only reserved for wealthy developed nations, and could benefit a larger audience globally, as it already has been doing for several years. Obtaining the goal of sustainable development within Africa and other developing countries' diverse communities requires that analysts and decision makers understand the characteristics of resource use as well as human conditions. Several GIS initiatives that have been implemented in Africa alone, where resources are generally limited, are highlighted in the following section. Off-shore public and private investments generally drive these initiatives as many developing countries have great potential for these investors, primarily because these countries have large amounts of natural resources.

6.4 GIS Projects in Africa

Many GIS projects have been active in Africa for some time. This section aims to highlight only a few to give an indication of the direction GIS projects are taking in Africa, considering its natural resources, wildlife, and trends toward urbanization. As with many global initiatives in developed countries, these projects also include examples of drought repercussions on wildlife, dam influences on urban and agricultural economies, burn clearing strategies on flora and fauna, and even the spread of infectious diseases. GIS can be used for researching the impact of population growth on a township's economic development, a school district's growth by grade level, or a nation's gross national product. As GIS has the ability to analyze the inherent overlapping spatial nature of sustainable development issues, the World Summit on Sustainable Development (WSSD) includes GIS as a key element in its plan of action for aiding Africa to design and implement more sustainable management strategies. The Geographic Information for Sustainable Development (GISD) prospectus states that one of the biggest challenges of African nations is to lead their economies to a sustainable developmental path without depleting its natural resource capital. This requires, among other things, proper management of natural resources, regulation and monitoring of air and water pollution, and managing agricultural practices such as the use of pesticides.

6.4.1 Zambia

The Ministry of Environment and Natural Resources is an initiative of the government of the Republic of Zambia for the purpose of analyzing existing natural resources and planning management strategies for future development. Zambia's Environmental Support Programme has developed the National Environmental Action Plan, which currently forms the policy framework for environmental intervention and management.

6.4.2 Uganda

Between 1995 and 2005, researchers in Uganda carried out a distribution census of mountain gorillas and other species in the Virunga habitat. GIS was used to map the spatial distribution of the different species and humans in the area. GIS showed that gorillas tended to avoid areas of high human impact but could tolerate low levels of human impact and gave insight into land-use and vegetation data for the Virunga region.

6.4.3 Kenya

Save the Elephants has helped the Kenya Wildlife Service conduct regular aerial and ground surveys of Kenya's elephant population. GPS collars provide insight into the elephant mind. Using GIS and GPS to track elephant movement patterns provide information required to decide on the best management options. The Laikipia Predator Project tracks lions, hyenas, wild dogs, and other large African predators outside protected areas using radio tracking by air and through satellite. GIS correlates predator movements and ecology resulting from alternative land uses that range from intensive agriculture to traditional pastoralism and commercial ranching. GIS indicated that lions avoid densely settled areas, confining themselves to commercial ranches that have very low human density.

6.4.4 Rwanda

Rwandan researchers used GIS to identify areas of conservation importance. From the data they collected, they were able to calculate species richness, diversity, and the number of Albertine Rift endemic species for each site. The rankings of sites helped to provide an overall zoning plan for the forest.

The Cheetah Conservation Fund in central Namibia is amidst the largest population of wild cheetahs remaining in the world. It strives to reduce conflicts between cheetahs and livestock ranchers through the development and implementation of sound livestock management practices. The agency uses its GIS program in mapping its cheetah study area including the habitat mapping of ranches in the vicinity.

6.4.5 South Africa

In the early 1990s, the South African government began formulating a long-term plan aimed at eradicating invading alien plant life. Today the Working for Water program employs approximately 20,000 people in a 20-year initiative to clear more than 10 million hectares of land. It is anticipated that this effort will release four billion cubic meters of water per year, or more than 7 % of the country's entire water supply, which is currently consumed by foreign plants. The project has also had a powerful socioeconomic impact in the country because the program is aimed at providing employment to the most marginalized in South African society, being women, youths, the disabled, and the single head of households.

6.4.6 Botswana

Time series modeling is also a useful GIS tool for water resource analysis. Paul Sheller of Botswana's Kgeikani Kwani project is using ArcInfo for time series modeling to analyze the effects of the severe droughts that plague Botswana. The model can also be used to determine the effects from dredging the Okavango Delta and predict the effects of regulated releases of water from a series of dams on the lower Boro River. GIS indicated the environmental impact these actions would have on the nature of the vegetation and wildlife including rare and endangered species that will be displaced from the area as their habitat is lost. GIS shows the channels that are most impacted and highlights the at-risk areas that need project engineers' attention.

As seen from these examples, the reach and influence of GIS can help not only the human population, but also animals. Animals are mostly endangered and put at risk due to human activities, and it is therefore our responsibility to work towards maintaining a balance in nature. To implement GIS to the benefit of humans or animals (or plants) is not necessarily an easy task. There are many underlying factors that need to be considered, be it technical, logistical, financial, or its management. The following section aims to provide some technical background on specifically GIS as a spatial information management system, with focus on computational geometry as a sub-discipline to program software to interpret results. Also presented in the following section are the types of geographical data and maps, the cardinal importance of a positioning service for location data, and a brief summary of remote sensing techniques.

6.5 Computational Geometry

Computational geometry is concerned with the computational complexity of geometric problems within the framework of analysis of algorithms. Computational geometry involves the study of algorithms for solving geometric problems and as in most cases, these algorithms are implemented on a computer due to the large number of calculations required. The emphasis is more on discrete and combinatorial geometry optimization through techniques such as the traveling salesman problem and the minimum spanning tree problem. In computational geometry, the focus is more on the discrete and combinatorial nature of geometric problems as opposed to continuous issues, meaning that most problems are specific to conditions and is not repeated multiple times (take for example the weather on a day and bearing in mind that the exact temperature, wind speed, humidity, ultraviolet radiation, dew point, etc., will rarely be precisely replicated the following day). GIS calculations deal more with straight or flat objects (lines, line segments, polygons) or simple curved objects such as circles, as opposed to a high degree algebraic curves generally reserved in complex three-dimensional modeling. Geometric algorithms are useful in data correction after data acquisition and input, data retrieval through queries, data analysis for map overlay and geo-statistics, and data visualization for maps and in some cases animations. There are several well-known computational geometry algorithms for geometric problems. This book does not aim to describe each of these techniques in detail, since a comprehensive outline and detailed descriptions can be found in (De Berg et al. 2008).

The most often used algorithms in GIS include convex hull problems, line segment intersection, Voronoi diagram computations (Charya and Gavrilova 2008), Delaunay triangulation, Minkowski addition, and rectangular range search. Additional methods apparent in some GIS calculations include Euclidean shortest path, polygon triangulation, mesh generation, and Boolean operations on polygons. Each of these problems can be categorized based on its computational time to solve a given problem. The basic principles of the commonly-used problems in GIS are given in Fig. 6.2 (De Berg et al. 2008).

In geometric query problems the inputs consist of the spatial information and the object query. These problems can become complex compared to the static problems given in Fig. 6.2, not only due to the added dimension, but also due to the nature of odd-shaped areas and structures generally found in maps and GIS. Four of the fundamental computational geometry problems in GIS are range searching, point location, nearest neighbor, and ray tracing. Additionally, dynamic problems allow incremental modification to the input data by the addition or deletion of geometric elements. From Fig. 6.2, the convex hull problem in (a) determines if a point in the Euclidean plane is convex, therefore if it contains the line segment connecting each pair of its points. This technique is generally applied in creating accessibility maps. Line segment intersection in (b), as the name suggests, identifies the lines within the Euclidean plane that intersect, useful in counting road intersections and residential blocks in GIS. A Voronoi diagram as in (c) partitions a plane into regions based on the distance between points in a subset of the plane. Lines are drawn equidistant between two points and perpendicular to the line connecting them forming a familiar honeycomb shape. These diagrams have several applications in GIS, some of them include finding nearest facilities such as hospitals or gas stations, spatial

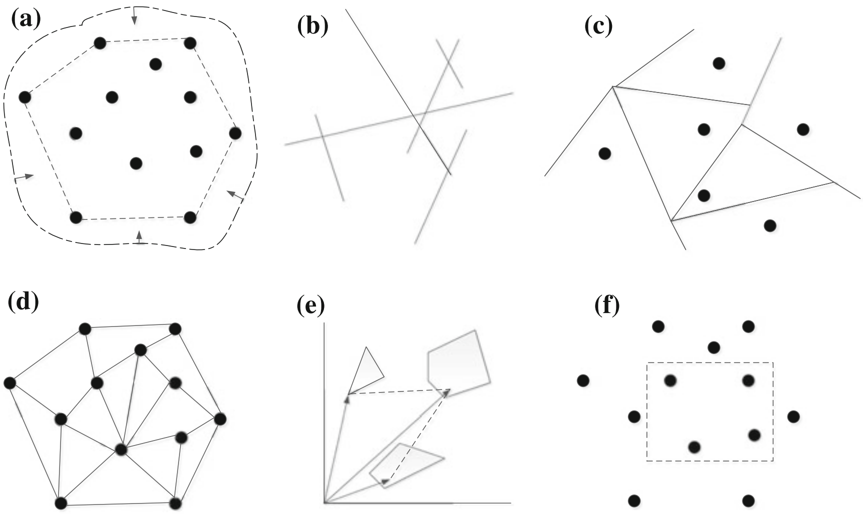


Fig. 6.2 Commonly found problems in GIS computational geometry. **a** convex hull, **b** line segment intersection, **c** Voronoi diagram, **d** Delaunay triangulation, **e** Minkowski addition, and **f** rectangular range search

interpolation, finding the largest empty circle used in sensor network distribution coverage, or simulating the growth of trees in forests to determine its optimal spatial distribution. Related to the Voronoi diagram is the Delaunay triangulation method in (d), which uses circumcircles and points in a subset to ensure no other points exist within the circumcircle. This method applies to many forms of geoscientific analysis that seek to collect data about spatial objects and domains such as aquifers, ocean currents, or atmospheric weather fronts which fill or enclose a three dimensional space. Minkowski addition, as shown in (e) adds two sets of vectors in a Euclidean plane and is often used in path planning for autonomous vehicles. Finally, rectangular range search techniques fit a set of data to a set of points and determines which points are located within a given region, as depicted in (f). In GIS this technique is useful for determining population densities in rural and urban areas.

These techniques are used to compute essential information from gathered data in GIS and are ideally presented on geographic maps along with their spatial information. This reduces the complexity of interpreting the data from spreadsheets or databases. A geographic map does have pre-defined characteristics such as scale and transformations, although a large variety of map types exist to display geographical data conveniently. The following section discusses several key points of geographic maps.

6.6 Geographic Maps

Maps are a fundamental part of GIS portraying collections of spatial geographical information as thematic (themed/relevant) layers and the art of map construction is called cartography. Types of maps range from topographic maps that show a variety of structures such as roads, land-use classification, elevation, rivers, political boundaries, and the identification of buildings. Specialized maps such as weather maps indicating low and high pressure systems which can become complex since the weather information and the map type must be precisely layered to convey accurate predictions.

GIS maps are used for communication and understanding large amounts of data in an organized way such as finding patterns, deriving new information using analysis, getting applicable status reports, compiling geographic information, communicating ideas, concepts, plans, and designs, and sharing geographic knowledge openly. Information conveyed on a map must be carefully transferred to avoid loss of information. The loss of information is however inevitable, as can be expected for example when transferring from a topographic map that abstracts three dimensional structures at a reduced scale to a two-dimensional map on a plane of paper. A technique to ensure accurate representations of the information through layer mapping is called translation or projection.

Map projection is required by GIS since the shape of the earth is roughly described as an ellipsoid, or more specifically a geoid since it has an uneven surface. The rotation of the earth causes the region near the equator to protrude outward which results in the shape that is slightly departed from a perfect sphere. The angular motion caused by the spinning of the earth additionally forces the polar regions to become slightly flattened. To represent information based on spatial location on a two-dimensional map on paper or computer screen, cartographers have developed a number of standardized transformation processes to minimize the distortion from such an operation. Different transformation processes focus on specific areas that require the most accuracy for a given application, such as distance, area, or bearing. Examples of projection standards include Mercator map projection, Gall-Peters projection, Miller cylindrical projection, Robinson projection, and Mollweide projection.

Since maps generally cannot be drawn to the scale of the area it represents, maps must be drawn on scales much smaller than the actual surface, this is where the loss of information occur. The scale represents a ratio between the distance of two points on the map and the actual distance it represents. Maps use representative fractions or ratios (e.g. 1/5,000,000 being a fraction and 1:5,000,000 being the ratio), or graphic scales which are illustrated to depict distances on the map in commonly used units of measurement such as kilometers or miles.

All geographic information is represented and managed using three primary GIS data structures, namely; feature classes, attribute classes, and raster datasets which are further extended for additional capabilities to manage data integrity. Typically, a GIS is used for handling several different datasets where each holds data about a particular feature collection that is geographically referenced to the earth's surface. In a GIS, homogeneous collections of geographic objects are organized into data themes such as parcels, wells, buildings, orthoimagery, and raster-based digital elevation models. A dataset is a collection of homogenous features for each theme. Geographic representations are organized in a series of datasets or layers. Most datasets are collections of simple geographic elements such as road networks, collections of parcel boundaries, soil types, elevation surface, satellite imagery, well locations, or surface water. Raster datasets are used to represent georeferenced imagery as well as continuous surfaces such as elevation, slope, and aspect. Table 6.1 lists some common GIS representations based on the theme and its geographic representation type.

Each GIS will contain multiple themes for a common geographic area. The collection of themes acts as a stack of layers. Each theme can be managed independently from other themes and has its own representation (Table 6.1). Since layers are spatially referenced they overlay one another and can be combined in a common map display. GIS analysis tools such as polygon overlay can fuse information between data layers to discover and work with the derived spatial relationships. Figure 6.3 presents a stack of layers in a GIS to create a common theme displayed on a single map.

In addition to working with datasets, users also work with the individual elements contained in the datasets in Fig. 6.3. These elements include individual features, rows and columns in attribute tables, and individual cells in raster datasets. The example in Fig. 6.3 displays a six-layer map structure each showing a theme and geometric representation type. Each layer is generated individually and transformed to fit the global scale, and can also be individually edited. Overlaying each layer one a single thematic map can provide valuable information regarding patterns depending on the application.

Table 6.1 Common GIS themes and representations

Theme	Geographic representations
Streams	Lines
Large water bodies	Polygons
Vegetation	Polygons
Urban areas	Polygons
Road centerlines	Lines
Administrative boundaries	Polygons
Well locations	Points
Orth-photography	Raster
Satellite imagery	Raster
Surface elevation	DEM raster
	Contour lines
	Elevation points
	Shaded relief raster
Land parcels	Polygons
Parcel tax records	Tables

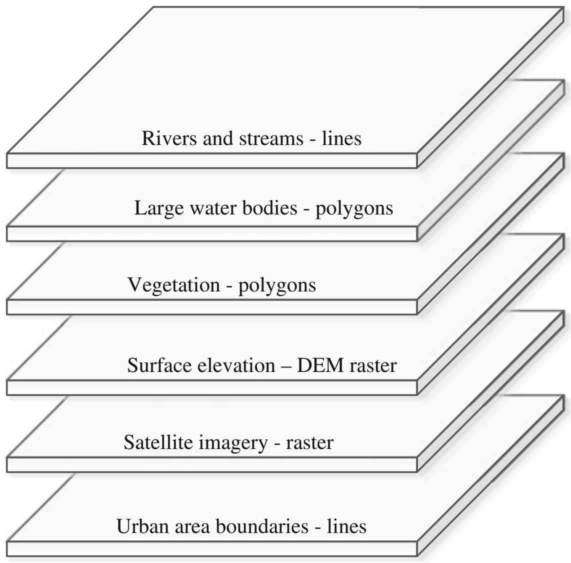


Fig. 6.3 GIS data layer from derived spatial relationships

The geographic data must be prepared to fit the geographic map, and all data types must be converted to a similar format if used in a single thematic map. The following section describes some common data transformation techniques used in GIS layered mapping.

6.7 Geographic Data

The key attributes of a GIS are the integration of geometric and thematic attributes of spatial objects. Geometric data describe the location of an object and thematic data describe the application-specific information of the object, generally gathered by sensors, surveys, or any other manual means of data gathering. GIS data capturing refers to the process of entering data into a GIS, generally in digital format as analogue to digital conversion was already applied by the time the data are entered into the system. Images taken by satellites and data tables can be used, or maps can be scanned in and uploaded to the system. The GIS takes the information from these sources and aligns the data according to scale to ensure it fits together in a single map. Additionally, a GIS must manipulate the projections of a map, a process to transfer information from the curved surface of the earth to a flat surface, which can result in minor distortion of the three-dimensional data. As a rule, any map can only show either the correct size of a land area or the correct shape, not both simultaneously.

Two types of geographic data are used in a GIS, vector data and raster data (Maling 1968). Vector data include points, lines, and polygons, and can be linked to a variety of data types such as traffic information, population levels, building altitudes, and street numbers. Raster data contain mainly color-variations generally found in aerial photos from planes, balloons, or satellites and are used to determine distribution patterns and distance analysis. Furthermore, continuous rasters are grid cells with gradual changing data. Examples are digital elevation models (DEM) and temperature data. Discrete rasters have distinct themes or categories, for example, land cover has discrete classes with clear boundaries. A GIS should ideally have hybrid technology, capable of processing vector and raster data types to allow altitude models and spatial statistics to be generated. As the name suggests, geographic data are dependent on spatial information and in most situations require knowledge of geometry, albeit very basic or complex. The amount of data to be processed limits the ability for an individual to analyze all the data and in most cases computerized software is used to do most of the heavy lifting.

The fundamental building block of a GIS is determining the distance between two points, represented as R. Pythagoras' theorem can be used in its basic form,

$$R = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{6.1}$$

where x_1 and y_1 are the coordinates of the first point in the Cartesian space, and x_2 and y_2 are the coordinates of the second point. The distance in (6.1) assumes two points in a two-dimensional plane as is generally the case for mapped GIS information if elevation information is not required. Three-dimensional information such as DEMs in a GIS requires the third dimension, z, to be included in the derivations. The effective distance between the two points in such a system would simply become

$$R = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{6.2}$$

To simplify the derivation of the concepts derived below, a two-dimensional system is considered. The representation of the two points in a two-dimensional space is given in Fig. 6.4 as reference for further derivations.

Determining the distance between two points as in (6.1) is important in GIS during spatial interpolation to smooth an interpolated surface through weighted averages. Although the equation itself is rudimentary, applying it to several thousand or even millions of data points requires assisted processing power. Many GIS applications are taking advantage of powerful graphical processing units as opposed to central processing units in modern-day computers as these devices are generally more equipped to manage geometric problems due to the nature of computer games. Distance calculations are important in clustering algorithms, polygon algorithms, and nearest neighbor routines for point pattern analysis and used to determine the competence of spatial sampling techniques. For point-in-polygon and polygon

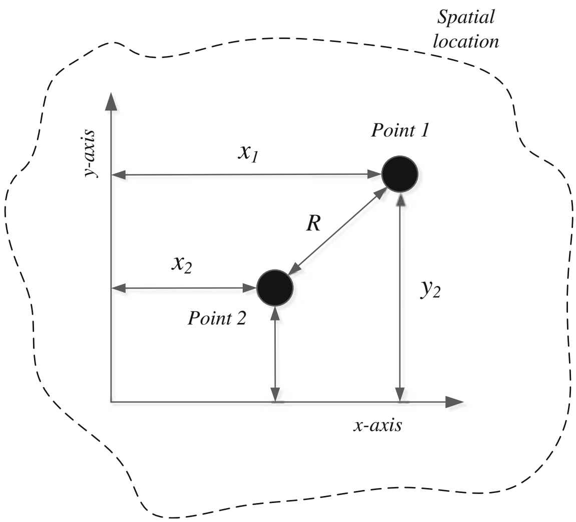


Fig. 6.4 Distance between two points using Pythagoras' theorem for two-dimensional Cartesian space

overlay routines, straight line intersection equations are fundamentally used. The intersection is of two lines is determined by the expression

$$x = -\frac{(a_1 - a_2)}{(b_1 - b_2)} \tag{6.3}$$

where a_1 and a_2 represent the y-intercept points of the two lines, and b_1 and b_2 represent the slopes of the two lines. Again, for convenience, a graphical representation of (6.3) is given in Fig. 6.5.

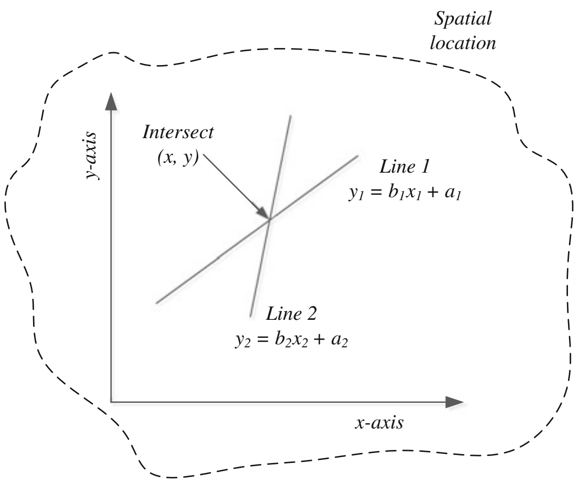


Fig. 6.5 Graphical representation to determine the intersection of two lines in two-dimensional Cartesian plane

The importance of (6.3) in GIS is highlighted if it is required to determine how frequently an observation falls within a predefined area, or polygon. The general operation to determine this is to draw a straight line from the observation point, and to count the number of times it intersects the line of the polygon. For low-complexity cases, if the amount of intersections is an even number, the point can be assumed to lie outside of the polygon. For odd numbers, the point should lie within the polygon. The complexity of the spatial area and polygon makes this approach somewhat crude, but it can be useful in some first-order calculations. A practical possibility would be to trace the direction of a pollutant downwind and determine if it exists inside a predefined susceptible area. Modeling the pollutant as a concentration, and equating the dispersion profile, could potentially give an estimate of the pollutant concentration inside the susceptible area. Another method to determine the interaction between two points or places is by using the gravity model. The interaction I can be determined by

$$I_{ij} = k\left(\frac{P_i \times P_j}{R_{ij}^2}\right) \tag{6.4}$$

where k is an empirical correction coefficient, P_i and P_j are the populations of the observed entity at the two points, and R_{ij} is the straight line distance between the two points as found in (6.1) or (6.2). The interaction quantity is a helpful guideline in determining the significance of introducing a new resource between two populated areas, with various applications from sensor placement in smart cities to feeding stations on farmlands. The gravity model gains statistical respectability through entropy maximizing models to determine the most likely distribution of movement between two points (Lui and Xu 2011). From the formal definition of entropy, consider P_{ij} in a decision matrix where there are n alternatives and k criteria in the decision matrix. The element of this matrix for the jth criteria is given by

$$P_{ij} = \frac{f_j(a_i)}{\sum_{i=1}^{n} f_j(a_i)} \tag{6.5}$$

where j = 1, 2, … k. The entropy E_j is calculated by

$$E_j = -\frac{1}{\ln n}\sum_{i=1}^{n} P_{ij}\ \ln\ P_{ij} \tag{6.6}$$

and E_j is a value between 0 and 1. The deviation degree d_i shows to what extent the jth sample contains useful information for the decision making process. If there is a small difference between criterion it points out that alternatives are indifferent based on this criterion and its effect in the final decision should be discarded. The deviation degree is determined by

$$d_i = 1 - E_i \tag{6.7}$$

and based on this result, the entropy, the weight of each sample is determined by

$$w_j = \frac{d_j}{\sum_{j=1}^{k} d_j} \tag{6.8}$$

In information theory, Shannon informational entropy expresses the minimum number of bits per symbol to encode the information in binary form or determine the randomness of a message based on the metric entropy. Entropy as used by Boltzmann utilizes entropy as a concept to measure disorder in thermodynamic system. In GIS water resource management, an example of using the entropy concept is to cope with uncertainties associated with hydrologic variables, hydrologic systems and their models, and the probability distribution function parameters. If information is gathered and processes and relationships are determined, it must also be mapped spatially to the earth, bearing in mind the earth is not a perfect spherical shape.

The almost spherical geometry of the earth requires spatial coordinates be converted to a two-dimensional coordinate system in the Cartesian plane. Various types of conversions are used in GIS and are known as projection equations. These types include analytical, direct or grid-on-grid, and polynomial transformations. The Cartesian coordinates (x, y) of a point on a map are functionally related to the position on the earth's surface and expressed in geographical coordinates (φ, λ) where

$$x = f_1(\varphi, \lambda) \tag{6.9}$$

and

$$y = f_2(\varphi, \lambda) \tag{6.10}$$

relate the Cartesian coordinates to plane coordinates used in maps, aerial photographs, or scanned imagery. Several assumptions regarding the shape of the earth can be made to simplify the projection equations. One such assumption is that the surface of the earth can be approximated by a smoothed spheroid as opposed to an irregular geoid. This assumption reduces the earth's approximate 9,277 m excursions of mountains and sea levels to less than 200 m, substantially closer to a perfect mathematical ellipsoid shape. Although modern day computers can facilitate the processing requirements to calculate transformation data based on the earth's actual surface irregularities, this assumption still holds for relatively accurate results. This assumption is also preferred if limited elevation data is available. Nautical charts generally use analytical transformation techniques such as Mercator's projection. For a tangent normal Mercator projection, the equations used are

$$x = R\lambda \tag{6.11}$$

and

$$y = R \ln\left[\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)\right] \tag{6.12}$$

where λ is the longitude expressed in radians and R is the radius of the earth expressed in millimeters at the scale of the proposed map. To express (φ, λ) in terms of (x, y), thus the inverse solution, the following equations are used;

$$\varphi = \frac{\pi}{2} - 2\tan^{-1}\left(\exp^{-\frac{y}{R}}\right) \tag{6.13}$$

and

$$\lambda = \frac{x}{R} + \lambda_0 \tag{6.14}$$

where λ_0 is the datum meridian from which the longitudes are measured. A datum is a reference system of the earth's surface against which positional measurements are made. Two categories of datum are used, horizontal and vertical datum. Horizontal datum is used to describe a point on the earth's surface as latitude and longitude, or in another coordinate system such as polar. Vertical datum is used to measure the elevation and underwater depths forming the geoid structure of the earth. The eccentricity of the spheroid is defined as a measure of how much the conic section deviates from a perfect circle. To determine the eccentricity ε of the earth's spherical shape having semi-axes a and b, the relationship is determined by

$$\varepsilon^2 = \frac{(a^2 - b^2)}{a^2} \tag{6.15}$$

which is equal to approximately 6.7×10^{-3}. The forward solution of Mercator's projection of the spheroid can be modified using semi-axes and eccentricity towards

$$x = a\lambda \tag{6.16}$$

and

$$y = a \ln\left[\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)\right]\left\{\frac{(1 - \varepsilon\sin\varphi)}{(1 + \varepsilon\sin\varphi)}\right\}^{\varepsilon/2} \tag{6.17}$$

and for the inverse calculation to find the latitude transcendental, an iterative solution is required in the form

$$\varphi = \frac{\pi}{2} - 2\tan^{-1}\left(\exp^{-y/a}\left\{\frac{(1-\varepsilon\sin\varphi)}{(1+\varepsilon\sin\varphi)}\right\}^{\varepsilon/2}\right) \tag{6.18}$$

where the first iterative solution is found by

$$\varphi = \frac{\pi}{2} - 2\tan^{-1}\left(\exp^{-y/a}\right) \tag{6.19}$$

To iteratively solve for (6.18), the result of (6.19) is used in the curled brackets of (6.18) to calculate the next value of φ until it converges. The longitude is obtained by

$$\lambda = \frac{x}{a} + \lambda_0. \tag{6.20}$$

The analytical method is independent of the size of the area to be mapped but it can be slow if large data arrays are used, even if using modern day processors. An alternative approach is the direct transformation, which does not require an inverse solution of the geographical coordinates (φ, λ) and is directly based on the rectangular coordinates of the same points on the two projections. Since it is rare that the axes of the source and target systems are exactly parallel and that the two systems have similar scales, this method is preferred. The direct transformation technique is commonly used in cartography to re-grid or re-plot a second grid on military topographical maps, in remote sensing, in modern analytical plotters from conventional aerial photography, and in multispectral scanner sensor acquired imagery used for geometrical and radio-metrical processed data. The transformation of a set of geocentric coordinates are transformed using the Bursa-Wolf seven-parameter equation (Hashemi et al. 2013), given by

$$\begin{bmatrix} X_T \\ Y_T \\ Z_T \end{bmatrix} = S\begin{bmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{bmatrix}\begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix} + \begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} \tag{6.21}$$

where X_T, Y_T, and Z_T are the coordinates in the target system, X_S, Y_S, and Z_S are the coordinates in the source system, α, β, and γ are the rotational angels about the x, y, and z coordinate planes respectively, dX, dY, and dZ are the translations of the origin, and S is the scale change between the source and target coordinate systems. The scale change parameter can be introduced as a parts per million value, δ, where

$$S = \left(1 + \delta \times 10^{-6}\right) \tag{6.22}$$

leading to

$$\begin{bmatrix} X_T - X_S \\ Y_T - Y_S \\ Z_T - Z_S \end{bmatrix} = \begin{bmatrix} \delta & \gamma & -\beta \\ -\gamma & \delta & \alpha \\ \beta & -\alpha & \delta \end{bmatrix} \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix} + \begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} \tag{6.23}$$

where α, β, γ, δ, dX, dY, and dZ are the seven unknown parameters. Frequently, when developing spatial databases for GIS, data is provided on map sheets that use unknown or inaccurate projections. To be able to register two datasets against each other, a set of control points must be identified and located on the source and destination system. There should be at least three control points identifiable of each map since three points provide six values to solve for six unknowns and these points should not be collinear, i.e. all on the same line, affine transformation can be used in such scenarios. The transformations can move lines into lines, while preserving their intersection properties and can move polylines into polylines and polygons into polygons. Every affine transformation can be expressed as a transformation that fixes the origin followed by a simple translation of the entire plane. Affine translation is divided into four categories, translation, scaling, rotation, and reflection. Routine uses of two-dimensional and three-dimensional affine transformations in GIS include map-to-display transformations, registering images and raster, changing three-dimensional viewpoints, modifying objects by rescaling, shifting, or rotation, and datum changes with three- and seven-point formulas. As an example, consider a point $\mathbf{x} = (x, y)$ in the two-dimensional Cartesian system. The point $\mathbf{x}$ can be represented as a column vector for example, given by

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{6.24}$$

and a two-dimensional transformation matrix, M, can be represented by

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{6.25}$$

and therefore the transformation matrix can be written as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{6.26}$$

An affine transformation can be written in the following generalized form.

$$\mathbf{x}' = \begin{bmatrix} ax + by + c \\ dx + ey + f \end{bmatrix} \tag{6.27}$$

where a, b, c, d, e, and f are scalar values. The four translation types can be representing by inspecting the values of these scalars. If a, e = 1 and b, d = 0, it follows that (6.27) becomes

$$\mathbf{x}' = \begin{bmatrix} x + c \\ y + f \end{bmatrix} \tag{6.28}$$

resulting in a pure translation as depicted in Fig. 6.6.

This result in (6.28) and Fig. 6.6 shows that translations can be handled separately, and performed sequentially, making it easier to understand if done by hand calculations, generally to test one set of data points before entrusting a computer program to handle to bulk of the calculations. Similar to a pure translation, pure scaling can be achieved for the case where $b, d = 0$ and $c, f = 0$, then (6.27) becomes

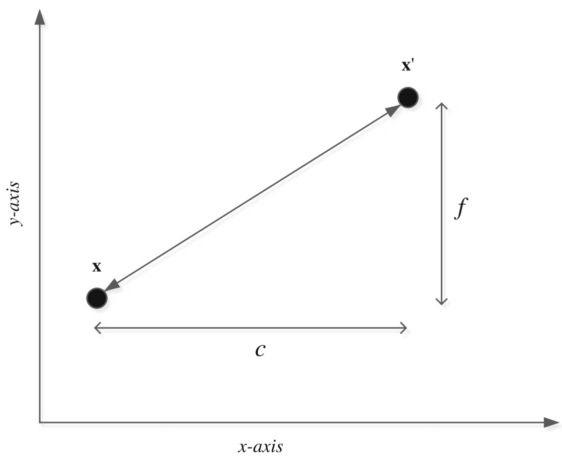
$$\mathbf{x}' = \begin{bmatrix} ax \\ ey \end{bmatrix} \tag{6.29}$$

which is a pure scaling transformation. To verify this result, the matrix representation is visually depicted in Fig. 6.7.

To achieve a pure rotation about the origin of the axes, suppose that the scalar values for a, e = cosθ, b = −sinθ, and c, f = 0 holds true, resulting in (6.27) translated to

$$\mathbf{x}' = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix} \tag{6.30}$$

Fig. 6.6 Affine translation resulting in pure translation between two points



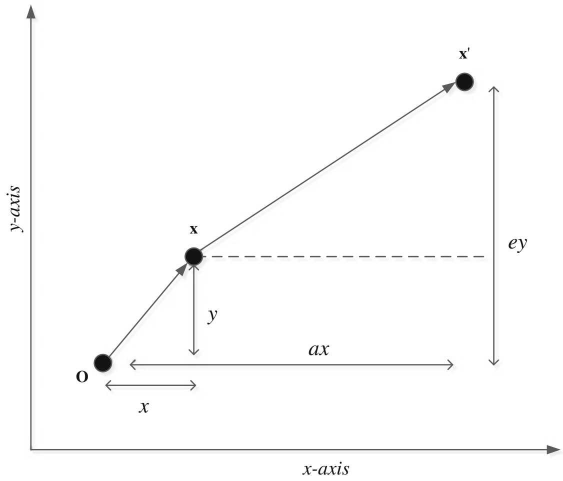


Fig. 6.7 Affine translation resulting in pure scaling between two points

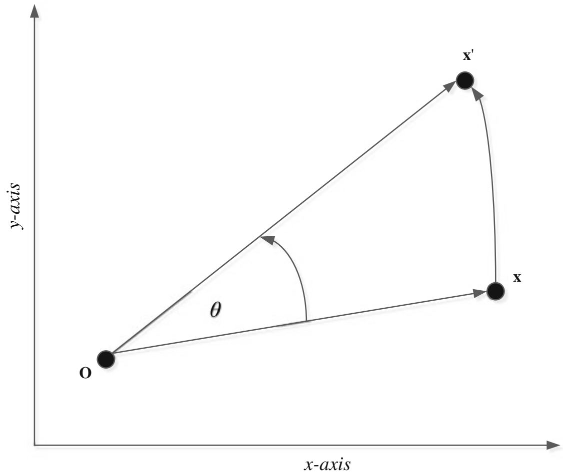


Fig. 6.8 Affine translation resulting in pure rotation between two points

which is derived through additional trigonometric transformations (such as the direction cosine matrix not shown here), and the result is shown in Fig. 6.8.

Finally, if a, e = 1 and c, f = 0, a pure shear translation is achieved, given as

$$\mathbf{x}' = \begin{bmatrix} x + by \\ dx + y \end{bmatrix} \tag{6.31}$$

as shown in Fig. 6.9.

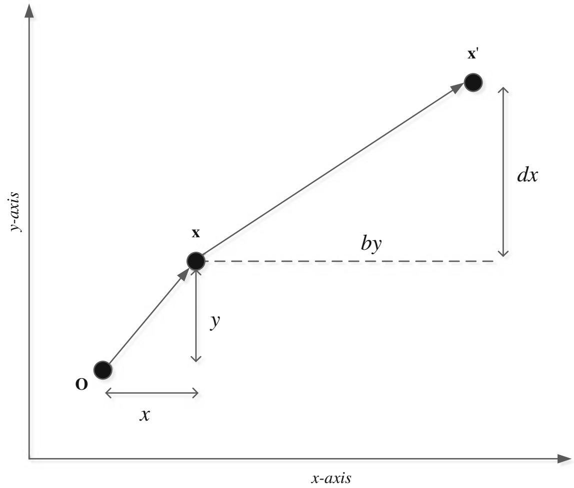


Fig. 6.9 Affine translation resulting in pure shear between two points

These affine transforms for scale, rotation, and shear are linear transforms and it is convenient to represent them in matrix multiplication of a point by using (6.26),

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} ax + by \\ dx + ey \end{bmatrix} = \begin{bmatrix} a & b \\ d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = M \begin{bmatrix} x \\ y \end{bmatrix} \tag{6.32}$$

Suppose a scaled matrix is represented by S, shear matrix by H, and the rotation matrix by R, then the transformation matrix can be written as

$$\mathbf{x}' = R(H(s\mathbf{x})) = (RHS)\mathbf{x} = M\mathbf{x} \tag{6.33}$$

through the associative multiplication theorem in matrix multiplication. This is possible since it was determined that a total translation of a system is a set of sequential transformation steps. To characterize the linear transforms, the three matrices are identified as

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \tag{6.34}$$

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{6.35}$$

$$H = \begin{bmatrix} 1 & h_x \\ h_y & 1 \end{bmatrix} \tag{6.36}$$

where s_x and s_y scale the x and y coordinates of a point, θ is the angle of counter-clockwise rotation around the origin, and h_x and h_y are the horizontal and vertical shear factors respectively. Non-homogeneous equations are used to represent all affine transforms in combined matrices to convert between two-dimensional and three-dimensional spaces. If a point in the two-dimensional Cartesian plane is again considered as $\mathbf{x} = (x, y)$, it is possible to create an identical three-dimensional vector by

$$\begin{bmatrix} x \\ y \end{bmatrix} \rightarrow \begin{bmatrix} x \\ y \\ w = 1 \end{bmatrix} \tag{6.37}$$

and by convention this third coordinate is the w coordinate to distinguish it from the generally used three-dimensional z coordinate. The two-dimensional scale, rotate, and shear matrices are also extended to this dimension such that

$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.38}$$

$$R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.39}$$

$$H = \begin{bmatrix} 1 & h_x & 0 \\ h_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.40}$$

and if the three-dimensional homogeneous matrices are multiplied by homogeneous vectors, the result is

$$\begin{bmatrix} a & b & 0 \\ d & e & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax + by \\ dx + ey \\ 1 \end{bmatrix} \tag{6.41}$$

which is the same result as the two-dimensional case with the exception of the w coordinate, which remains 1. If the parameters c and f are placed in the resulting matrix, then

$$\begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax + by + c \\ dx + ey + f \\ 1 \end{bmatrix} \tag{6.42}$$

which allows translations as a linear operation in homogeneous coordinates. The translation matrix can also be added to the series of translations as

$$T = \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix} \tag{6.43}$$

where dx and dy are the translations in the x and y direction respectively.

As an example, consider a 2 × 2 matrix centered at the origin. The example of the four points (−1, 1), (1, 1), (1, −1), and (−1, −1) centered around the origin is presented in Fig. 6.10.

To perform a total transformation on the coordinates in Fig. 6.10, the following procedure can be followed, and the sequential nature of the transformation provides a good guideline of the required steps. If it is required to rotate this square by 45° about its center and move (translate) the square so that its center is at (3, 2). No scaling is performed in this transformation. The matrix multiplication M is given by the multiplicative properties of the translation (T) and rotational matrix (R) as

$$R_{45} = \begin{bmatrix} \cos 45 & -\sin 45 & 0 \\ \sin 45 & \cos 45 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.44}$$

which results in the rotation given in Fig. 6.11 as

The transformation is given as

$$T_{(3,2)} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.45}$$

and the multiplication matrix is determined as

$$M = T_{(3,2)} R_{45} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.46}$$

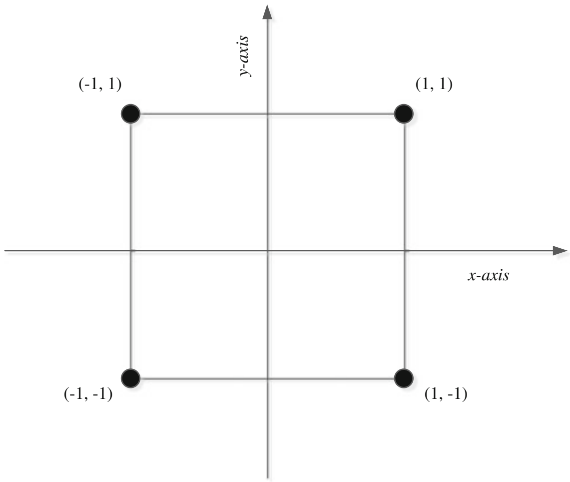


Fig. 6.10 Affine translation original matrix of four points centered at the origin

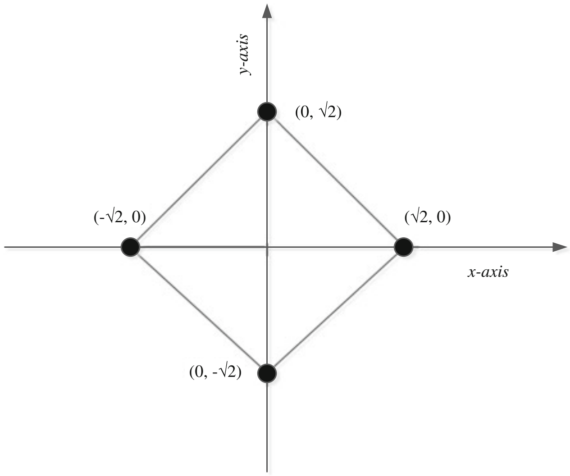


Fig. 6.11 Affine translation of center point towards (3, 2) and 45° rotational

therefore

$$M = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} & 3 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 2 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.47}$$

To find the translated coordinate of $(x, y) = (1, 1)$, it is determined by

$$M\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} & 3 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 2 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2+\sqrt{2} \\ 1 \end{bmatrix} \tag{6.48}$$

and the center point (origin) $(x, y) = (0, 0)$ coordinate is translated to

$$M\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} & 3 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 2 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \tag{6.49}$$

which is verified in Fig. 6.10.

The following section discusses the fundamentals of satellite-based positioning systems (Fig. 6.12).

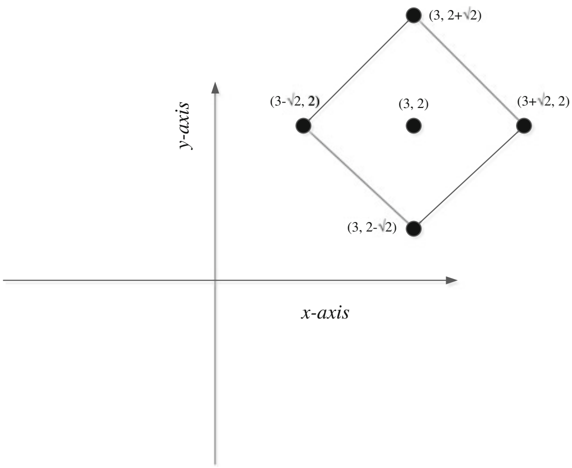


Fig. 6.12 Affine translation of center point towards (3, 2) and 45° rotational

6.8 Global Positioning System

A satellite navigation system with global coverage are referred to as a global navigation satellite system (GNSS) such as the US Navstar global positioning system (GPS), GLONASS from Russia, China's regional (but expanding) BeiDou Navigation Satellite System and Galileo; the UN's GNSS in its initial deployment phase and expected to be fully operational by 2020. GPS is a fundamental enabling technology for GIS spatial information. GPS was originally designed for military applications during the Cold War in the 1960s. In 1973, the Pentagon proposed a global positioning satellite system known as Navstar. The Navstar network of satellites is used for both military and civilian applications. The first eleven GPS satellites forming the Block I constellation were launched in 1978 with only Navstar 7, which was launched on the 18th of December 1981, failing to reach orbit. Since then, there have been several generations of Navstar satellites and the latest generation is expected to have a lifetime of 15 years in orbit. As of July 2015, there are 70 GPS satellites in orbit, at various altitudes, not all being operational. The minimum number of satellites that needs to be operational within a constellation is 24 in six orbital planes, translating to at least six satellites within the user's line of sight at any given time since four satellites are required to calculate the position of the user.

The reliability of GPS data through three dimensional triangulations is however an important consideration in building GIS. GPS consists of three functional parts, space-based satellites, control of the manufacturing, launching, and orbital positioning, and off course the user, be it designed for military or civilian applications. GPS does have its limitations, although generally is it considered a reliable technology. One such limitation (apart from selective availability of full accuracy that adds intentional time varying errors of up to 100 m, but has been disabled in

May 2000) is the requirement for line of sight between the user and the orbiting satellites. This generally becomes a problem in urban areas, where buildings and large structures shield the signal from the satellites, or create multipath effects as the signal bounces from the object causing echo signals. A clear view of the sky is preferred to synchronize with orbiting satellites. Recent technologies have introduced additional methods to decrease lock time with satellites and to use cellular telephone base stations as first-order location estimation.

To determine the latitude and longitude of a user on earth, three GPS satellites measure the distance from the user to each of the satellites. The GPS knows the location of the satellites and the instant the reading was taken, and these distances intersect at one point, which is the location of the GPS antenna. Additionally, a fourth satellite is required to determine the altitude/elevation of the GPS receiver antenna of the user. The distance between the user and the satellite is determined either by code based or carrier phase based techniques, Kalman (2002).

Code based systems generate a pseudo-random code and compare the time that the signal was sent from the satellite with the time it was received at the antenna, and determines the distance by multiplying the time interval with the speed of light. This technique is generally accurate only to a few meters due to ionospheric variations. Improved accuracy can be attained by differential GPS, requiring the existence of a base station, which is a GPS receiver collecting measurements at know latitude, longitude, and elevation, assuming the base station antenna location in precisely known. The base station can store measurements that can be used for post-processing or broadcast corrections in real-time. GIS generally applies the corrections in post-processing to enable a large multitude of hardware compatibility. Differential GPS is depicted in Fig. 6.13.

From Fig. 6.13, the reference receiver or base station that stores the measurement for post-processing or real-time corrections is shown on the right. This stationary receiver knows its own location precisely and calculates the timing error (reverse calculation) based on this knowledge. This receiver then transmits a signal back to the roving receiver (top left of Fig. 6.13) with the proposed corrections. This roving receiver then sends the information to the GPS satellites used to calculate the user's location. Since these satellites are relatively close (few 100 km) to this receiver and the signals do not pass through the atmosphere and ionosphere, the signals do no undergo further degradations and all GPS satellites can use this information to adaptively correct timing errors.

Generally, in geodetic control systems, carrier phase based receivers are used and are capable of sub-centimeter differential accuracy. This type of receiver calculates the distances to visible satellites by determining the number of full wavelengths and the partial wavelength. If the amount of wavelengths for a transmission is known, the range is determined by multiplying by the amount of full wavelengths with the period of the carrier signal. A baseline distance and azimuth between any pair of receivers operating simultaneously can then be determined and with one receiver placed at a point with known latitude, longitude, and elevation, this baseline can be used to determine the location of the other receiver. This technique is more expensive to implement compared to code based methods and is more

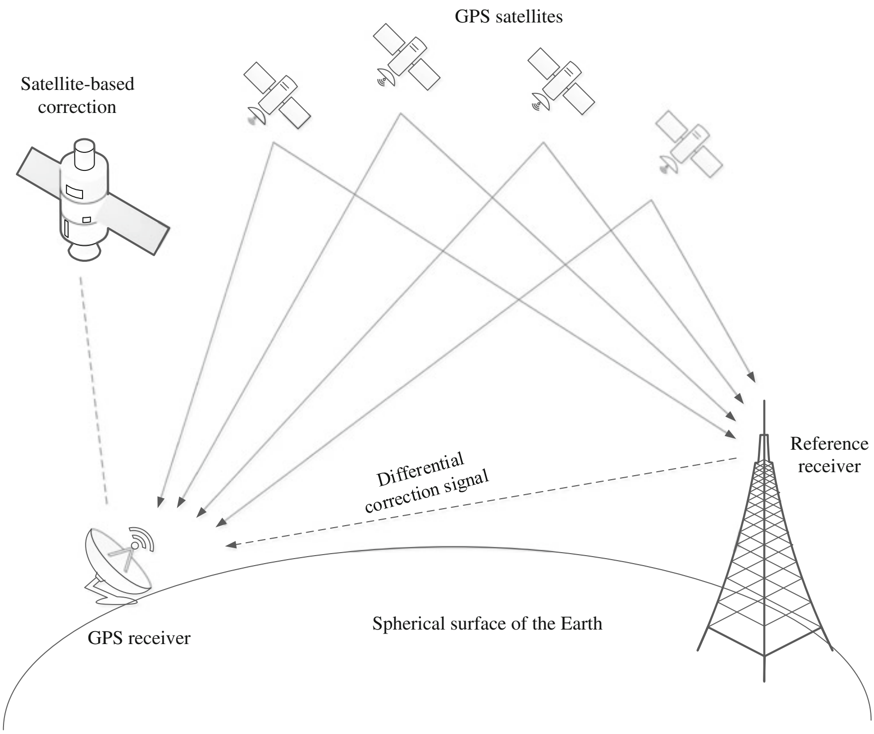


Fig. 6.13 Principle of differential GPS

commonly used in machine control, military applications, and precision farming that require a high degree of accuracy.

To mathematically determine the three dimensional coordinates (latitude, longitude, and elevation) of a GPS receiver, it has been noted that spatial information from at least four satellites are required (Thompson 1998). Suppose the satellites are located at points in space R_i (i = 1, 2, 3, and 4) with coordinates (x_i, y_i, z_i) and the signal is transmitted from the satellites to earth at time T_i. If the signals are received at the receiver on earth at time T_r measured in nanoseconds, the error in clock time is represented by the difference (again in nanoseconds) between the two times Δt_i where

$$\Delta t_i = T_i - T_r \tag{6.50}$$

and the receiver allows for the mean effects of the passage through the atmosphere of the Earth and computes the distances in meters from the center of the earth as

$$d(\Delta t_i, \varepsilon) = c\Delta t_i \tag{6.51}$$

where ε represents the possible error in the clock times, generally as a result of electronic component tolerances that deviate from the atomic clock, and c is the speed of light at approximately 300 × 10^6 m.s^{-1}. The estimated distance between the satellites from the possible error is known as the pseudo range. The receiver position can be located in four large spheres, given by the coordinates (x_r, y_r, z_r). In most situations there will be only one sensible value of ε that allows the spheres to have a point in common. The location of the receiver is determined by numerically solving the system of equations given by

$$d(\Delta t_1, \varepsilon)^2 = (x_r - x_1)^2 + (y_r - y_1)^2 + (z_r - z_1)^2 \tag{6.52}$$

$$d(\Delta t_1, \varepsilon)^2 = (x_r - x_2)^2 + (y_r - y_2)^2 + (z_r - z_2)^2 \tag{6.53}$$

$$d(\Delta t_1, \varepsilon)^2 = (x_r - x_3)^2 + (y_r - y_3)^2 + (z_r - z_3)^2 \tag{6.54}$$

$$d(\Delta t_1, \varepsilon)^2 = (x_r - x_4)^2 + (y_r - y_4)^2 + (z_r - z_4)^2 \tag{6.55}$$

where the coordinates are then converted (transformed) to spherical coordinates of latitude, longitude, and altitude above sea level. These equations can be solved through numerical analysis such as the Newton-Rhapson method combined with calculus tools to approximate its tangent line, and elementary algebra to determine the intersection with the y-axis and follows an iterative solution. To revert the non-linear equations to linear algebraic solutions, it is particularly useful to use a least-squares solution.

If data from only three satellites are received, the altitude could still be determined by substituting the surface area of the earth for the missing fourth sphere. If a receiver is stationary and the GPS coordinates are continuously recorded, the results usually vary slightly. These variations are primarily due to random errors in the clock time, selection of different satellites over time, and the effects of variations in the atmosphere. Additionally, selective availability enforced by the DoD also contribute to changes in location data although the receiver is stationary.

For GIS, several key issues must be addressed when considering GPS as an appropriate tool for capturing data. The primary requirement in the accuracy of the data depends on the application. For high accuracy site-specific analysis that requires accuracy within a meter, high-quality code based differential GPS receivers should be used. For property boundary mapping which could require accuracies of 10 cm or lower, carrier phase differential GPS techniques can be used. Every GIS database must be referenced to a base map or layer, and the reference datum of the various data layers must be similar. In GIS, if three or more distinctive points can be located on both the satellite image and on the earth's surface, GPS receivers can be used to collect accurate geographical coordinates at these locations. The rest of the image can then be adjusted to better represent the real-world coordinates. GPS receivers can also be used in GIS to investigate unusual reflections or scattering from surfaces on the earth by enabling the user to navigate directly to the area of interest.

It is evident that GPS is a key contributing factor of GIS, and henceforth of remote sensing applications. Various remote sensing techniques are available, and applications thereof are practically infinite. The following section describes some key factors to consider about remote sensing.

6.9 Remote Sensing

Remote sensing utilizes airborne sensors to collect information about an area of interest. The sensors can be placed on satellites, airplanes, drones, unmanned aerial vehicles (UAVs), balloons, and even remote controlled planes. A multitude of applications exist, from forest fire detection, grid management, earthquake detection, weather patterns, urbanization trends, and many more. Applications also include information gathering in smart cities for sustainable monitoring techniques and used by various research institutions to analyze trends and patterns. Data collections can be implemented using passive or active techniques. Passive sensors such as spectral imagers detect natural radiation emitted from sensed objects due to its temperature variations based on exposure from a natural source, such as the sun. In active remote sensing such as radar, energy is emitted by an additional device and the returned signal is measured and analyzed based on the changes that occurred during its travel.

Passive sensors utilize the three fundamental energy interactions of incident radiation from predominantly the sun as a blackbody source, Hassebo (2012). These interactions include the reflected energy E_r from the object, absorbed energy E_a into the object, and transmitted energy E_t from the surface of the object into the environment. The incident energy formula previously mentioned is written in the form

$$E_i = E_r + E_a + E_t \tag{6.56}$$

where E_i is the total incident energy on the object from the radiative source. Passive sensors detect specifically the reflected energy E_r in (6.56), where the portion of reflected and total incident energy is called the spectral reflectance ρ, given by

$$\rho = \frac{E_r}{E_i} \tag{6.57}$$

which is a quantity specifically characterizing the spectral properties of the material that the object consists of. Passive sensors measure this natural energy at specific wavelengths. The frequency spectrum of the measurement is determined by the application and at which frequencies it should see or discard (optical windows). If a sensor is able to distinguish between multiple spectra, it is known as a multispectral sensor.

In contrast to a passive sensor, an active sensor generates its own energy that is directed at the target and measures the changes in the signal as the signal returns to the sensor. The sources of energy vary, it can be visible light for example, the flash of a photographic camera (note that a camera that does not use a flash is in essence a passive sensors), or any other source within the electromagnetic spectrum.

Radio detection and ranging (radar) for example generates radio waves that are not attenuated by clouds. Near-infrared lasers are used as a source in topographic light detection and ranging (Lidar) sensors. Green light is used in underwater depth profiling in lakes or oceans as it can penetrate water without significant attenuation. Sound navigation and ranging determines location and velocity information of objects.

The advantage of active sensors is that it can be used during nighttime and daytime, since it does not require natural light as a source and can generate any spectral source required for its application. A problem that might occur is additive noise in the fundamental frequency from stray electromagnetic waves that can distort a signal and cause inaccurate and unusable images or data. Passive sensors are less expensive since they do not require additional light sources and are used the most frequently in earth observation techniques and GIS, although cost-reduction from technology advancements has made active sensors more attainable. They are however dependent on natural sources of light that could be blocked or attenuated by natural objects and should be ideally operated during daytime and in warm climates. A generalized representation of a passive sensing infrastructure is shown in Fig. 6.14.

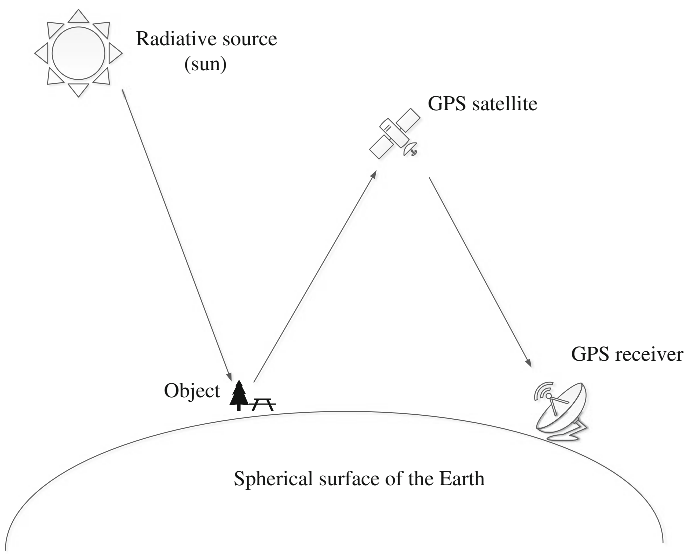


Fig. 6.14 Passive remote sensing infrastructure using the sun as a natural radiating body

From Fig. 6.14, it can be seen that the radiative source, the sun, emits natural energy towards objects on earth. This change in the object's temperature is transferred to the sensing elements on the aerial device, in this case a satellite, and transmitted back towards a ground station where the data is logged, typically using GIS or alternative methods. Any sudden changes in the temperature characteristics of the object, apart from the natural warming and cooling resulting from the strength of the source as time progresses, can be noted and investigated. An active sensor infrastructure works on a similar concept and is depicted in Fig. 6.15.

In Fig. 6.15, the sun as a radiative source is omitted from the figure to illustrate the light source is present together with the sensing element on the aerial device. The light source directs its energy at the object, knowing the parameters of the energy it emits, it receives an altered signal back from the object. The information is then transferred to the ground station where it is analyzed and interpreted, and in GIS it is mapped as layers.

Remote sensing, whether utilizing passive or active sensors, does not have a single definition per se, and the applications are vast. Table 6.2 lists several applications of remote sensing that are used in the industry and research, with many more applications that exist, only limited by the creativity and necessity of its users.

As seen in Table 6.2, the number of applications for remote sensing vary greatly and the general theme is monitoring and preventing events that negatively influence large numbers of people or animals, therefore risk management. The number of applications for remote sensing grows every year as local research institutions, industries, and military institutes require more information and better data gathering techniques to improve their effectiveness and as sustainable smart cities are becoming more prevalent. Technologies are also increasing in reliability and decreasing in cost, making these techniques more and more viable for even small institutions to maintain and integrate into current static systems. The keys to successful applications are secured continuity, secured access, global coverage, and

Fig. 6.15 Active remote sensing infrastructure using additional sources of energy on the aerial structure

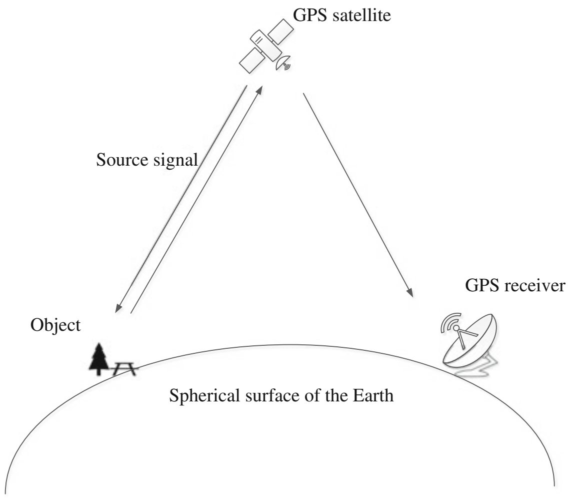


Table 6.2 Remote sensing applications for sustainable cities, rural areas, and pollution monitoring

Application	Description
Active volcanoes	Monitoring volcano activity for safety and pollution purposes
Air quality	Monitoring air quality in the lower atmosphere in large cities
Algae monitoring	Determining nitrogen and phosphorous levels in lakes from algae growth
Archaeology	Penetrating earth's surface to uncover archaeological findings
Autonomous vehicles	Safer transport by eliminating human error during driving
Drinking water	Spatial satellite images to create base maps for clean drinking water
Epidemiology	Prediction of disease spreading
Fishery sustainability	Monitor ocean temperature and color which indicate specific fish species and control fisheries
Glacier melting	Analyzing and monitoring the melting of the polar glaciers
Habitat sustainability	Habitat monitoring for animals such as pandas
Human impact change	Comparing areas that changes due to human intervention (for better or worse)
Hydrological modeling	Identify areas that are prone to flooding
Interferometry	Measuring landscape deformation can prevent building collapses
Inventorying rural roads	Pothole detection, washboard analysis, and crown conditions or rural roads
Land cover	Analyzing land use in rural and urban areas
Landslides	Inventorying potential landslides to prevent small earthquakes and erosion
LiDAR	Digital surface models, digital elevation modes, and light intensity models
Mapping services	Aerial imagery of destinations and navigation services
Measuring gravity	Accurate measurements of gravity by satellites
Military surveillance	Enemy territory and terrorist activity
Mineral deposits	Extracting mineral deposits in mines with hyperspectral sensors
Natural disasters	Weather forecast for natural disaster prevention
Navigating ships	In addition to GPS, locate risks and natural events for ships
Ocean currents	Determining safe routes and swimming areas by analyzing ocean currents
Oil spills	Marine life and surrounding environment protection
Photogrammetry	Contour mapping, surface models and volumetric surveys for GIS
Polar bear counting	Controlling and monitoring populations of hard to reach animals
Precision farming	Effective use of fertilizers through precision farming
Sediment transport	Tracking anthropogenic factors on aquatic systems
Soil moisture content	Water cycles, weather forecasting, drought, and floods
Solar energy	Optimizing solar energy output by identifying ideal locations

(continued)

Table 6.2 (continued)

Application	Description
Telecommunication planning	Optimize capacity requirements for increasing use of mobile devices
Traffic congestion	Reduce traffic congestion by early detection and alternative route suggestions
Urbanization	Tracking the shift from rural to urban growth
Vehicle emissions	Control and monitor vehicle emissions in urban areas
Water supply	Planning lift irrigation to improve agriculture water supply
Wind energy	Optimizing wind energy output by identifying ideal locations
Wind speed and direction	Measurement of wind profiles for surfers or other recreational activities

affordability. Sharing of data and information is largely encouraged in the GIS and remote sensing communities, as many industries can contribute to ensuring open-access of geometric information. Not all institutions are for example able to launch their own satellites into space to contribute to spatial monitoring, or even launch drones or UAVs and these companies can only contribute to the larger goal, being environmental sustainability, through sharing of information by larger and more equipped players. The remote sensing community generally welcomes the practice of information sharing and an increase in these practices is beneficial.

In remote sensing, radiometric characteristics describe the information content within an image. Each time an image is acquired from the sensor, its sensitivity to the magnitude of the surrounding electromagnetic energy determines its radiometric resolution. This quality describes the system's ability to discriminate between slight variations in transmitted energy. A smaller radiometric resolution translates to an ability to detect smaller differences in reflected or emitted energy. Digital resolution refers to the number of bits comprising each digital sample. Images are displayed in a range of grey tones, with black representing the binary number 0 and white representing the maximum value 255 in an 8-bit system, since $2^8 - 1 = 255$. Four parameters can be used to describe any spectrometer, these include spectral range, spectral bandwidth, spectral sampling time, and the signal to noise ratio. Spectral range refers to the spectral band of wavelengths that the sensor is sensitive to, or in other words, the band that the sensor can see. Spectral bandwidth is the width of an individual spectral channel in the spectrometer, where a narrower bandwidth translates to a narrower absorption that the spectrometer can accurately measure. Spectral sampling is the distance in wavelength between spectral bandpass profiles for each channel in the spectrometer as a function of wavelength. Finally, the signal to noise ratio refers to the ratio of the transmitted or reflected signal to the additive noise generated during its path, and the ability of a spectrometer to distinguish between the two and still be able to process the information signal. Additionally, the temporal resolution specifies the revisiting frequency of a satellite sensor for a specific location. The different spatial, temporal and spectral resolutions are the limiting factor for the utilization of the satellite image data for different applications. Due to technical constraints, satellite remote sensing systems can only offer a high spatial resolution with a low spectral resolution and vice versa.

In GIS, lidar is a powerful tool for probing the earth's atmosphere through laser transmission using elastic and/or inelastic scattering techniques. The fundamental functional subsystems of lidar are the transmission, receiver, and electronic manipulation circuit subsystems. The transmission subsystem uses a pulsed or continuous wave laser light source having a narrow bandwidth, high repetition rate, high peak power, and small divergence. Optics are additionally integrated to further reduce divergence and control beam polarization. Harmonic generators can be used for wavelength selective generators by utilizing the harmonic content of the output signal. The receiver subsystem consists of an optical telescope to gather and focus the backscattered radiation and optics to improve the signal on the detector by focusing strong polarized signals. Additional components such as mirrors, collimated lenses, aperture, neutral density filters, interference filters provide careful filtering of background and ambient radiation. The electro-optical components of the receiver convert incident light to electrical energy. The electronic subsystem handles data acquisition, analog to digital conversion, radar support circuitry, and control systems for polarization discrimination to track azimuth angles to improve the signal to noise ratio. Software is also used, especially in GIS, to manipulate and post-process data to be displayed as logical content to the user. Lidar systems are classified using roughly nine main criteria.

- The physical process it uses, such as Rayleigh scattering, Mie scattering, elastic and inelastic backscattering, absorption, or florescence.
- The types of lasers used, for example Die or neutral density Yag lasers.
- The objective of the measurements, such as determining aerosols, cloud properties, temperature, ozone, humidity, or wind profiles.
- The atmospheric parameters that the system can measure, for example atmospheric density, gaseous pollutants, or atmospheric temperature profiles (horizontally or vertically).
- The wavelength used by the laser source, infrared, ultraviolet, or visible light are some examples.
- The configuration of the lidar system, if it is monostatic, biaxial, coaxial, vertically pointed, or scanning lidars.
- The measurement mode can also be altered depending on the application, if it measures an analog signal or gathers data digitally.
- The platform type influences the cost of the system, it could be stationary in a laboratory, mobile in vehicles, or in situ on balloons, aircrafts, drones, or satellites.
- The number of wavelength measurements also play a role in choosing a lidar system, options are single or multiple wavelengths, also increasing the cost if multispectral systems are preferred.

6.10 Conclusion

The key aspects of GIS and its subsystems, remote sensing, and a brief look at some applications within the sustainable environment and potential risk management of natural or man-made disasters and hazardous conditions are presented in this chapter. Some key highlights include considerations of theoretical and practical implementations during the planning phase of using such systems to monitor, analyze, and predict behavior that could increase productivity or provide valuable information on these hazardous conditions. In GIS, thematic map layering is an important aspect to visually represent gathered information. Any GIS is dependent on several hardware and software limitations, and knowing these limitations before planning to use such a system could potentially save cost and time. Research institutions also actively use GIS and openly available data from various sources to improve algorithms and techniques used in industry and on governmental level.

References

Charya, P. B., Gavrilova, M. L. (2008). Roadmap-Based Path Planning: Using the Voronoi Diagram for a Clearance-Based Shortest Path. *IEEE Robotics & Automation Magazine*. June 2008.

De Berg, M., Cheong, O., Van Kreveld, M., Overmars, M. (2008). Computational Geometry. *Published by Springer-Verlag*. Third Edition, ISBN 978-5-540-77973-5.

Hashemi, A., Kalantari, M., Kasser, M. (2013). Direct Solution of the 7 Parameters Transformation Problem. *Appl. Math. Inf. Sci.* 7(4), 1375-1382, 2013.

Hassebo, Y. (2012). Remote Sensing: Advanced Techniques and Platforms. *Chapter 14*. June 2012.

Kalman, D. (2002). An Underdetermined Linear System for GPS. *The College Mathematics Journal*. 33(5), 384-390, November 2002.

Kelly, P. (2008): Current Problems and Issues in Using SDI in Mega cities. Report of FIG Working Group 3.2: Spatial Data Infrastructures in Mega cities, unpublished.

Liu, F., Xu, Y. (2011). Comprehensive Evaluation of Marine Sediment Quality Based on Attribute Recognition Model with Entropy Weight and GIS. *IEEE 2011 International Conference on Multimedia Technology (ICMT)*, 3468-3471.

Maling, D. H. (1968). Coordinate Systems and Map Projections for GIS.

Thompson, R. B. (1998). Global Positioning System: The Mathematics of GPS Receivers. *Mathematics Magazine*. 71(4), 260-269, October 1998.

Chapter 7
Ubiquitous Computing: Distributing Mobile Computing to Build a Global Network of Things

7.1 Introduction

The term ubiquitous computing, also known as pervasive computing, describes a growing trend towards embedding microprocessors into everyday objects to allow communication to and from these devices, effectively rendering them smart and ideally invisible to people because of their small size and unobtrusive placement. In this chapter, ubiquitous computing and pervasive computing are used interchangeably. The predecessors of this field of technology are distributed systems, followed by mobile computing. The key in ubiquitous computing is using set standards of input and output interface, communication protocols, microprocessor design language and user interface to allow large wireless networks of devices to operate seamlessly. This approach enables integration of low-cost sensors and actuators to create an IoT. Ubiquitous computing depends on the merging of wireless technologies, advanced microelectronics and the internet to create smart devices and hence WSNs able to gather information and communicate inconspicuously. The relatively cost-effective smartphone era has opened up various new possibilities in combining processor power, sensors and actuators, and the internet towards ubiquitous computing. Open-source operating systems such as Android® and academic and commercial variants of Unix® platforms allow developers to take full advantage of these devices' hardware and software layers. Devices and application can be distributed even to developing countries where low-cost smart phones such as the Google® Android® One® is offered, predominantly in India, Pakistan, Bangladesh, Nepal, Sri Lanka, Indonesia, the Philippines and Turkey (at approximately USD $80). Ubiquitous computing facilitates better decision-making based on real-time information on the environment, although it can also present a security risk and several privacy concerns. This chapter aims to describe ubiquitous computing systems and their applications, specifically in developing countries, and discusses the advantages and disadvantages of such systems.

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_7

Pervasive computing, like any other technology, has roots in other forms of networking and communication, still used abundantly, but is slowly moving away from these systems and forming its own individual identity. These systems, briefly discussed in the following sections, are distributed systems and mobile computing. Figure 7.1 represents a summarized history of computing leading to pervasive computing.

As illustrated in Fig. 7.1, starting from the left, the initial computing paradigm started with large mainframe computers around the late 1960s using single-server resources and physical data distribution from a central location. Enhancements in technology and miniaturization of electronic components led to personal computers becoming feasible in terms of size and cost, presenting the popular desktop computer in the late 1980s and early 1990s. Desktop computers could be connected through wired networks to the internet, and were able to access databases to extract information from the internet. Peer-to-peer sharing became more prevalent during this time, mainly through local area networks. Distributed computing aimed to take better advantage of networking resources and started an era where practically any computer device, including mobile devices, could be connected to a network, whether through a wired or wireless connection. Locally hosted servers accessible through cloud computing, grid computing and mobile computing started becoming extremely popular and feasible in terms of cost, the amount of data that could be stored locally and the speed of distributing the data through fixed-line communication. The trend continued towards the 2010s (and still does) and processing power on mobile devices became comparable to earlier desktop computing. In addition, the node size of microprocessor technology shrunk to sizes below 32 nm, leading to processors with billions of transistors capable of digital processing with footprints

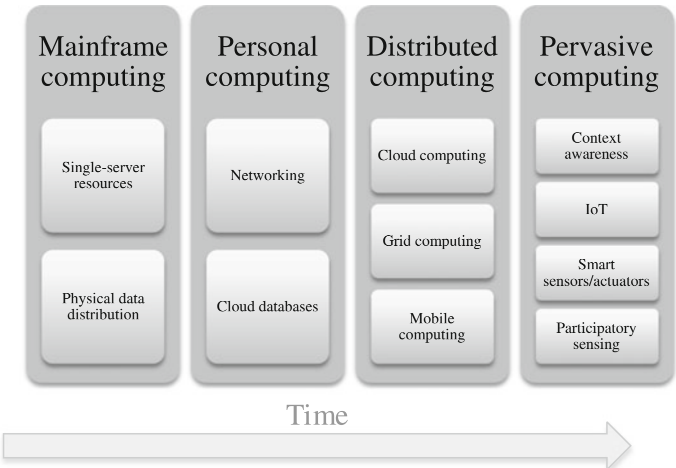


Fig. 7.1 Computing paradigms starting with mainframe computing (*left*) and leading towards pervasive computing (*right*) and beyond

of only a few millimeters. Connectivity also gained much needed traction due to the potential of these mobile devices, and low-energy and high-bandwidth protocols emerged to facilitate WSNs and ultimately the IoT. Applications of the IoT are still being developed at fast rates, arguably too fast for standards to adapt, but the exciting future of the IoT continues to evolve worldwide. In developing countries, however, technical advancements have been adopted at slower rates, understandably since resources are limited and priorities lie elsewhere (food and water). Large corporations such as Google® and Facebook® are, however, aiming to bring these advancements to developing countries, mostly to increase the quality of life of the populations through increased knowledge and information-sharing capabilities. Using technologies such as IoT and WSNs can potentially help developing countries to create smart and sustainable cities, ideally preparing the growing megacities to become sustainable and shifting resources towards expanding and growing the economies using ubiquitous computing to facilitate the process.

Although mainframe and desktop computing also played a key role in developing pervasive computing, the following sections focus on distributed and mobile computing, as these paradigms are somewhat more closely related to pervasive computing and its roots. Pervasive computing combines the advantages of spatially distributed and mobile devices to create unobtrusive networks of data-gathering devices for a practically unlimited number of applications.

7.2 Distributed Computing

Technology advances in communication protocols and miniaturization of devices have led to many similarities between distributed, mobile and pervasive computing and it is therefore more convenient to refer back to the historical invention of each, or at least the time where each system started merging into its own entity. Barbeau (2002) presented the following explanation of distributed systems: “A distributed system includes resources, resource managers, and clients. A resource may correspond for instance to a printer, a window on a software application or a data element. Telecommunications networks are the infrastructure on which distributed systems rely. Concretely, each resource is located on a network node and can be used remotely from other nodes using telecommunications. A resource manager is a piece of software responsible for the administration of a type of resource. It has a telecommunications interface through which users access and update the resources. A manager also enforces access policies associated with each type of resource”. Although this definition was presented in 2002 it is not technically historical, but still dates from a time before smart phones and pervasive computing as it is known and used today, and in view of the speed of technology advancements it could be considered a long time ago. The foundations of distributed computing were researched between the 1970s and 1990s when networking of devices created a conceptual framework involving more than one device that could communicate and distribute information between them, as highlighted in Satyanarayanan (2001).

Clustering of computers to create a single resource with processing power comparable to supercomputers at much lower cost also drove the field of distributed computing from 1993 onwards (Hill 2003). The networking and internal communication between these devices were some of the biggest challenges to perfect, and in a way still are today, since transferring large amounts of data adversely influence the system's performance to a larger degree than the available (abundant in many cases) processing power. Probably the largest difference between distributed computing and pervasive computing is the type of networked nodes. Distributed computing focused largely on sharing resources between multifunction devices such as personal computers, through internal networks or the internet. This limits its capability to integrate with smart cities because of its relatively static approach. Pervasive computing, as will become apparent, also connects various nodes, but the word multipurpose is becoming less prevalent in describing such a network, where application-specific devices are becoming a more accurate description. Evidently, the bulk of journal and magazine articles based on distributed computing appeared in the late 1990s and early 2000s, before mobile computing started heading up the research fields.

Factors that have been adapted all the way through to mobile and pervasive computing include the large dependence on remote communication, whether wireless or wired, high fault tolerance, high availability, remote access and control and emphasis on security of information. Additional techniques and themes particular to distributed computing include coordination, locality, parallelism, symmetry breaking, synchronization, service discovery and uncertainty.

An interesting view on distributed computing came from L. Peter Deutsch from Sun Microsystems in 1994; he and his team identified eight fallacies in distributed computing. Not only is it worthwhile analyzing these fallacies as being almost inevitable in any system, but in addition it serves as a guideline during the design and implementation phase of any new infrastructure. These fallacies, found relatively commonly in literature, are:

- The network is reliable: consider hardware and software redundancy and weigh the risks of failure against the required investment for critical and non-critical application reliability. Non-critical applications might not require large degrees of redundancy and fast response times, lowering the investment costs, whereas critical applications such as railway management software require extreme attention to reliability.
- Latency is zero: the time it takes for data to move between devices is practically never zero and high latencies can decrease overall throughput, since the system must wait for the data to reach its destination. This is particularly evident in mobile connections and generally less problematic in wired (Ethernet) connections.
- Bandwidth is infinite: the amount of data that can be transmitted in a specific timeframe can be a limiting factor in high-bandwidth applications such as transmitting high-definition images or videos. Bandwidth limitations could be perceived as less of an issue in developed countries with advanced

infrastructure, but in developing countries the problem still exists, with internet speeds far below the global average. For example, Niger and Burkina Faso in North Africa have download/upload speeds of 0.6/0.2 and 0.84/0.29 Mbps respectively, compared with Singapore having 97.67/78.69 Mbps download and upload speeds, according to the Internet Society.

- The network is secure: Kaspersky® Lab products reported (Garnaeva et al. 2014) that for the period between November 2013 and October 2014 over 6 billion threats on Apple®, Windows®, and Android® applications were identified. It reports that 38 % of user computers were subjected to at least one web attack per year. In essence, a network can never be considered fully secure. This is a large driving force for specialists to ensure a safe and risk-free environment for users, increasingly more evident also in pervasive computing and the IoT. Norton® Antivirus reported in 2013 that attackers were turning to the IoT in new attacks; baby monitors, security cameras, local routers, smart televisions, automobiles with internet capability, points of sale, and medical equipment were identified as popular targets. Table 7.1 lists the number of attacked users per country and the risk of malicious software infections per country.

From Table 7.1, it is evident that the number of attacked users and even the risk of infection for users are not specifically linked to developing or developed countries. India, Vietnam and Kazakhstan feature in both instances, but so do Germany and the UK. Attackers generally target countries with a large number of users, such as developed countries, since a small percentage of success could potentially lead to a large number of infected devices. In developing countries, security is sometimes lackluster, and penetrating a device or network could be easier. Attackers are also testing their malicious software in developing countries before moving the attacks to more lucrative markets. Therefore, the level of infection and the number of successful attacks are generally higher in developing or poorer countries, as is evident from Table 7.2.

Table 7.1 Kasperksy® Lab products report on the number of attacked users per country and risk of malicious software infection per country, for the period between November 2013 and October 2014

Number of attacked users (out of total)			Risk of malicious infections	
#	Country	Percentage (%)	Country	Risk (%)
1	Russia	45.7	Vietnam	2.34
2	India	6.8	Poland	1.88
3	Kazakhstan	4.1	Greece	1.70
4	Germany	4.0	Kazakhstan	1.62
5	Ukraine	3.0	Uzbekistan	1.29
6	Vietnam	2.7	Serbia	1.23
7	Iran	2.3	Armenia	1.21
8	UK	2.2	Czech Republic	1.02
9	Malaysia	1.8	Morocco	0.97
10	Brazil	1.6	Malaysia	0.93

Table 7.2 Kasperksy® Lab products report on the level of infection for the period between November 2013 and October 2014

	Top 20 countries by level of infection(%)	
1	Vietnam	69.58
2	Mongolia	64.24
3	Nepal	61.03
4	Bangladesh	60.54
5	Yemen	59.51
6	Algeria	58.84
7	Iraq	57.62
8	Laos	56.32
9	India	56.05
10	Cambodia	55.98
11	Afghanistan	55.69
12	Egypt	54.54
13	Saudi Arabia	54.37
14	Kazakhstan	54.27
15	Pakistan	54.00
16	Syria	53.91
17	Sudan	53.88
18	Sri Lanka	53.77
19	Myanmar	53.34
20	Turkey	52.94

According to Table 7.2, the level of infection (success rate) of malicious attacks paints a different picture compared to the risk of infection given in Table 7.1. It is evident from Table 7.2 that many developing or poorer countries have the highest level of infection globally and attackers are aware of the lack of antivirus and firewall software in these countries.

Distributed computing still has a large footprint in antivirus software malware attacks and infections, according to both Kaspersky® and Symantec® Corporation (Symantec Corporation 2014), but the IoT is quickly becoming ever more prevalent in attacks. Another fallacy defined by Deutsch is that the topology of a network will never change.

- Topology does not change: Operations teams will inevitably replace or upgrade components in a network, potentially requiring changes to routing topologies and adding additional storage and processing power that also require addressing and routing changes within the system. Assuming that a network topology never changes can ultimately cost the user and the operations team large sums of money because of lack of initial planning to future-proof a system.
- There is only one administrator: Enterprise systems and large-scale pervasive systems usually have several administrators assigned according to expertise. Databases, web servers, networks, and Unix® systems are a few examples of specialty disciplines. Assuming that one administrator is required, and more importantly, able to manage a system throughout its lifetime, is a fallacy;

administration requires careful planning and risk management during implementation. Developed countries generally have resources available to train system administrators, but in developing countries these resources can be limited or non-existent and infrastructure that is not maintained leads to system faults, resulting in unusable and rejected topologies.

- Transport cost is zero: Transport cost of a network is defined by the transport of data as bits, which adds to bandwidth usage and increased latency and the cost of transmitting through a telecommunications network. Planning for the running costs of a networked system must be prioritized, especially in developing countries where data transport tends to be an expensive commodity, even in 2015. According to the Internet Society, in the Central African Republic, for example (worst statistic in 2015), fixed-line broadband access in the country is estimated at approximately 2194 % of the average GDP per capita, compared to 0.32 % in Macao, China, topping the list as the most affordable. In the Gambia in West Africa mobile broadband cost is 164.28 % of the average GDP per capita (most expensive country), and in Austria, the cheapest, it only accounts for 0.13 %. Any sustainable city or rural area with smart systems for example, is heavily dependent on transmitting the gathered data through networks such as the internet; This can result in exorbitant costs and can render such systems non-feasible.
- The network is homogeneous: James Gosling added an eighth fallacy: a network is generally not homogeneous at application level and it should not be assumed that proprietary protocols will be available to interconnect all parts of the network; choosing components based on standard protocols are commonly overlooked and underestimated. This is again a fallacy where lack of planning is generally only identified later in the system's lifetime when upgrades and additional technologies are planned. Having non-proprietary technology at this point could potentially lead to high costs to effect overall system compatibility for newer technologies; again, in developing countries, this could lead to abandonment of the project entirely.

In summary, the definition of distributed systems might have been structured and accepted decades ago, but the basis of this technology still exists and is relevant for future technologies and topologies such as pervasive computing. The fallacies that plagued distributed systems are still evident in current technologies, perhaps in somewhat different forms, but in essence the same considerations should be accounted for.

Distributed computing followed mobile computing. Smaller computers in the form of laptops, handheld computers, tablets and powerful mobile phones have led to computing no longer requiring physically large and static systems, but mobile components that operate for adequate amounts of time on mobile power sources. A brief outline on mobile computing is presented in the following section.

7.3 Mobile Computing

Nomadic computing, another term for mobile computing, is the use of multi-purpose portable computing devices in conjunction with mobile communication technologies to enable users to access the internet and data from anywhere in the world. From this definition, it is clear that mobile computing does not yet refer to application-specific and fully ubiquitous or pervasive computing, but is rather an additional building block leading to pervasive computing. According to Satyanarayanan (2001), mobile computing could in 2001 already be grouped into five main categories, namely mobile networking, mobile information access, support for adaptive applications, system-level energy savings techniques and location sensitivity. These categories are again closely related to those of pervasive computing, but their applications and use were/are not centered on application-specific devices, distinguishing them from ubiquitous computing in this sense.

Mobile computing suffers from similar inherent limitations as distributed computing and the initial forms of pervasive computing, these being limited range and bandwidth, security concerns, power consumption, transmission interference, potential health hazards (direct and indirect), and practical limitations on the human interface layer due to size reduction of components. Mobile computing did, however, pave the way for ubiquitous access to information, the internet, and bridged the communication gap with instant messaging and the voice-over-internet protocol. Several advantages emerged from mobile computing, including flexibility, mobility, robustness (especially wireless and satellite-based communications that are not affected by earth-based disasters), scalability, ease of installation, low cost and relatively low complexity during infrastructure planning. Mobile computing, in essence, embodies the convergence of mobile phones and desktop computers towards devices such as tablets and netbook computers. The mobile computing paradigm has been superseded by pervasive computing, small application-specific nodes or sensors that are spatially distributed to create a large network of things, gathering data and aiming to improve the quality of life of humans through information sharing.

7.4 Evolution of Pervasive Computing

Pervasive computing started off with numerous products and applications characterized by goals of mobility and ad hoc networking between devices. This phase focused on devices with smaller footprints and more powerful processing capabilities and having at least one method of communicating through radio signals, either between devices or to a central server. A large driving factor of this phase was the miniaturization and integration of various functions within devices, generally designed for alternative purposes, such as smart phones and tablets. These integration techniques converted the devices, already capable of data processing, into smart devices capable of gathering information. Such devices remain somewhat disjointed, as manufacturers design their products arguably in isolation and invent new standards and protocols specifically for their range of devices. The capabilities of these devices are further tailored for specific tasks employing a range of sensors and modules on the device itself, thus designing an application with what is available rather than designing a device that is application-specific. This practice has been very commonplace, since although these powerful computing devices were relatively cost-effective, designing for a single application could not yet be justified and users expected these devices to perform various tasks on a daily basis. A transition phase was noticed during this time, as processor technology increased drastically and the processors and storage became cheaper.

The embedding of pervasive computing into a large number of existing components is the driving force to achieve truly ubiquitous computing that employs various technologies, practically unnoticed by the general population. Relatively homogenous components can be transformed into application-driven nodes and communication devices capable of accurate sensing and data gathering. One factor that is considered to slow down the process of achieving such a goal is the pre-treatment and/or new procedures required to transform these everyday objects into smart devices while keeping the manufacturing costs down to a minimum, since generally many devices are required to achieve truly ubiquitous applications. In developed countries the pursuit of ubiquitous computing is strong and abundant resources are allocated to achieving this goal, whereas in developing countries the uptake on ubiquitous computing has been slow.

To distinguish between these two phases (multipurpose versus application-specific computing), take the smartphone as an example. Smart phones are capable of achieving many tasks effortlessly with their variety of sensors, actuators and communication radios. Applications for these phones are practically countless, and designing an application to serve as a smart alarm that wakes the user based on sleeping patterns can be implemented without the requirement of additional hardware. However, smart phones, especially high-end phones, can become expensive (>USD $500) and many people prefer to use them only as communication devices in the classical sense, making phone calls and delivering messages. The processing power in such a device is more than enough to perform everyday tasks easily. However, a second phase of ubiquitous computing devices can take advantage of older generation processor chips and their lower retail price, and implement the same functionality in a low-cost bedside alarm system that performs similar functions, with perhaps the drawback of higher power consumption due to the older generation processors. Spatially distributing a number of these devices throughout the household could potentially increase the functionality, such as detecting a user's morning routine and adapting the household functions, at a fraction of the cost of multi-purpose devices. This argument drives current ubiquitous computing to levels that can enable city-wide (and extended towards rural areas) monitoring of various environmental, cultural and structural parameters with many more applications when considering scalable cloud technologies through the IoT, as proposed by Xu and Helal (2015), to form truly sustainable cities, households and agricultural lands.

A relatively simple to design and cost-effective technique to achieve ubiquitous computing is distributing active or passive RFID tags throughout an environment to read and transmit specific information about the environment at predetermined intervals or if certain actions are performed. RFID technology has been around for quite some time and miniaturization of these technologies and increasing the data processing capabilities of each node are leading to ubiquitous computing.

7.5 RFID in Ubiquitous Technology

It could be argued that RFID is the backbone for the ubiquitous technology environment, supported also by technologies such as Wi-Fi, Bluetooth®, ZigBee®, and ad hoc sensor networks. Sensing devices such as RFIDs connected through wireless communication systems can capture, process and disseminate useful information on a population's environment. RFID tags, especially passive tags, present low-cost infrastructure that can be spatially distributed over large areas. Although these tags have several advantages, they can also have weaknesses. One such weakness, especially in multi-tag and multi-reader applications, is the phenomenon of false negatives where a tag is present, but not detected (Floerkemeier and Lampe 2004). This can happen owing to several erroneous states, such as collisions on the air interface, tag detuning, tag misalignment and metal or water in close proximity of the system. RFID tags also require a reader to generate incident energy on the tag, which limits the spatial range of these devices. Planning and proper testing of such systems are crucial; implementing a system and disregarding these activities can prove detrimental to overall efficiency and accuracy. Pervasive computing is just as dependent on its operating software as it is on its hardware. Software engineering for pervasive devices has become a popular and widely adopted discipline and some of the principles are discussed in the following section.

7.6 Software in Pervasive Computing Applications

Pervasive computing applications have created new software engineering disciplines and design practices exclusively to achieve real-time operation and efficiency. Software applications targeting specific hardware devices are becoming more commonplace, as opposed to designing software for existing frameworks. Interdisciplinary techniques between software engineering and low-cost system-on-chip designers are becoming more prevalent and the disjointed nature of these disciplines is decreasing with the availability of affordable technology nodes. Older technologies, such as 0.35 and 0.6 μm, are more accessible to smaller startup companies to design integrated hardware with accompanying software in open-source and standardized programming languages. Stewart (2001) called this accompanying software miniature software, capturing the physical nature of the underlying hardware and the new approaches required to the software engineering. Hardware limitations, such as memory storage size, processing power, energy limitations and physical integration areas, have forced software engineers to focus on overcoming these limitations that are not associated with desktop-targeted applications. These desktop application languages, such as Java, COBRA, wireless network protocols and UML-based modeling, are too large simply to be adapted for miniature devices. Embedded processors designed by large corporations such as Microchip®, Texas Instruments®, Motorola®, and Intel® are shifting their focus towards architectures that specifically address the limitations of pervasive computing to increase system reliability through awareness of software limitations. The introduction of small-scale operating systems, such as TinyOS®, ambientRT®, eCOS®, and computation/communication-intensive applications significantly increase the energy consumption of the processor component of WSN nodes. Singh et al. (2014) reviewed and listed several routing protocols designed specifically for ubiquitous WSNs. These protocols are adapted from their article and listed in Table 7.3.

Pachouri et al. (2010) proposed several schemes for future low-power operating systems that are network-based and efficient in running networking applications primarily through cloud computing. The benefits of cloud computing in developing countries are crucial, as the bulk of the processing can be done off-site using powerful servers, reducing the implementation costs and long-term maintenance at these locations, although a steady internet connection is required. The two proposed schemes in Pachouri et al. (2010) address the limitations of ubiquitous computing and aim to use intelligent processing techniques to overcome them. The first schema is designed to follow a procedure to:

- calculate the power of each device’s subsystems and its operating modes,
- schedule algorithms to activate these systems efficiently without affecting the user’s efficiency, and

Table 7.3 Routing protocols for WSNs adapted from Singh et al. (2014)

Category	Representative protocols
Location-based	MECN, SMECN, GAF, GEAR, Span, TBF, BVGF, GeRaF
Data-centric	SPIN, Directed Diffusion, Rumor Routing, COUGAR, ACQUIRE, EAD, Information-Directed Routing, GradientBased Routing, Energy-aware Routing, Information-Directed Routing, Quorum-Based Information Dissemination, Home Agent Based Information Dissemination
Hierarchical-based	LEACH, PEGASIS, HEED, TEEN, APTEEN
Mobility-based	SEAD, TTDD, Joint Mobility and Routing, Data MULES, Dynamic Proxy Tree-Based Data Dissemination
Multipath-based	Sensor-Disjoint Multipath, Braided Multipath, N-to-1 Multipath Discovery
Heterogeneity-based	IDSQ, CADR, CHR
QoS-based	SAR, SPEED, Energy-aware routing

- use information retrieved from performing a set of instructions to calculate the overall power consumption and adapt the scheduling algorithms accordingly.

The second proposed schema follows a somewhat different procedure, summarized by a set of actions during an operation, given by:

- graphing the power consumed and the number of instructions computed per unit time,
- locating the points on the graph where the number of instructions have exceeded the optimum range of computations, and
- scheduling the process accordingly to meet efficiency and operations per time unit specifications.

As with desktop processors initially pushing clock speeds to increase performance (200 MHz processors in 1995 versus 3 GHz processors in 2003), engineers are now rather researching smarter ways to perform a set of instructions per clock cycle to increase performance but without increasing power consumption (through additional transistors). Structuring the pervasive computing system model towards human-centered devices and operating systems ensures devices that are not only capable of data-gathering, but also ensures compatibility with older systems, easy-to-use graphical interfaces and efficient networking. This system model is described in the following section.

7.7 Human-Centered Pervasive Computing System Model

The human-centered pervasive computing system model can be directly related to the open systems interconnection (OSI) model, the main difference being its focus on pervasive/ubiquitous computing. The OSI model is a conceptual model that characterizes and standardizes the communications functions of telecommunication or computing systems without taking into account their underlying internal structure and technology, therefore somewhat of a broad overview. Each layer can be discussed in varying degrees of complexity, depending on the purpose of the discussion (design-centered versus informative). The OSI model is defined by seven layers, namely (from lowest to highest) the physical, data link, network, transport, session, presentation and application layers. A similar approach, defined by Miao and Yuan (2005), characterizes five main layers for pervasive computing with human-centered focus, namely (again from lowest to highest) the human core, human-machine interaction, pervasive device, pervasive access, and pervasive network layers. This layer structure is represented in Fig. 7.2.

According to Fig. 7.2, starting from the top, the pervasive network layer consists of all network connections that are capable of connecting to the internet directly or indirectly. Generally, a network layer is responsible for logically addressing every device on the network, routing of data through a series of interconnected networks and devices, packet encapsulation, fragmentation and reassembly of large (above

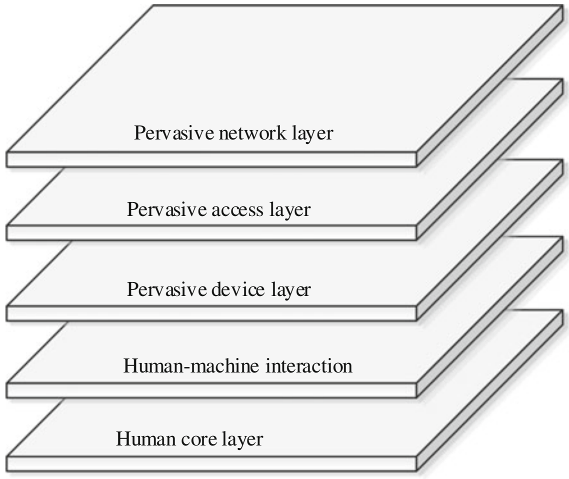


Fig. 7.2 Human-centered pervasive computing system model

the specified limit) messages, error handling and diagnostics between devices that are logically connected. The server and gateway connections and protocols are especially important in the pervasive network layer.

The pervasive access layer can be compared to the data link layer in the OSI model. In the OSI model, this layer comprises wireless or wired network protocols such as Ethernet, Token Ring and the IEEE 802.11 set of standards. In pervasive computing, these protocols are mainly based on wireless, low-power, relatively short-range, and medium throughput protocols, such as Bluetooth® and the IEEE 802.15.4 set of standards. Miao and Yuan (2005) also categorize this layer as predominantly dealing with pervasive network connection issues such as service discovery and management, security and privacy, integration of physical and information space and context awareness on the network side. The key tasks of the access layer are therefore logical link control for the establishment and control of logical links, media access control to control a device's access to the network, data framing to encapsulate the final message in a transferable frame, addressing the hardware address, and error detection and handling of errors that occur at the lower levels of the network stack, such as cyclic redundancy checks.

The pervasive device layer in context refers to the physical devices that are spatially connected to form a distributed pervasive infrastructure. These devices include direct and indirect nodes, devices that are visible (non-enclosed), practically invisible to humans (unnoticed rather than invisible would be a more apt description), and active and passive devices. All hardware must have some relation to the physical layer in order to send data over the network; however, physical devices generally implement additional layers (such as the pervasive access layer) to achieve communication. The pervasive device layer, compared to the physical layer of the OSI model, is responsible for the definition of hardware specifications such as the cables, connectors, radio transceivers and network interface cards, whether wireless or wired (between receiver and server for example). This layer is also responsible for encoding and signaling to transform data bits to transferable analogue signals, data transmission and reception, and the topology and physical network design in terms of its spatial limitations. The pervasive device layer, according to Miao and Yuan (2005), consists of physical low-power sensors and actuators, smart devices and embedded user interfaces.

The pervasive human-machine interaction layer is the layer of interaction between the user and the networked nodes as well as the network-connected devices that facilitate the interface between the user and the presentable data, generally performed by smart phones or application-specific interface devices. The OSI model's presentation layer is comparable to the pervasive human-machine interaction layer, since it facilitates the presentation of the data in a form that is readable and understandable to the user. In addition, in pervasive networks, the format in which the data are received by the receiver responsible for displaying the results also forms part of this layer. Specific types of data-handling issues in this layer include translation between different types of hardware and manufacturers' products, compression to improve data throughput, especially between low-power and low-bandwidth sensor nodes, and encryption to facilitate a degree of security of the transmitted data.

Finally, the human core layer is associated with the application of the infrastructure. The type of application has several impacts on the characteristics of the system, including its spatial distribution, number of nodes required, throughput requirements, accessibility, physical size, network up-time and communication protocols. Examples include applications in health care in rural areas, smart homes and buildings in potentially smart cities, transport, automated toll services and mobile commerce, to name a few. These applications also determine the type of networking and computing infrastructure that best fits the implementation. Such decisions include choosing either mainframe-based or timeshared computing resources, personal computing, network computing such as client-server or peer-to-peer systems, and consideration of the pervasive computing requirements and available resources. The theme of the application largely determines the implementation of each layer, where redundancy, efficiency, throughput, security, accuracy, latency, and presentation of the data all play significant roles in the final implementation. Since the number of applications of pervasive computing is practically unlimited, a few thematic guidelines are presented in the following section, which can be adapted for more specific requirements.

7.8 Pervasive Computing: Thematic Guidelines

Information and communication technologies can provide citizens of cities, especially large cities where populations have grown to a point where individual information-sharing becomes impractical, with efficient services to improve people's quality of life and provide sustainable infrastructure. Smart devices and the IoT have contributed largely to establishing smart cities; however, the fast pace of emerging technologies has left little leeway for implementing standards and cross-platform interoperability. Various international conferences and workshops in the last decade have attempted to address these issues and also to share interesting applications and new knowledge within the discipline. The International Workshop on Pervasive Internet of Things and Smart Cities, the International IEEE Workshop on Pervasive Systems for Smart Cities, and the International Workshop on Smart City and Ubiquitous Computing Applications, in conjunction with the IEEE Symposium on Computers and Communications are three examples. To date (2015), relatively little information in the form of journal articles from ISI-accredited journals is available on the topic of application-specific pervasive computing standards, indicating a field that is currently being explored and paving the way for future standards and innovations. The IEEE Internet of Things Journal is a joint publication of the IEEE Sensors Council, the IEEE Communications Society, the IEEE Computer Society and the IEEE Signal Processing Society which publishes on the latest advances on the various aspects of the IoT and launched in 2014. Pervasive computing is viewed less as a discrete field of technology than as an emerging application of information and communication technology that is integrated into the everyday world. Additional seminars and workshops that aim to further the field of application associated with ubiquitous computing are listed in Table 7.4.

The conferences listed in Table 7.4 are relatively new and have only been held for a few years. For example, the IEEE International Conference on Smart Cities will host its first conference in October 2015; its theme was quoted as: "*As cities worldwide are affected daily by changes in population, climate, congestion, and more, challenges continue to arise…ISC2 will focus on these ever-changing factors and the investments needed to fuel sustainable economic development.*" Similarly, CROWDBENCH 2016 lists its main goal as "*The primary goal of this workshop is to synthesize existing research work, in ubiquitous crowdsourcing and crowdsensing,*

Table 7.4 Worldwide conferences related to ubiquitous computing

Conference	Conference name/description
AH	Augmented Human International Conference with Proceedings in the Association for Computing Machinery International Conference Proceedings Series
WCCI	IEEE World Congress on Computational Intelligence
ISC2	IEEE International Conference on Smart Cities
PAKDD	The Pacific-Asia Conference on Knowledge Discovery and Data Mining
DMCities	Workshop on Data Mining for Smart Cities
IJAIA	International Journal of Artificial Intelligence and Applications
DataCom	International Conference on Big Data Intelligence and Computing
ECSA-SSS	International Electronic Conference on Sensors and Applications—Smart Systems and Structures
CROWDBENCH	Workshop on Benchmarks for Ubiquitous Crowdsourcing at IEEE PerCom

for establishing guidelines and methodologies for the evaluation of crowd-based algorithms and systems." Common themes in these conferences, addressing the future of smart cities, sustainability, and ubiquitous computing, include:

- smart city theory, modeling and simulation,
- intelligent infrastructure,
- sensors and actuators,
- smart economy development,
- open data and big data analytics,
- safety and security systems,
- smart health and emergency management,
- smart environment and policy development,
- citizen engagement and smart governance,
- connected vehicle technologies,
- smart mobility and transportation,
- IoT and smart cities,
- intelligent vehicle-to-infrastructure integration,
- smart integrated grids,
- environmental capital reduction,
- digital city and smart growth,
- smart traffic system operations,
- smart home and smart buildings,
- smart city implementation,
- pedestrian and bicyclist safety; mobility systems,
- smart city for special needs,
- smart manufacturing and logistics, and
- environmental monitoring technologies.

To meet the demands of an integrated system that exists everywhere and can go unnoticed by most users, some characteristics should be noted. These include miniaturization of the nodes and processing equipment, embedding and transformation of smart objects, networking through radio signals, ubiquity, which relates to the perceived invisibility of the devices, and context awareness relating to the information these sensors gather. From a ubiquitous computing perspective, Fig. 7.3 highlights several common themes that address the implementation of ubiquitous services.

Figure 7.3 gives a general structure of thematic applications based on ubiquitous computing towards achieving smart cities using the IoT. It includes among others:

- smart city systems and platforms,
- experiences from real-world experiments and deployments,
- smart city applications, for example within the public transport or environmental control areas,
- end user involvement,
- acceptance and user support aspects,
- participatory sensing and processing,

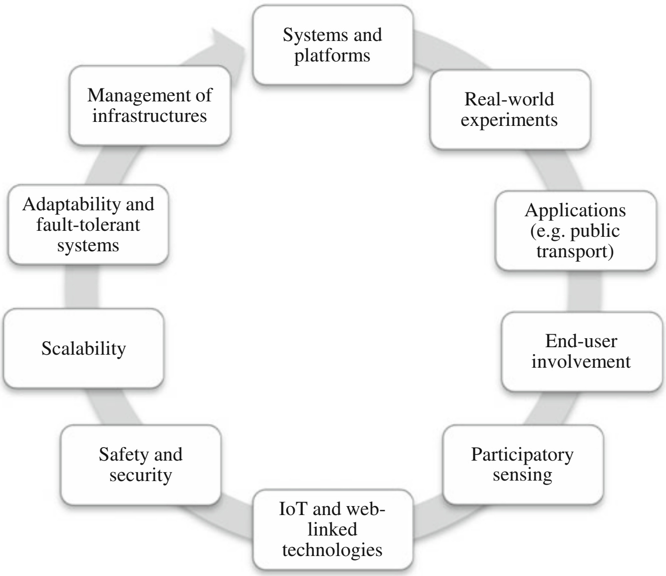


Fig. 7.3 Thematic guideline for ubiquitous applications

- IoT; web and linked data technologies for smart cities (Tragos et al. 2014),
- sensors and actuators and their networking safety, security and privacy,
- scalability,
- adaptable, recoverable, and fault-tolerant smart city systems management, and
- configuration and deployment of smart city infrastructures.

Furthermore, specific applications for urban analysis in developed and developing countries include pervasive systems for the monitoring of:

- economic activity,
- human behavior,
- mobility patterns,
- resource consumption,
- parametric urban design tools,
- walkable neighborhoods,
- urban energy, water, food, and waste simulation,
- responsive technologies, and
- typology of mobility nodes.

The above applications or themes are becoming increasingly easier to monitor through situated social context, the ability to know the social behavior of local populations combined with geographical information, largely attributed to the growing trends in social networking. The situated social context paradigm is discussed in the following section.

7.9 Situated Social Context

Social interaction through the internet has become one of the largest industries to date, with websites such as Facebook®, Twitter®, Instagram®, and messaging applications such as Whatsapp® having millions (even billions) of active monthly users. In developed and developing countries, there are many similarities in the scope and prospective applications of having millions of active users able to communicate instantly, but there are also significant differences in the use of these sites worldwide due to resource limitations. To highlight the potential of current social networking sites and applications, as of December 2014, the ten largest social networks ranked in terms of active users were:

In Fig. 7.4, it is notable that among the ten largest social networks worldwide, two are Chinese websites (Qzone and Sina Weibo) and one a Russian site, Vkontakte, which might be somewhat unknown to the rest of the world. This clearly indicates that large user-based websites also depend on the size of the country, its population, and its accessibility to the internet. Developing countries, such as African countries, are struggling to build large user database websites to shape their local populations' habits, not because of lack of people, but because of the lack of affordable and accessible technologies in poorer countries. This limits the impact of social ubiquitous computing in these nations and technology distribution must first be addressed before such endeavors can gain traction. That being said, technology distribution in many poor countries is not a priority; clean drinking water and a constant food supply through classical means of foraging are still far more important in these countries.

Combining large user databases based on social and geographical context with ubiquitous computing and real-time monitoring can have several advantages (and disadvantages in terms of security and safety of users and their information). To classify and design ubiquitous systems that combine location and social information

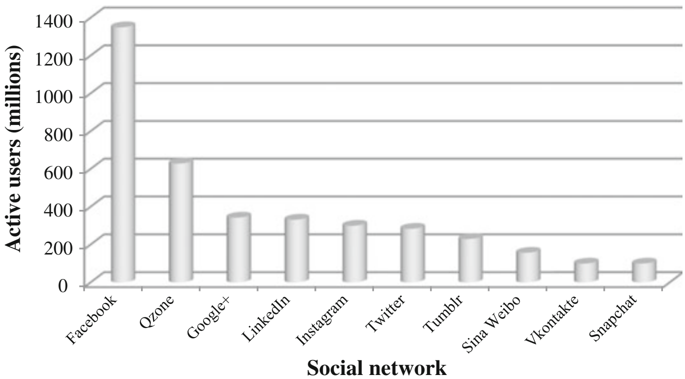
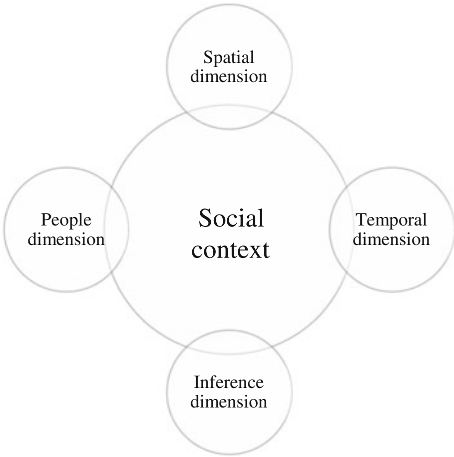


Fig. 7.4 Social network rankings in terms of active users, end of 2014 (data obtained from the Social Times)

Fig. 7.5 Four dimensions of social context (Endler et al. 2011)



for means of user interactions towards situated social context computing, the social context in itself must be clearly defined. Endler et al. (2011) aim to provide a definition of social context through a framework to classify people-to-people-to-geographical places into four dimensions, namely the spatial, temporal, inference and people dimensions, summarized in Fig. 7.5.

The spatial dimension defines the extent of the geographical distance between users and its relevance becomes evident through the choice of establishing communication and interaction between social groups. In the spatial dimension, Uteck (2013) conceptualizes space in three key categories. These are territorial space, which is generally owned (leased) by individuals, groups, or organizations, personal space that humans define as their own physical space related to their privacy, and public urban space, which has a temporary quality but to which any individual has occupancy rights, shown graphically in Fig. 7.6.

The spatial dimension defines the reach and technology investments of social computing and larger populations with fewer technology limitations tend to foster users that share similar geographic locations and hence promote intercultural communication.

The temporal aspect of social interaction, which guides peer discoverability and maintains social establishments and user activities, is determined in the temporal

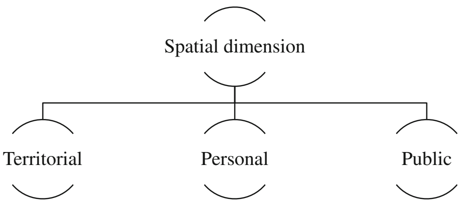


Fig. 7.6 Conceptualization of the spatial dimension (Uteck 2013)

dimension. Temporal in its classical sense is defined by a relation between objects where time is the only limitation. In social computing, the temporal aspect is directly related to the goal of its users/peers. Goals range from short-term applications where an exchange of goods or services is required, to long-term goals and relationships with peers and/or vendors, for example. Geographical information in a social context is again important in this dimension, as the distance between users would affect the adoption and reach of the application.

The inference dimension derives logical conclusions from results of experiences known to be true (or untrue). In the context of social computing, the inference dimension can describe the means used to infer social interactions based on known logic of the users' predisposition. In actuality, this refers to the two main types of communication between users, either as peer-to-peer direct communication or through a centralized service. The tradeoffs between peer-based communication and centralized data processing include expensive server infrastructure able to process large amounts of data versus user-based processing, which in many cases is the less expensive option, but is not be viable if users cannot afford mobile devices capable of such data processing.

Finally, Endler et al. (2011) identify the people dimension, which in the social computing context determines the granularity of people forming the entirety of the social context. Social computing can be targeted at individuals communicating with each other, groups with similar interests and cultural similarities, or at anonymous communities that share information with no regard for a user's identity or background.

Through defining situated social context, pervasive social computing can be explored in further detail, where social interactions and geographic information present their value to ubiquitous computing with enormous databases of user information and personal preferences.

7.10 Pervasive Social Computing

The introduction of social networking and the growing popularity and availability of smart mobile devices have shifted the means of interaction that describe the physical context in which people network and communicate. The pervasive social context can be defined as an integration of social media mining and pervasive sensing. Social media mining is a process of extracting statistical data patterns from social media data and pervasive sensing is linked to the sensing capabilities of mobile devices. The result of the pervasive social context is more meaningful and more targeted social and economic interaction between individuals, groups and organizations. Pervasive social computing entails systems that move the focus from users' physical environment to their social environment and interactions toward individuals that share social interests. Pervasive devices that monitor a user's social interactions, combined with sharing data on social networks, lead to pervasive social computing, potentially a very powerful tool in statistical analysis of people's wants and needs, often invaluable to large corporations to establish their target audience. Pervasive devices in their current form refer mostly to mobile devices such as smart phones and other smart devices, whereas future generation devices would be sensors and microelectronic devices implanted into humans to track their social interactions. In an extract from (Forrester 2006), the following suggestion on social computing is made: "*Easy connections brought about by cheap devices, modular content, and shared computing resources are having a profound impact on our global economy and social structure. Individuals increasingly take cues from one another rather than from institutional sources like corporations, media outlets, religions, and political bodies. To thrive in an era of Social Computing, companies must abandon top-down management and communication tactics, weave communities into their products and services, use employees and partners as marketers, and become part of a living fabric of brand loyalists*".

Examples of pervasive social computing include monitoring the patterns of places an individual recently visited, detecting unusual crowding in certain places, mining landmarks from social content for tourist suggestions, common preferences of local groups and co-located users, sharing of live events, and awareness of closest and most meaningful contacts and service providers (Schuster et al. 2012).

Zhou et al. (2011) categorize pervasive social computing into five interactions, namely behavioral sensing, human-machine interaction, community involvement, content awareness and physical environment sensing. These interactions are depicted in Fig. 7.7.

Human social intelligence can largely be attributed to behavioral awareness (Fig. 7.7) of social and prosodic clues during conversation and other non-verbal interactions. Computing devices are yet to fully incorporate behavior awareness as a key concept in decision-making and understanding behavioral patterns of humans. Voice recognition and pitch analysis have evolved in recent years and research conducted to identify changes in a user's voice based on the current condition

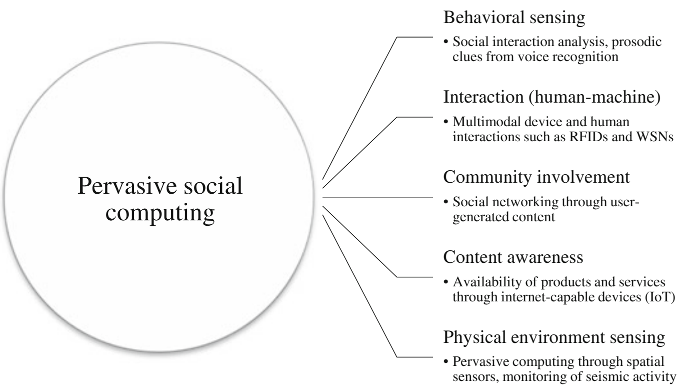


Fig. 7.7 Logical interactions and examples of pervasive social computing

through acoustic processing (Vicsi and Szaszák 2010) has been presented. Behavior analysis has important applications in various fields, including,

- consumer analysis,
- marketing strategy,
- business intelligence,
- customer relations,
- intrusion detection,
- fraud detection,
- event analysis,
- web usage and preference analysis,
- risk analysis, and
- group decision-making.

It is important to consider the fact that behavior in itself is ubiquitous; it happens everywhere and anywhere and practically goes unnoticed. The ability to record and store behavior data is a powerful tool in the applications mentioned above. The study of behavior informatics addresses the use of advanced technologies to evaluate behavior change interventions. The in-depth analysis of human behavior has increasingly been recognized as a crucial means for disclosing interior driving forces, causes and the impact on businesses of dealing with many challenging issues such as behavior modeling and analysis in virtual organizations, web community analysis, counter-terrorism and stopping crime (Cao 2010). Cao (2010) also represents an empirical behavioral model by extracting attributes that represent the general properties of behavior, given by the:

- subject: The entity (or entities) that issues the activity or activity sequence,
- object: The entity (or entities) on which a behavior is imposed,
- context: The environment in which a behavior operates; context may include the pre-condition and post-condition of a behavior,
- goal: The goal represents the objectives that the behavior subject would like to accomplish or bring about,
- belief: Belief represents the informational state and knowledge of the behavior subject about the world,
- action: Action represents what the behavior subject has chosen to do or operate,
- plan: Plans are sequences of actions that a behavior subject can perform to achieve one or more of its intentions,
- impact: The results led by the execution of a behavior on its object or context,
- constraint: Constraint represents what conditions affect the behavior; constraints are instantiated into specific factors in a domain,
- time: When a behavior occurs,
- place: Where a behavior happens,
- status: The stage where a behavior is currently located; for instance, status may refer to passive (not triggered), active (triggered but not finished yet) or done (finished); in some other cases, status may include valid or invalid, and

- association: Other behavior instances or sequences of actions that are associated with the target; behavior associates possibly exist when a behavior has an impact on another, or behaviors are related through interaction and business processes to form a behavior network.

Sustainable cities and smart systems encourage changes in behaviors that prompt new choices and activities, creating a shift in the way one consumes and shares energy and, hence, contributing to the development of low-carbon urban economies and societies, a driving force in urban informatics. Urban informatics uses data to understand better how cities work. This understanding can remedy a wide range of issues affecting the everyday lives of citizens and the long-term health and efficiency of cities, from morning commuting, emergency preparedness, or air quality monitoring, to ultimately create fully sustainable cities. The increasing ubiquity of digital technology, internet services and location-aware applications allows for seamless transitioning between the visible and the invisible infrastructure of cities: road systems, building complexes, information and communication technology and people networks. Building on the need to address emissions and the potential of a smart infrastructure for engagement and management, a core premise is that people make inefficient decisions because of poor information. With better information, behavior change will follow, reducing CO_2 emissions and increasing quality of life.

To make human-computer interaction more intelligent, interaction awareness is required to recognize the type of interaction-related information and adapt to the user's requirements or needs in interaction. Typical human-machine interactions include visual-based, gesture-based, touch-based, speech-based, and sensor tag (RFID) action-based interfaces. Urban areas enhanced with ubiquitous sensors, actuators, and other computational resources allow users to interact, adapt, improve, and contribute to information.

Five dimensional areas of research for pervasive social computing are social signal processing, multimodal human computer interaction, social networking, social media and pervasive computing.

7.11 Social Computing Challenges

Since social computing websites can in certain cases be required to manage millions of people, inherent problems and issues appear from the impractically large active user counts, all feeling (and rightfully so) that they should receive individual attention from the creators of the social network. Nevertheless, companies aim to provide services to all their users, and some might succeed in doing so, but there are additional issues that arise when such social sites are operating and serving millions of users. Some examples of such issues are:

- uncertainty/trustworthiness between the client and the organization due to a feeling of unimportance among the millions of other users (Prandi et al. 2015),
- social data mining for information sharing where consent has not been given,

- security, privacy, and compliance concerns about user information,
- custody of data, relationships and record keeping,
- IT governance and predictability,
- lack of robust reliability and scalability of the physical systems, hardware and related software,
- lack of standards, physical and ethical, plaguing new technologies in their infancy phases,
- usability across a workforce with various levels of comfort in working with social computing,
- productivity and minimization of distractions for end-users,
- time-to-market and investment costs, and
- fraudulent accounts.

The issues and risks that plague social networking are amplified when incorporating an additional hardware layer through pervasive social computing. The additional hardware is another potentially hackable device where sensitive information regarding user details might be stored. To fully integrate smart and sustainable cities using pervasive social computing, the trustworthiness of the network holding a user's information should be considered the highest priority, after network stability and reach.

In conjunction with pervasive social computing, which focuses on users' social interests and interaction to determine their preferences and geographic habits, pervasive service computing operates in a similar manner but with the focus shifted to users' consuming habits, which are discussed in the following section.

7.12 Pervasive Service Computing

Zhou et al. (2010) define pervasive service computing as a service that aims to support its users' everyday activities through a combination of fundamental technologies, service collaboration and service coordination. Their definition of a service is based on the hypothesis that a service is provided by consuming a resource, which in turn is owned by an entity (individual or group), referred to as peers. Peers with similar rights to control a service belong to a peer group. Taking this definition into consideration, Zhou et al. (2010) list the following characteristics of pervasive service computing:

- Service oriented: Employing an architectural computer software pattern for applications that provide a service to other components through communications protocols such as wireless networks. This technique ensures compatibility independent of manufacturer, product type, or technology, assuming the communications protocols match.
- User requirement: Automatic system management and information reporting by a ubiquitous system can only be performed if the user requirement is completely defined. Service peer networks can share information and applications designed

to interpret the data and provide a service to their users in the form of analyzed data.

- Pervasive principles: Service peers and web services interact constantly to ensure that accurate and real-time data are not only efficiently gathered, but stored for historical reference. Applying proper pervasive principles in designing a networked infrastructure is critical to guarantee service delivery to the user in a pervasive and continuous manner.
- Shared understanding: A pervasive system is only as effective as the presentation of the gathered data. Users of service-oriented pervasive systems vary with discipline and do not have only a technical background. Providing semantics-based information exchange can increase the effectiveness and understanding of the data and increase the reach and effect of the system through all disciplines (such as medical staff, human resources, architecture, town planners, etc.)
- Decentralization of information: Independent information sharing and peer collaboration without underestimating privacy concerns are essential in effective pervasive service computing. Service peers are generally likely to have their own policies and less rigid agreements about peer collaboration, and actively encouraging such infrastructure can largely benefit services. Shared resources, as with any other application (not only technology-based), usually leads to larger contributions of people, although they could potentially also lead to misuse if not managed at all.
- Trustworthiness: Following on the above characteristic, trustworthiness is a factor that contributes to the integrity of shared information. If data are shared, peers should place trust in each other regarding the validity of the data, especially if risky transactions are involved in the process. Privacy control, optimized web services and peer relations form the backbone of trust in pervasive computing.
- WSNs: An ubiquitous service computing infrastructure also requires spatially distributed sensors and devices, generally associated with WSNs that interact with each other, which is critical in planning pervasive service computing.

Even though a great deal of research is conducted into pervasive services, no generic model has been developed yet that would support building advanced pervasive services and adapting them to the situation and goals of the user. Bottaro et al. (2007) list five requirements for service-oriented computing frameworks to be effective in the context of home-gateways, namely automated service availability, uniform treatments of remote and local services, therefore hiding network heterogeneity, automatic service adaption in case of interface fragmentation, dynamic service ranking and evaluation of interests, and service continuity, which includes evolving content. Amazon® launched the Amazon® Dash® on 31 March 2015, a small pervasive service electronic device that aims to make ordering of products easier and faster. The device contains an embedded button; if this is pressed, a specific product is delivered to a user's doorstep within 48 h. The device has received mixed reviews from various sources and highlights some of the main problems with pervasive service computing. The products on offer by Amazon® Dash® include washing powder, razor blades, nappies and various household devices. Reviewers of Dahs® generally agree that the device, although innovative and user-friendly, is generally expensive (taking service delivery into account), the product range is limited (which could arguably change if more products are added in time), and a 48-hour delivery time is considered relatively impractical, especially for products offered as “instant relief”, for example headache pills. These issues are difficult to overlook even in developed countries such as the USA where the product was launched, but even more so in developing countries where the same issues would exist, but would be likely lead to such a device not being feasible at all. In addition, in developing countries service delivery is also limited, especially in rural areas, and delivery times would probably exceed 48 h. These shortcomings of pervasive service computing are leading to further research being conducted in the field to develop and manage services automatically and efficiently to support pervasive services adaptively.

Pervasive computing (services and social) in developing countries is primarily aimed at improving people’s quality of life through addressing their basic needs rather than offering luxurious service delivery of household items. Examples of such services, followed by general challenges in these countries, are presented in the following sections.

7.13 Pervasive Computing in Developing Countries

Pervasive computing generally aims to promote applications that can benefit its users in many different ways. Since the applications of pervasive computing are practically infinite, some examples that aim to improve the quality of life in developing countries are highlighted in this section.

Smart classrooms have undeniably found their way into many schools and educational institutions worldwide, especially in the United States and in Europe, to facilitate situated learning (Kurti et al. 2007). Student performance in a learning environment is often associated with the availability of resources, and this should not be a limiting factor in any country. The dynamic availability of learning resources worldwide can be extremely beneficial and many developed countries are using smart devices in the classroom to take advantage of the resources provided by the internet. Currently, educational pervasive computing mainly refers to the availability of the internet at any time and everywhere, through tablets, smart phones, e-readers, or education-specific devices. Students are therefore permanently able to access additional information if they do not understand all the concepts explained during lectures. Teachers are encouraged to move resources towards online content. In the work of Oluwagbemi et al. (2014), trends of smart classrooms relevant to developing countries are outlined and examples given include flipped classrooms. This is a type of blended learning that moves instructional content online, therefore outside the classroom, whereas traditional tasks such as homework are moved into the classroom through hands-on activities. Other methods of teaching that are considered less expensive compared to alternatives where all students must have access to tablets with internet capabilities, are real-time analysis of student's learning experiences of online content, tablet computing through governmental grants that provide a limited number of tablets per classroom and massive open online courses aimed at unlimited participation and open access via the internet, generally designed by specialist institutions and distributed freely. Oluwagbemi et al. (2014) also mention their vision of the future of smart classrooms, involving virtual reality, 3D printing, virtual assistants and educational gaming. Suo et al. (2009) developed a system called the Open Smart Classroom on Open Smart Platform to enable intercontinental and intercultural classes with advanced features, such as a better experience of mobile device cooperation with the ability of easy deployment. The authors envision a future learning system where classrooms connect and collaborate with one another in an open network, while students and teachers from different cultural background, using different languages in different countries, have classes together. It aims to achieve this infrastructure through an open architecture for integrating increasing human-computer interfaces and mobile devices, open service-oriented channels for easy deployment of different requirements and for improving system capability, and an open network in which different classrooms bind with one another.

Moore et al. (2012) present a low-cost alternative using open-sourced software (developed by the authors) and hardware to monitor seismic activity for earthquake early warning systems in countries such as Japan where earthquakes occur relatively regularly. The authors aimed to build a network of embedded computing devices capable of forming a broadcast group across a range of different technologies, including the standardized IEEE 802.15.4-based network protocols to achieve a pervasive computing network. The aim is to deploy their solution in developing countries that lack infrastructure, to integrate more sophisticated technology. Pervasively distributed accelerometer sensors present a more cost-effective solution compared to classical seismic sensors and seismograms to monitor seismic activity, with various nodes distributed and communicating with a central server through low-cost and low-power communication protocols. Data collected by each device can be sent directly to the main network through custom software.

Information kiosks provide touch-based persistent online interactive content, for example agriculturally relevant information such as weather patterns or medical information for the local community. Rural communities and poorer countries might not have individual access to the internet at all times and lack the opportunity to access information at any time; this is where information kiosks become essential. Different types of content-sharing kiosks exist, including information kiosks as mentioned above, advertising kiosks, service kiosks and entertainment kiosks. Priority-wise, information kiosks are most feasible in rural communities. Moore et al. (2015) present a low-cost, offline information kiosk that hosts a large database of information. Information on such a device, if not persistently connected to the internet, can be uploaded and upgraded intermittently, or during short periods when internet access in available. Presenting the local community with articles on contraceptives, life-threatening diseases, traveling information, farming techniques and various other relevant topics can help educate the community without them requiring internet-capable devices and sufficient data from their respective service providers. These kiosks can be designed using relatively low-cost open-source hardware and software and placed strategically in rural areas to serve their purpose. Regular maintenance and new content can be performed by the technical administrators, providing the users with relevant and up-to-date information.

Effective wireless communication in crowded cities and developing countries introduce a framework for sustainable improvement to public transport systems, as proposed by Magdum et al. (2015), where arrival time prediction and indication of seat availability are achieved through pervasive computing. Developed countries are using these systems in many cities and states, commonly paired with mobile applications on smart phones. Technologies that can be combined to achieve a pervasive network are, for example, ZigBee® and GSM networks, where a multitude of low-cost sensors are placed along the route and monitor the passing of buses and other transport vehicles and the availability of seats with pressure sensors on the seats. This information can then be sent through the GSM network to prospective commuters. Updated information can be displayed either at the bus stops on centralized interface devices, or provided as online content to users who can access the information from their own devices. In developed countries this system might sound trivial and might already exist, but developing countries present additional challenges not associated with these regions.

As with various additional examples, including the case study presented by Agushinta et al. (2011), ubiquitous computing in developing countries is still in the development stage and improvements to the infrastructure are required to take advantage of pervasive computing applications that are available anywhere. The study divides ubiquitous computing into three categories, namely ubiquitous mobile applications, ubiquitous web applications and ubiquitous payment system applications, each offering advantages and challenges when planned in developing countries. Pervasive computing challenges in developing countries, based on previous experience described by several authors, are presented in the following section.

7.14 Pervasive Computing Challenges in Developing Countries

Pervasive computing as a computer science on its own is not an easy task to distribute in developed countries and even more so in developing countries. Apart from the technical complexity that large-scale infrastructure presents, additional unexpected challenges inevitably arise. These challenges might include unforeseen technical difficulties, environmental aspects and cultural facets playing a role in successful integration. The implementation of a ubiquitous system by skilled professionals is attainable, assuming resources (time, skilled workers, and financial incentives) are abundant, but maintaining and keeping such a system active could be the biggest challenge. Brewer et al. (2006) outlined a practical approach to challenges their team faced during ubiquitous infrastructure implementations in India, Ghana and Cambodia and highlight these issues. The published work divides the difficulties experienced into three categories, firstly those of a technical nature, secondly, unforeseen environmental conditions, and thirdly, cultural aspects that make widespread adaptability difficult.

The technical challenges Brewer et al. (2006) mention are specifically relevant if considering that their projects were predominantly based in developing countries. The team might not have experienced these difficulties if the projects had been implemented in developed countries. The publication refers to the following challenges, which they categorize under technical challenges:

- Equipment failures: In developing countries, the immediate environment/climate around electrical and electronic equipment is generally not controlled. This leads to components being pushed to their limits in a wide variety of operating conditions, especially large variations in temperature and humidity. Many systems are also constantly exposed to dirt, dust, water, high winds and direct sunlight/ultraviolet rays. Even test equipment such as laptops and handheld computers may be exposed to harsh conditions when traveling between sites, causing failure of these devices. Developed countries are generally considered to have more controlled environments where technologies are implemented, specifically to protect the equipment itself. This, however, requires a large initial investment, which developing countries struggle to meet.
- Power outages: Practically all electronic equipment is in some way dependent on power from the grid. In pervasive computing and WSNs, servers and client-side computers still require alternating current to operate, even if individual nodes and sensors can be powered by renewable energy. This being said, developing countries generally have low-reliability and low-quality power available from the grid. Low reliability refers to frequent power outages, planned or unplanned, causing hardware and software to enter unknown states that must be reset by the administrator or skilled workers. Low-quality power is also a reality in developing countries, often plagued with high-voltage spikes of 1000 V or more (on an average 220 V line), or brownouts, where the power drops below its nominal value for periods of time. These issues affect the power sources of the equipment, which can thus be damaged by even uninterrupted power supplies and have their lifetime significantly reduced, adding to the total cost of the project owing to increased maintenance and repair costs.
- Software infections and reinstallation: Users of personal computers in developing countries are more prone to malicious software and virus attacks on operating systems, not necessarily due to the information these computers hold, but due to the lack of firewall and antivirus security. This effectively means that once a system or infrastructure has been fully deployed, lack of skilled computing personnel means that as the computers are inevitably eventually infected,

the entire system will fail and render the hardware and distributed nodes inoperable. In developing countries there is a definite lack of people educated to handle these tasks and more often than not a system will be offline once infected and only come back online once skilled workers are directed to repair it, which could take months.

- Remote management: Connecting remotely to a computer and managing updates or performing repairs seems trivial if the system is set up properly and a reliable connection is available. However, a reliable connection and a proper setup (see software infections above) are not a given when connecting to a computer in a developing country where mobile and fixed-line broadband are limited. Once a system has entered an unrecoverable state, it would require either a skilled worker to be sent to the site, which becomes costly, or a call to a support engineer to rectify the problem. This could work if the site technician is well informed, but in many cases it would not be a viable option because of the complexity of the problem and lack of technical ability.

Furthermore, apart from these technical difficulties especially prevalent in developing countries, environmental challenges pose further problems when it comes to installing and managing systems aimed at improving the quality of life of the local residents. These challenges include logistical issues and environmental disasters often not planned for and causing long-term damage to systems. Several of these challenges are listed below:

- Transportation: Traveling to and from rural areas can be difficult at times. Lack of scheduling, bad road conditions and unreachable and remote areas are some of the problems that may be encountered in developing countries. In some areas, traveling by road is directly dependent on the weather since rain (and snow) can render a road impassable. Getting equipment to such sites can often be delayed or entirely impossible, which makes maintenance and installations unreliable.
- Customs and shipping: Advanced electronic equipment can be alarming for uninformed border control officials who might mistake crucial equipment for threatening devices such as bombs or sabotage modules. If traveling in developing countries, certified paperwork is not always sufficient to convince officials of the intended use and delivery could be delayed or worse, equipment could be confiscated by ignorant personnel.
- Local purchase and manufacturing: If complex technology is used in any system such as pervasive computing networks or WSNs in rural areas, a high possibility of malfunctioning equipment exists. Replacing this equipment is probably not possible through local vendors and technicians, and proper planning to ensure backup equipment must be done. This increases the cost of the expedition, not only to procure the equipment, but also to transport it. In many cases, performing tasks such as working on network cables becomes near-impossible, since local vendors are unable to supply even the simplest equipment required. Finding mechanical workshops able to machine supporting equipment such as brackets could be difficult, causing further delays in a project if the necessary equipment must be shipped and delivered from international vendors.

- Urban danger: Many poor neighborhoods around the world are plagued with gang violence fueled by drug use. The streets and areas around these neighborhoods are patrolled by non-official authorities and it is a risk to move in these areas without guards or police. Surveying and understanding the needs and requirements in these areas become an impossible task if crime lords and thugs mistake the intent of researchers, technicians or even students. If a technologically advanced system is integrated in a high-risk area, it could be removed if it is perceived as a violation of the local infrastructure and continued maintenance in these areas also becomes a risk.
- Natural disasters: In overcrowded developing countries, especially in cities, a natural disaster can cause major irreparable damage to the infrastructure. This is also true for developed countries; there is little to be done to stop the havoc that can be inflicted by natural phenomena. However, priorities in the aftermath of a large disaster are different in developing and developed countries. Developing countries might have to spend and invest a large portion of their resources in getting the local community back on its feet by focusing on rebuilding housing and restarting the local economy. Technologically advanced monitoring systems would be a lower priority at such times and delay or possible cancellation of repairs to these systems is probable.

The final category that Brewer et al. (2006) list, highlighting this as the most challenging of all, is cultural challenges, especially if the team is unaware of local traditions or beliefs. Cultural challenges often encountered include staff incompetence, tampering, theft, corruption and illiteracy.

- Staff and training: Configuration of a system and a network of components to communicate and share information wirelessly (or wired) is usually done by skilled workers. Once installed and running, local staff must be trained to maintain the system and rectify any possible faults that may occur over time. In developing countries, there is a general lack of staff capable of performing these tasks and basic networking and computing knowledge is limited. Many projects have failed in the long term because of these problems. Furthermore, these problems make it difficult to distribute working systems globally and to gather information not only for local residents, but also for worldwide distribution.
- Tampering and theft: Copper cables and computer equipment are considered valuable commodities in many countries worldwide. Theft of equipment results in failures of entire networks and drives the cost of replacing equipment to sometimes impractical levels. Educating the local communities on the long-term effects of tampering or stealing equipment is necessary, but is not always done. This is not necessarily due to lack of interest in sharing knowledge, but often to underestimating the scope of potential theft in these countries. In South Africa, for example, considered a developing country with industrialized tendencies, USD $160,000,000 a year is lost in the state-owned enterprise responsible for providing the country with power. A large part of this loss is due to bad planning in earlier years, but the situation is exacerbated by corruption in the authorities currently in charge (2015).

- Corruption: Contracting in developing countries can be difficult and upholding a contract agreement even worse. Assuming that local contractors will uphold their part of the agreement or will not make any sudden changes (such as dramatically increasing the cost) could potentially lead to poor relations between the local vendors or contractors and the investment company aiming to provide technological solutions to the population. In many cases it is required to contract local people to perform tasks such as machining supporting equipment or borrowing/transporting heavy equipment, but deciding on a price for such an effort or placing full trust in the ability of suppliers might not be as easy as it sounds.
- Illiteracy: Finally, similar to incapable staff and training, many people around the world, especially in developing and poor nations, are not able to read or write at adult level. This causes many additional difficulties during project development and sometimes requires long-term investment in the population to raise standards up to a point where the local community can contribute to the project.

Focusing on healthcare for a moment to justify these concerns, in an abstract from Brown and Adams (2007), it is well described that ubiquitous healthcare is an emerging area of technology that uses a large number of environmental and human-centered sensors and actuators to monitor and improve patients' physical and mental condition. This technology faces some challenging ethical questions, ranging from the small-scale individual issues of trust and efficacy to the societal issues of health and longevity gaps related to economic status. It presents particular problems in combining developing computer/information/media ethics with established medical ethics. Ethical issues are categorized in terms of privacy, agency, equity and responsibility for errors. The privacy aspect refers to the ownership of medical information and who should have access to it, especially if classed as sensitive information. The agency aspect is focused on the behavioral changes that patients could present, knowing exactly what medical limitations their bodies are encountering in every-day life—posing a question, is too much information (especially regarding a person's body) sometimes a bad thing? Equity would also present a crucial ethical challenge, as has happened for many years in many different and non-related fields. In developing countries, the health and wealth gap between the rich and poor are much more prominent compared to most developed countries, but in either of these cases, would a wealthier person be given priority if it comes to ubiquitous healthcare? Finally, electronic equipment will always be susceptible to erroneous operation, and medical mistakes and wrong diagnoses can lead to large civil court cases where the medical institution can suffer financial setbacks and ruined reputations. Informed consent for ubiquitous health monitoring is not yet clear in many medical institutions and can only be finalized as more cases are reported.

All of the above issues are practical challenges experienced around the world, but unfortunately in developing countries, these challenges are more prominent. This section of the book does not aim to proclaim that people in developing countries are illiterate or untrainable; on the contrary it emphasizes that developing countries have just as much to offer as any industrialized and developed country. It has, however, been found through experience that these issues exist and though they do not make technological advancements impossible, implementing them is certainly somewhat more challenging compared to integrating similar systems in an already-established infrastructure. It all boils down to the amount of money that can be invested in such a project.

Brewer et al. (2006) continue to make five simple recommendations when planning to integrate technological solutions in developing countries. Without delving into too much detail on these, which are presented in the original material, the recommendations are summarized as:

- Proper planning with sufficient windows for alterations from the original plan—ensure that planning is done well, but always be prepared for sudden changes in any part of the project.
- Preparedness for delays: Lackluster regard for time in developing countries generally leads to delays. Ensure that these delays can be managed, even if workers must be sent back and forth between sites, but aim not to abandon a project only based on deadlines not being met.
- Having a plan B: This deals with redundancy, robustness, and expecting the worst.
- Simple UIs: Once delivered, maintenance and operation are in the hands of the local workers. Ensure easy-to-use interfaces.
- Local partners: Build strong and long-term relationships with local partners.

When distributing pervasive systems in developed or developing countries, sufficient documentation, proper planning, alternative resources and skilled workers to ensure success are crucial. Determining the preferences and requirements of the local community is critical, since deploying a pervasive system in an area that does not necessarily require such infrastructure can lead to ultimate failure of the project.

7.15 Conclusion

This chapter highlights the evolution of computing towards ubiquitous/pervasive computing that allows constant access to information and statistics about the environment everywhere. Ubiquitous computing has evolved as a result of device miniaturization and other technology improvements that are making devices smaller and more powerful than ever. Smart phones and tablets are receiving a lot of attention currently for their powerful processing capabilities and wide variety of communication, sensor and actuator options, already built into many of these devices. Ubiquitous computing aims to have more application-specific devices focused on a single task and with abundant processing capabilities to perform each task seamlessly. Crucial to the operation of these devices is their ability to communicate wirelessly between devices and central nodes and to distribute information in ways that are easily accessible and standardized.

Developing countries can benefit immensely from pervasive computing and large open areas and agricultural lands require some kind of monitoring capability, ideally provided by ubiquitous systems. Various practical barriers limit the adoption rate of these systems and infrastructures in developing countries and these limitations were also mentioned in this chapter. Overall, ubiquitous computing has many advantages in the sustainability and future integration of smart environments and as a new and growing technology its potential is unquestionable.

References

Agushinta, D., Kusuma, M., Junatas, B., Trihasta, D. (2011). Case Study: The Condition of Ubiquitous Computing Application in Indonesia. *Ubiquitous Computing*, ISBN: 978-953-307-409-2.

Barbeau, M. (2002). Mobile, Distributed, and Pervasive Computing. *Handbook of Wireless Networks and Mobile Computing*. 581.

Bottaro, A., Gérodolle, A., Lalanda, P. (2007). Pervasive Service Composition in the Home Network. *Advanced Information Networking and Applications.* IEEE, 596–603, 2007.

Brewer, E., Demmer, M., Ho, M., Honicky, R. J., Pal, J., Plauché, M., Surana, S. (2006). The Challenges of Technology Research for Developing Regions. *Published by the IEEE CS and IEEE ComSoc*. University of California, Berkeley.

Brown, I., Adams, A. A. (2007). The Ethical Challenges of Ubiquitous Healthcare. *International Review of Information Ethics*. 8(12/2007), 53–60.

Cao, L. (2010). In-Depth Behavior Understanding and Use: The Behavior Informatics Approach. *Information Sciences*. DOI: 10.1016/j.ins.2010.03.025.

Endler, M., Skyrme, A., Schuster, D., Springer, T. (2011). Defining Situated Social Context for Pervasive Social Computing. *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*. 519–524, 2011.

Floerkemeier, C., Lampe, M. (2004). Issues with RFID Usage in Ubiquitous Computing Applications. *Pervasive Computing Lecture Notes in Computer Science*. 3001, 188–193.

Forrester Research. (2006). Social Computing: How Networks Erode Institutional Power, and what to do about it. Forrester Published Report, February 13, 2006.

Garnaeva, M., Chebyshev, V., Makrushin, D., Unuchek, R., Ivanov, A. (2014). Kaspersky Security Bulletin 2014. Overall Statistics for 2014. Published 8 December 2014.

Hill, M. (2003). Distributed Computing: An Unstoppable Brute Force. *SANS Institute InfoSec Reading Room.*

Kurti, A., Spikol, D., Milrad, M., Svensson, M., Petterson, O. (2007). Exploring how Pervasive Computing Can Support Situated Learning. *Pervasive Learning (2007).* 19–26.

Magdum, N., Maldar, A., Patil, S., Tamhankar, S. (2015). A Low Cost M2 M Architecture for Intelligent Public Transport. *2015 International Conference on Pervasive Computing (ICPC).* 1–5.

Miao, Z., Yuan, B. (2005). Discussion on Pervasive Computing Paradigm. *2005 IEEE Region 10 TENCON*. 1–6, 2005.

Moore, J. P. T., Bagale, J. N., Kheirkhahzadeh, A. D., Komisarczuk, P. (2012). Fingerprinting Seismic Activity across an Internet of Things. *2012 5th International Conference on New Technologies, Mobility and Security*. 1–6.

Moore, J. P. T., Younger, R., Abdelnour-Nocera, J., Rosunally, Y. Z., Kheirkhahzadeh, A. D., Bagale, J. N. (2015). A Sustainable Information Kiosk Driven by Sound. *2015 Sustainable Internet and ICT for Sustainability (SustainIT)*. 1–2.

Oluwagbemi, F., Misra, S., Omoregbe, N. (2014). Pervasive Computing in Classroom Environments and Applications. *2014 Global Summit on Computer and Information Technology (GSCIT)*. 1–6.

Pachouri, A., Sharma, M., Tewari, T. (2010). Green Operating System: Future Low Power Operating Systems. *2010 International Journal of Computer Applications*. 1(21), 77–80.

Prandi, C., Ferretti, S., Mirri, S., Salomoni, P. (2015). Trustworthiness in Crowd-Sensed and Sourced Georeferenced Data. *The 2nd International Workshop on Crowd Assisted Sensing Pervasive Systems and Communications*. 402–407.

Satyanarayanan, M. (2001). Pervasive Computing: Vision and Challenges. *IEEE Personal Communications*. 8(4), 10–17, 2001.

Schuster, D., Rosi, A., Mamei, M., Springer, T., Endler, M., Zambonelli, F. (2012). Pervasive Social Context—Taxonomy and Survey. *ACM Transactions on Intelligent Systems and Technology*. 9(4), 39:1–39:21, March 2012.

Singh, S. K., Kaur, K., Aggrawal, A. (2014). Emerging Trends and Limitations in Technology and System of Ubiquitous Computing. *International Journal of Advanced Research in Computer Science*. 5(7), 174–178, September 2014.

Stewart, D. B. (2001). Miniature Software for Large Pervasive Computing Applications. *Invited Paper at New Visions for Software Design Productivity Workshop*.

Suo, Y., Miyata, N., Morikawa, H., Ishida, T., Shi, Y. (2009). Open Smart Classroom: Extensible and Scalable Learning System in Smart Space Using Web Service Technology. *IEEE Transactions on Knowledge and Data Engineering*. 21(6), 814–828, June 2009.

Symantec Corporation. (2014). Internet Security Threat Report 2014. Volume 19.

Tragos, E. Z., Angelakis, V., Fragkiadakis, A., Gundlegard, D., Nechifor, C. S., Oikonomou, G., Pöhls, H. C., Gavras, A. (2014). Enabling Reliable and Secure IoT-based Smart City Applications. *The First International Workshop on Pervasive Systems for Smart Cities*. 111–116.

Uteck, A. (2013). Reconceptualizing Spatial Privacy for the Internet of Everything.

Vicsi, K., Szaszák, G. (2010). Using Prosody to Improve Automatic Speech Recognition. *Speech Communication*. 52(5), 413–426, May 2010.

Xu, Y., Helal, S. (2015). Scalable Cloud-Sensor Architecture for the Internet of Things. *IEEE Internet of Things Journal*. PP(99), 1–14, 2015.

Zhou, J., Gilman, E., Ylianttila, M., Riekki, J. (2010). Pervasive Service Computing: Visions and Challenges. *2010 IEEE 10th International Conference on Computer and Information Technology (CIT)*. 1335–1339.

Zhou, J., Sun, J., Athukorala, K., Wijekoon, D., Ylianttila, M. (2011). Pervasive Social Computing: Augmenting Five Facets of Human Intelligence. *Journal on Ambient Intelligent Human Computing*. 1–13.

Chapter 8
Designing and Planning for Sustainable Urbanism with Focus on Developing Countries

8.1 Introduction

Sustainable urbanism in broad terms could be defined as integrating walkable and easily traversable urbanism incorporated with high performance buildings and a high performance and currently relevant infrastructure. This definition summarizes the concept of sustainable urbanism adequately, and touches on two important considerations. Firstly, it refers to walkable and easily traversable areas, implying that an urban dwelling should have easy access to a host of local services. Travelling through a city should be relatively effortless, especially to important public areas such as hospitals or police stations. Humans are naturally programmed to avoid complex and time consuming actions, and a sustainable urban area should consider this. The second important factor mentioned in the definition is high performance buildings. High performance is very subjective and in context it ranges from improved structural design and low-cost and effective building techniques to the technologies incorporated within the building, such as structural monitoring, environmental impact assessments, and the distribution of resources such as Internet connectivity, electrical power, and water within all areas of a building. For tall buildings, additional means to minimize effort of accessing higher levels are also considered during smart and sustainable planning. Erecting a building such as the 828 m high, 154 story Burj Khalifa in Dubai and not considering effective means to reach the top would effectively render it useless, or unsustainable. The third term, high performance infrastructure, refers to the same type of planning compared to planning for individual buildings and structures, but on a larger scale such as city-wide planning. Interconnecting buildings, roads, pavements, and accessibility to services, should be prioritized during planning of a self-sustaining city. It is seldom that a city is built from the ground up, and generally sustainable urbanism planning is based on the expansion of cities from large numbers of people inhabiting it, but the same principles apply, granted there are several limitations to

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_8

identify and overcome in this process due to existing structures and ageing infrastructures.

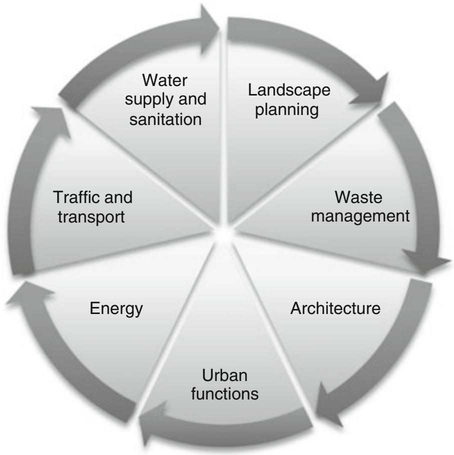
Three key components should be prioritized in any sustainable city. These not only include high performance technologies that simplify information distribution, but also several social, environmental, and economic factors. The three factors are summarized in Fig. 8.1, being the environment, the local community, and the growth of the economy.

From Fig. 8.1, the three key factors to a successful and vibrant sustainable city are highlighted, and the importance of each of these are arguably equally distributed. If looking at economic prosperity for example, in a sustainable city, economic prosperity does not only refer to increasing profits and expanding markets. The key is to achieve prosperity, but not at the expense of the environment or social harmony. Planning should include creating jobs and markets that contribute to reduced waste generation, conserves energy, prevent pollution, conserve natural resources, and educate the city's workers. In doing this, several additional goals are automatically achieved; advancing human capital, developing critical skills, and encouraging collaboration between individuals and industries. Secondly, from Fig. 8.1, environmental and ecological integrity, as the title suggests, aims to reduce emissions of greenhouse gases, keep the air and water supplies clean and risk-free in a way that future developments can easily adjust to. To achieve both economic prosperity and environmental integrity, the local population should work together towards a common goal. A socially and culturally stable population is crucial to encourage collaboration even if people do not share the same religious or political beliefs. In preserving harmony and tolerance, livable neighborhoods and sustainable goals can be achieved. Deeper discussion into these three key elements are presented in this chapter.

It is also possible to identify several other areas or categories in city function that can be adapted for not only sustainable, but also towards a smart city. Although it is practically impossible to list all possibilities to make a city smart, these categories

Fig. 8.1 Three key considerations for any sustainable city

Fig. 8.2 Seven categories that are generally addressed when planning or expanding sustainable cities



serve good purpose as a guideline of the sectors that can benefit from technological advancements in knowledge generation and sharing. The categories or sectors are briefly outlined in Fig. 8.2.

From Fig. 8.2, it should be note that each category has equal priority. Each category or sector can function independently (although not advised), but also contribute to all neighboring sectors, even more if transformed to a smart infrastructure. Smart technologies and sustainable practices have many advantages such as helping distribute fresh water supplies, providing efficient sanitation and smarter cost-effective architectural designs, managing waste water without disrupting residential areas, planning and placing urban functions conveniently, reducing overall energy consumption, generating alternative energy, and improving transportation routes and public transport.

It must also be highlighted that sustainable urbanization is not an easy goal to achieve, and there are several limiting factors that influence the success of such an objective. Developing countries in general have intense limiting factors that make it difficult to reach sustainable levels in all or any of the above categories. Countries ideally should have common goals and values they aim to achieve, for example a commitment to the future of its cities and rural areas, a business and prosperous perspective, research tenacity, encouraging creativity of its citizens, flexibility, and connectivity. Developing countries have to cope with fast rates of urbanization making it difficult for the city to adapt at this pace. Also, large groups of its population live in poverty and in slums and sometimes seek unrealistic expectations. These countries also have an aging population, who might be unwilling to accept change.

A characteristic of practically all areas of land on earth, with of course some exceptions such as the polar regions and desserts, is that these areas have some degree of sustainable resources that can be utilized to grow and develop a sustainable and economically stable city. It is however important to faction in the fact that no life can be supported without first ensuring that the basics needs of man are well looked after. One such awareness endeavor, is the United Nations Millennium Development Goals (MDGs) established in 2000, and the results of this summit are discussed below with focus on sustainable development.

8.2 United Nations Millennium Development Goals (2015 Update)

The MDGs (United Nations 2015) were established in 2000 following the Millennium Summit of the United Nations (UN) in 2000. Eight MDGs were set out during this summit, and at the time all 189 (now 193) United Nations members and an additional 23 international organizations committed to help achieve these goals. The eight MDGs that were developed aimed to:

1. eradicate extreme poverty and hunger,
2. achieve universal primary education,
3. promote gender equality,
4. reduce child mortality,
5. improve maternal health,
6. combat HIV/AIDS, malaria, and other diseases,
7. ensure environmental sustainability, and
8. develop a global partnership for development.

These goals are ambitious and the aim of this summit was also to raise awareness and promote global collaboration towards common goals (Zhan et al. 2012). It provides a clear development framework for all institutions part of the initiative to achieve a sustainable and fair future for all nations, especially developing and poverty stricken countries. Within the scope of this book, MDG 7, which aims to ensure environmental sustainability, is of particular interest. All other MDGs are equally important, and the overall aim of a better life for all should remain priority. Regardless, the UN released a report in July 2015 regarding the status of the MDGs to date. Table 8.1 lists the achievements reached of MDG 7.

From Table 8.1, it is noted that great progress has been made in ensuring sustainable development worldwide. This chapter is dedicated to sustainable urbanism, and it is evident throughout the chapter that adequate water supply, low values of pollution, and social and economic stability are integral in achieving sustainability as it forms the backbone of human survival. These goals are not specific to sustainable urbanization and it holds true for generally any human settlement, regardless of its size or geographical location. The aim of this section is to highlight the importance of these issues throughout the discussion of achieving a sustainable urban development, where technological advancements and basic necessities must combine to create smart and livable cities. Resources must be used carefully and wisely in any settlement since it forms the pillar of the local economy. These potentially sustainable resources, whether non-renewable or renewable, are discussed in the following section.

Table 8.1 United Nations Millennium Development Goal 7 achievements between 1990 and 2015

Issue/goal	Achievement (1990 versus 2015)
Drinking water	In 2015, more than 90 % of the global population is using an improved drinking water source, compared to around three-quarters in 1990. About 2.6 billion people have gained access to improved drinking water since 1990 and 1.9 billion people have gained access to piped drinking water.
Sanitation	Worldwide 2.1 billion people have gained access to improved sanitation. The proportion of people practicing open defecation has fallen almost by half since 1990.
Ozone depletion	Ozone-depleting substances have been virtually eliminated since 1990, and the ozone layer is expected to recover by the middle of this century.
Slums	The proportion of urban population living in slums in the developing regions fell from approximately 39.4 % in 2000 to 29.7 % in 2014.
Protected areas	Terrestrial and marine protected areas in many regions have increased substantially since 1990. In Latin America and the Caribbean, coverage of terrestrial protected areas rose from 8.8 % to 23.4 % between 1990 and 2014.

8.3 Sustainable Resources

Resources are fundamental to any economy and by using and transforming it, the wealth of past, present, and future generations is quantified. The largest difficulty is considering future generations. Scientists warn of the threats to the earth's environment resulting from human activity, pointing to the destruction of biodiversity and the disappearance of entire species as a result of a scarcity of resources and the forceful spread of modern lifestyles. Current usage of resources is depleting sources fast, without proper planning for future generations, and specifically in developing countries, the need to use these resources now outweighs the necessity to conserve it for the future. Natural resources like land and other raw materials can be found as naturally occurring substances. The value of these deposits is usually dependent on the amount available for extraction (supply and demand). Not only are resources being depleted at alarming rates, the impact on the environment around the excavation sites leads to the elimination of natural habitats. The European Commission proposed a thematic strategy on the sustainable use of natural resources in Europe to raise awareness and combat malpractice in resource use. The objective of the strategy is to reduce the environmental impacts associated with resource use and to do so while supporting a growing economy. Such a strategy drives sustainable resource management in the right direction, its enforcement however is a challenging objective. Generally, if the amount of available material for extraction as well as the ease of the extraction techniques makes exploitation commercially viable then the value of the land increases and human nature takes control of the situation, and in many cases, this unfortunately refers to our destructive nature. If natural resources (or sometimes called natural capital) is used up faster than it can naturally be replenished, which is the case with fossil fuels, which can take millions of years to form, metals, and fresh water sources, it creates an unsustainable ecosystem destined for perpetual extinction.

Natural resources are divided into two broad categories, namely non-renewable resources, and renewable resources. Technically these two categories are identical, both in essence can be replenished, the crucial difference being the time it would take to renew a depleted (or stagnant) resource. Non-renewable resources generally refer to sources of natural materials that could take thousands or millions of years to form by complex processes of heat, pressure, organic activity, and weather conditions within the earth's core. Examples are earth minerals, metal ores, fossil fuels such as coal, petroleum, and natural gas, nuclear fuels (uranium), and groundwater in certain aquifers. Conversely, renewable energy refers to sources of energy that are considered environmentally friendly by harnessing energy of natural processes. These sources are called renewable since using them does not deplete the resource creating the energy. Many large-scale renewable sources exist, but small off-grid applications are becoming more important in rural and remote areas where energy is a crucial commodity for human development. Renewable sources such as tidal and wave power, wind power, solar power, radiant energy, geothermal power, biomass, compressed neutral gas, and to some extent nuclear fission, are considered to be sources that cannot be depleted. In October 2014, it was reported that developing nations were adding capacity from renewable energy projects at nearly twice the rate of developed countries. The surge reflects the economic advantage that cleaner technologies have in emerging markets, which are scaling up energy installations to match the demands of expanding populations and economies. The survey of 55 countries, including China, Brazil and South Africa, found that combined renewable projects grew by 143 % from 2008 to 2013. Wealthier nations, by comparison, saw renewable source energy generation increase by 84 %, according to the report by Climatescope. Renewable projects included solar panels, wind turbines, geothermal facilities, and biomass plants but excluded hydroelectric facilities.

Some renewable sources however, can be depleted if not managed properly. These include large areas of enriched soil, fisheries, and forests. The time to renew the energy source is not thousands or millions of years, but if overused, these sources can be depleted and could take large investments and significant time to restart any form of harvestable life in these areas. Developing countries are especially guilty in miss-managing these trades, and tend to ignore future repercussions. Ocean fisheries off the coast of Indonesia and Malaysia for example use dynamite to kill thousands of fish to simplify the harvesting process, however also destroying areas of coral reefs, home to millions of fish and other organisms, permanently. Malaysia and the Philippines are also known for large scale deforestation due to overexploitation and diminishing the resource base and fruitful soil. These processes inevitably lead to practically permanent damage to ecosystems, pollution of air, water, and soil in the surrounding areas, and destruction of the natural habitat of plants and animals.

The phrase "tragedy of the commons", first described by biologist Garrett Hardin in 1968, describes how shared environmental resources are overused and eventually depleted. The theory of tragedy of the commons can easily be explained through various examples. One such example, especially relevant in urbanization, is traffic congestion. Highways are built to enable citizens to travel at relatively high speeds, therefore reducing the time needed to travel large distances. The tragedy here is that the more people that make use of highways due to its convenience, the more congested the highways become, effectively rendering the main purpose defeated and again increasing the commuting time to levels possibly worse than before. Another example is in renewable sources where fisheries frequently suffer from tragedy of the commons. Shared resources, being the fish, are eventually depleted as individual fishermen and large corporations hone in on fruitful areas, initially designated for sustainable and limited fishing. The growth in population and demands of all humans contribute to these fisheries becoming depleted, as supply and demand cannot be balanced. Other examples include deforestation, pollution, and depleted groundwater resources from high demands of the global population. An estimated 5.2 million hectares of forest were lost globally in 2010. Overexploitation of marine fish stocks led to declines in the percentage of stocks within safe biological limits, down from 90 % in 1974 to 71 % in 2011. Underprivileged low-income families' livelihoods are more dependent on natural resources, and these families often live in the most vulnerable areas, suffering the most from increases in the price of food (due to lower supply) in its processed state.

The solution? Unfortunately, there is no official solution to the problem of tragedy of the commons. Hardin (1968) points out that this problem is an example of a problem with no technical solution and states that "a technical solution may be defined as one that requires a change only in the techniques of the natural sciences, demanding little or nothing in the way of change in human values or ideas of morality". Therefore, any potential solution requires that we, as a society, change our values of morality. As expected, this is not an easy feat, and relying on a global change in values of morality is probably not something that will be achieved unless a global crisis enforces it. Developing countries do still have a significant advantage in developing sustainable cities with resource depletion in mind, but as history dictates, the process is bound to unravel as insufficient practices to feed masses of people become priority above sustaining the environment. Developed nations already have large infrastructures in place, not all being sustainable, and expansion must be done with sustainability in mind, often leading to redesigning current infrastructures. Referring back to Fig. 8.2, the categories of creating a sustainable city must be carefully planned, and if responsible development is practiced, many risks and environmentally insufficient practices can be eliminated.

8.4 Responsible Development

Responsible sustainable resource development refers to the management of air, land, water, and biodiversity and the efficient and fair governing of these resources. The term responsible development is another example of a very open-for-discussion term, with various sources stating different practices to achieve a common goal. The

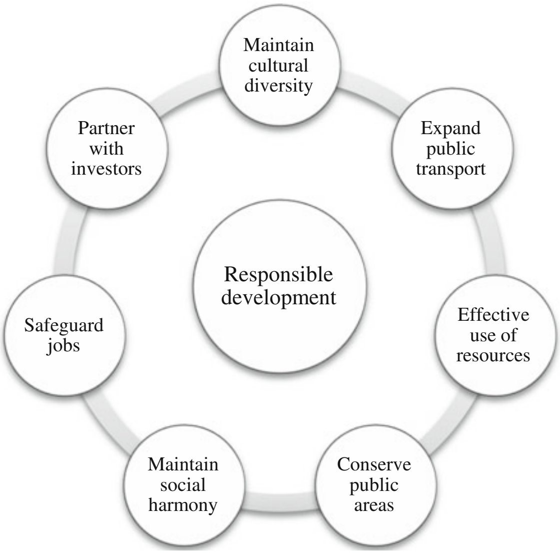


Fig. 8.3 Categories of responsible development in sustainable cities

vast number of topics, disciplines, and applications, each consists of its own responsible development drivers, but there are some common objectives that lead to successful integration into smarter and more sustainable cities that provide fast and reliable information to its citizens. Developing countries can use these objectives as guidelines, and manipulate each objective for their applicable applications during their own urban developments. These key categories are highlighted and described in Fig. 8.3.

The first objective, again not prioritized but chosen arbitrarily as the first objective, is maintaining cultural diversity and the idea of conservation of an individual's or a group's identity. From (UNESCO and UNEP UNESCO and 2002) cultural diversity can be defined as the capacity to maintain the dynamic of transformation in all of us, whether individuals or groups. There exists a wide range of distinct cultures worldwide. Defining a particular culture can prove more difficult to distinguish than what might appear at first sight, leading to certain cultures not being understood by the majority and eventually being outcast. Moreover, awareness of this diversity has become relatively commonplace, being facilitated by the globalization of exchanges and the greater receptiveness of societies to one another to drive social sustainability. While this greater awareness in no way guarantees the preservation of cultural diversity, it has helped to give the topic greater visibility. Cultural diversity emerged as a key concern at the turn of the new century. There are predictions that globalization and the liberalization of the goods and services market will lead to cultural standardization, reinforcing existing imbalances between cultures. Cultural diversity, beyond the mere fact of its existence, has aesthetic, moral and instrumental value as the expression of human creativity, the embodiment of human strivings and the sum of humanity's collective experience. In the contemporary world, characterized by the social interactions linked to the speed of new communication and transportation technologies, the growing complexity of these interactions, and the increasing overlap of individual and collective identities, cultural diversity has become a key concern, amid accelerating globalization processes, as a resource to be preserved and as a lever for sustainable development (UNESCO 2009). Cultural diversity is not a static body that only requires conserving; it is a baseline for incessant and amalgamating dialogue between expressions of identity. This identity must be upheld if diverse groups of people are to work together in any environment. During the Johannesburg Declaration on Sustainable Development, it was stated that "we (South Africa) are determined to ensure that our rich diversity, which is our collective strength, will be used for constructive partnership for change and for the achievement of the common goal of sustainable development." This statement summarized the potential strength in cultural diversity through promoting dialogue and cooperation among the world's civilizations and peoples, irrespective of race, disabilities, religion, language, culture, and tradition. As positive as this message was conveyed during the summit, it must not be ignored that all of these differences in people's beliefs and habits could negatively impact any nation if not managed. The concept of sustainable development, based on a clear understanding of the role of biological and cultural diversity in maintaining ecological systems, cannot be viewed exclusively through an economic viewpoint that only considers technological progress. Globalization tends to create a context favorable to interdependence, often to the detriment of the least developed countries and without consideration for the diversity of cultures. Challenges arising from globalization are making it increasingly important to redefine the relationship between culture and development or as a perspective between cultural diversity, biological diversity and growth.

Referring back to Fig. 8.3, the second category of responsible development is the (sustainable) expansion of public transport. If read out of context, expanding public transport might seem superfluous as opposed to being essential. This is in fact not true, as public transport has many ripple effects in a community. Sustainable transportation can enhance economic growth, promote trade opportunities, and improve accessibility. Sustainable, reliable, and safe transportation achieves better integration of the economy while respecting the environment. (Yigitcanlar and Kamruzzaman 2015) identifies nine commonly considered options for sustainable urban transport in cities in developing countries, inclusive of megacities. This article continues to focus on smaller cities with inhabitant populations less than one million people to support these statements and highlight how the options can be implemented on a smaller scale from smaller initial investments, and eventually be transferred to megacities. The options to consider are:

- the quality of the road infrastructure and conditions to and from cities,
- the availability, reach, and effectiveness of railway-based public transport,
- the availability, reach, and quality of road-based public transport,
- support and maintenance for non-motorized travel modes, such as cycling or walking,

- implementation and planning for technological solutions to increase efficiency and capacity of existing routes,
- campaigns to raise awareness towards alternative travelling modes to avoid congestion of the most-used methods,
- competitive pricing over the entire range of travelling modes to encourage variety,
- vehicle access restrictions to maintain roads and railways, and
- effective control of land-uses to sustain and maintain the surrounding environment.

To highlight technological solutions from a perspective of pollution control, many developed countries are looking towards alternative-fuel type vehicles, such as electric or hybrid vehicles and, predominantly in developed countries, intelligent transportation systems (traffic surveillance, commercial vehicle tracking, electronic ticketing and toll collection services, bus management systems, and traveler information services to encourage ride-sharing). Technology is used to decrease various transport-related issues, such as air and noise pollution, the dependency of oil imports, relieving traffic congestion, and reducing the number of fatal accidents. Although technology solutions can be effective in combating these concerns, the cost-effectiveness and attainability in developing countries are generally difficult to justify. Alternatively, natural gas transport vehicles are somewhat easier to implement in such scenarios, as can already be witnessed in South America (Brazil and Argentina) and several heavily populated Asian countries (India, China, and Pakistan). Financial incentives or specific mandates from government towards the use of natural gas can drive the initiatives towards cleaner energy consumption in the transportation sector. Electric cars, again a costly alternative for developing countries, will take time to penetrate these markets as viable alternatives to internal combustion engine vehicles and non-motorized travel. The overall consensus is that sustainable transport is significantly more attainable when employing a variety of transportation modes as opposed to accounting on a single solution. Infrastructure upgrades and improvements by governmental bodies can help to alleviate the already stressful situations on many roads, and with support of the international community, key considerations in sustainable development include:

- Promoting transport sustainability by integrating economic, social and environmental considerations in decision-making within the transport sector;
- Developing transportation systems which are responsive to development needs and reduce negative environmental impacts;
- Managing transportation demand and facilitate the flow of and access to goods;
- Reduce noise pollution from transport and make use of better vehicle technology if financially feasible;
- Improved inspection and maintenance schemes for vehicles;
- Cleaner conventional fuels, as well as the development and promotion of alternative fuels;
- Involving the private sector with financial initiatives towards improvements in efficiency and emission control of each mode of vehicle;

- Promoting information technology to provide real-time information and status updates on potential issues or risks on local roads;
- Facilitating research and development and technological innovation in the transport sector at university level;
- Promoting links between different modes of transport with an aim to make more efficient use of existing infrastructure;
- Promoting access to efficient, safe, affordable and environmentally sound public transport systems, including for rural, remote, urban and inter-urban transport services;
- Undertaking information distribution aimed at promoting road safety;
- Maintaining and promoting access to affordable transport systems, and examining the potential for increasing reliance on low-cost, readily available modes of transport, including non-motorized transport;
- Promoting gender-sensitive planning and planning for the aged and disabled for transport services and systems, and increasing participatory, inclusive transport planning approaches which address social needs of the local community;
- Promoting public participation in transport decision-making involving all stakeholders and access to information and enable consumers to make informed choices;
- Encouraging the planning for and provision of safe infrastructure for cycling transport.

The third objective, as described in the section pertaining to sustainable resources, is the effective use of natural resources. Renewable and non-renewable resources must be harvested and used responsibly, since depleting these resources is generally detrimental to the entire economy. Important consideration concerning the effective use of resources is given in the sustainable resources section within this chapter.

Cities may no longer be essential for a technologically advanced and a prosperous civilization, but they embody a sense of identity, history, public space, and social diversity. Humans crave social interaction, generally in relaxed and comfortable environments. Public space is an area where all social classes can share and develop culture, religion, political stand, hobbies, interests, and engage in general conversation. Conserving of public spaces is another category that plays an important part in responsible development, not only from an environmental standpoint, but also in upholding morale and cultural heritage of local populations. A sustainable city must aim to provide universal access to safe, accessible, and green public spaces. Urban public spaces are needed to sustain the productivity of city workers, social cohesion and inclusion, civic identity, and quality of life. Uncontrolled rapid urbanization in developing countries creates disorderly settlement patterns with dangerously low shares of public space. Many cities in developed countries are also experiencing a dramatic reduction of public space. Public spaces and urbanization are at battle for habitable space and rising prices of real-estate in cities are a driving force of this battle. The importance of public spaces should not be underestimated and its land value, or potential value, should not be used as measure of its worth. Public spaces are generally not as easy to design as it might appear, or at least effective public spaces that draw people of various cultures to feel relaxed and comforted. Erecting a public space that does not have an approachable aesthetic for any culture or social group is pointless, and the Project for Public Spaces lists several guidelines when planning a public space within urban areas. The guidelines include:

- Approach the local community to provide a historical perspective, insights into how the area functions holistically, and identifying meaningful events and structures within the nearby area. In doing so, the community will immediately feel part of the area and will be drawn to participate in events or as means of relaxing and unwinding without feeling threatened.
- People tend to avoid overly grandiose places for fear of rejection as well as dark, dreary, or seemingly abandoned areas for fear of safety. Bearing this in mind, the design of any public space should be somewhere in the middle, an inviting area without posing grand structures that might seem exclusive by the general community. People should feel welcome and willing to express themselves without the fear of rejection.
- Observe other public places, areas that have successfully drawn diverse crowds and equally so, areas that have failed to meet the standards and demands of city-goers. Possibly approaching communities in popular public places and surveying them about their likes and dislikes of the particular area could add valuable insights. This might however estrange citizens from these areas as they prefer not to be engaged in market-research or any business related activities during their off-time. Careful and considerate observing of these areas can prove useful in planning for new public places in urban areas.
- Public places and open-area infrastructures also require collaboration between several entities such as local service and goods institutions, historical museums, schools and universities, or any other institutions that have been operating within the area for an extended period of time. The collaboration from these institutions can offer valuable insight into the preferences of the local community.
- There should always be a sense of pride in the people using or passing the public places, generally achieved through the inclusion of momentous occasion memorabilia or historic event artifacts or statues. Citizens will find pride and accord with such places. Additionally, historical objects could be cornerstones for the local community to show to guests and inform them about the local history, with a sense of pride instilled in them if the public place is well maintained.
- Large initial investments for public places are generally not the ideal way to start designing and implementing these areas. A commitment for large initial investments and resources will take time. It is proposed to start small, adding luscious grass and colorful flowers, benches, and light physical activity equipment, and expand on these as the community starts embracing the area.
- Budget should not be deterring factor of erecting any public place for the enjoyment of the community. Starting small is one way of limiting the overall budget, but additional finances can be obtained through private and public

investors, fundraisers, and government grants. Many cities have recently adopted methods of overnight urban ambush initiatives where non-profit organizations gather old and non-used materials and convert public places into friendly and inviting gatherings for the community, expecting no incentive for their initiative.

From the above discussion it is easy to see that inviting public areas contribute to sustainable development of urban areas. This also aims to invest in retaining social harmony between different cultural and religious groups, and maintaining interaction between them not only in recreational activities but also in the workplace.

According to Fig. 8.3, responsible development also requires the safeguarding of existing jobs and creating new sustainable jobs, particularly during the development and planning phases of these cities. In developing countries, new jobs are generally created during these phases, but the sustainability of these jobs are not guaranteed after the work is completed. Enterprises, workers, and governments should not be passive bystanders in the transformation of economies. They are considered essential agents of change, able to develop new ways of working in sustainable enterprises that safeguard the environment, create good jobs, and instigate social inclusion. The following extract is taken from the UN Industrial Development Organization report on Inclusive and Sustainable Industrial Development: "… the share of manufacturing value-added jobs created in developing countries has almost doubled in the past 20 years, from 18 % in 1992 to 35 % in 2012. The structural transformation that occurs when economies move from a high reliance on agriculture and natural resource extraction to activities that foster local value-addition and related services has a dramatic development impact. It unleashes dynamic and competitive economic forces that generate employment and income, facilitate international trade, and use resources more efficiently." The report focuses on ways to improve living standards and sharing the benefits of growth equitably. One way to achieve this is if decent employment opportunities arise in various segments of the workforce. Manufacturing industries and related services sectors can potentially absorb large numbers of workers and provide stable jobs with respectable benefits and with this increase the prosperity of communities. Especially relevant in developing countries, increased participation in international trade also improves local working conditions through the need to comply with international standards and greater access to modern technologies and best practices. The report poses the question of what kind of industrialization should be prioritized to maximize the interactions with the global development agenda towards sustainable development and prosperity.

Finally, from Fig. 8.3, investment is a primary driver of such growth and prosperity. Rallying investments and ensuring that it contributes to sustainable development objectives should be a priority for all countries, and for developing countries in particular. According to the UN Conference on Trade and Development (UNCTAD) Investment Policy Framework for Sustainable Development (Zhan et al. 2012), current generation investment policies strive to create synergies with wider economic development goals or industrial policies and achieve seamless integration in development strategies. These policies aim to raise responsible investor behavior, incorporate principles of corporate social responsibility, and ensure policy effectiveness in the institutional environment within which they operate. These points are crucial in any development seeking financial investors for long-term backing and support. Developing countries and economies in transition are now primary foreign direct investment (FDI) destinations, and their importance as FDI recipients continues to increase. In 2010, for the first time, developing countries received more than half of global FDI funding, in part as a result of the decrease in investment opportunities in developed countries due to slower (but not stagnant) economic growth. New types of investors are also emerging, examples being state-owned enterprises and sovereign wealth funds. Private equity specialized firms are being established to seek investment opportunities in renewable and alternative energy, water and waste management, energy efficiency, environmental emission reduction markets, sustainable forestry and agriculture, and low carbon/sustainable products and materials. UNCTAD proposes a comprehensive framework for sustainable development consisting of a set of core principles for investment policymaking, guidelines for national investment policies, and guidance for policymakers on how to engage in the international investment policy system, in the form of possibilities for the design and use of international investment agreements.

To summarize, Fig. 8.4 highlights the key concepts described in this section towards responsible sustainable development in urban areas, applicable to developing and developed countries, and for any application aimed to create a sustainable ecosystem.

In Fig. 8.4 and relating to Fig. 8.1, the main considerations for responsible development are summarized. These include, as shown, the three key categories;

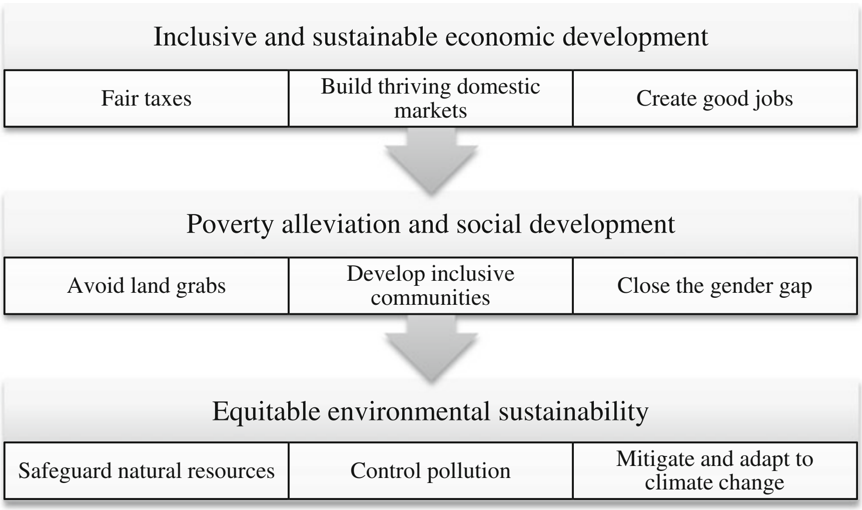


Fig. 8.4 Key concepts for responsible sustainable development

- providing an inclusive and sustainable economic development plan, which addresses fair taxes for its citizens to better the infrastructure of the city, creating thriving domestic markets, and decent and sustainable jobs,
- social development and the alleviation of poverty are required to build an effective and efficient working force. The buying or leasing of large pieces of land, especially prevalent in developing countries by domestic and transnational companies, governments, and individuals should be avoided to enable fair usage of this land to the local population. Communities must have an inclusive aura about them and encourage working together towards a common goal, and also in closing the gender gap, again lacking in developing countries, must be prioritized, and
- an equitable environmental sustainability must be maintained to ensure that natural resources that form the backbone of the economy are well managed and responsibly used. Natural resources must therefore be safeguarded, pollution in the area limited by respectable regulating bodies, and climate change must ideally be limited but also understood, and cities must be able to adapt to these inevitable changes.

Any decisions made towards creating a smart and sustainable future for a city, will have consequences, whether being positive or negative in its nature. These decisions must not be made without researching and understanding the causality and environmental impact assessment studies, which are crucial during the decision making process.

8.5 Impact Assessment

Impact assessment, of any nature, can be defined as a systematic process of identifying future consequences of a current or proposed action, related to the idea of causality (cause and effect). Well-governed cities and towns that involve their citizens in environmental decision-making will lead to better planning for the future and help to ensure the sustainability of the systems which are created. In developing a smart and sustainable city, the impact assessment for future generations include several key topics that can be identified and quantified. These topics are summarized in Fig. 8.5.

From Fig. 8.5, impact assessment is for instance required to determine the accountability of decisions about a city's future, cost effectiveness, practicality, certainty, credibility, flexibility, community participation, and transparency. The European Commission report on making cities attractive and sustainable lists several techniques to ensure a democratic and well-managed city or town. These include reducing corruption, having a transparent and accountable decision making process, and putting local resources to its best use for cost-effectiveness and practicality,

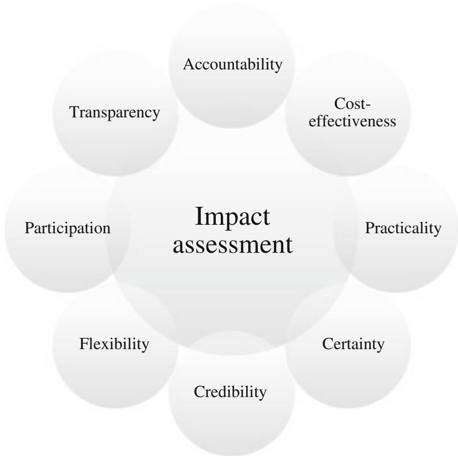


Fig. 8.5 Key concepts for sustainable development impact assessment

therefore maintaining a credible and flexible image. Five key concepts are mentioned to achieve this image, these being the use of integrated environmental management systems (IEMS), green public procurement, participatory urban planning, assessing environmental impacts, and constantly tracking progress. IEMS in its simplest form the integration of four essential actions during environmental management, namely a cyclic relationship between planning, doing, checking, and acting upon any decision. Green public procuring is a voluntary instrument that allows public authorities to achieve environmental targets by opting for greener products and services. Participatory urban planning is about involving citizens in urban planning to help ensure sustainable economic development and it plays a vital role in providing well planned cities. As citizens are deeply affected by urban planning, authorities need to ensure that they are involved and provided with a medium for expressing their opinions. The last two concepts, impact assessment and progress tracking, are imperative but sometimes overlooked due to time and budget constraints, but having a clear indication of the impact of planned and acted on decisions and tracking the progress throughout the development cycle and beyond are crucial.

First designed for water resource development, the Battelle method (Wagh and Gujar 2014) can easily be adapted for urbanization impact assessment. The underlying principle lies in dividing environmental impacts into four major categories: ecology, pollution, aesthetics, and human interest. These categories are further divided into thematic data as shown in Table 8.2.

Furthermore, the Battelle method suggests that the thematic data be divided into environmental indicators, each given a total weighting factor consisting of individual thematic components. The expanded Battelle environmental classification method with corresponding weight factors is shown in Table 8.3.

The Battelle method as shown in Table 8.3 groups thematic information to determine the parameter importance unit (*PIU*), which is a total number with a maximum

Table 8.2 Four major categories of environmental impact assessment

Category	Thematic data	Category	Thematic data
Ecology	Species and population	Pollution	Water pollution
	Habitats and communities		Air pollution
	Ecosystems		Land pollution
			Noise pollution
Aesthetics	Land	Human interest	Educational
	Air		Historical
	Water		Cultural
	Biota		Atmospheric
	Man-made objects		Life patterns

Table 8.3 Battelle method environmental classification and corresponding relative weight factors

Ecology (240)		Environmental pollution (402)	
Terrestrial species and populations		*Water quality*	
Browsers and grazers	14	Basin hydrologic loss	20
Crops	14	Biochemical oxygen demand	25
Natural vegetation	14	Dissolved oxygen	31
Pest species	14	Fecal coliforms	18
Upland game birds	14	Inorganic carbon	22
Aquatic species and populations		Inorganic nitrogen	25
Commercial fisheries	14	Inorganic phosphate	28
Natural vegetation	14	Pesticides	16
Pest species	14	pH	18
Sport fish	14	Stream flow variation	28
Water fowl	14	Temperature	28
Terrestrial habitats and communities		Total dissolved solids	25
Food web index	12	Toxic substances	14
Land use	12	Turbidity	20
Rare and endangered species	12	Air quality	
Species diversity	14	Carbon monoxide	5
Aquatic habitats and communities		Hydrocarbons	5
Food web index	12	Nitrogen oxides	10
Land use	12	Particulate matter	12
Rare and endangered species	12	Photochemical oxidants	5
Species diversity	14	Sulphur oxides	10
Aesthetics (153)		Human interest/social (205)	
Land		*Education/scientific*	
Geologic surface material	6	Archeological	13
Relief and topographic character	16	Ecological	13

(continued)

Table 8.3 (continued)

Aesthetics (153)		Human interest/social (205)	
Width and alignment	10	Geological	11
Air		Hydrological	11
Odor and visual	3	Historical	
Sounds	2	Architecture and styles	11
Water		Events	11
Appearance	10	People	11
Land and water interface	16	Religions and cultures	11
Odor and floating material	6	Western Frontiers	11
Water surface area	10	Cultures	
Wooded and geological shoreline	10	Indians	14
Biota		Other ethnic groups	7
Animals (domestic)	5	Religious groups	7
Animals (wild)	5	Mood/atmosphere	
Diversity of vegetation types	9	Awe/inspiration	11
Variety within vegetation types	5	Isolation/solitude	11
Man-made objects		Mystery	4
Man-made objects	10	Oneness with nature	11
Composition		Life patterns	
Composite effect	15	Employment opportunities	13
Unique composition	15	Housing	13
		Social interactions	11

of 1000 (240 (ecology) + 402 (environmental pollution) + 153 (aesthetics) + 205 (human interest/social)), that is used to calculate the environmental impact unit (*EIU*). This value can be determined by

$$\sum EIU = \sum (EQ_i)_1 \times PIU_i - \sum (EQ_i)_2 \times PIU_i \tag{8.1}$$

where $(EQ_i)_1$ is the environmental quality for indicator i accounting for the specific project conditions (a number between 0 and 1 where 0 indicates very poor quality and 1 indicates very good quality), and $(EQ_i)_2$ is the environmental quality for indicator i without the project conditions (also between 0 and 1 with similar scale).

Environmental metrics are not the only concern when planning for sustainable ecosystems, social metrics have become a high-profile metric as its effects have become noticeably prevalent in sustainable communities. Additionally, economic metrics have since the earliest of human settlements been at the forefront of any successful human ecosystem. These metrics are discussed in the following section.

8.6 Social and Economic Metrics

Historical research on sustainability has been mainly limited to environmental and economic concerns (Colantonio and Dixon 2009). However, in recent years, social sustainability has gained increased recognition as a fundamental component of sustainable development, beginning to receive political and institutional endorsement within the sustainable development itinerary, and the sustainable urban regeneration treatise. Social sustainability concerns how individuals, communities, and societies live among each other and set out to achieve the objectives of development models, which they have chosen for themselves, also taking into account the physical boundaries of their dwellings and our planet. Social and economic metrics include four identifiable components, which a society thrives to achieve, being, in order of perceived importance, wealth, health, specificity, and social improvement. Figure 8.6 depicts these components.

Money is an intricate driving force in any economy and socially considered a benchmark of success. Although money does not guarantee happiness, the fact is that this way of approaching life is not destined to change, and should rather be acknowledged. The smallest circle in Fig. 8.6 is therefore wealth and social welfare, not the largest contributor in an economy, but definitely a contributing factor. Atkinson and Pearce (1993) made a comment that if (total) wealth is related to social welfare, then changes in wealth should have implications for sustainability. An optimal economy aims to enforce intensification of social welfare, conceivable through maximizing present value utility. Resource depletion, stock pollutant damages, and variations in human capital accumulation change social welfare.

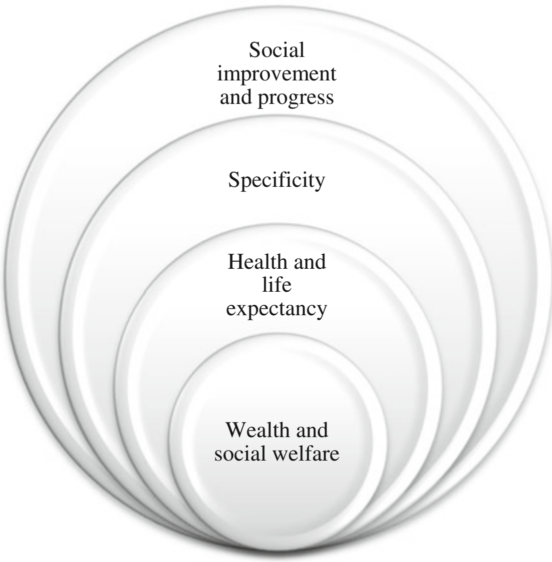


Fig. 8.6 Social and economic metrics in sustainable development

The goals of sustainable development cannot be achieved when there is a high prevalence of incapacitating diseases and population health cannot be maintained without ecologically sustainable development. Poor health continues to be a restriction on development efforts. In some cases, the process of development itself is creating conditions where (as a result of economic, political and social disorder, environmental degradation, and uneven development resulting from corrupt actions) human health suffers. More than 200 million people (in 2015) live in countries with an average life expectancy of less than 45 years and in some of the poorest countries of the world, one in five children still fails to reach the age of five, mainly due to infectious diseases from poor environmental conditions. Short life expectancy and high mortality are large contributors to slow development and unsustainability since a person's health directly affects their morale and ability to perform to set standards.

Additionally, from Fig. 8.6, specificity in social and economic developments must be shared with communities. Without specificity, communities tend to shy away from new developments and might seek refuge in different areas where there are less unknowns they have to deal with. People generally tend to avoid situations where there is high uncertainty and are specific-goal oriented, providing individuals with specificity is important to build trust, have an open and transparent image, and assure people to accept change.

Finally, people are social creatures and seek social interaction in their lives. This is arguably one of the largest driving forces towards economic success for any urban or rural settlement and has been underestimated historically. People should be free to explore social development while ensuring their own sustainable future. Social improvement and therefore social progress lead to improved political, economic, and environmental progress. As societies become more and more comfortable with their surroundings, they aim to improve it, or at least to sustain it. The idea of progress is the impression that advances in technology, science, and social organization can improve the human condition and contribute to social improvement. This notion can be somewhat fabricated in certain circumstances, and the assurance of new technologies should be upheld to maintain trustworthiness between communities and developing organizations. Social improvement must be supported by social sustainability to allow citizens long-term security in their jobs, living arrangements, and transportation alternatives. Social sustainability takes on many forms, some of which are highlighted in Fig. 8.7.

Figure 8.7 promotes several social sustainability criteria required for a population to flourish and what are generally expected from people to feel secure in their surroundings. These criteria include, and is not limited to, good education and skills development opportunities, decent employment, attention to health and safety, housing, freedom to express their own identity and culture, community participation and empowerment during decision-making processes, social capital and cohesion, a good quality of life, and openness to demographic changes. If these criteria are lacking within a community, people tend become demotivated and counter-productive, with several adverse effects that may accompany this behavior, as seen in Fig. 8.8.

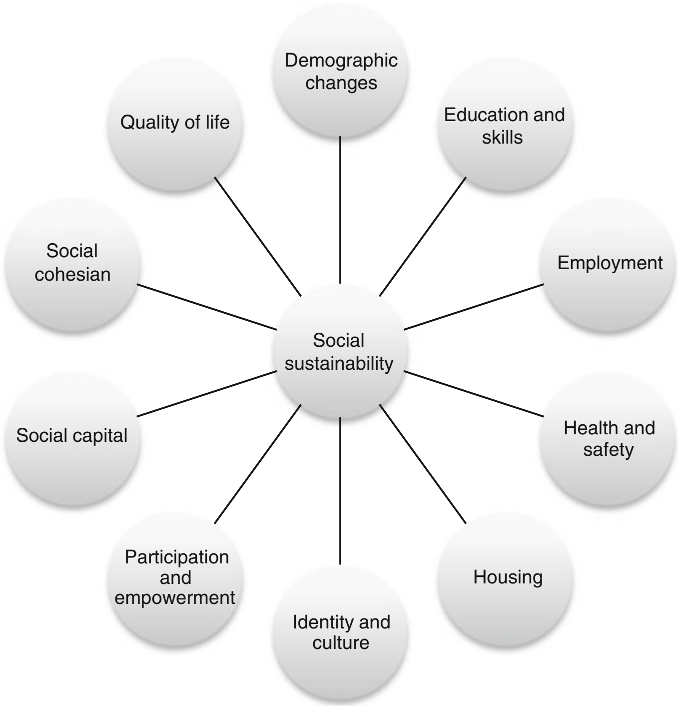


Fig. 8.7 Critical areas for social sustainability of local communities during urbanization

Fig. 8.8 Adverse effects of incompatible and unsustainable social and cultural communities

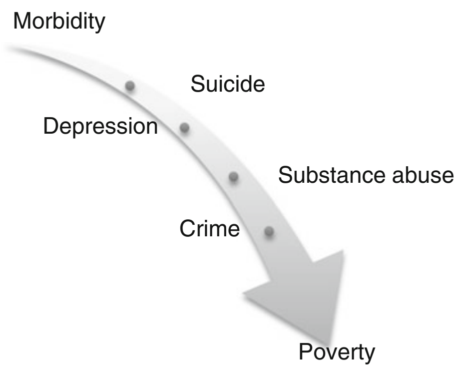


Figure 8.8 highlights some of the (call it mainstream) adverse effects that generally contribute to unsustainable, inefficient, and unappealing places for humans to associate with. These include higher than usual morbidity rates, suicide, depression, substance abuse, increased crime, and inevitable poverty. The Independent Newspaper in July 2015 awarded cities across the world scores depending on lifestyle challenges faced by the people living there. Importantly, each city was

scored on its stability, healthcare, culture and environment, education, and infrastructure. Noticeable in this list, is the prevalence of developing countries (specifically in Africa and The Middle East), with the ten worst cities to live in being Damascus in Syria, Dhaka in Bangladesh, Moresby in Papa New Guinea, Lagos in Nigeria, Harare in Zimbabwe, Algiers in Algeria, Karachi in Pakistan, Tripoli in Libya, Douala in Cameroon, and Tehran in Iran.

Population dynamics and demographic considerations play a large role in the look, feel, and inner workings of large cities. The size, structure, and distribution of populations and spatial and temporal changes influence urban populations to a large extent. A discussion on these considerations is given in the following section.

8.7 Demographic Considerations

The growth of the world population and production, as well as the combined unsustainable consumption patterns places increasingly severe stress on the life-supporting capacities of our planet. These interactive processes affect the use of land, water, air, energy, and other resources. Rapidly growing cities, unless well-managed, face major environmental problems. The increase in both the number and size of cities calls for greater attention to issues of local government and municipal management. The human dimensions are key elements to consider in this intricate set of relationships and they should be adequately taken into consideration in comprehensive policies for sustainable development. Such policies should address the linkages of demographic trends and factors, resource use, appropriate technology dissemination, and development. Population policy should also recognize the role played by human beings in environmental and development concerns. There is a need to increase awareness of this issue among decision makers at all levels and to provide both better information on which to base national and international policies and a framework against which to interpret this information.

In order to integrate demographic analysis into a broader social science perspective on environment and development, interdisciplinary research should be increased. International institutions and educational institutions should further develop their scientific capacity, taking full account of community experience and knowledge, and should freely distribute the knowledge gained from multidisciplinary approaches. Better modeling capabilities should be developed, identifying the range of possible outcomes of current human activities. This is especially important for the related impact of demographic trends and factors (per capita resource use and wealth distribution) as well as the major migration tendencies from increasing climatic events and cumulative environmental destruction. Socio-demographic information should be developed in a suitable format for interfacing with physical, biological, and socio-economic data and technological advancements should be at the forefront of distributing this information. Compatible spatial and temporal scales, cross-country, and time-series information, as well as global behavioral indicators should be established to encourage learning from local communities' perceptions and attitudes.

Existing strategies for sustainable development have generally recognized demographic trends and factors as elements that have a critical influence on consumption patterns, production, lifestyles, and long-term sustainability. Countries must improve their own capacities to assess the environment and development implications of their demographic trends and factors. Developing countries need to formulate and implement policies and action programs where appropriate. Policies should be designed to address the consequences of population growth built into population momentum, while at the same time incorporating measures to bring about demographic transition. They should combine environmental concerns and population issues within a holistic view of development whose primary goals include the alleviation of poverty, secure livelihoods, good health, quality of life, improvement of the status and income of women and their access to schooling and professional training, as well as fulfilment of their personal aspirations, and empowerment of individuals and communities (REF). Recognizing that large increases in the size and number of cities will occur in developing countries under any likely population scenario, greater attention should be given to preparing for the needs for improved municipal management and local governing.

An effective consultative process should be established and implemented with concerned groups of society where the formulation and decision-making of all components of the programs are based on a nationwide consultative process drawing on community meetings, regional workshops, and national seminars, as appropriate. The poor and underprivileged should be priority groups in this process.

Relatively poor economies will initially have a high ratio of labor to capital resources, which implies that the benefits to productivity of an additional investment in capital will be greater than in a relatively wealthy country. Over time, returns to human and physical capital will diminish to the point where income levels converge. Using the (unrealistic) assumptions that domestic savings ratios, the pace of labor force growth and the level of technical progress are equal across countries, productivity growth in poor countries will outstrip that of wealthier countries. As a result, the poorer countries will enjoy a convergence of capital/labor and capital/output ratios as well as income levels over the long run. As this process of convergence proceeds, the growth rates enjoyed by the poorer economies will be relatively large.

Malthusian theory is the term used to describe the position of the 19th century political economist, Reverend Thomas Robert Malthus, in his arguments about how and why population changes. In his works (An Essay on the Principle of Population) he seeks to explain the natural dependencies of population fluctuation. Ultimately, he reminds us that humans are essentially tied to their immediate environment. "Even the most highly developed society can be eradicated by a severe enough drought or plague. earth will eventually reach its carrying capacity, rendering further population growth impossible". Malthus' theory points out that human population grows at a geometric rate, or exponentially with each generation. It also points to the difference between this geometric rate of growth for human

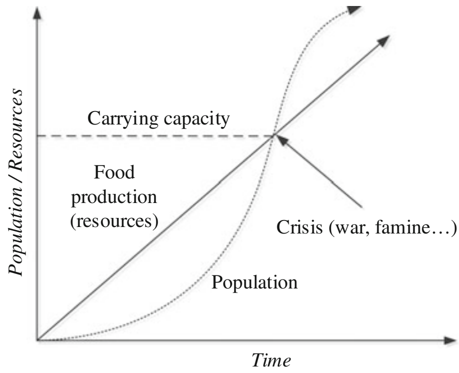


Fig. 8.9 Malthusian demographic transition theory of population

populations and the arithmetic rate of food production, which means that with each generation, the food supply will only increase by the same set number. A graphical representation of the Malthusian demographic transition theory of population is given in Fig. 8.9.

A commonly used phrase in the discussion of population growth is demographic transition, which describes a progressive movement from high birth and death rates to low birth and death rates. The demographic transition theory argues that population growth is tied to a society's level of technology. As the society advances in its usage of technology, in every industry from healthcare to crop production, its birth and death rates shift, directly impacting the population numbers and growth rates. Phase one of this demographic transition model is associated with pre-industrial society with high birth- and death-rates, effectively creating a stable population with relatively slow growth. Closely tied to the nature and the limitations of geographical resources, this stage was observable for most of human history. Phase two generally concerns developing countries. These countries have a high birthrate and the death-rate significantly decreases with reduced infant mortality. Children under five are no longer dying at such high rates, leading to a larger youth population in a given community. A large population of youth generally points to a developing nation. This imbalance, forming the basis of demographic transition, can be temporary and occurs as the result of a sudden improvement in the quality of life. Recently, large investment opportunities in developing countries in Africa have boosted local economies and the adoption of westernized medicine has led to a vibrant and young and full of potential continent.

The earth however does not have infinite resources to sustain all of the populations and species that exist on earth itself and growing populations should especially take care in planning for their future. It is therefore important that we have an understanding of environmental impact metrics that determine the sustaining capabilities of the earth, not only for this generation, but also for generations to come. The following section discusses certain macro-effects that are prevalent in urbanized areas, and could potentially lead to unsafe and unhealthy environments for growing cities.

8.8 Urban Macro-Effects

Urban thermo-physical and geometrical characteristics, anthropogenic moisture, and heat sources in urban areas such as people emitting energy as heat, cars, buses, trains, and general weather conditions all contribute to urban areas having distinctively different climates compared to the surrounding rural and open areas. It was already reported in 1976 and 1981 that that northern hemisphere urban areas annually have an average of 12 % less solar radiation, 8 % more clouds, 14 % more rainfall, 10 % more snowfall, and 15 % more thunderstorms than rural areas in close proximity. The temperatures in heavily urbanized areas can be 2–4 °C higher compared to its rural counterparts, a phenomenon referred to as the urban heat island. Urban heat islands can be categorized into three types;

- urban canopy layer heat island which is the layer of air closest to the surface (roads) in cities, extending upwards to approximately the average building height,
- urban boundary layer heat island which ranges in thickness of between 1 km by day and a few hundred meters by night, and
- urban surface heat island which refers to the heating effect near the surfaces (roads) in the city.

Figure 8.10 represents the three types of heat islands and the transitioning zones between urban and rural area atmospheres.

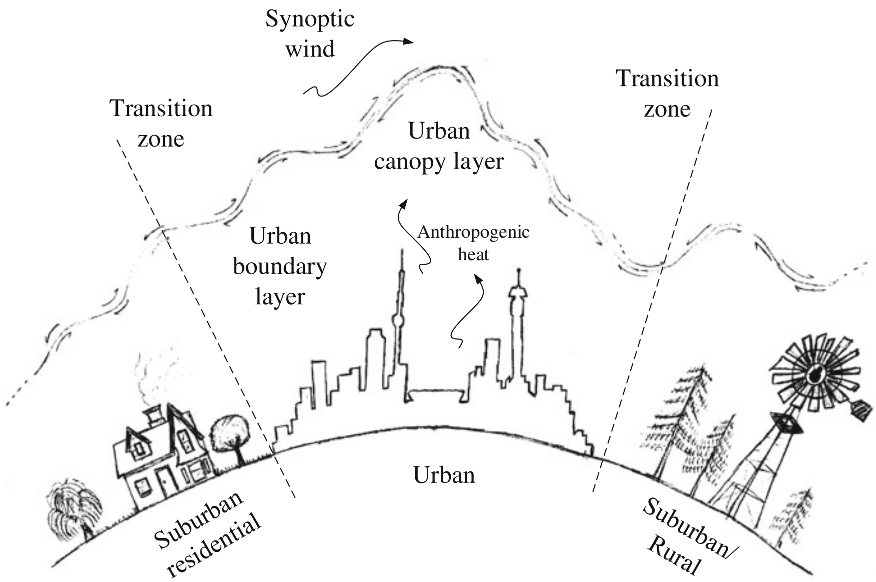


Fig. 8.10 Urban heat island effects and transition zones

The types of heat islands vary in their shape, time characteristics, and underlying physical processes that contribute to their forming. Canopy layer and surface heat islands can directly be measured using thermometers and weather stations, whereas boundary layer heat islands are generally measured by sensor networks attached to aerial devices such as balloons or other aircraft.

Urban areas are densely populated, meaning there are a lot of people in confined spaces. Urban areas are also densely constructed, meaning buildings are constructed in close proximity to each other. When there is no more room for an urban area to expand, engineers build upward, creating skyscrapers. All this construction means waste heat and heat that escapes insulation has nowhere to go. It lingers in and between buildings in the urban heat islands. Nighttime temperatures in urban heat islands remain high. This is because buildings, sidewalks, and parking lots block heat coming from the ground from rising into the cold night sky. Since the heat is trapped on lower levels, the temperature is warmer.

Albedo is the percentage of incoming radiation reflected off a surface. The earth's climate system remains in equilibrium as long as the amount of absorbed solar radiation is in balance with the amount of terrestrial radiation emitted back to space. Earth's albedo values are very important in shaping local and global climates through the radiation budget, determined as the difference between the amount of absorbed shortwave radiation and the outgoing longwave radiation. An albedo of 1 means that 100 % of incoming radiation is reflected and no radiation is absorbed; and an albedo of 0 means that 0 % of incoming radiation is reflected meaning all radiation is absorbed. Generally, lighter colored surfaces have a greater albedo effect. A way to mitigate urban heat islands, i.e., the higher temperatures in cities compared to those of their surroundings, and their negative impacts on cooling energy consumption is to use high albedo materials on major urban surfaces such as rooftops, streets, sidewalks, school yards, and the exposed surfaces of parking lots. Table 8.4 lists several material albedos (ρ) parameters that are prevalent in urban areas.

Table 8.4 General effects of heat islands and their positive of negative connotation

Material	Albedo (ρ)
Fresh asphalt	0.04
Black acrylic paint	0.05
Worn out asphalt	0.12
Conifer trees	0.08–0.15
Deciduous trees	0.15–0.18
Bare soil	0.17
Green grass	0.25
Desert sand	0.40
Fresh concrete	0.55
Fresh snow	0.80–0.90
Old snow	0.40–0.60

Table 8.4 therefore lists the ground reflected components of the sky dome affected by the albedo of reflective surfaces. As seen in Table 8.4, desert sand, fresh concrete, and snow have relatively high albedo parameters and reflect solar radiation the most. These materials also have lower heat storage capacity during the night. Asphalt, black acrylic paint, vegetation, and bare soil absorb solar radiation during the day and release this heat during the night. The heat island effect is amplified by these materials, although the vegetation heat emission during the night is a natural process and not considered detrimental to the urban environment and considered a good natural heat source during colder winter months. The large absorption of heat onto darker buildings and other urban surfaces and the effects from multiple reflections inside urban canyons effectively reduces albedo in urban areas. The albedo on a spherical coordinate system used to study the impact of albedo on a microclimate is represented by

$$\rho = \frac{\iint_{\lambda_1}^{\lambda_2} K \uparrow \cos\theta_z d\omega d\lambda}{\iint_{\lambda_1}^{\lambda_2} K \downarrow \cos\theta_z d\omega d\lambda} \tag{8.2}$$

where K is the radiant intensity in W m^{-2}, θ_z is the zenith angle, ω is the solid angle, and λ is the specific wavelength. The upward and downward arrows indicate reflected and incident radiation respectively. In (Taha et al. 1992) wavelengths between 0.28 and 2.8 μm, shortwave and near-infrared wavelengths, are discussed since this band is considered relevant in studying the response in surface temperature of building materials to solar irradiation. Albedo also depends on the ratio of isotropic to specular reflection and is generally considered as equal ratios (50 % for both types of reflection). The difference between albedo of specular and diffuse surfaces vary with zenith angle, where a smaller zenith angle generally relates to smaller differences between the two. Effective albedo also varies with the radiant intensity from reflective surfaces based on the density of the buildings within an area. Surface reflectance has been derived through the use of satellites and remote sensing technology. The International Satellite Cloud Climatology Project established as part of the World Climate Research Programme has been collecting surface and atmospheric reflectance data since 1983. A traditional technique for estimating the earth's albedo is observation of the moon's ash-grey light, earthlight reflected from its dark hemisphere. Several terrestrial factors affect albedo, these include soil type and texture, soil moisture, vegetation types, soil and vegetative color, micro-topography, and macro-topography. Albedo in soil is generally strongly affected by mineral salt content such as sodium chloride and magnesium chloride. Vegetation has distinct characteristics to reflect and absorb sunlight where leaf aspect and color play an important role in its albedo. Conifer forests for example have much lower albedo compared to angiosperm or broadleaf forests. Topographic factors can be sub-divided into macro- and micro-effects. Macro-topographic effects refer to areas of steep slope that produce lower albedo since the angle of reflection results in incoming radiation being reflected towards further absorption paths and in effective increases the path length on the incoming radiation waves. Micro-topographic effects rely on small crevices in the material that allow multiple reflections within the same surface and reducing the radiation intensity at its final destination.

Urban heat islands also contribute to decreased air and water quality than the neighboring rural areas. Urban areas generally have more pollutants from vehicles, industry, and people and these pollutants can be blocked from being absorbed into the larger atmosphere by tall buildings and structures. Lower water quality can also have detrimental effects as warmer water enters local water streams and endanger aquatic species adapted for cooler water environments.

The spatial characteristics of a heat island are due to the horizontal lines of equal temperature forming a shape that resembles an island, which loosely follows the shape of the urbanized region and is surrounded by cooler areas. A sharp rise in temperature occurs in the canopy layer at the boundaries of rural areas with warmest temperatures occurring in the city center. The boundary layer has a shape resembling a plume and warmer temperatures are transported downwind of the city.

To alleviate heat islands there are two general options. The first is during the design and erecting of the buildings. Building materials with less heat capacity can be used in construction of large buildings to decrease the amount of energy stored and released during nighttime from these structures. Buildings can also be placed less densely to reduce the amount of trapped air between structures, but this is not necessary feasible, due to population demands. Roadways can be designed to minimize the amount of dark asphalt roads, although this is not always feasible due to transportation demands, but pavements for cycling and train tracks already decrease the number of asphalt roads. These techniques are generally less feasible especially in already established cities, and more viable techniques to reduce heat island effects are required. In essence, these techniques are based on distributing vegetation throughout strategic places within the urban area. Vegetation provides important shade and cooling through evaporation among other advantages. For warm climates, planting trees around buildings shade urban surfaces such as roofs and walls and reduce their temperature, leading to reductions in energy consumption of air conditioning areas inside the building due to the penetrating radiation from outside structures. Strategically placed trees also shade asphalt roads and parking lots, which generally absorb large amounts of solar radiation during the day, and releases this energy during the night. Additionally, shading vehicles in parking lots reduce the evaporative emissions from the heated gasoline, which also contribute to increased greenhouse gases. Rooftop gardens contribute to significantly cooler rooftops as a significant fraction of the absorbed radiated energy is converted to evaporated water as opposed to heating the roof and henceforth the surrounding environment. Essentially, vegetation is a cost-effective method to decrease urbanized heating effects and do not require re-design of urban structures and it improves air quality as greenhouse gases and pollutant emissions are naturally absorbed. The effects of heat islands can be summarized in Table 8.5.

From Table 8.5, it is seen that heat island effects for certain categories are considered beneficial. Human comfort is affected positively in winter months since the surround environment is generally warmer, also leading to less energy consumption for heating equipment. This evidently leads to less ice and snow forming

Table 8.5 General effects of heat islands and their positive of negative connotation

Effect	Winter months	Summer months
Human comfort	Beneficial	Adverse
Energy use	Beneficial	Adverse
Air pollution	Adverse	Adverse
Biological activity	Beneficial	Beneficial
Ice and snow	Beneficial	Beneficial

that can block roadways and increase traffic congestion. Vegetation also generally thrives in warmer climates, and additional heat during winter months can contribute to faster growing of plants and trees. It is however important to note that these positive effects are not really considered sustainable or by any means a good idea. Not only is the variability of the effects considered detrimental in activity planning, but the fact that pollution and densely populated areas are used to increase temperature should not be advised.

8.9 Environmental Impact Metrics

Human activities consume resources and produce waste, and as our populations, especially urbanized populations, grow and consumption increases, it is essential that we understand nature's capacity and limitations to meet these demands. The ecological footprint (EF) has emerged as one of the world's leading measures of human demand on nature (Ewing et al. 2010). Simply put, EF addresses whether the planet is large enough to keep up with the demands of humanity. "If everyone lived the lifestyle of the average American we would need five planets." EF has two distinctive properties that define it. The first is bio-capacity, which signifies the planet's biologically productive land areas such as forests, pastures, croplands, and fisheries. These areas in their natural form can potentially absorb much of the waste we generate, especially our carbon emissions. EF represents the productive area required to provide the renewable resources humanity is using and to absorb its waste. Today (2015) humanity uses the equivalent of 1.5 planets to provide the resources we require and to absorb our waste. This effectively translates to an 18 month period that the earth would take to regenerate the resources we use. Moderate UN scenarios suggest that if current population and consumption trends continue, by the 2030s, we will need the equivalent of two earths to support us. National governments using EF are able to assess the value of their country's ecological assets, monitor and manage their resources, identify the risks associated with ecological shortfalls, set policies that are enforced due to ecological risks, and measure progress.

Sustainable human development will occur when all humans can have fulfilling lives without degrading the planet, plainly put, if we put back in the earth the same amount that we use daily. Good urban design can help to reduce carbon emissions, in particular by reducing wasteful transport patterns (Smallwood et al. 2007).

Environmental and resource risks are multi-faceted, interconnected and increasing in severity over time. They can impact economies in a number of ways: Nations are exposed to scarcity and/or price volatility of non-renewable resources such as fossil fuels, metals, minerals, and renewable resources such as water, soil, and biomass. Countries can experience a long-term loss of income due to the overuse and degradation of bio-productive assets such as cropland and forests. Climate change can also affect cropland productivity and cause financial losses through drought and extreme weather events. A nation's dependence on fossil fuels and use of carbon-intensive technologies can lead to stranded assets as market pressures drive a transition to a lower carbon economy.

The global effort for sustainability will be won, or lost, in the world's cities, where urban design may influence over 70 % of people's EF. High footprint cities can reduce this demand on nature greatly with existing technology. Many of these savings trickle down to the consumer and make cities more affordable. Since urban infrastructure is a long-term commitment, EF infrastructure decisions determine the future of any city.

8.10 Conclusion

A sustainable city is dependent on a multitude of factors, most importantly economic, environmental, and social indices that influence the well-being of the population. Several techniques are discussed in this chapter to achieve sustainability of an expanding city since urbanization is at an all-time high in recent years. The fundamental planning criteria towards a sustainable city are summarized in Fig. 8.11.

In Fig. 8.11, the fundamental areas that require attention during city planning are addressed. The key areas include, but is not limited to, energy efficient buildings, adapting to climate change, reducing carbon footprint, an increase of integration of renewable energy, efficient traffic management systems, smart transportation system, smart technologies in crucial parameter monitoring, smart health services, community involvement, and designing to decrease heat island effects and city albedo. If these elements are adhered to, a city can potentially flourish, but inevitably with an increase in population, pollution in a city is bound to increase. To reduce the pollution generated in a city, the following aspects must be monitored and managed constantly. These aspects are graphically depicted in Fig. 8.12.

Pollution can be controlled in a large city, if every person takes note of his or her carbon footprint. Small individual changes in the lifestyle of large communities can lead to significant decrease in generated waste. Noticeable improvements are gained from lower vehicle emissions through careful driving and ride-sharing, increases in non-tailpipe emission by walking or cycling, reducing household energy usage, reducing industrial energy usage, controlling water consumption, reducing waste water generation, efficient recycling techniques, and individuals lowering the amount of wasted power, water, and food. Efficient transport can reduce global CO_2

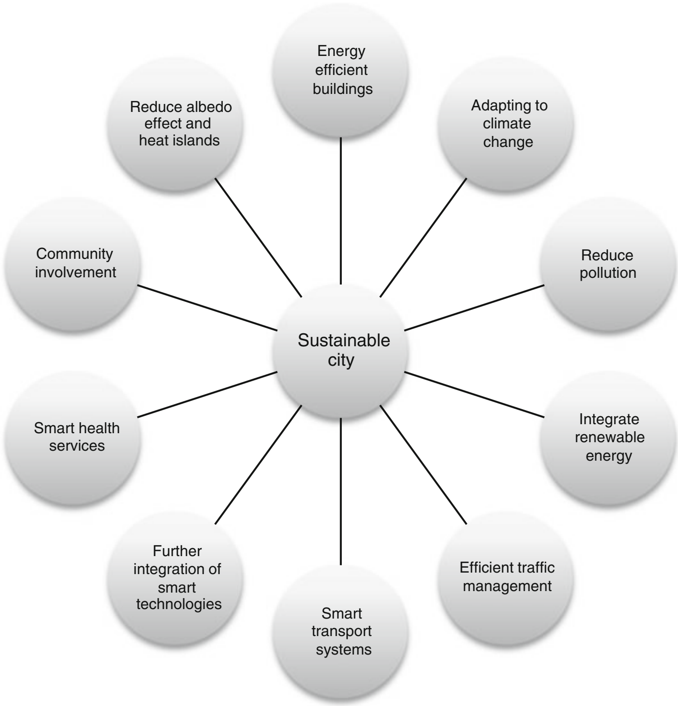
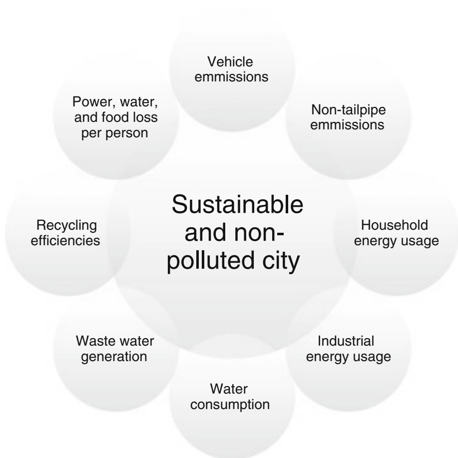


Fig. 8.11 Key planning areas towards sustainable and smart cities

Fig. 8.12 Environmentally conscious lifestyle considerations for sustainability and pollution reduction in urban areas



emissions, small particle air pollution, noise and injuries. Food production is a major contributor to global emissions. Reducing total global consumption of animal products (meat and dairy foods) can lead to reduced CO_2 and methane emissions produced by animals. Waste is an important contributor to carbon emissions. Reducing waste can lead to big emission savings and lower land fill requirements, with consequent reductions in air and land pollution. Waste not only discharges CO_2 and methane into the atmosphere, it can also pollute the air, groundwater and soil.

References

Atkinson, G., Pearce, D. (1993). The Concept of Sustainable Development: An Evaluation of its Usefulness Ten Years After Brundtland. Centre for Social and Economic Research on the Global Environment (CSERGE). ISSN 0967-8875.

Colantonio, A., Dixon, T. (2009). Measuring Socially Sustainable Urban Regeneration in Europe. *Oxford Institute for Sustainable Development (OISD)*. October 2009.

Ewing, B., Moore, D., Goldfinger, S., Oursler, A., Reed, A., Wackernagel, M. (2010). Ecological Footprint Atlas 2010. Oakland: Global Footprint Network.

Hardin, G. (1968). The Tragedy of the Commons. Science. 162(3859), 1243–1248, December 1968.

Smallwood, C., Baker, N., Carr, S., Eveleigh, H., Hoeksma, S., Smith, M., Stewart, J., Ward, M. (2007). Valuing Sustainable Urbanism: An Overview of the Report. *The Prince's Foundation for the Built Environment*. ISBN 978-1-906384-00-5.

Taha, H., Akbari, H., Sailor, D., Ritschard, R. (1992). 'Causes and Effects of Heat Islands: Sensitivity to Surface Parameters and Anthropogenic Heating', Lawrence Berkeley Lab report. 29864, Berkeley, CA, 1992.

UNESCO and UNEP. (2002). Cultural Diversity and Biodiversity for Sustainable Development. *A jointly convened UNESCO and UNEP high-level Roundtable held on 3 September 2002 in Johannesburg during the World Summit on Sustainable Development*.

UNESCO. (2009). Investing in Cultural Diversity and Intercultural Dialogue. *UNESCO World Report*. ISBN 978-92-3-104077-1.

United Nations. (2015). The Millennium Development Goals Report 2015. ISBN 978-92-1-101284-2, Retrieved July 9, 2015 from http://www.un.org.

Wagh, C. H., Gujar, M. G. (2014). The Environmental Impact Assessment by Using the Battelle Method. *International Journal of Science and Research*. 3(7), 82–86, July 2014.

Yigitcanlar, T., Kamruzzaman, M. (2015). Planning, Development and Management of Sustainable Cities: A Commentary from the Guest Editors. Sustainability, 7, 14677–14688, doi:10.3390/su71114677, 2015.

Zhan, J., Bolwijn, R., Dupriez, Q., Karl, J., Ripinsky, S., Turk, E., Weber, J., Alschner, W., Brahms, A., El Kady, H., Rosert, D., Van Giffen, T. (2012). Investment Policy Framework for Sustainable Development. *United Nations Conference on Trade and Development (UNCTAD)*. Doha, 21–26 April 2012.

Chapter 9
Tools and Facilitators Towards Successful Planning for Sustainable Cities

9.1 Introduction

Planning for sustainable cities using, for example, microsensing networks requires thorough understanding and awareness of the issues that plague cities and rural areas, in developed and developing countries alike. These issues must first be identified and comprehended before investments in technology can be made to address them. A sustainable city must factor in various requirements of its population and prioritize these needs based on the current availability of necessities such as food, water and transportation. Population growth in cities, especially in developing countries, is the highest it has been in history and the affluence of people in cities is creating additional problems in terms of distributing resources. A balance must be found between supplying resources by classical means and integrating smart alternatives to assist in their distribution, usually biased because of financial and human capital constraints and speed of implementation. Each chapter in this book discusses an important factor in creating sustainable cities, ranging from statistical information on population growth to the identification of critical technologies that convert cities to smart cities using sustainable microsensing networks spatially distributed throughout a city or any other area that requires monitoring. An important factor to consider where any human settlement is situated, large or small, is pollution in the immediate area and the effect it has on humans, animals and plants. Microsensing networks generally aim to increase the quality of life for humans as well as for the environment. These networked devices monitor and analyze the environment for specific changes that might result in risks for inhabitants. Pollution is largely due to anthropogenic activity and combined with natural events that lead to air and water pollution, the final result could be devastating. Smart cities enable the distribution of information about the environment to as many citizens as possible. Giving individuals and organizations knowledge about their own habitat is a way of practically ensuring automatic sustainability through community awareness. People tend to react to situations where their own lives are

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_9

in danger, whereas a smaller percentage will also act upon identifying risks to the lives of others (human, animal, or even vegetation). The first obstacle to overcome, however, is distributing knowledge/information for people to use.

Technological advancements, especially miniaturization of electronic components, have shifted the paradigm of ubiquitous computing. More and more devices are integrated into networked structures and working together towards a common goal. There are inevitably various teething problems with some newer generation technologies, duly noted in the implementation of smart and sustainable cities, but the potential of these technologies is undeniable. GIS and remote sensing have advanced in leaps and bounds in recent years and their effectiveness has increased as a result of advancements in processing power, miniaturization, storage capabilities and wireless communication. Together with ubiquitous computing and energy-harvesting techniques, devices are more capable of operating independently of the grid, and computing anywhere is becoming practically viable in many situations. Even in developing countries where technology sometimes lags behind state-of-the-art implementations, devices are becoming increasingly more affordable, with processing power, integrated sensors and actuators that enable basic ubiquitous and mobile systems to monitor the environment. Developed countries are also increasingly investing in international markets where there is potential for smart-city conversions and sustainable practices that can benefit both parties financially, economically and socially.

This chapter aims to expand on the topics discussed throughout the book but with increased focus on identifying potential problems, highlighting possible causes and providing possible solutions, or at least giving readers the ability to draw his or her own conclusions and, it is hoped, offering innovative solutions based on the information given. Novel solutions to any problem can only be found if enough information on a topic is available.

Population growth in urban areas is at an all-time high as people seek higher-paying jobs and increased social interaction, and move away from agricultural and rural settlements. Rapid urbanization has detrimental effects, but also advantages, and the following section aims to highlight these.

9.2 Population Growth

During urbanization, especially at high rates, the basic needs of humans must first be addressed before anything else. Some countries use the fast growth of their urban areas to their advantage and implement think-forward infrastructures for their citizens, although many countries find it difficult to cope with the growth rates. If a city struggles to maintain its increasing population, the effects of overpopulation are more evident and rectifying some inherent problems over the long term becomes increasingly difficult, especially for utilities and health/safety-related incidents. The growth in the earth's population, not only in cities, but everywhere, is due to several reasons, as highlighted below, and this leads to overcrowded and overpopulated habitats.

Decline in mortality rate
Many factors can increase the mortality rate for short periods, such as epidemics or natural disasters: however, the ones that increase the birth rate do so over a longer period. The development of agriculture provided the ability to sustain nutrition without hunting. This created the first imbalance between birth and death rates, which was exacerbated by other developments as well.

Improved medical care
Technological improvements following the industrial revolution further increased the lifespan of people. Science produced more effective and sustainable means of producing food, whereas medical science led to new discoveries that defeated a large range of diseases, including measles, mumps, rubella and polio, through the invention of vaccines. The increase in food supply combined with the lower mortality rate boosted overpopulation.

Large families
In developing countries, families tend to have between four and ten children during their fertile lifetime. In many countries more children are regarded as a sign of a successful life and parents aim to prove their significance through large families, leading to overpopulation in these countries. Moreover, in developing countries families tend to be large to increase the number of providers of money and food. Families that are poverty-stricken or have experienced natural disasters which have claimed many of their belongings, tend to increase their family size for these reasons. Family planning and education on responsible intercourse are limited or even non-existent in several developing countries and leads to large families living below the poverty line and increasing the fertility-span of women. Unfortunately, these citizens are unable to grasp the negative effects of overpopulation and if they are not educated to understand these effects, they will continue such lifestyles and have larger families.

Fertility treatment
Technological advances have also given parents that were previously unable to conceive children alternative ways of doing so. Modern medicine has also ensured that female pregnancies are safer through fetal assessments using specialized obstetric and gynecological imaging technology. Delivery by caesarian section also lessens the risk of mortality in new-born babies if natural birth is deemed unsafe for the mother or the child, depending on the situation. Babies born prematurely have increased chances of survival due to advances in medical science.

Immigration
The density of the population in developed countries with good medical care, education, security and employment increases and such countries become overcrowded as people immigrate to them to seek a better quality of life. Not only does this leave their country of origin with fewer skilled workers, it also strains the employment opportunities in the destination country. Immigrants take up additional resources in a country such as food, water, clothing, energy usage and houses, which

may lead to an overall shortage of these resources if not planned for in advance. Planning to mitigate these risks is difficult because statistical data regarding the amount and movement of new immigrants need to be determined and analyzed.

Overpopulation does present advantages, albeit debatable and difficult to quantify. A larger population may arguably benefit cities in becoming sustainable in the long term. The following paragraphs summarize the advantages of growing populations.

Increase in market size
The size of markets will increase, which will enable organizations to take greater advantage of economies of scale. In smaller countries start-up companies often fail since the local population is too small to sustain a profitable market. Products are generally tested first in the country of origin before being exported, since unknown brands in international markets rarely succeed. Extra demand from a growing population stimulates investments and can lead to the introduction of new technologies.

Increased labor force
An increase in the local population through immigration, combined with a higher birth rate, especially in developing countries, can be beneficial in the long term. Through immigration, the labor force is automatically increased, augmented by an increase in the birth rate. A larger number of children increase their dependency in the short term, but the children can be trained and introduced into the workforce by already established skilled workers. Developed countries generally do not encourage immigration of unskilled workers on a large scale and often tend to control immigration laws to accept fewer and professionally trained people.

Enhanced diversity
Population growth adds diversity in communities and cities around the world. Societies can learn about different cultures, form inter-racial relationships, and share different viewpoints of life with one another. A growing population, regardless of ethnicity, also encourages relationship-building, which is one of the core priorities in human happiness.

More innovation
Population growth leads to more minds capable of innovating and designing new technologies, sustainable practices, human resources and medical advances. The idiom "two heads are better than one" holds true in this argument and rapid growth in population could potentially lead to new innovations if the population can be properly educated and trained towards innovative and critical thinking in disciplines such as environmental sustainability.

Focus on reducing pollution
Local pollution eventually begins to fall as countries develop and their populations increase because of new discoveries that reduce pollution, and also because a larger share of the incomes of richer countries is spent on controlling the output of pollutants, e.g. by offering incentives to combat air/water pollution.

The disadvantages of overpopulation and unsustainable growth rates are somewhat more tangible, and are summarized in the following paragraphs.

Burning of fossil fuels
The demand for fossil fuels and other non-renewable resources grows with a higher population as well as with economic development. Larger populations tend to produce greater pollution both locally and globally as more cars are driven, industrial output rises and more homes burn fuels. The focus towards reducing pollution in growing populations becomes more important to lower and/or limit energy usage and waste production which can lead to air/water pollution. Implementing energy-saving techniques in households and in organizations can reduce the amount of energy use per capita and decrease the demand on fossil fuels.

Depletion of natural resources
As the world population grows above seven billion people, natural resources are taking a lot of strain to support the lifestyles of all individuals. Natural resources under severe pressure from current rates of consumption include water, oil, natural gas, coal and rare earth elements such as scandium and terbium used in powerful magnets in wind turbines.

Waste production
The amount of garbage that humans throw away every day is rising with population growth and this rate cannot be sustained unless transformational changes in handling solid waste are implemented worldwide. The results of this hold serious consequences for public services, government budgets and the space consumed by landfills. Landfills and uncollected waste contribute to climate change through the production of methane, a potent and poisonous greenhouse gas. Techniques to reduce generated garbage in cities and rural areas include better transportation and storage systems that reduce food waste, construction strategies that encourage the reuse of materials and policies such as disposal fees and recycling programs.

Food shortages
If a country is overpopulated and its agricultural productivity is low, it runs a risk of not being able to feed its entire population without considering importing food at additional expense. One billion people worldwide, one out of every seven people alive, go to bed hungry. Every day, 25,000 people die of malnutrition and hunger-related diseases. Almost 18,000 of them are children under 5 years old. The food shortages causing this are not only witnessed in rural areas with a limited economic drive, but also in overcrowded and overpopulated cities that struggle to meet the demands of the fast-growing population.

Damage to natural habitats
Population growth leads to damage and sometimes extinction of animal life, oceanic life and other natural ecosystems. Solid and liquid waste entering water bodies through streams or by any other physical means can endanger and kill life in the area, since animals cannot distinguish these waste products from food sources

and often ingest poisonous or solid waste that they cannot digest and from which they eventually die.

Overcrowding
Overcrowding in cities increases the pressure on housing and social capital, and causes traffic congestion and delays during everyday commuting, effectively lowering the quality of life of citizens by wasting time. Constant struggles in overcrowded cities to access amenities and local services, long queues, shortages of resources and generally long waiting times lead to tense environments and can further degrade the quality of life.

Unemployment
Pressure is placed on employment opportunities if more people of working age enter a city through immigration. Governments generally focus on educating and training the local population, but this must be prioritized, which often does not happen in many underdeveloped countries. A rise in unemployment is generally associated with a rise in crime as people (must) steal from others to obtain food or other resources, as they are unable to purchase these items themselves.

High cost of living
As a country becomes overpopulated, the difference in supply and demand of housing, medical treatment, food and water becomes bigger, leading to higher prices for these commodities, which increases the gap between the wealthy and the poor even more. These problems are generally associated with developing countries.

Water shortages
The population has grown to exceed the amount of usable water for both agriculture and drinking. Lack of safe drinking water contributes to epidemic diseases and high child mortality rates. Many rivers are so heavily used for irrigation that they run dry before reaching their mouths, degrading or destroying fish habitats and downstream agriculture.

Air quality
The risk of lung cancer and asthma from air pollutants worldwide is largely due to automobile transportation sources. The risk of breathing in poisonous and polluted air is higher for pregnant women, since it can harm their unborn fetuses during their critical developmental phase. As cities and suburbs continue to grow at a fast pace, pollution emitted by commuters is growing exponentially.

Deforestation
Deforestation is the permanent destruction of forests in order to make the land available for other uses, such as housing and urbanization. Deforestation also occurs when timber is harvested to create commercial items such as paper, furniture and homes or, for example, to create room for cattle ranching, and when ingredients from plants are used to manufacture commercial products such as bath oils. Common methods of deforestation are burning trees and cutting; burning trees

increases the CO_2 count in the air, leading to further pollution and climate change. Deforestation also leads to the loss of species and soil erosion.

Conflict
A strain on resources in a country, especially due to immigration, often leads to conflict and even wars among communities, races and religious groups for control of these resources. Most wars break out because of conflict over resources. Increased diversity in a country, overpopulation and overcrowding, combined with competition for scarce resources such as water and food, lead to hostile environments.

Technology is becoming the preferred means to implement strategies to combat overpopulation and to distribute not only information, but also physical aids to educate populations about the risks of overpopulation, especially within their personal framework (family). The sustainability of a family, providing its members with food, water, shelter and an education, is heavily dependent on parents' awareness of risks, and educating them first about technological relief can prove beneficial for future generations. Rural and hard-to-reach dwellings in developing countries are becoming increasingly easier to reach and monitor in response to advancements in GIS, remote sensing, product distribution/transport and the lower cost of devices that aim to combat overpopulation, such as contraceptives.

Universal access to safe contraceptives
Family planning allows parents to plan and have children when they are financially and emotionally stable and ready to raise a family. Many parents find themselves having children in circumstances that are not ideal for them or their new-born child, being financially strained or working towards a career and having to put this on hold to care for the child during his or her critical first few years. Governmental policies to provide proper family planning and provide effective contraceptives/birth control to married and unmarried couples can assist in decreasing overpopulation, but more importantly, to decrease the number of children born into less-than-ideal circumstances.

Education and career opportunities
It is generally believed, though the theory is controversial and unproven as yet, that people who have obtained some secondary school education have fewer children and have children later in their lives compared to uneducated people. This might be because they have better knowledge of family planning through education, or (somewhat less controversial and more believable) because they are encouraged to build a career as skilled workers to provide better for their prospective families and are therefore postponing having children at a young age and generally having fewer children. This idea is advantageous because it controls overpopulation, but it also improves the chances of children being born into better conditions, and increases their chances of survival and their contribution to the economy. Lynn and Harvey (2008) did find some correlation between intelligence and fertility rates in a country. Shatz (2008) extended the research on a broader international scale and found that there is in fact a strong tendency for countries with lower average

national IQ scores to have higher fertility rates. It could therefore be argued that better education and more job opportunities in a country do decrease the birth rate and therefore reduce overpopulation in these regions, hence globally. Smart phones, tablets+, and low-cost laptops have become increasingly more attainable even in poorer countries, and these devices are ideal to distribute knowledge through the internet.

Eradication of gender bias and false social norms
Gender bias and gender discrimination reflect prejudice based on the gender of a person, particularly documented to affect women and younger girls. Gender bias had existed for many years and is only slowly starting to become less prevalent worldwide, although in many countries, especially developing countries and communist nations, gender prejudice still exists. Gender stereotyping casts women's purpose as only child-bearing. In nations where this is prevalent it is not uncommon for women to have more than five children at the demand of men, having little right or choice to do otherwise. This leads to overpopulation and again to children being born into non-ideal circumstances, reducing their chance of success later in their life. Eradicating the idea that persists in some nations that a larger number of children equates to greater success for the parents based on their "belongings" could slow overpopulation worldwide.

Age-appropriate sex education
Exposing students to educational programs that details topics such as puberty, intercourse, abstinence, birth control and respect for other's sexual rights can potentially prevent unwanted pregnancies, reduce birth rates and lead to fewer children being born into environments where parents are not yet ready to care for them. Children who are not aware and not educated about responsible intercourse tend to be unaware of contraceptives and are often embarrassed to seek information on the topic. Making sex education mandatory in schools and to larger groups of children can alleviate the embarrassment of these children to talk about their issues and ask important questions. Multi-purpose devices such as tablets can serve as visual aids in sex education and in a smart city environment, or even in rural areas, the availability of these devices should be prioritized, even if they are sparsely distributed.

Awareness of population and the environment at school level
Environmental education and sustainability practices in schools have been emerging as an important trend in 21st century education and a growing body of research and practice indicates that green schools can save money, improve health and boost academic achievement. School education in most countries focuses on presenting information on a more factual basis, without giving students the opportunity to explore wider issues such as environmental sustainability. Environmental sustainability is arguably a crucial topic for the education of future generations. One strategy is using energy efficiency projects in schools to teach children about sustainability, thus making it part of their learning experience. Schools do not necessarily have to create lessons dedicated solely to the

environment and energy to teach children about these issues. This knowledge can be diffused into core subjects such as mathematics, science and even literacy. Access to the internet is again crucial to support such a movement.

Tax concessions
An effective method of controlling the birth rate in a country is to provide tax benefits to families that have fewer than a certain number of children, similar to a government-sponsored children's fund for parents having their first child, and decreasing the incentive for additional children. Additional tax benefits could be offered to parents who choose to adopt less fortunate children rather than having their own. This procedure could also be viewed as less of a "punishment" compared to programs that impose fines on families with many children, as implemented in China's one-child policy in 1980. Reducing the tax margin on devices sold to schools and tertiary institutions to educate the local population could be implemented as well, giving a much needed boost to the ubiquitous availability of knowledge through technology.

Adjustment to an aging population
Controversially, limiting birth rates also leads to a shift in the population's demographics. Countries/regions where governments or institutions enforced birth control now have an aging population and a majority of elderly dependents. Governments and private investors can assist an aging community by raising the retirement age, but would also have to consider preventive healthcare. This might include lifestyle education, nutrition education to improve the quality of life and working abilities of aging people, age-appropriate fitness programs at reduced cost, safe driving assistance and special planning and training for public safety workers to increase their skills to assist older adults. In addition, tax assistance and property-tax relief could provide relief to older adults who inevitably have a reduced income. Adjusting the economy towards an aging population while encouraging birth control might alleviate some of the problems currently experienced as a result of the large population growth.

Overcrowding in cities can lead to a decrease in the quality of life of their citizens. These effects are not only visible in the short term, such as traffic congestion and higher cost of living, but also permeate to future generations. Smart cities can relieve the strain of cities having to provide in the needs and requirements of citizens through more informed decision-making processes. Distribution routes can be planned better and alternatives determined in real-time. Furthermore, constant monitoring of critical resources enables utilities and governments to act pre-emptively upon potential risks. Smart city technology is, however, in its infancy, especially when considering the standards and protocols required to unanimously implement smart infrastructures globally. Smart city adoption is also relatively slow in emerging markets and various barriers inhibit the growth of technology infrastructures at a sustainable pace compared to the growth of populations. Several key points regarding the advantages of smart cities and the barriers that limit their growth are highlighted in the following section.

9.3 Smart Cities

An intelligent city refers to a widespread knowledge network in a city to advance and encourage economic development through empowering its citizens. Smart cities focus on the integration of advanced technology and smart devices with the built environment to monitor and report on specific events. As these technologies evolve, they merge rapidly to fully incorporate social and behavioral components with distributed miniature devices, all connected in one large network and working together. Smart cities use information and communication technology (ICT) to engage citizens, render city services and enhance the urban environment. Figure 9.1 is adapted from the work of Bucchin et al. (2015), and summarizes the categories of city developments and ICT implementation to achieve smart cities capable of serving their populations with knowledge about their environment.

The future of smart cities is closely related to the distribution of the IoT, which is estimated will have 26 billion devices connected to it by 2020. Innovations in IoT are driving private and public organizations towards developing platforms and applications that address the needs and requirements of their respective communities. Smart cities consisting of large numbers of wirelessly interconnected devices

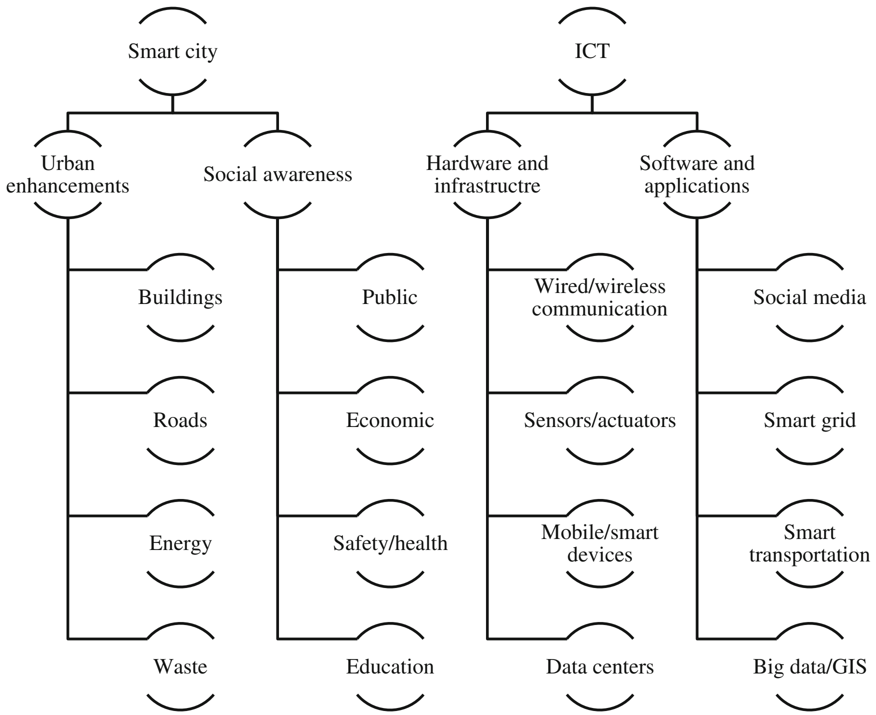


Fig. 9.1 Categorization of city-specific requirements and incorporation of ICT to achieve smart cities (Bucchin et al. 2015)

can provide citizens with real-time information about their environment and are considered a critical step in planning the future of sustainable urbanization. Cities are becoming smart through various initiatives, but the main drivers towards smart cities through IoT, WSNs combined with GPRS (Al-Ali et al. 2010), and to a degree RFID technologies are:

- planning for the long-term growth of cities based on a current high rate of urbanization where people from rural developments move to cities to seek job opportunities,
- developing sustainable energy resources to become inherently less dependent on the burning of fossil fuels and the increasing price of oil,
- upgrading, improving and expanding the existing infrastructure of the city (especially distribution of resources), and
- addressing the safety and security of the city's population through technology advancements.

These drivers towards smart cities have a positive effect on increasing the uptake of technologies to distribute information. Already in early developments of smart cities, several advantages and streamlining in the service and trade industry have been observed, including:

- In the services department, operational efficiency and asset tracking become feasible through cost-effective IoT components. Vehicle, container, and public transport tracking effectively monitors distribution and service routes, enables real-time analysis of delays and gives the ability to propose alternatives if required.
- Limiting the generation of greenhouse gases by using alternative and renewable energy sources has several environmental advantages for humans and other species on earth through decreased pollution.
- Smart managing of operations such as temperature and humidity regulation, automatic energy adjustment through smart lighting and other household appliances enhances the quality of people's lives and improves energy saving.

Although several smart city projects have been launched worldwide, the adoption of fully integrated smart cities has been slow. There are several reasons for the slow uptake, but two main factors have been identified:

- Firstly, the definition of a smart city is rather abstract. The term might have completely different meanings for developing and developed countries, for example, where each will interpret the definition based on its community requirements and economic scale. The interpretation of smart technologies also differs based on the availability and affordability of devices. In certain developing countries that experience extreme poverty, device technology could lag behind what developed countries are using and deeming state-of-the-art. The technological gap creates difficulty in standardizing smart city implementation worldwide and countries tend to implement infrastructure based on the available technology. This results in many countries being slow to adapt to smart city integration as they are "waiting" for newer and more affordable technologies to arrive.

- Secondly, the fact that this is still an emerging market where many devices are in the development, prototyping or testing phase and the sustainability of these technologies is questioned contributes to the slow uptake of smart cities worldwide. Investors are wary of aggressive markets and risk finance when it comes to smart city investments, as the prospect of viable returns on initial investments are still doubtful. Individual organizations such as Tesla® are leading the race towards sustainable technologies, but even these companies face pressure and sometimes impractical expectations from investors.

There are three basic principles that cities should incorporate when planning their long-term vision for immediate smart and sustainable operation and for future-proofing their investments, which can be broadly categorized as:

Integration
In general, most cities already have functioning infrastructures, although improvements can always be made to any system. Cities are encouraged to investigate current smart solutions in a variety of environments and initiatives and aim to adapt these technologies into their own services, as opposed to conceptualizing new systems. A large amount of active research on sustainable technologies, ICT and the IoT is undertaken by various educational institutions and commercial organizations and more often than not, these technologies can be adapted to help cities become smart.

Practicality
Cities are encouraged initially to invest in smart solutions that offer practical and financially viable ways to deal with real problems and needs of the community, as opposed to systems that might be considered novelties. Communities tend to accept new technologies if their value-added benefits are clear and easy to understand. For example, the Uber® taxi service, which uses smart phones and geographic information to inform clients of nearby drivers, had been adopted worldwide in a USD $40 billion industry as of December 2014. One can compare Uber® to, for example, the Google® Glass, which as a novel technology failed to penetrate the market as a high-volume selling product, mostly because of its limited practicality, high selling price and the fact that it is a non-essential commodity.

Synergy
Taking on a smart city project is an extremely large endeavor, in view of the size of the population of a city, especially megacities, and the necessity to serve all citizens, not excluding anyone. This is a mammoth task for any individual, organization or government to take on and should generally not be attempted without the coordination and participation of local businesses, the community and/or international organizations. Large industry players in the smart city and sustainable environment include IBM®, Cisco®, Microsoft®, Google®, Facebook®, Apple®, and Oracle®, and approaching such companies with practical and sustainable application initiatives could prove beneficial to both parties and future smart city developments in more countries.

Apart from the slow uptake of smart cities worldwide, additional obstacles that have been identified as limiting development and growth are summarized below:

Funding
Budgeting in most cities prioritizes statutory and essential services above technological advancements. Transforming cities requires considerable investment. Debt crises have adverse effects on many municipal budgets. Not only are investment requirements considerable, but a large number of cities have neither the means nor the credit rating to find cheap sources of funding. Austerity measures and decreasing tax revenue risk delaying the de-carbonization of cities, which is a core requirement for decreasing greenhouse gas emissions. This has negative repercussions for industries in the low-carbon sector, and it will ultimately have an adverse impact on the economy, as energy, transport and ICT are core economic sectors.

Risk of investment
Smart technologies in general are still in the early stages of development and must first be proven to be viable alternatives to longstanding services. Companies and governments are therefore expected to invest large sums of money in a technology that is yet to be proven lucrative over the long term, without being guaranteed a return on the initial investment. Companies are understandably averse to investing under such circumstances and it often takes courageous entrepreneurs to kick-start a new technology or development.

Expertise
ICT and technology-related skilled workers are still being trained and developed in cities to understand the implications of smart technologies better. Skilled workers are therefore still relatively scarce, taking into account the general size of a team required to implemented full-scale systems.

Broadband
The IoT and hence smart cities require wireless bandwidth to operate. In many cities, providing all users with uninterruptable and constant broadband connectivity for cell phone calling and data usage is still a challenge, and it is not yet feasible for providers to commit to additional bandwidth for large numbers of new devices and generating profit.

Departmental collaboration
In many cities, funding and resources for apparently similar departments are still obtained and managed separately. Smart cities incorporate a variety of departmental services to work together in providing urban information and this separation in departmental finance makes it difficult to allow smart technologies to be incorporated across all platforms.

Privatization
Many utilities such as gas, electricity and water are privatized and cities have little impact on decisions made by these utilities on the use of smart technologies. Private companies are generally focused on revenue and rarely committed to investing in high-risk future endeavors.

Privacy
Open data platforms and integrated health services require cross-platform availability of considerably sensitive data. People are, understandingly so, averse to providing personal information if the information is to be shared on large networks where breach of information security could occur. Trustworthiness in these systems must first be established before the community can be expected to embrace them.

Public commitment
Reaching a city-wide audience through smart technologies is difficult considering the demographics and sometimes limited access to the internet in poorer countries. For a smart city platform to be successful, it requires the backing and support of the entire community and it takes time to reach out to the majority of the population in a city to ensure the investment.

Overpopulation and rapid population growth in cities do lead to problem-solving and new and innovative ways to achieve smart city status to improve the lives of communities. These systems are generally tested in smaller areas of a city and extended city-wide and even internationally if found to be viable and cost-effective. However, population growth also leads to increases in air and water pollution. Pollution can have devastating effects on an environment and affect humans, animals and plants. Ubiquitous computing, GIS, remote sensing and other spatially distributed technologies often aim to lessen the burden on the environment by monitoring areas prone to pollution. This is not always enough, since a general shift in human behavior towards respecting the environment is required to ensure a sustainable future. Part of solving the problem of pollution, either through technology or behavioral adjustments, is understanding the types of pollution and its sources. Educating the population on how pollution is generated and presenting ways to curb pollution in cities and rural areas can largely benefit the environment and future generations. These aspects are discussed in the following section.

9.4 Pollution

Air pollution becomes harmful to humans, plants, and animals when it accumulates in high concentrations in the air. Long-term exposure to air pollution increases the chances of cancer and causes damage to the immune, neurological, reproductive and respiratory system. There are many causes of air pollution in rural and urban environments, all contributing to the overall density of pollution on earth. Urban and industrialized air pollution are generally more concentrated and visible in the sky. The worst causes or air pollution are:

Burning of fossil fuels
CO_2 is a good indicator of the amount of fossil fuels burned and polluting the air. CO_2 is not normally considered a pollutant, since it is a natural constituent of air. In excess concentrations, however, CO_2 leads to adverse effects on the environment

through heating of the earth's atmosphere, called the greenhouse effect. Additional potentially poisonous greenhouse gases emitted from the burning of fossil fuels include NO, NO_2, SO_2, CH_4 (methane), PM and fluorinated gases.

Agricultural activities
The extensive use of pesticides, insecticides and fertilizers generates by-products such as ammonia, which is a hazardous gas in the atmosphere and can potentially cause water pollution.

Industrialization
Factories and other manufacturing industries strain the environment by producing large amounts of greenhouse gases, hydrocarbons, organic compounds and unwanted chemicals that are harmful when humans and animals are exposed to them for extended periods of time, which often happens to (rural) people living in the vicinity.

Mining
Mines conduct operations such as drilling, hailing, collection and large-scale transportation associated with high energy usage and excessive gas combustion that are sources of air pollution. Mining sites are often outside the boundaries of urbanized areas, but in some cases these sites are very close to urban areas and cause additional pollution for the inhabitants.

Indoor air pollution
Urban dwellings are generally more polluted compared to rural dwellings, as reported by Hulin et al. (2010) when accounting for indoor air pollution. Everyday exposure to indoor air pollution is associated with a higher risk of childhood asthma, even in low concentrations.

Natural air pollution
Not all air pollution is anthropogenic; some are generated naturally in the environment through radioactive decay, forest fires, dust and active volcanoes, for example.

The effects of air pollution lead to several hazardous and unpleasant scenarios and it is critical to minimize these effects for long-term sustainability of a healthy environment in cities and in rural areas. The most distressing effects are those that influence human and animal health. Additional effects such as haze and smog that hang over many cities and industrial areas, which negatively affect people's morale, can be limited by practices similar to those used to decrease hazardous air pollution. The most likely effects of air pollution due to urbanization and industrialization are:

- health hazards to humans and animals caused by respiratory and heart problems,
- global warming as a result of greenhouse gases emitted mostly from coal-based (fossil fuel) sources, and
- acid rain formed when water and nitrogen/sulfur oxides react, damaging farmlands and making rainwater undrinkable.

Other effects of air pollution include:

Algae
A high concentration of nitrogen can form thick unwanted layers of algae on water surfaces, which block natural sunlight and limit oxygen intake in water, leading to potentially dangerous environments for animal species.

Wildlife
Similar to adverse health effects on humans, wildlife can be harmed by air pollution if toxic chemicals enter the respiratory systems of animals. City-generated air pollution, agricultural activities and transport of people and goods towards rural areas contribute to air pollution affecting wildlife and farm animals.

Depletion of the ozone layer
A thinning of the ozone layer has been experienced because of air pollution, especially the presence of chlorofluorocarbons and hydro-chlorofluorocarbons in the atmosphere. A purpose of the ozone layer is to protect the earth from harmful UV rays and this process becomes less effective as the ozone is depleted.

Haze
Fine particulate matter in the air causes haze, the encounter of sunlight with pollution particles in the air, obscuring the clarity and hence visibility, color and texture of the air from its ideal state. Haze is aesthetically unappealing and can be carried downwind, affecting large areas on earth.

Crops and farming damage
Ground-level ozone can lead to reductions in agricultural crops, reduce the growth and survivability of plants and increase susceptibility to diseases and pests. In addition, acid rain and increased UV through ozone depletion contribute to damage to crops.

Limiting and/or eradicating air pollution is not an easy task and requires the cooperation of industries, agriculture, technology, governments, and individuals in urban and rural areas, in developed and developing countries. Smart cities generally aim not only to improve the quality of life of their inhabitants through information-sharing, but more importantly, to secure a sustainable future by addressing pollution generated in the city and surrounding areas. There is a practically unlimited number of ways to reduce people's carbon footprints, depending on their situations, but some important methods of decreasing air pollution are:

Carpooling, public transport, or cycling
By having more people using one vehicle, carpooling reduces each person's travel costs such as fuel costs, tolls and the anxiety of driving. Reducing the demand for oil and gas by reducing the number of vehicles on the road daily can effectively lead to less air pollution in congested cities. Carpooling is also a more environmentally friendly and sustainable way to travel, as sharing journeys reduces carbon emissions, traffic congestion on the roads and the need for parking spaces. Authorities often encourage carpooling, especially during periods of high pollution or high fuel

prices. In an effort to reduce traffic and encourage carpooling, some governments have introduced high-occupancy vehicle (HOV) lanes in which only vehicles with two or more passengers are allowed to drive. HOV lanes can create strong practical incentives for carpooling by reducing travel time and expense. In some countries it is common to find parking spaces reserved for carpoolers.

Efficient use of energy
Many daily routines require lighting, air-conditioning, the preparation of food and entertainment. Energy-saving appliances have enjoyed a large market share in recent years and pave the way for sustainable energy savings in everyday life.

Recycling
Incentives to encourage recycling of glass and plastic, and innovative and interesting ways to reuse items provide another effective means to reduce a person's carbon footprint.

Clean energy, solar, hydro, and wind power
Energy from renewable sources is an active and promising research field, with advances being made every year in the efficiency of renewable energy devices. Residential and commercial buildings can benefit from using renewable energy through the installation of solar panels to alleviate dependence on the power grid.

Energy-efficient devices
Incandescent lamps have mostly been replaced by LED/fluorescent alternatives, boasting an average of 80-90 % savings in energy per light. Households and companies are encouraged to replace inefficient devices with low-energy alternatives.

Electric cars (e.g. Tesla®)
Although this technology is in its infancy, major strides have been made towards electric cars replacing the internal combustion engine. At the forefront of this technology are companies such as Nissan®, Chevrolet®, BMW®, Mitsubishi®, and Tesla®, which are reviving interest in electric cars and providing more than adequate alternatives to the range and refueling time advantages of the internal combustion engine, which is responsible for the larger part of air pollution in urban areas.

Raising awareness
Many people are not aware of methods and techniques that are available to reduce their energy consumption and hence to limit air pollution. Governments and private companies are encouraged to raise awareness and improve the education of citizens to limit and combat air pollution.

Taxes, laws, and regulations
Industries that generate large amounts of air pollution through manufacturing processes should be (and are) regulated by international regulating bodies such as the Environmental Protection Agency to govern air pollutants released into the atmosphere.

Similar to reducing air pollution in urban, industrialized and rural areas, water pollution has also become prevalent in many overcrowded places, leading not only to unpleasant sites but also to potential health hazards. Among the different types of water pollution, urban water pollution is increasing as cities grow larger and effective distribution of waste products becomes more difficult. Agricultural water pollution through run-off of pesticides and chemicals used for farming is somewhat easier to regulate compared to urban water pollution, purely based on the scale of the population. Sources of urban water pollution include:

Sewage
Untreated or poorly treated sewage is low in dissolved oxygen and high in pollutants such as fecal coliform bacteria, nitrates, phosphorus and chemicals that have various detrimental effects on the environment and human and animal life.

Urban run-off
Industrial and residential run-off from localized sources in streets have cumulative effects when entering water bodies such as rivers, increasing the chances of cancer and other threatening diseases to large parts of the population dependent on the water sources. When rain falls or snow melts, the runoff washes pollutants of streets, parking lots, construction sites, industrial storage yards and lawns. Urban runoff carries a mixture of pollutants from cars and trucks, outdoor storage piles, muddy construction sites and pesticide spills. Efficient systems of ditches, gutters and storm sewers carry the polluted runoff to nearby lakes and streams, bypassing wastewater treatment systems.

Wastewater
Wastewater is a term that is used to describe waste material that includes industrial liquid waste and sewage waste that are collected in towns and urban areas and treated at urban wastewater treatment plants. It also refers to water that comes from single houses in the countryside and treated on-site in either septic tanks or individual wastewater treatment systems such as domestic wastewater treatment systems.

Chemical waste
Chemical waste from factories and local industries, even the mismanagement of smaller quantities of domestic chemicals used in cleaning products, can enter water streams in a similar way as urban run-off and these polluted water streams can have detrimental effects on the water quality in urban and surrounding areas.

Radioactive waste
Radioactive waste can vary greatly in its physical and chemical form. It is found in various forms such as solid, liquid, gas, or a mixture called sludge. Any given radioactive waste can be primarily water, soil, paper, plastic, metal, ash, glass, ceramic, or a mixture of many different physical forms, including chemical forms with varied consistency. Each chemical has a distinct biochemical action; for example, iodine seeks out the thyroid gland, strontium clumps to the bone and teeth (calcium), and cesium is distributed throughout the soft tissues. Radioactive waste

can contain radionuclides of very light elements, such as radioactive hydrogen (tritium), or of very heavy elements, such as uranium.

Oil pollution
Apart from relatively small amounts of oil entering urban water streams from run-off, harbor cities have experienced major crude oil spills from tankers, offshore platforms, drilling rigs and wells, causing long-term damage to the environment. Oil can continue to seep into city waterways decades after the spills have occurred, causing residual effects in urban and rural areas.

Plastics
Plastics that are inadequately disposed of from cars, households and over-full rubbish bins, or are carried by wind, pollute the environment and have become common in many parts of the world. Illegal dumping of plastic and overflowing of containment structures play a significant role. Polyethylene, polyvinyl chloride and polystyrene are largely used in the manufacture of plastics. Excessive molecular size seems to be mainly responsible for the resistance of these chemicals to biodegradation and their persistence in the soil environment for a long time. Plastic pollution is mostly regarded as an aesthetic nuisance over the short term for humans, but can be fatal for animals mistaking the items for food, or becoming entangled in elastic plastics.

Invasive species
Physical and biological disruptions of aquatic systems caused by invasive species alter water quantity and water quality. Invasive species are non-native species that threaten the diversity or abundance of native species because of their uncontrollable population growth, causing ecological or economic impacts. Common methods of accidental introduction of non-indigenous species in marine environments include the release of ballast water of oceangoing vessels, illegal fish stocking and improperly cleaned equipment.

Thermal pollution
Thermal pollution is defined as a sudden increase or decrease in the temperature of a natural body of water such as a lake, river, pond or an ocean by human influence. This generally arises when factories and industrial plants use water to cool down equipment and channel the water, with altered temperature, into streams. Aquatic forms of life have evolved to be very temperature-dependent and often cannot regulate their body temperature to rapidly changing conditions, resulting in potential fatal consequences for these species.

Sediment
Sediment is the loose sand, clay, silt and other soil particles that settle at the bottom of a body of water. Sediment can come from soil erosion or from the decomposition of plants and animals. Wind, water and ice help carry these particles to rivers, lakes and streams. The most concentrated sediment releases come from construction activities, including relatively minor home-building projects such as room additions

and swimming pools. Sediment entering water streams degrades the quality of water for drinking, wildlife and farming crops.

Air dispersion models are used to provide an estimate of a concentration or deposition of a pollutant emitted from an industrial process (point source) or a road (line source). Output from dispersion models is often used to predict the contribution of a new or existing process or the level of pollutants at specified points. The modelled outputs of concentrations and depositions can then be compared with environmental limits and human health air quality limits. Typical applications of these models include:

- IPPC authorizations,
- odor modeling,
- water run-off modeling,
- environmental impact assessments, and
- appropriate assessments under the Habitats Directive.

Air and water pollution dispersion models generally follow a certain structure in their development and application to determine the benefits of the model and the risks inherent to pollution in the immediate and surrounding areas. Adapted from the European Consortium for Modeling Air pollution and Climate Strategies (EC4MACS) methodology, the generalized structure is summarized as tasks that are performed:

- scenario development to include assessment of demand,
- quantification of emissions under a baseline scenario,
- quantification of emissions under scenarios involving further abatement than that specified in the baseline scenario,
- quantification of the costs of abatement in the new scenario,
- modeling the dispersion of emissions, including the formation of secondary pollutants,
- estimation of exposure of susceptible receptors,
- application of exposure-response functions,
- monetary valuation of impacts,
- comparison of quantified costs and benefits,
- consideration of the effect of uncertainties, and
- benefits, known and possible impacts.

Taken from the EC4MACS website, "*The EC4MACS toolbox is now ready for scientific and economic analyses to inform the revision of the Thematic Strategy on Air Pollution in 2013 and the European Climate Change Programme on climate strategies beyond 2012. The EC4MACS toolbox informs about the costs and benefits of the various policy options to reduce greenhouse gas emissions and to further improve air quality in the European Union while maximizing the benefits to EU energy, transport and agricultural policies.*" The EC4MACS consortium involves the major European institutions that work on the modeling of air pollution and climate strategies, including:

- The International Institute for Applied Systems Analysis, Austria.
- The Coordination Centre for Effects at RIVM, Bilthoven, Netherlands.
- The E3 M-Lab of the National Technical University, Athens, Greece.
- The Laboratory for Thermodynamics of the Aristotle University of Thessaloniki, Greece.
- The Institute for Agricultural Policy, Market research and Economic Sociology of the University of Bonn, Germany.
- EuroCARE GmbH Bonn, Germany.
- Ecometrics Research and Consulting, AEA Technology, MetroEconomica, UK.
- L'Institut National de l'Environnement Industriel et des Risques INERIS, France.

Two EU joint research centers also participate in EC4MACS:

- The Institute for Environment and Sustainability of the Joint Research Center in Ispra.
- The Institute for Prospective Technological Studies of the Joint Research Center in Seville.

The inevitable statistical unknowns that are generally estimated through practical experimentation and previous experience are important to consider during air and water pollution modeling. Developing countries can benefit from institutions and organizations in developed, well-funded nations that have performed many of these experiments or are willing and able to assist in developing new models and estimates for these regions. Dispersion models can become mathematically complex and computer-intensive to solve, and modeling opens up new disciplines in developed countries to study, research and distribute this knowledge and gain expertise to assist developing countries seeking assistance.

Pollution dispersion models are only as good as the data from which they are derived. GIS and remote sensing provide various means to gather information about a specific area, including weather patterns, human interactions, natural disaster risk monitoring, water level, pollution and many other sources. A crucial component in any monitoring system is accessibility to power sources. Energy harvesting from ambient sources provide a sustainable way to gather and store energy for use in these devices, regardless of the application. Several energy-harvesting techniques are available, such as radiation harvesting, wind and hydro power, RF energy harvesting and movement (piezoelectric harvesting). The requirements, advantages, potential applications and disadvantages of energy harvesting are summarized in the following section.

9.5 Ambient Sources and Energy Harvesting

One of the biggest market drivers for energy harvesting is low-power communications and devices. Any company that makes products for portable devices is eligible to benefit from this market opportunity. Many semiconductor companies, for instance, are already making ICs that could fit this space, or they could modify an existing product line.

Successfully developing energy-harvesting technology requires expertise from various disciplines of physical sciences, including:

- energy capture (sporadic, irregular energy rather than sinusoidal),
- energy storage,
- metrology,
- material science, and
- systems engineering.

In developed countries, the expertise is being developed at a relatively adequate pace and graduates are entering the market to assist in developing energy harvesting as a sustainable means of device integration required by technologies such as the IoT. In developing countries, however, there is a lack of new graduates entering the market because of a lack of school-level education, funding, certified teachers and limited exposure to technology. International organizations can aid developing countries in developing their human capital to contribute to sustainable educational practices, which could in future offer courses and learning environments for students willing and interested in investing in science and technology. Globally, energy harvesting is currently receiving a fair share of interest and is a growing market with ample potential.

Energy-harvesting technology-enabled devices and even wearable technology present the potential for massive transformation in many industries. The more noticeable ones include consumer electronics, communications and military applications. Early adopter industries of wearable technology include clothing, healthcare, sport and fitness. Smart watches that are wristwatches able to run (parts of) an operating system and perform calculations and display vital information such as incoming messages and calls, and information about the wearer's daily activities are becoming more commonplace in the market. The socio-economic relevance of energy-harvesting technologies is relatively straightforward, with a projected market potential of USD $3.3 billion forecast for 2020 and abundant opportunities for job creation. Moreover, wireless sensor networks powered by energy harvesters are relatively easy to install, require minimal maintenance and reduce energy consumption, while the data they generate can be used to optimize all kinds of processes, from wildfire detection in rural areas to building automation in urbanized environments.

Energy harvesting has several other advantages, many of which are application-specific, but the overall implementation of such a technology has distinct advantages for consumers and businesses. Some of these examples include:

Convenience
Devices operating on battery-powered sources must be recharged at relatively short intervals, ranging from daily to weekly, depending on the device and its usage. Energy harvesting aims to improve convenience by stretching the intervals between recharge events, or eliminating the recharge process completely. The advantage to

the consumer is convenience because less planning is required around the operation time of the devices. In poorer countries where energy from the grid is sometimes limited, users would be able to reduce their dependence on the national grid and have devices that operate regardless of the availability of conventional power.

Backup energy sources
The reliability and consistency of many electronic systems are crucial, especially in hospitals. Crucial monitoring devices require constant power sources that cannot not always be guaranteed from the national grid, especially in developing countries. Backup power sources through effective energy harvesting can prove vital in these devices' operation in emergencies such as blackouts.

Mobility
People tend to shy away from technologies that are static and dependent on a single power source, or that have very limited operation due to energy restrictions. WSNs and generally the IoT thrive on mobility to set them apart from conventional technologies, and are another driving factor in the growing market of wearable devices. Communication protocols are reinvented and modified to use less energy and IC manufacturers are decreasing their power requirements and the physical footprint of their products, leading to increased mobility and adding to the convenience of these devices. Researchers, explorers and medical staff operating in rural and sometimes difficult-to-reach places require mobility of their equipment to complete tasks associated with their jobs effectively and efficiently. Being tied to technology that is bulky and requires constant attention inhibits their ability to operate in the field. Energy harvesting in these small devices, using various forms of harvesting techniques, enables longer operation times, thus precluding the need for batteries or connections to power lines.

Business practices
Initial investment and operating costs for consumers are reduced in terms of device packaging, cost of development, disposal, longevity and reuse. Device installation and long-term maintenance cost are also reduced through the omission of charging devices, additional cabling and energy cost. Corporations can inevitably save costs on distributing these devices and focusing only on the development of these device as opposed to the supporting circuitry and dependence on other technologies. Smaller startup companies in developing countries can therefore enter the market with lower overheads and increase their chances of success in an already active market segment where competing with European and American companies is often difficult for these companies. Generating local revenue through locally owned and operated companies is economically encouraged for long-term sustainability.

Spatial independence
One of the advantages of increased mobility in these devices is their spatially independent features. WSNs can be placed in conventionally hard-to-reach places, in rural areas with limited power sources, and in areas where there is less human activity and therefore less potential for tampering.

Radiation, wind, water and movement are not the only large players in the energy-harvesting market; rainwater harvesting can also be seen as a separate entity with large potential, especially in countries where rainfall is abundant or of high intensity. Although close to three fourths of our planet consists of water, not all of it is suitable for use. The water in the oceans and seas cannot be used as drinking water and little of it can be used for other purposes. As a result, there is a constant shortage of water that is good for either drinking or home and industrial use. In developing countries this poses a problem, as in certain areas there is already a shortage of drinking water, low rainfall, and geographical limitations to accessing large water bodies, therefore good use of the limited amount of available water should be prioritized. Rainwater harvesting is a process or technique of collecting, filtering, storing and using rainwater for irrigation and for various other purposes, most importantly as drinking water in many rural areas. In an urban setting, harvesting is generally accomplished through infrastructure adjustments that use storage equipment such as storage tanks. This is why rainwater harvesting can be regarded as a form of energy harvesting, where a naturally occurring event is captured, stored and used. Rainwater in its purest form, if not contaminated already and precipitated as acid rain, is free from pollutants as well as salts, minerals, and other natural and man-made contaminants. There are several advantages to rainfall harvesting that make this technique attractive in urban as well as rural areas, including:

Maintainability
The technology required to store rainwater is of relatively low complexity. The implementation cost is generally lower than that of water purification or groundwater pumping systems. The non-renewable properties of drinking water further highlight the necessity of implementing such systems in areas that have limited access to drinking water.

Utility bills
For most practical reasons, rainwater is free for anyone to use and is not charged by utilities through standard means of billing. Anyone with a rainwater storage facility can harvest rainwater for use later, either for drinking water or for agricultural purposes, or if an excess is available, for any other purpose.

Irrigation
Combining rainwater-harvesting systems with low-cost irrigation systems can be extremely beneficial if an abundance of rainwater is collected during the rainy season in a country. Implementing capturing systems at higher altitudes means that the water can automatically be distributed downwards to ground level, which decreases operational cost compared to more expensive groundwater pumping systems. Rainwater, as mentioned, is also free of chemicals and contaminants and can be very effective to irrigate crops and gardens.

Versatility
Rainwater can be used for many purposes, practically for anything that utility-supplied water is used for. In urban areas where the sizes of homes are

generally smaller compared to open areas, rainwater harvesting can be used for a variety of applications, such as toilets, gardening or washing.

Unfortunately, rainwater harvesting is not fool-proof and is subject to several disadvantages that plague its uptake and effectiveness. Some of these disadvantages are listed below. Some can be overcome, but some are nature-specific and difficult and sometimes impossible to ignore, making rainwater harvesting a less viable option in some areas.

Unpredictability
Rainfall varies significantly between countries and rainfall patterns can be drastically different in areas separated by only a few hundred kilometers. In countries with very low annual rainfall, depending on rainwater for critical operations is not advised and alternatives should be explored in these areas. In Aoulef in Algeria, for example, the average annual rainfall is less than 12.19 mm. This makes it a very dry area in an already dry country, making rainfall harvesting practically impossible.

Initial investment
Although deemed less expensive than for example water purifiers and pumping systems, the initial cost of storage tanks and accompanying distribution ducts can be high, depending on the size of the area the system will be serving. Similar to most other energy-harvesting techniques, such as solar panels, the initial cost is sometimes a deterrent, and the promise of recovering the cost over a certain period (generally a number of years) might not be enough to convince the consumer to make the initial investment. In developing countries especially, people tend to be much less enticed by future benefits and sometimes only view their immediate expenditure as the deterrent to investment in these technologies.

Maintenance
As with most consumer products, regular maintenance is required to ensure long-term sustainability. If not looked after well and maintained regularly, natural contaminants can decrease the quality of especially warm and static water. Such contaminated water becomes a breeding ground for various species, and the cost of maintenance, or even the commitment to regular maintenance at the time of purchase, could also deter many users from adopting such a system. Government grants in rural and underdeveloped countries could help sustain these systems, but priorities are generally focused on the classical methods of water supply, since these are more sustainable (less dependent on rainfall).

Storage limitations
The physical size of the storage system, the amount of rainfall it can gather and the size of the dwelling all contribute to limiting the amount of rainfall that can be stored and used later. Urban areas generally have very little additional space, especially if the dwellings are predominantly apartments, and the size of the potential storage tank might be too small for sustainable use. In larger areas, such as farming land or smallholdings, the size of the storage tank required might be too large, making the initial investment and maintenance costs impractical.

These disadvantages are not only applicable to rainfall harvesting, but also applicable to many other renewable energy-harvesting techniques. In general, energy harvesting of naturally occurring energy is an unpredictable and initially expensive exercise and still requires research and improvements to convince users in urban and rural areas to invest in it. Only now (2015) are companies such as Tesla® able to convince people that electric cars, for example, can be sustainable and a viable alternative to the internal combustion engine, although electric cars have been around since the 1880s. People are still weighing up price, performance, convenience and long-term sustainability when thinking about adopting a new technology and the fact is that these technologies are still expensive (Tesla® Model S costing > USD $70 000) and can also practically be considered less than ideal (charging station availability and charging time). Technology advancements and steady increases in market share contribute to the wider acceptance of these alternatives. Technological developments such as the IoT and Big Data and the increasing scarcity of resources in volatile markets are also driving energy-harvesting technologies, along with the rate of climate change and its impact on society. The issue of climate change on its own is an intensely debated and vastly researched topic and geographically dependent. Each country/region generally conducts its own studies of the effects of climate change and aims to rectify some of these issues through technology and constant monitoring of sites through systems such as WSNs and the IoT. In brief, the following paragraph highlights some effects experienced in the USA, based on region and local climate.

On Nasa®'s website describing the effects of climate change on the planet, it lists several impacts that are currently visible throughout the USA and will continue to affect the particular regions and inevitably spread to other regions as well. These effects are also visible in many other countries, based on the similarity of their geographic location to the four mentioned areas in the USA. These effects, according to the Third National Climate Assessment Report 2, released by the US Global Change Research Program, are:

- Northeast (includes dense cities, sparsely populated towns and extends from the coast to inland plateaus and mountains): Heat waves, heavy downpours, and sea level rise pose growing challenges to many aspects of life in the northeast. Infrastructure, agriculture, fisheries and ecosystems will increasingly be compromised. Many states and cities are beginning to incorporate climate change in their planning.
- Northwest (bordered by the Pacific Ocean to the west and Canada to the north, with mild climates on the west of the mountain ranges and sunny and dry climates on the east): Changes in the timing of streamflow reduce water supplies for competing demands. Sea level rise, erosion, inundation, risks to infrastructure and increasing ocean acidity pose major threats. Increasing wildfires, insect outbreaks and tree diseases are causing widespread tree die-off.
- Southeast (a region with generally warm and wet climates with mild and humid winters): Sea level rise poses widespread and continuing threats to the region's economy and environment. Extreme heat will affect health, energy, agriculture

and more. Decreased water availability will have economic and environmental impacts.

- Midwest (hot and humid summers and cold winters, since the region is far from the temperature-moderating effect of the oceans): Extreme heat, heavy downpours and flooding will affect infrastructure, health, agriculture, forestry, transportation, air and water quality and more. Climate change will also exacerbate a range of risks to the Great Lakes.
- Southwest (a wide range of elevation and climate types across the region with drier and desert climates in the south and cold weather with more precipitation and snow near the mountains in the north): Increased heat and drought (and insect outbreaks) are all linked to climate change and have contributed to an increase in the amount of wildfires. Declining water supplies, reduced agricultural yields, health impacts in cities due to heat, and flooding and erosion in coastal areas are additional concerns.

More information on the effects of climate change globally is presented in the final section of this chapter. To tie up with energy harvesting, some of the advantages and limitations of GIS, remote sensing and ubiquitous computing are presented in the following section.

9.6 GIS, Remote Sensing and Ubiquitous Computing

GIS is attracting growing interest worldwide and the realization of its economic and strategic benefits is part of the large driving force towards its growth. The generalized advantages of GIS are discussed below.

Increased profit margin
Increasing profit through cost savings by implementing GIS can be witnessed in three areas of the market. Firstly, GIS provides geospatial information to its users, enabling them to plan operations better initially to deal with possible risks, or to use information to their advantage. Better planning often leads to more seamless completion of tasks and reduces the time required, equating to an increase in profits. Secondly, mid-operation, real-time information updates allow users to update and manage tasks constantly and avoid potential risks that may occur during operation. Often re-routing of product distribution due to traffic congestion leads to time savings and again to an increase in profits through saving on petrol, man-hours and vehicle maintenance cost. The third effect of better planning can be witnessed indirectly in increased customer satisfaction, ultimately leading to better relations with clients and a more trustworthy company profile.

Strategic decision-making
The abundant information available from GIS makes strategic planning and decision-making more accurate. Implementing a GIS that extracts information about a specific location provides more information about the site than can

necessarily be gathered through classical approaches. Natural resource extraction is an example of GIS providing additional information on the potential extraction sites, and the user is able to gather enough information on each site to plan and execute endeavors that are more successful. An open-information architecture and approach can benefit developing countries, since the availability of the information is not limited, although the physical hardware required to implement GIS might not be feasible for organizations in these countries. Using the information provided can therefore increase the chances of success for developing nations' companies that are trying to make a name and contribute internationally.

Better understanding
The interpretation of raw data is generally difficult in any situation. In any area, displaying raw data in a logical and easily readable format is a discipline in which entire teams are generally dedicated to the process. This is an important, sometimes underestimated procedure that enables anyone looking at the manipulated results to make informed decisions, regardless of his or her background, education and technical expertise. It is often found that providing visually enticing results to local communities in developing countries, not necessarily educated in science, can provide valuable feedback based on their long-term experience of living in the area. Weather patterns and drought, for example, can be presented visually to the community, and the potential means to implement countermeasures can be contributed by the locals.

Maintaining records
Geographical accounting encompasses recording and maintaining information for future reference for various reasons. When dealing with rezoning, population census, land ownership or administrative boundaries, for example, the storage of records for future generations is essential. GIS allows data records and reports to be kept safe in the long term through tools that support GIS applications and workflows. If information has not been accessed over a long period, its interpretation may change or it may be forgotten, whereas GIS provides easily understandable visual cues to lessen the effort of understanding historical data, and map-based visualization is a standard that has not changed over time (understandably so, since there are only so many ways to use a map), benefiting future generations.

Official communication
Governments and administrative bodies use GIS to communicate the status of a community to one another and local citizens. GIS information provides effective means to organize and present geographical knowledge, which directly influences local populations, and governmental distribution of the information generally increases the trustworthiness of the information supplied. Utilities, for example, can distribute information about water usage and encourage citizens in geographically separated areas to control their consumption of non-renewable resources. It is generally found that people tend to react strongly to visually appealing information such as "red-zones" (dangers/risk/high usage) compared to supplying them with

tables of raw data, and GIS already provides these graphical representations as standard procedure.

The integration of digital remotely sensed data and cartographic information (thus, remote sensing as a separate discipline) into GIS for visual representation leads to more complete and accurate data displayed on GIS maps. Historically, the difference between remote sensing and GIS was purely based on the fact that digitally gathered data through remote sensing served as the input to GIS, which generated the appropriate outputs. The dependence of GIS on remote sensing has become more integral with technology improvements, since placing sensors on aerial devices such as planes, balloons, or satellites has become more commonplace, eliminating the need to gather data on geospatial analysis physically. Davis and Simonett (1991) define the basic principles/properties of a remote sensor as having;

- spectral coverage (spectral band locations),
- spectral resolution (spectral bandwidth),
- spectral dimensionality (number of spectral bands),
- radiometric resolution (quantization),
- an instantaneous field of view,
- an angular field of view,
- a point spread function, and
- a temporal response function.

As noted from these requirements of a remote sensing system, they are thematic to the digital world. Remote sensing can therefore in broad terms be defined as the discipline in which digital data are remotely (without being in physical contact with the monitored component) collected and used in conjunction with GIS where these are presented. Similar to GIS, the applications of remote sensing include:

- agriculture,
- urban planning,
- resource monitoring,
- natural disaster assessment,
- epidemiology, and
- archeology

and the information provided by either active or passive remote sensing equipment is generally categorized as

- 2D/3D geometry,
- descriptions of land cover,
- evidence of land use,
- color through spectral reflectance,
- surface texture,
- surface roughness,

- moisture content,
- vegetation biomass,
- cloud/smoke density and range, and
- wildlife ecology.

Advantages of remote sensing include:

- improved vantage points (synoptic views),
- broad spectral sensitivity,
- increased spatial resolution,
- 3D perspective through aerial imagery,
- comparability of data,
- digitally storing records,
- rapid data collection,
- quantitative analysis,
- cost savings, and
- sharing/integrating layers in Google® Earth.

GIS and remote sensing do also have their fair share of problems when it comes to integrating systems. Inherently, digital systems, data gathering and data distribution often present issues and risks due to human error, component malfunction or inadequate maintenance. Identifying these issues is important in any environment, since errors can propagate through a larger system and have adverse effects for the end-user.

Errors and uncertainty
Digital equipment is prone to errors, depending on the amount of redundancy in the programming code. Errors such as bit-flips, systematic errors, random errors, reading errors, calibration errors and data uncertainty about the probability of spatial data errors occur in digital processing and must be accounted for in GIS and remote sensing as well. Typical errors experienced in GIS include positional and topological errors from the GPS and DEM accuracy, and attribute inaccuracies from the radiative sensing equipment. Identifying the errors might not always be a simple task, as error propagation through the system is generally only picked up at a late stage in the data analysis. Inconsistent data and visual maps generally identify the presence of errors earlier in the data-gathering phases, generally in the remote sensing application. Errors could also propagate through the GIS if incorrect or inconsistent graphical representation techniques are implemented. Supplying developing countries with raw information or GIS spatially distributed visual maps that are incorrect, inaccurate or inconsistent can lead to long-term effects, especially if these countries use the information directly to implement precautionary measures at high initial cost.

Processing power
GIS and remote sensing equipment may require significant processing power to analyze and display raw data in visual maps. Developed countries overall invest

large sums of money in equipment for various applications, including GIS and remote sensing. Developing countries, however, tend to prioritize more essential equipment for basic utilities and spend less on processing equipment for GIS and remote sensing. The downside of this is that these countries are then dependent on the already manipulated data supplied by more developed institutions, do not have much say in the final presentation and cannot determine whether the data are incorrect.

Skilled workers
A similar investment problem as with physical processing power is investment in human capital, sometimes lacking in developing countries. GIS and remote sensing require skilled workers such as engineers and scientists that interface between the equipment and the presentation of the data. This is generally acceptable during the initial phases of such a system's implementation, but can become problematic for longer-term maintenance. Local talent must first be developed from school level, where the level of instruction is often already unsatisfactory, before workers can be trained in using and maintaining the equipment. The electronic systems and programming requirements of this type of infrastructure are generally relatively complex and impossible to simplify, since the accuracy, trustworthiness and relevance of the data are heavily dependent on the quality of the system and its operators.

Broadband requirements
Digital distribution of information to developing countries requires good quality broadband internet access that can sustain large amounts of data from international organizations to developing countries and vice versa. Mobile or fixed-line data connections are not always a given in rural areas, especially in developing countries, and distributing the digital data can be problematic in these cases. Since some countries already cannot afford fully implemented GIS and remote sensing equipment, they rely on receiving the data in a timely manner, or in the case of emergency systems, the data must be distributed as soon as possible. Not having a reliable network that can manage incoming information at high bandwidth requirements could limit the use and effectiveness of such a system. Visually manipulated data, such as layered maps, are generally large formatted files, even if compressed, and sending the raw data might again not be feasible, since a lack of skilled workers to process the information at these sites is also an issue, which was addressed in the previous paragraph.

Omission errors
The resolution and minimum detectable signal of digital equipment determine the equipment's ability to distinguish between disturbances and naturally occurring phenomena, which might not be identified as a risk or might fall in the low-impact category. Often GIS and remote sensing equipment are calibrated to a level where actual anthropogenic or natural events that pose a risk to the community are not

picked up and overseen by the system. These are termed omission errors and can become problematic in emergencies or in early identification of disasters. Constant changes and calibration alterations are required in these systems to teach them to distinguish between events, and if risks are overlooked or underestimated, or if a lack of skilled workers inhibits these checks, GIS and remote sensing equipment cannot be fully efficient and functional.

Geo-referencing historical maps
Historical maps of GIS data and even older generation remote sensing equipment operating according to outdated or obsolete standards present difficulties in the thematic layering of data onto newer generation digital maps. Old maps, generally created by hand, must be scanned in and digitized to be compatible with modern programs, and these operations often introduce scaling and projection errors of spatial-temporal data. The scale of a map and geographical rendering of area representation are very important aspects, since the information content depends mainly on the scale set and resulting geolocation information of the map's representations. In order to digitize a map, the map has to be checked within theoretical dimensions, then scanned into a raster format, and the resulting raster data have to be given a theoretical dimension by a rubber sheeting/warping technology process. The potentially introduced errors are mainly due to human error and are often difficult to pick up or rectify, since they are not dependent on tried-and-tested automated computer programs.

Certain aspects of ubiquitous computing can be similarly categorized with GIS and remote sensing in terms of their technological roots, meaning that the ubiquitous computing devices inherently suffer from disadvantages such as limited energy resources, low rates of adaptability, risky initial investments and relatively high cost for fully integrated pervasive systems. Their advantages are, however, worth noting and exploring the future of ubiquitous computing in developed and developing countries can be extremely powerful, as it provides practically invisible computing, interactions between architecture such as smart buildings, enhanced decision-making and an idealistic convergence of digital technologies.

GIS and remote sensing are more focused on reporting on environmental changes than ever before and technology advances are enabling these technologies to predict climate changes much more accurately. Satellite-based remote sensing is becoming relatively less expensive and educational institutions and small companies are turning towards technologies such as CubeSat to launch their own remote sensing infrastructures. Open-source data and the availability of resources such as cluster computing, even in some developing countries, enable more researchers to access valuable information about the environment. However, climate change is affecting the world and changing historical natural patterns. Implementing technology to combat climate change requires sufficient information about the sources (pollution), effects and possible solutions (technology and information sharing).

9.7 Environmental Impact of Climate Change

According to the IPCC, an increase in greenhouse gases in the atmosphere is likely to boost temperatures over most land surfaces, though the exact change will vary regionally (Mills et al. 2006). Possible but more uncertain outcomes of an increase in global temperatures include an increased risk of drought and increased intensity of storms, including tropical cyclones with higher wind speeds, a wetter Asian monsoon and possibly more intense mid-latitude storms. The effects of climate change are widespread and climate-specific. Geographically, regions are affected differently by climate changes, but the general consensus is that climate change is altering the environment for the worse. The general effects of climate change that have been identified are summarized below.

Higher temperatures
Heat-trapping gases emitted by power plants, automobiles, deforestation and other sources are warming the planet. Warmer temperatures affect oceans, weather patterns, snow and ice in the polar regions, plants and animals. Increased air temperature does not affect these environments in isolation and spreads between ecosystems. Warmer temperatures in oceans, for example, force aquatic animals and coral reefs to adapt at much faster rates than what these species are genetically programmed to do, through evolution for example, and can have fatal consequences for these species. Increases in air temperature also affect humans directly as more frequent and more severe heat waves occur, causing fatigue and sometimes death in humans if there is no access to air-conditioning or other means of cooling the immediate environment, often seen in developing countries. In a hot environment, the body relies mainly on two mechanisms for cooling off and maintaining a safe temperature: evaporation of sweat from the skin and increased blood flow to the skin. These processes can place strain on the heart and lungs. Excessive heat can result in a range of adverse health impacts, including heat cramps, heat edema (swelling), heat syncope (fainting), heat exhaustion and life-threatening heat stroke. High humidity presents an additional threat by hindering the evaporation of sweat, which can cause body heat to accumulate and further increase the risk of health problems. In severe cases, for example in the polar regions, animals such as polar bears have no way of escaping the rise in temperature and could suffer fatal consequences if they are unable to adapt to warmer temperatures.
In order to protect public health and combat climate change, governments must immediately enact legislation to reduce global warming pollution and implement comprehensive heat preparedness plans. This should include the creation of health warning systems about heat, establishment of cooling centers, mobilization of emergency response systems and public education and outreach.

Changing landscapes
Accompanied by rising temperatures and a shift in weather patterns, trees and plants are moving to areas that are more suited to their sustainability, generally closer to the polar regions and towards higher altitudes where the air is colder and the

oxygen/carbon monoxide content is better regulated. The moving vegetation over a long term will leave behind dry and barren areas prone to erosion, and decrease the oxygen content where human settlements have been established for many years (generally close to natural resources such as water, food and vegetation). Animals will also suffer as the vegetation moves towards cooler areas and natural and anthropogenic barriers might block the movement of these animals, depriving natural sources of food and water. Animals are also adapted to very specific environments, not based on only the vegetation, but also on the elevation, humidity, weather patterns and natural shelter, further limiting their options of moving towards more lush areas.

Wildlife at risk
Climate change in essence is happening faster than many animals' ability to adapt. Species that are already in danger of extinction experience increased pressure to adapt and survive in a changing environment. Examples of animals under severe pressure of climate change are tigers, snow leopards, Asian rhinoceros, orangutans, African elephants, polar bears and Antarctic penguins. Animals such as the American pika are forced to move to higher altitudes each year by the rise in air temperature, and a time will come when the pika cannot move any higher and will enter the endangered species list. Wildlife depends on healthy habitats and requires constant and predictable temperatures, fresh water, plentiful food sources and shelter to raise its young. Climate change and global warming are affecting all of these requirements and are extremely detrimental to most wildlife.

Rising seas
Thermal expansion in water, as warmer water in the ocean takes up more space compared to colder water, leads to oceans rising. In addition, glaciers in the polar regions are melting into the ocean and adding to the volume of the oceans, also leading to rising sea levels. Rising sea levels have many detrimental effects, the most alarming being the threats to dense coastal populations (such as Calcutta in India, with a population of over 14 million people), eroding shorelines and the destruction of entire ecosystems such as mangroves and wetlands. Coasts are sensitive to sea level rise, changes in the frequency and intensity of storms, increases in precipitation and warmer ocean temperatures. Moreover, increasing atmospheric concentrations of CO_2 are causing the oceans to absorb more of the gas and becoming more acidic. This rising acidity could have a significant impact on coastal and marine ecosystems. These ecosystems will have no warning and no way of adapting to the rise in sea levels and might suffer serious consequences.

Increased risk of drought, fire, and floods
Areas affected by drought are more prone to fire and even floods if heavy rain falls in eroded areas. These natural events cause devastating disruptions in agriculture, water supply, human health, and animal and plant ecosystems. Initially, warmer temperature might make many crops and vegetation grow faster, as plants generally prefer warmer climates. However, in the case of farming crops, the faster growth and warmer temperature may reduce yields, since the seeds will have a shorter time

to grow and mature. Increases in dead and dried crops also increase the risk of fire and decrease the oxygen produced by converting CO_2, because the ratio is disrupted. Where extreme droughts have not been experienced historically, but occur in response to climate change, this can eliminate entire crops and have detrimental effects on those dependent on agriculture.

Stronger storms and increased storm damage
Global warming could affect storm formation by decreasing the temperature difference between the poles and the equator. The temperature difference fuels mid-latitude storms, which affect the earth's most populated regions. Warmer temperatures could increase the amount of water vapor that enters the atmosphere. The result is a hotter, more humid environment. Warmer temperatures also increase the energy of the climatic system and may lead to heavier rainfall and intensified storms in some areas. It is generally expected that climate change will increase the frequency of heavy rainstorms, putting many communities at risk of floods. Higher ocean temperatures are the key contributor to increased weather activity, since hurricanes and tropical storms get their energy from warm water. As sea surface temperatures rise, developing storms will contain more energy. At the same time, factors such as rising sea levels, disappearing wetlands and increased coastal development threaten to intensify the damage caused by hurricanes and tropical storms.

Heat-related illness and disease
Climate change may increase the spread of infectious diseases, mainly because warmer temperatures allow disease-carrying insects, animals and microbes to survive in areas where they were historically unable to thrive because of cold weather. Diseases and pests that were limited to tropical conditions, such as mosquitoes that carry malaria, may find hospitable conditions in new areas that experienced increased temperature and humidity. Researchers have long debated whether malaria will move as a result of climate change. Studies are showing that the disease does migrate to higher altitudes when temperatures rise, which threatens the many millions of people who live in historically malaria-free regions of the high-altitude tropics. Adapted from the WHO, Table 9.1 lists some infectious diseases that could occur as a result of diverse environmental changes.

Table 9.1 Examples of possible infectious disease shift/migration attributable to possible environmental changes in response to climate change and global warming

Environmental change	Example of disease
Shifting dams, canals, and irrigation patterns	Schistosomiasis, malaria, helminthiasis and river blindness
Agricultural intensification	Malaria, Venezuelan hemorrhagic fever
Urbanization, urban overcrowding	Cholera, dengue fever, cutaneous leishmaniosis
Deforestation	Malaria, oropouche, visceral leishmaniosis
Reforestation	Lyme disease
Ocean warming	Red tide
Intensified precipitation	Rift Valley fever, hantavirus pulmonary syndrome

Changes in infectious disease transmission patterns are a likely major consequence of climate change. Understanding the environmental impacts of climate change and combining this with knowledge of preferred habitats for vectors of infectious diseases will enable accurate predictions of the spread of potentially hazardous diseases and early detection and preventative measures can be put in place to avoid large-scale epidemics. Institutions such as the WHO are actively planning and predicting such shifts in diseases and spreading the knowledge and required resources to fight the potential risks in developing countries. This is crucial to ensure sustainable habitats in rural areas and equally important, in urbanized areas.

9.8 Conclusion

Designing a sustainable city that is able to conduct comprehensive monitoring of the requirements of a population in an ever-growing infrastructure can be a challenging task and requires proper planning and skilled laborers. Sustainability can only be achieved if technology operates seamlessly and communicates information about the environment without interruptions, giving users the ability to act upon any situation that requires it. Population growth and urbanization have increased in the last decade to rates that are difficult to sustain and manage. Technology is the preferred catalyst in the sustainability and management of population growth and the goal is to create cities where communication between buildings, roads, transport systems, and other monitoring devices such as pollution is free-flowing and always accessible. Developing countries are in some ways struggling to cope with the increases in urban populations and are more than willing to adopt technology to aid them in managing their citizens. GIS, remote sensing and ubiquitous computing have reached levels that make them attainable to a larger audience, although initial teething problems are still evident in their early adoption.

This chapter highlights the causes, effects, and potential solutions when adapting to sustainable technologies in growing populations. The theme is that knowledge is power, and that providing people with sufficient background information about related issues encourages innovation towards new and exciting applications of pervasive systems in cities and in rural areas where applicable. It serves as an introduction to the following chapter, which presents case studies of circumstances where the integration of technology has assisted cities to become increasingly aware of their surroundings by using technology infrastructure as their baseline. Understanding the difficulties that overpopulation and overcrowding create, especially pollution, gives developed and developing countries the advantage of knowing how to address pressing issues in their communities. Prioritizing the requirements and needs of the community is an undeniably important step towards sustainability, as government and private investors must realize the potential of the technology for them to trust these endeavors and develop confidence in them.

References

Al-Ali, A. R., Zualkernan, I., Aloul, F. (2010). A Mobile GPRS-Sensors Array for Air Pollution Monitoring. *IEEE Sensors Journal*. 10(10), 1666–1671, 2010.

Bucchin, M. T., Barnett, B., Haake, D. G. (2015). Smart Cities and Sustainability Initiative. American Planning Association (APA). Report, April 2015.

Davis, F. W., Simonett, D. S. (1991). GIS and Remote Sensing. *Geographical Information Systems: Principles and Applications*. 14, 191–213, 1991.

Hulin, M., Caillaud, D., Annesi-Maesano, I. (2010). Indoor Air Pollution and Childhood Asthma: Variations Between Urban and Rural Areas. *Indoor Air*, 20(6), 512–514, December 2010.

Lynn, R., Harvey, J. (2008). The Decline of the World's IQ. *Intelligence*, 36, 112–120, 2008.

Mills, B., Heidel, K., Chung, C-F. (2006). Alternative Earth-Based and Space-Based Techniques for Mitigating Global Climate Change: What Can We Learn by Examining Them? *2006 IEEE EIC Climate Change Technology*. 1–10, 2006.

Shatz, S. M. (2008). State IQ and Fertility Rates in the United States. *Mankind Quarterly*, 49(1), 2008.

Chapter 10
Case Studies: Developing Countries Committed to an Ecologically Aware Future

10.1 Introduction

The definition of a smart city cannot be structured in such a way that the concept can be measured on a scale or provide a definite yes or no answer to the question: Is this a smart city? The implementation and perception of smart cities vary with the technologies used and applications they aim to achieve, as well as the development status of a country and the amount of money spent on achieving such status. For these reasons, an evidence-based metric is required to quantify a city's performance with respect to being smart or not, or determining its progress. To define such a metric, three broad categories can be regarded as contributing factors. These categories are, in no particular order, social performance and the quality of life of the city's people, environmental factors and overall pollution generation and efforts to reduce it and economic performance due to enhanced business practices, predominantly through savings in energy consumption and the effective use of new technologies to streamline processes and business practices (Arcadis 2015). These categories are depicted in Fig. 10.1 and used as reference throughout the chapter.

To elaborate on these three categories, examples of typical applications in each category (profit, people and planet) are given in Fig. 10.2.

The lists given in Fig. 10.2 present a broad guideline on matters to be considered in smart city development and can and should be expanded during the implementation phases of new technology to improve processes in a city, such as the distribution of information or the limiting of pollution-generating industries. Building on these parameters, Saha and Paterson (2008) in addition provided guidelines for such a process by dividing the process into distinct categories. There are typically four aspects of sustainable urban development that could be used as a general guideline to plan and implement sustainability in cities, as well as in suburban and rural areas where applicable. These four categories, as identified by Saha and Paterson (2008), are defined and expanded on in the following section.

J. Lambrechts and S. Sinha, *Microsensing Networks for Sustainable Cities*,
Smart Sensors, Measurement and Instrumentation 18,
DOI 10.1007/978-3-319-28358-6_10

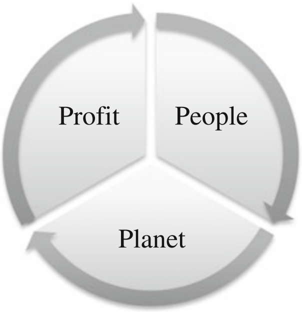


Fig. 10.1 Evidence-based metric to quantify a city's performance in becoming ecologically aware (Arcadis 2015)

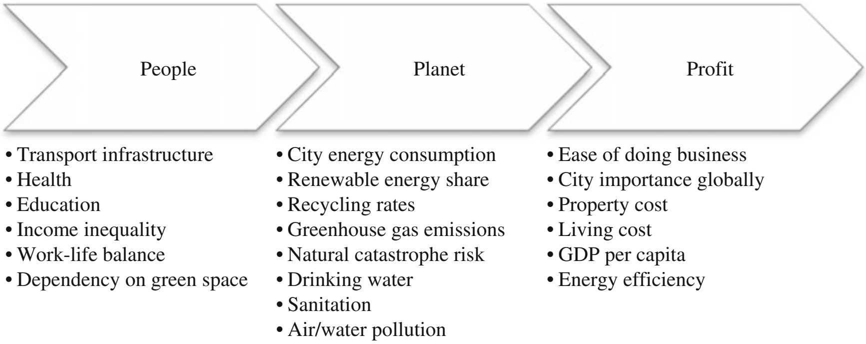


Fig. 10.2 Typical applications of the three evidence-based metrics when quantifying a city's performance (Arcadis 2015)

10.2 Categories of Sustainable Development

These four categories identified by Saha and Paterson (2008) as mentioned previously are energy, economy, social and governance sustainability. Energy focuses on environmentally sustainable urbanization and the measures that can be taken to implement its sustainability through energy savings and energy-efficient measures. These measures can be divided into energy efficiency, pollution prevention, and economic and social sustainability ones. Sub-sections of each measure are also provided and outlined as:

Energy efficiency: Taking into account several factors, such as the availability of alternative energy, energy conservation, smart and green building practices and renewable energy. These practices in a developing or developed city can add to its sustainability profile and provide valuable insight into best practices to lower pollution and the cost of maintenance.

- Alternative energy offered to consumers at reasonable prices to ensure competitive alternatives for users that choose to implement one or more energy options at home, businesses or industrial facilities.
- Energy conservation efforts in all forms of energy generation to reduce the amounts of energy wasted on a daily basis. Smart metering is one way of keeping track of energy usage in residential and industrial parks and provides users with real-time statistics on their energy footprint. This could potentially lead to users making active decisions to reduce their footprint, as they are aware of the cost savings associated with less wasted energy.
- Environmental site design regulations to encourage energy-efficient and sustainable practices for various disciplines, especially in the construction business. In general, it is easier to design and plan a building or any other structure with smart capabilities and sustainable technologies/materials compared to implementing such practices in already existing structures.
- Green building or sustainable building programs using environmentally responsible and resource-efficient design, construction, operation, maintenance, renovation and even demolition that add to a reduction in a city's energy footprint. Fast-growing megacities often have many construction projects running at any given time and the amount of resources, especially energy, in these projects is significant in the city's total energy usage.
- The use and promotion of renewable energy by city government to lessen the burden on coal-based non-renewable energy generation. Renewable energy sources have received substantial attention and interest in recent years. Techniques for improving their efficiency and the initial cost of implementation are major driving forces in their adoptability in cities and suburban neighborhoods. Developing countries are struggling to adopt renewable energy at rates that are required to reduce worldwide pollution, but more and more countries are subsidizing these practices.

Energy efficiency is mainly driven by climate change and the price of non-renewable energy to consumers, regulated by the supply of and demand for coal. Climate change might seem the buzzword of recent years, but its effects have been obvious and have been studied for almost 200 years. French scientist Jean-Baptiste Joseph Fourier (primarily known for his investigations of the Fourier-transform and its application to heat-transform and vibrations) calculated that the temperature of the earth, as a function of its distance from the sun, the only radiation source, should be colder than it was. This phenomenon led him to believe that the earth is generating its own heat and pointed to the first usage of the term "greenhouse effect" in 1827. From that point onwards, many scientists and researchers investigated the greenhouse effect and with improvements in technology its effects can be tracked more accurately. A summarized timeline of how the importance of climate change to society has changed through the years, compiled by Local Governments for Sustainability, founded in 1990 as the International Council for Local Environmental Initiatives (ICLEI ICLEI 2008), is presented in Fig. 10.3.

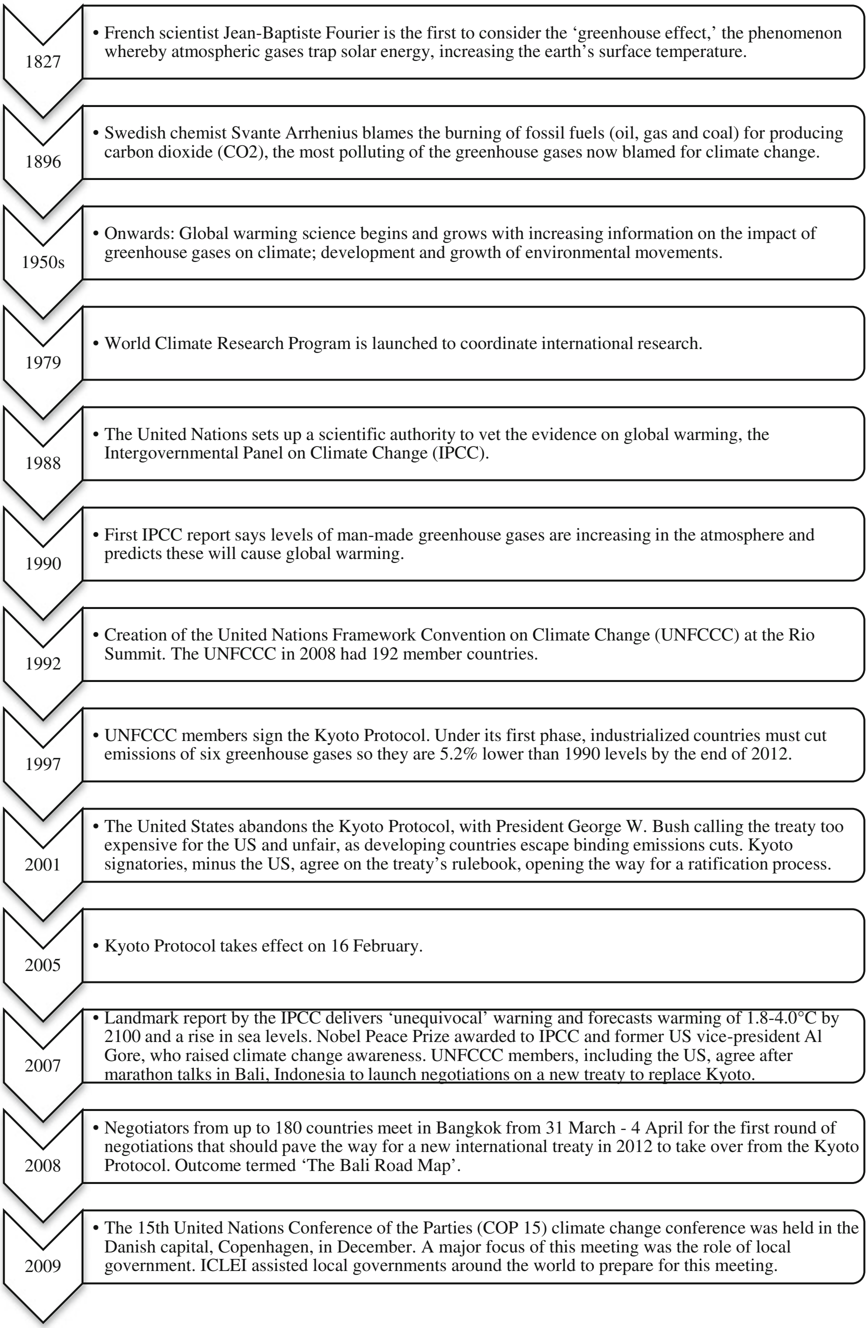


Fig. 10.3 Timeline of climate change since perceived in 1827 (ICLEI 2009)

Pollution prevention and reduction: This category provides guidelines for city planning in terms of reducing not only air and water pollution but also physical pollution by waste products such as plastics. Landfills are increasing in size in many cities and recycling programs aim to reduce the land designated for landfills, which could potentially become unpractically large. Considerations in this category include curbside recycling, environmental education and the purchase of environmentally friendly products and water-quality protection programs.

- Curbside recycling programs in urban and suburban areas remove household waste using appropriate containers and methods, outlined by local governments and municipalities. Recycling is crucial to reduce the footprint of waste products, which inevitably end up in landfills around cities. These landfills are being shifted closer to inhabited areas as cities expand beyond their historical borders.
- Environmental education programs for the community to inform them of the advantages of waste management and the disadvantages of waste products that cannot be recycled.
- Purchasing of environmentally friendly products and ensuring ease of access and distribution of these products by local suppliers to the community to encourage more green procurement by individuals and organizations. People often tend to purchase products that suit their needs based on availability at local markets and suppliers and if environmentally friendly products are offered at the same outlets at competitive prices, their adoption would be significantly faster.
- Water-quality protection programs to maintain the quality of water sources used to supply local communities with drinking water. Water as a commodity must be protected not only by municipalities but also by individuals as they realize the importance of maintaining safe and clean drinking water.

Encouraging recycling and the purchasing of environmentally friendly products significantly reduces energy usage, air pollution from incineration, water pollution from landfilling and the consumption of non-renewable raw materials. These practices should be prioritized in sustainable city development. Technology can again aid the monitoring and statistical analysis of the pollution sources through WSNs and RFID systems and provide citizens with information about the products they use and the hazardous footprint these could have.

Also of considerable concern in the pollution prevention and production planning category are issues such as open space planning, transportation and planning to protect the environment by decreasing pollution.

Open space and natural resource protection are factors that have a significant impact on the morale of the population. Measures to erect new public and open spaces and maintain and protect existing spaces are required to ensure city sustainability through caring for citizens. Two guidelines that may be proposed during planning are given below:

- Implementing environmentally sensitive area protection of green spaces and conserving these areas either through government interventions or through educating the local communities on maintaining them can reduce the cost and

effort of a single body to protect the area and distribute the responsibility throughout the local community.
- Initiating open space preservation programs that demarcate areas as types of heritage sites that may not be harmed or transformed ensures that the areas are kept as they are, with sufficient legal backing against proposed changes, such as urbanizing these areas.

Transportation planning can also contribute to prevention and reduction of pollution, especially air pollution, through various initiatives such as:

- Optimizing the operation of inner-city public transit (buses and/or trains) and distributing the types of transportation based on physical routes, energy efficiency, and the spending capabilities of the community (i.e. these should be as cost-effective as possible).
- Transportation demand management, which allows citizens to choose alternative methods or transport based on real-time demand. Technology plays an important role in transportation demand management and scheduling. Mobile applications for Android® and iPhone® devices have evolved towards complex scheduling and management assistance in many developed countries and are being adopted in developing countries as well.

Tracking progress on protecting the environment through awareness of energy usage and an ecological footprint, supported by technology advancements, can reduce pollution due to exuberant energy usage.

- Ecological footprint analysis in cities and even in rural areas should be available to the entire community to monitor and see the status in real-time. This will encourage users to decrease their current consumption or make wiser choices in terms of energy and transport use. Energy harvesting for small appliances can moreover contribute to overall energy savings and more efficient use of naturally available energy sources.

Economically sustainable urbanization: Ensuring that urbanization is economically sustainable will ensure viable future urbanization developments and the necessary funding and/or backing from local governments. Measures that support successful economic backing include:

- Smart growth measures during urban planning and transportation to concentrate growth in compact walkable urban centers to avoid sprawl, promoting compact, transit-oriented, walkable, bicycle-friendly land use.
- Agricultural protection zoning, especially near urban areas, to separate farming and related activities from other land uses, such as residential and urban dwellings.
- Reclamation programs for previously contaminated land areas (brownfields) when industrial or commercial processes in these areas cease.
- Cluster or targeted economic development, which addresses predefined areas of urban or rural development based on current requirements and issues.

- Sharing of resources such as information, water, energy, infrastructure and natural resources between businesses and communities to create sustainable eco-industrial park development.
- Land recycling or infill development to repurpose land areas in urban developments for economic use of existing infrastructure.
- Purchase of development rights and/or transfer of development rights to enable owners of land to make economically aware changes to existing infrastructure within bounds set by the state or local law authorities.
- Tax incentives for environmentally friendly development, especially to encourage the uptake of renewable energy sources.
- An urban growth boundary and/or urban service boundary to control urban sprawl by mandating that the area inside the boundary be used for higher density urban development and the area outside be used for lower density development.

A key to economic sustainability is future-proofing new and novel developments and investments by allowing citizens to contribute to sustainability while making a decent living in the process. People are intrinsically less willing to undertake long-term commitments if there are few or no incentive advantages, whether financial or otherwise. Therefore, it is important to consider promoting local employment and industries that give citizens a chance to earn money, recognition, or any other incentives. This can be achieved by addressing the following measures:

Promoting local employment and industries through:

- business retention and expansion programs to create and keep new jobs,
- empowerment and enterprise zones from the local community, and
- local business incubator programs to encourage skilled and unskilled workers to apply their skills in sustainable practices.

The third category is *socially sustainable urbanization* and it plays an important role in ensuring citizens have abundant choices and alternatives to engage socially in the community. People have different interpretations and expectations about community involvement and not all individuals prefer interaction in large groups. The thematic focus of involvement groups also plays a considerable role. Measures to provide various alternatives for social interaction and support systems for family members and neighboring individuals include:

- Affordable housing and provision of accommodation that considers all classes of people in any given urban or suburban environment. People should be given alternatives in terms of housing based on their income and not be excluded from urban living because of overly expensive houses and other types of living quarters.
- Daycare services that provide safe and stimulating environments for family members, typically young children, should be readily available to service sector and low-income employees to maintain peace of mind for working parents that their children are in safe and stimulating environments.

- Prevention of homelessness and intervention programs through tax-based initiatives that preemptively account for such circumstances, with additional focus on eradicating fraud and corruption in the administration of these issues.
- Compelling property owners and developers to provide public benefits such as affordable housing, preservation of historically important buildings and areas and green spaces through inclusionary and incentive-related zoning.
- Responsible distribution of employment opportunities and the workforce population in a given geographic area to optimize the jobs-housing balance, especially in fast-growing urbanized areas.
- Agreeing and committing to the provision of a decent living wage in order not to take advantage of laborers who are willing to work for less than the minimum wage, thus ensuring fair distribution of wages.
- Access to mass transit with local income subsidies for large groups of people working in close proximity and sharing travel routes.
- Neighborhood planning for effective future sustainability in various disciplines, including transport, energy, food and water distribution.
- Sustainable food systems and food security programs for low-income families to give everyone access to healthy and fresh food supplies.
- Increased interest in women- or minority-oriented business community development programs to encourage larger workforces in urban areas that generally require less physically challenging work compared to, for example, farming or other rural jobs.
- Opportunities for youths and anti-gang programs, again to promote opportunities to create a larger and more sustainable workforce.

Finally, a well-defined and properly managed governance and institutionally sustainable urbanization procedure should be in place to address any issues that will inevitably be raised when large groups of people from different backgrounds and ethnicities live together, to improve sustainability within a community. Governance of these groups and events is required and a few measures to implement it, together with proactive planning for possible risks, include:

- Social dispute resolution to avoid large-scale civil clashes.
- Public participation to promote innovative and entrepreneurial activities.
- Regional co-ordination in communities to build social cohesion.

These categories give relatively straightforward guidelines for sustainable development planning that are still open to interpretation. Equally important as guidelines are indicators that provide a benchmark for measuring the success of implementation. Such indicators (also serving as general guidelines and variables based on each city's applications and needs) are presented in the following section.

10.3 Indicators of Sustainable Cities

Monitoring and evaluating the success of sustainable cities are as important as working towards a common goal and therefore require a set of indicators, although generally defined in relatively broad terms, to provide guidelines to quantify the success (or failure) of implementation. These indicators, as presented by CIDA (2012), include categories for the economy, the environment and social elements, and can be used to describe achievements. Each of these three categories can be subdivided and used by developed and developing countries as a general baseline to verify the status of their future goals and incentives. These sub-categories, adapted from CIDA (2012), are briefly outlined as:

Economy: The economy category includes factors such as employment statistics, employment opportunities and the monetary value of these opportunities, economic growth and foreign investments. To summarize these points, consider the following highlights:

Unemployment rates, number of jobs and the ability to follow a career in sustainable development that offers competitive remuneration. In this sub-category, the following statistics reflect the status of success in a country or city:

- Underemployment/employment/unemployment rates provided annually, with a common goal to reduce unemployment and create new and sustainable jobs.
- Percentage of green jobs in the local economy that can sustain its labor force through competitive remuneration packages.
- Average duration of the labor force's professional education, which determines the skills versus education of the labor force and should ideally be balanced to provide each person with lucrative alternatives.

The economic growth of any country is important for various reasons. Essentially, it provides new opportunities for citizens, either in existing markets, or in new and innovative markets that allow them to compete globally. Technology plays a large role in developing sustainable practices in communities, especially in the information age where microsensing networks are becoming more prevalent in distributing information to inform and educate citizens about their environment. Parameters that are considered during transformation and growth of a country are:

- The growth in production and gross values added by its citizens, thus the annual GDP growth rate.
- The market value of all products and services produced by the labor force of a country, therefore the annual GNP growth rate.
- Net export growth rates (the percentage increase of a country's total exports minus the value of its total imports per annum).
- Foreign direct investments, growth in controlling ownership in a business enterprise based primarily in other countries.

Environment: In the environmental category, free and abundant access to green spaces and fast and easily accessible transport contribute largely to people's morale and cooperation. Information distribution through ubiquitous systems is also becoming integral in current and future sustainability. People are inclined to contribute more to a goal if their own frame of mind can be positively influenced as a result of their own conscious decisions and those of others sharing similar goals and interests. To aid in quantifying the success of sustainability and the morale of citizens, the environmental category is subdivided to include green spaces, low pollution, easy access to transport, good quality drinking water, air quality, technological advancements to increase the quality of life and recycling initiatives to reduce not only pollution but also improve aesthetics in a city. The following is relevant in this regard:

Green spaces provide citizens with fresh air and natural surrounding with a generally calming effect and are required in any sustainable city to reduce greenhouse gases. They give citizens the chance to get together in a natural, relaxing and ideally untouched space. To measure this, reflect on the:

- Percentage of preserved areas/reservoirs/waterways/parks in relation to total land area occupied by buildings and other man-made structures. It is difficult to define a percentage to be adhered to strictly, since many external factors play a role in defining available green spaces. For example, London in the UK has approximately 47 % of the space in the city allocated as green space.
- Percentage of trees in the city in relation to the city area and/or population size. Trees contribute to good practice in modern cities and are especially effective when placed next to roads and walkways.

Green spaces in a city contribute to the mood and morale of its citizens directly, but an indirect contributor to human wellbeing is the reduction of greenhouse gases. Reducing greenhouse gases and increasing energy efficiency in a city play a major role in reducing toxins in the air and in water bodies. Measuring a city's carbon footprint is a relatively easy and cost-effective way of rating its current and future sustainability and indicates whether action should be taken to decrease its carbon generation. Indicators include:

- Total amount of greenhouse-gas emissions per city and per capita.
- Percentage of total energy consumed in the city that comes from renewable sources, ideally to reduce the dependency on non-renewable sources such as coal and move to cleaner energy generation.
- Responsible agricultural practices that do not promote gestation farming that releases large amounts of toxins into the ground by leading waste from animals to nearby water bodies.

Greenhouse gases in cities and rural areas have many (primarily man-made) sources. Mobility in a city can sometimes be underestimated as a major factor in decreased morale, inevitable tension between citizens and heightened pollution. Traffic congestion and long travel times contribute significantly to these issues and should be addressed in any sustainable city. Pollution is another prominent adverse effect of traffic congestion that affects citizens' health, which can potentially decrease the efficiency and size of the active labor force. City sustainability is measured, in terms of transport, by:

- Transportation mode alternatives (percentage of each mode of transportation, including private commuting by cars or motorcycles, public transport such as buses and trains, bicycles and safe and accessible walkways for pedestrians).
- The average travel time and cost influence a person's decision to follow a career or submit a job application; these aspects should be justifiable. In developing countries people tend to spend many hours commuting, often by walking and using public transport services. The cost of getting to their destination each day often outweighs the benefits of their salary or other incentives. In some countries, transport to a job alone can cost a person up to 30–40 % of his or her daily salary, making these circumstances unsustainable.
- Future sustainability through planning new and energy-efficient, renewable, mass-transport systems in a city can reduce travelling times for its citizens but more importantly, reduce air pollution caused by millions of combustion engine vehicles operated each day.

The sustainability of a city therefore depends on transport alternatives that offer its citizens various options to get to work daily, but also affect the pollution generated by internal combustion engine vehicles.

The water quality in a city is as crucial as the air quality. The quality and availability of water in a city are crucial for many reasons, the main one being that humans and animals require water to live. The integrity of an ecosystem is typically assessed through its ability to provide goods and services on a continuous basis. To ensure a growing economy and future sustainability, clean drinking water is essential in city planning and should be prioritized above all other factors. In many countries, even in European countries, tap water is not potable and must be boiled first. In most cases this does not present serious health concerns and many countries do in fact have clean and drinkable tap water, which should be standard throughout.

- Water quality is an expression used to portray the chemical, physical and biological characteristics of water, generally in terms of suitability for a particular or designated use (such as drinking or washing). Parameters of water quality that are typically monitored are fecal bacteria, temperature, dissolved oxygen, algae and turbidity. Various WSNs exist that can monitor these parameters in real-time and provide a sustainable city's inhabitants with information on the status of their drinking water.
- The total amount of water available for each person in a city or in rural areas. Billing citizens for water usage at reasonable prices ensures sustainable development without people having to forage for drinking water; they can instead focus on growing the economy in other areas of interest. Supplying water free of charge is counter-productive, as people will then inevitably waste water on non-essential activities, even though water is already a commodity that requires substantial treatment to prepare it for consumption.

- The chemical, physical, biological and radiological characteristics can be defined by a water quality index and municipalities should adhere to keeping these scores above the minimum required for drinking purposes.
- The proportion of the population with access to adequate and safe drinking water should ideally be close to 100 %, but this is not always attainable for various reasons. Surveys and audits in large cities and expanding urban areas can help to determine any deficits in drinking water availability and give a good indication to governments and municipalities of how to improve distribution to all citizens.

The air quality of the environment should ideally be carefully monitored, using advanced and accurate technologies in a sustainable and smart city. Quantifying air quality is a relatively easy process (in terms of PM quantities, CO levels, and nitrous oxides) and contributes to the smart capabilities of a city. The accuracy of the measurements may vary with the quality of the measurement equipment used; poorer countries might opt for lower-cost devices. Initial investment in the installation and distribution of these systems is a large contributor to the overall cost. Sensors and nodes can be replaced as the technologies become more attainable and less expensive, if sustainable and modular designs are incorporated from the beginning. Indoor and outdoor air quality monitoring are equally important, as people tend to spend large portions of their day indoors during working hours and at night, with large portions also spent outside for recreational purposes. Air quality considerations include:

- Both outdoor and indoor air could affect human health, attitudes and productivity. It is important to maintain the quality of outdoor air, since all life forms depend on it, and since the quality of indoor air is dependent on that of the outdoors.
- Levels of PM (PM10 and PM2.5 measured in mg m^{-3}) can be monitored using, for instance, WSN and other microsensing networks, where information and data on the current state of the air quality can be distributed to the local population (and globally) using the internet or local networks.
- The air quality of the environment contributes not only to clinically verifiable diseases such as cancer and Legionnaire's disease, but also to various syndromes such as fatigue, lethargy, dizziness, lack of concentration, respiratory tract irritation, headaches and eye irritations. All these contribute to a general decrease in human productivity and overall wellbeing.

In any sustainable city—whether planned during the initial phases or during upgrading of the current infrastructure—waste management, product reuse and effective recycling are critical to ensure a clean and waste-free environment. Waste monitoring and statistical analysis of its generation with recycling in mind can prove pivotal in creating a sustainable ecosystem. Considerations that affect procedures and initiatives to manage it include:

- The rate of recycling products versus the manufacturing of new products on the market.

- The net volume of solid waste generated in and around a city or in rural areas. The net volume of generated waste depends primarily on the size of the population and their usage habits. Adequate control over these processes should be implemented to reduce overall production of waste. Reducing landfills through recycling and responsible use of products that are harmful to the environment would improve the sustainable future of a city.

Social: The final category of indicators that quantify the successful integration of sustainable methods in a growing city addresses social cohesion and the quality of life of its citizens in measurable terms. These indicators include neighborhood statistics, housing facilities and availability, the quality and abundance of public spaces, local and international educational facilities, sanitation and the general and mental health of the population. These factors are briefly expanded on, with the first focus point being neighborhood services:

Neighborhood facilities and compactness are easy ways to quantify smaller areas within a city based on their performance and sustainability. Megacities with upwards of 10 million citizens can be difficult to monitor and it is generally preferred to divide the city into several smaller neighborhoods and assess each of these in terms of:

- Access to local/neighborhood services within a short distance. This generally requires that services (such as police stations, fire departments, pharmacies, hospitals and other critical services) be distributed according to the geographic limitations of a city. It must be easy for citizens to reach emergency services and local supply chains.
- Crime rates in neighborhoods tend to be dependent on the wealth of citizens, the security measures of local residences and the location of poorer neighborhoods with respect to the wealthier areas. Crime rate statistics can give valuable information about a population, especially with regard to employment rates and civil acceptance in communities.
- Measures of income distribution and inequality in smaller areas are more manageable opposed to city-wide indicators. Variable tax structures and services offered, and even lower prices of services and products in poorer neighborhoods, can provide a needed shift in the balance of income distribution throughout a city.

The housing in cities is another indication of social standing and distribution of wealth and potential for a sustainable future. In most countries, especially in developing countries, there are significant differences in the types and sizes (and inevitably the cost) of housing between the wealthy and the poor. These differences are easy to distinguish just by visually comparing neighborhoods and give a reasonable perspective of a country's development status. The most significant indicators are:

- Percentage of social/affordable/priority housing distributed in large cities. In developing countries where the employment rate is generally low, the availability of affordable housing is crucial to sustain citizens and provide them with

shelter while they seek better working conditions or work towards building long-term careers.
- Breakdown of the housing sector by property type (owner-occupied/rental, single occupant/couples/family/multifamily) is another indication used to determine the wellbeing of a city statistically and indicate where improvement in the quality of life is required or encouraged.

From a mental wellbeing perspective, high-quality and abundant public space indicate the efforts of local governments and authorities to improve morale and provide citizens with designated areas for relaxation and other recreational activities. Indications of such efforts are:

- The percentage of roadways that are kept in good condition and maintained thoroughly.
- The percentage of green space (public parks) coverage in relation to city area and/or population size.

Education is arguably one of the most important social factors when considering smart, sustainable, environmentally friendly and technologically advanced cities. It is crucial for the local population to understand the causes and effects of pollution and ways to improve their own health and quality of life. Following the industrial revolution, the technological revolution (also called the second industrial revolution or the information age) has proven itself a large influence for current and future generations. To take full advantage of this revolution, education is seen as a crucial building block to achieve sustainability for current and future generations. City planning and infrastructure development should prioritize the placement, availability and tuition cost for all citizens. Indicators that provide some idea of a country or city's developmental state in terms of its educational system are:

- The number of schools with environmental education programs that provide students with opportunities to take up studies related to sustaining and looking after the environment.
- The adult literacy rate, which is the largest factor or indicative parameter of a country or city's education. Improving the literacy rate of developing countries is the first step towards educating the population about environmental issues and eradicating bad practices such as gestational farming.

Sanitation can also be considered an indicator of social development in growing cities. At its core it represents human dignity and basic human rights and if addressed properly, it can speed up social development (UNWATER 2008). Most importantly, a city committed to a sustainable future is characterized by:

- The percentage of the population with access to water-borne or alternative sanitary sewage infrastructure. The relevant statistics are generally obtained through municipal records or surveys in communities to gather information about the status of the sanitary infrastructure.

- Sanitation that aids progress toward gender equality. Women and younger children in poor countries with no sanitation are exposed when having to attend to personal hygiene, often subjecting them to harassment and humiliation.
- Proper sanitation, which can promote social inclusion and therefore encourage social sustainability. Urban slum dwellers are often surrounded by human waste and garbage that marginalize these communities and deprive people of social interaction and participation.
- Improved education in rural and poor countries lacking proper sanitation at schools and other institutions. Parents of children attending schools in poor communities without adequate sanitation often do not allow the children to attend schools, especially during a child's puberty when changes in his or her body could be embarrassing for the child if there is no personal space and working sanitation away from home.
- Measures to address health concerns in schools and educational institutions. These issues are problematic for parents, since many undeveloped countries and even cities do not provide clean facilities for their students or employees to wash their hands or follow other personal hygiene routines, which often leads to illnesses and epidemics spread through communities, decreasing people's efficiency and having long-term adverse effects on education.

Finally, the health of the population of a city determines its active workforce, its potential workforce, and the mental and physical wellbeing of its citizens. Two of the most valuable indicators of a city's health profile are:

- Mortality rate and life expectancy.
- The percentage of the population with access to health care services.

Questions municipal planners can ask to assess their readiness for smart city solutions include:

- Are they facing multiple service-delivery problems that put a strain on existing systems?
- Is the municipality laying down roads and sewerage systems, and creating public transport systems that are interlinked with broadband capability?
- Are police, hospitals, emergency systems and clinics connected to provide citizens with the best service, or could they benefit from working together in an interconnected system?
- Do they find pockets of excellence in how they operate in problem areas that would benefit from integration?
- Is the infrastructure in place to help grow universities and businesses, and foster centers of entrepreneurship and new economic sources of growth?
- Do they have the vision and political will to create a smart city strategy?

An article on the UK website, Policy Exchange, in March 2015 highlighted some questions that policymakers were asking themselves, but deemed these incorrect questions to ask for several reasons. The views in this article are interesting and accurate up to a point, and some of the information is shared in this book. Firstly, the identified incorrect questions that should be avoided are listed as:

- How do we prove the business case for investing in smart city infrastructure?
- How do we ensure smart city technology will be future-proof?
- What are the technical components of a city-as-a-platform (a platform that supports and interconnects all the digital functionality in a city that serves operating requirements and engages its citizens)?
- What smart city solutions should cities buy and how will local authorities afford them?

Glancing at these questions generally asked by policymakers reveals some misconceptions. These questions are directly aimed at future-proofing large and expensive new developments and infrastructure to create a “super” smart city. This could indeed be a goal, but it would be a long-term goal and chances are it would not be realizable in the near future and also be costly. The financial aspect would involve not only lack of funding or interest if timelines are not met, but also keeping up to date with technology improvements in the current technological age. Planning for a sustainable and smart city could be the first mistake. Of course planning will always be required, but the level of planning and amount of resources required by these planning structures are not always feasible in developing countries, not even in some developed countries. The following answers expand on the “incorrect” questions posed above:

- Q: How do we prove the business case for investing in smart city infrastructure?
- A: This can be done at a lower scale. Engineering is fundamentally based on small prototypes and testing an idea or innovation in a safe environment before approving these prototypes for mass production. Similarly, any business case to encourage investment in a smart city should first be tested on a smaller scale, such as in certain pre-identified neighborhoods.
- Q: How do we ensure smart city technology will be future-proof?
- A: This depends on how future-proof is defined. Are the current prototype ideas modular in design, or “what-you-see-is-what-you-get” devices? Future-proof design is about modularity and adaptability for inevitable changes that will occur in topology and standards. It is the job of the trained and skilled professional to foresee these possible changes and develop a system that can easily be adapted if required. A future-proof design is difficult to ensure, but it generally involves a product that is essential to humans.
- Q: What are the technical components of a city-as-a-platform?
- A: Municipalities should follow a logical procedure to identify the core required interactions and define the data that powers these interactions. They should subsequently identify the software, tools and skills needed to create an all-encompassing platform and finally identify the data producers (such as police stations and medical facilities), as well as the data consumers (citizens) and ways of distributing and maintaining the data.

- Q: What smart city solutions should cities buy and how will local authorities afford them?
- A: Depending on the applications, equipment is expensive and needs to be maintained constantly. Developing countries could consider buying equipment that is older but still performs the required tasks, or buy second-hand equipment from developed nations that are looking to upgrade/replace their current systems.

As seen again from these questions and proposed answers, it is difficult to define answers that only relate to and are applicable to one country and for specific applications. Skilled workers, engineers, architects and people employed in environmentally conscious disciplines all need to work together in implementing sustainable technologies that suit the needs of a community and have the financial backing and support of local authorities and private investors, as well as the local community.

Based on several of these guidelines and indicators proposed in the previous sections, sustainability and the smart-ranking of cities are being measured by institutions such as Arcadis NV, established in 1946 and reporting globally on various topics such as climate change and rapid urbanization growth. The Sustainable Cities Index (SCI) is one such index that rates cities based on several factors classifying the top sustainable cities worldwide.

10.4 Sustainable City Rankings

The SCI explores the three demands of people, planet and profit (see Fig. 10.1) to develop an indicative ranking of 50 of the world's leading cities. The top 50 sustainable cities as indicated by Arcadis (2015) are listed in Table 10.1.

An extract from the Arcadis (2015) sustainable city index report is highlighted: "The Sustainable Cities Index not only benchmarks individual places today but offers a roadmap for future improvements outlining specific areas for attention. Most importantly, this index offers a fundamental truth: cities have unique qualities based on their histories, geographic contexts, and level of development. In highlighting a city's character—its strengths and weaknesses—The Sustainable Cities Index provides a platform for public and private decision-makers. Fundamentally, however, it provides guidance and allows thoughtful decision makers to "Measure what can be measured and make measurable what cannot be measured," as per Galileo's sage advice."

Frankfurt (voted the most sustainable city in 2015 by Arcadis (2015), as seen in Table 10.1) is widely recognized as a major international financial and trade center, as well as transportation hub. Frankfurt's high ranking in the SCI comes from a long track record of proactively taking action to improve its sustainability.

Table 10.1 Arcadis (2015) top 50 sustainable cities using the SCI criteria

Ranking/City			
1	Frankfurt	26	Kuala Lumpur
2	London	27	San Francisco
3	Copenhagen	28	Los Angeles
4	Amsterdam	29	Dallas
5	Rotterdam	30	Santiago
6	Berlin	31	Sao Paulo
7	Seoul	32	Mexico City
8	Hong Kong	33	Dubai
9	Madrid	34	Abu Dhabi
10	Singapore	35	Shanghai
11	Sydney	36	Istanbul
12	Toronto	37	Johannesburg
13	Brussels	38	Buenos Aires
14	Manchester	39	Beijing
15	Boston	40	Rio de Janeiro
16	Paris	41	Doha
17	Melbourne	42	Moscow
18	Birmingham	43	Jeddah
19	Chicago	44	Riyadh
20	New York	45	Jakarta
21	Houston	46	Manila
22	Philadelphia	47	Mumbai
23	Tokyo	48	Wuhan
24	Rome	49	New Delhi
25	Washington	50	Nairobi

In 1990 the city created its own energy agency and it is a founding member of the Climate Alliance of European Cities, pledging to reduce its CO_2 emissions continuously by 10 % every five years, resulting in a 50 % cut by 2030. Since 1990, Frankfurt has already decreased its CO_2 emissions per capita by 15 % while increasing its economic power by 50 % and office space by 80 %. Frankfurt's new master plan, "100 % Climate Protection", goes even further. By 2050, 100 % of Frankfurt's energy will originate from renewable (and mainly local) sources, causing a 95 % decrease in greenhouse gas emissions. Frankfurt plans to achieve this by increasing energy efficiency and decreasing the demand side by half in residential and office buildings, the transportation sector and in communications. Frankfurt was one of three finalists for the European Green Capital award in 2014. Furthermore, the city has been recognized as the European City of Trees 2014—not only is every tree registered and monitored; the information is also publicly available online. Frankfurters can also enjoy Germany's largest city forest of more than 8000 hectares, or one third of the city. The adjacent green belt that runs around the city from the Main river bank, is not only a close recreational area but also the reason why Frankfurt is such a compact city of short distances. The pleasant nature and compactness of the city explains why 15 % of all commuting is already done by bicycle.

On the bottom end of the rankings, in a commendable 50th place (13 places behind Johannesburg in South Africa), Nairobi has been named the most intelligent city in Africa in 2014 and 2015. It is also one of the only two African cities (the other being Johannesburg) that appear on this list. Nicknamed the "Silicon Savannah", Kenya has ushered in a new era of ICT in the last decade (IBM 2012). Its capital, Nairobi, is considered a pivotal urban center of the ICT revolution in East Africa. However, the city faces critical transportation and logistics challenges as it continues to grow and expand. According to the government of Kenya, the population is set to quadruple from 3.1 million in 2014 to 12.1 million by 2030. Traffic-related congestion, high road fatalities and an unregulated mass transit network not only pose distinct challenges to equitable growth, but also opportunities, as technology-for-transport startups are offering an abundance of digital mobility services. Media criticism focuses on Nairobi's high crime rates and question how it can be named a smart city considering this, but it is argued that the high crime rate in Nairobi is the source of creative and innovative solutions that try to combat it. Similar circumstances in Johannesburg are witnessed, where many security-related startups have been appearing in recent years. In Kenya, some of the innovative projects include the use of mobile money to pay fees by the Nairobi county government, thus decreasing payment inefficiencies and introducing a layer of transparency over the process. Also, several innovation and incubation centers have been integrated with the city's urban culture. These include iHub's inception in 2010, serving as a middleman between investors and entrepreneurs, 88 mph, which has invested in 36 companies in Kenya and South Africa during 2011–2014 and university innovation centers, which have become the breeding ground for new business ideas.

Many other notable smart city initiatives have been identified worldwide where developing and developed countries are striving to meet the demands of smart and sustainable technologies to increase the quality of life of their citizens. The following section highlights a few of these endeavors based on location. Developing and developed cities are discussed to highlight the difference in the approaches adopted by these cities. These differences are generally attributed to the available funding, but more importantly to the crucial requirements and needs of the communities and the problems they face on a daily basis. Crime and education in developing countries are receiving the primary focus in building sustainability, as these are issues that must first be overcome before focusing on non-essential upgrades and innovations.

10.5 Worldwide Smart City Initiatives

10.5.1 South Africa—Human Capital in ICT, Information Distribution and Transport

Johannesburg (the provincial capital and largest city in South Africa) recently unveiled its approach, a program referred to as COJEDI, to educate "digital interns" and prepare them for careers in ICT. The program is a big part of the early phase of its smart city plan and is expected to train 1000 students in the IT industry to ensure the city has the talent it needs (IT News Africa 2015). The COJEDI program focuses on driving innovation and entrepreneurship through critical thinking, preparing students for employment and empowering them to participate significantly in the mainstream of the ICT sector and the growing economy. As part of the program, the city is collaborating with Cisco, Microsoft, FiberCo, the Technology Innovation Agency and Nunnovation Africa Foundation.

City officials recognize the need for professionals with the right technological capabilities to help make the smart city plan a reality. There also is a more immediate need: the city's communities now have extensive broadband connectivity but there are too few trained professionals to manage and maintain it efficiently. Students selected for the program will have the opportunity to see smart city connectivity in action in other countries.

The city has already installed a number of Wi-Fi hotspots to improve internet access and plans to install 1000 more in the forthcoming years. Many residents are already separating their waste, part of a plan to cut the amount sent to landfills by 20 % by next year.

Another key smart city initiative involves overhauling its public transportation. So far, the city has converted two buses to run on compressed natural gas and diesel, and will convert another 30. The converted buses have 90 % lower carbon emissions, but also cost less to operate, since the fuel comes from local waste and crops.

10.5.2 South Africa—Local Municipal Energy Conservation Efforts

Sustainable Energy for Tshwane (SET) is an interdepartmental committee coordinating energy and climate change initiatives in the Tshwane municipality in South Africa (ICLEI 2008). It is led by the Social Development Department's Environmental Health Unit. Other participating units include Energy and Electricity (newly named to include all energy sources), Transport, Environmental Resource Management, Housing, Local Economic Development, Integrated Development Planning, Spatial Planning, Waste Management and Water and Sanitation.

The establishment of this committee was made possible through top-level direction from both councilors and senior managers. Through support from Sustainable Energy Africa, a local non-governmental organization (NGO), the municipality became part of the Sustainable Energy for Environment and Development program, which focuses on building capacity in municipalities concerning energy issues.

Initially the committee experienced some challenges. The Social Development Department was frustrated by lack of commitment in the committee, some departments did not incorporate SET into their own lines of decision making and the relevance of the energy issues was not always immediately apparent to departments such as the Housing Department.

10.5.3 Italy—Training Programs in Theoretical Physics Aimed at Creating Environmentally Aware and Therefore Sustainable Communities in Developing Countries

Another initiative to promote sustainable development in developing countries, specifically on the African continent, is undertaken by the Abdus Salam International Centre for Theoretical Physics (ICTP) in Trieste, Italy. The ICTP has a long tradition of scientific capacity building in Africa. Over the last few decades, ICTP has supported numerous activities throughout the continent, including training programs, networking and the establishment of affiliate centers. The center's Trieste campus has welcomed more than 15,000 visits from African scientists since 1970, providing advanced research and training opportunities unavailable to scientists in their home countries.

Since ICTP's inception, one of its main goals has been the development of scientific and technological capacity in Africa. ICTP carries out this task in various ways:

- In 2014, ICTP had 728 visitors from Africa (26 % women) for an average stay of 50 days each.
- ICTP founded and is sponsoring six affiliated centers, two projects and two networks, as well as the ANSOLE network and a doctoral program in African universities.
- ICTP organizes and sponsors more than 10 scientific events in Africa every year.
- In 2013, 21 least-developed countries (according to the United Nations) of Sub-Saharan Africa were represented at ICTP, including Ethiopia, Senegal and Sudan.

In September 2015 the ICTP opened applications for three post-doctoral positions in earthquake studies available in its Earth System Physics section. Two of these positions were funded by the GENERALI Group, a major participant in the global insurance industry. The third position was funded by a framework of an international collaborative effort following the Nepal Gorkha earthquake in April 2015, which claimed more than 9 000 lives and injured more than 23 000 people.

10.5.4 North Africa—Consultation Towards Smart Initiatives

The Smart Cities Initiative for North Africa (SCI-NA) is an NGO whose mission is to help public and private institutions to implement innovative, citizen-centric smart initiatives for sustainable economic and social development. SCI-NA aims to develop smart initiatives through expertise, innovation, consultation and communication. SCI-NA also organizes the International Summit for Smart Cities in North Africa, which started in 2014 and is an annual gathering of influential policy makers, leading academics and sector experts aiming to find innovative and actionable solutions for urban living in North Africa. According to SCI-NA, Morocco leads the way in working towards numerous major reforms in urban planning, clean energy, water security, agricultural development and transportation.

10.5.5 China—Successful Smart Integration by Provincial Governments

Rizhao, which means City of Sunshine, is a city of 3 million people in northern China (ICLEI 2008). It has over a 500,000 m^2 of solar water heating panels; 99 % of households in the central districts use solar water heaters and more than 30 % do so in the outlying villages. Moreover, almost all traffic signals, streetlights and park illuminations are powered by photovoltaic solar cells, 6000 households have solar cooking facilities and more than 60,000 greenhouses are heated by solar panels, reducing overhead costs for farmers in nearby areas. At his appointment in 2001, Mayor Li Zhaoqian recognized that Rizhao, with a lower per capita income than most of its neighboring cities, would have to focus on increasing the efficiency and lowering the cost of solar water heaters. The Shandong Provincial Government provided subsidies for this. Instead of funding the end users, as in most industrial countries, the government funded the research and development activities of the solar water heater industry. Solar devices now cost the same as electric alternatives, while saving the users energy costs. Mayor Li Zhaoqian and the Rizhao municipal government have adopted several measures and policies aimed at popularizing clean energy technology, including the Regulations on Implementing Solar Energy and Construction Integration that standardize the use of solar energy (particularly solar water heaters) in new buildings.

The achievement of solar-powered Rizhao was the result of an unusual convergence of three key factors: a government policy that encourages solar energy use and financially supports research and development, local solar panel industries that seized the opportunity and improved their products and the strong political will of the city's leadership.

10.5.6 United Kingdom—Strategy to Lower Carbon Emissions Due to Energy Usage

The London Energy Partnership (LEP) was established as an independent body to provide coordination and synergy between the many groups, organizations and networks working on energy issues in London. It provides a vehicle for the delivery of the city's energy policy. Until the formation of the partnership, London lacked an adequate mechanism to enable broad collaboration required to deal with these crosscutting issues. Through a consensual process with energy stakeholders, the LEP guided the development of the city's 2004 energy strategy. In 2007, the city developed a climate change action plan to strengthen action in this area. The plan aims to achieve a 60 % cut in CO_2 emissions by 2025. The LEP is responsible for implementing and reviewing the city's energy strategy and action plan.

10.5.7 Brazil—Focus on Water and Sanitation Distribution

The Companhia de Água e Esgoto do Ceara (CAGECE) in the northeast of Brazil, in partnership with Alliance to Save Energy, aimed to improve the distribution of water and access to sanitation services, while reducing operational costs and environmental impacts. Over four years, CAGECE saved 88 GWh of energy, improving efficiency each year. Before CAGECE instituted its energy-efficiency program, it provided access to 442,400 households. Four years later, the utility provided 88,000 new connections over the original baseline, while decreasing total energy consumption and costs and maintaining water consumption levels. Four years of official data show savings of over US $2.5 million, with an initial investment by CAGECE of only US $1.1 million. As a result of this 127 % return on investment after four years, CAGECE was initially approved for financing by the energy-efficiency fund of the governmental Brazil Fight against Electricity Waste Program to work with the World Bank to implement further efficiency measures. The alliance helped develop five projects, including replacing motors with high-performance motors, maximizing pumping efficiency, suspending pumping during peak hours, increasing the capacity of the current pumping stations and compiling specifications for energy efficiency. If implemented, these projects would add a saving of 7 million kWh per year, with a total investment of US $2 million by the PROCEL and the World Bank. The cost/benefit analysis predicts a payback period of 3.5 years. However, the financing opportunity was lost because funds were obligated to pass through the state energy utility in Ceará (COELCE) and the legal departments of COELCE and CAGECE could not come to an agreement. Further intervention included automation of operations, rewinding and replacement of motors, maximizing existing pump systems' efficiency and increasing storage capacity to allow the shutdown of pumps during peak hours. An operations procedure manual was created to serve as a reference for daily performance to operations crews and CAGECE management. CAGECE established an operational control center for the water supply system of metropolitan Fortaleza. The objectives of the automation of the water supply system of Fortaleza were to optimize operations to reduce energy costs, improve system management by centralizing control, speed up recognition of and response times to maintenance needs using sensors, acting through controlling devices and generating system diagnostics using historical records of operational data.

10.5.8 Joint European Initiative—Financial Research Conducted in Europe to Support Smart and Sustainable Cities

The Joint European Support for Sustainable Investment in City Areas (JESSICA) is a policy initiative of the European Commission developed jointly with the European Investment Bank and in collaboration with the Council of Europe Development Bank. JESSICA aims to support sustainable urban development and regeneration through financial instruments, combining European Structural Fund resources with other public and private sources of finance to create revolving investment funds such as Urban Development Funds to invest in public-private partnerships and other projects included in integrated Plans for Sustainable Urban Development.

This study seeks to understand the potential for JESSICA financial instruments to support smarter and more sustainable cities to help reach EU 2020 objectives. Commencing in September 2011, the study included desktop research and stakeholder interviews to understand more about how smart city initiatives arise and evolve, and how projects are procured and financed, to understand the potential role JESSICA could play better. Four cities were studied in depth as case studies: Malmö, Barcelona, Manchester and Amsterdam. The study also included a workshop in December 2011 with industry and policy makers, and a European conference to explore the business models for smart and sustainable city projects in March 2012.

10.5.9 Sweden—Best Practice in Mixed Use of Environmental Adaption in Urban Areas

The Western Harbor project in Malmö (Hirst et al. 2012) is an ecological development in the south of the city in a former dockyard area, which has been evolving over the last two decades. The development will ultimately be mixed use, with commercial, residential, educational and retail establishments. The aim is for the project to be an international example of best practice in environmental adaptation of a dense urban area. The City of Malmö initiated the development scheme, which started with B001—City of Tomorrow, including housing, offices, shops and services on the site of a previous industrial park. B001 was the first large demonstration area for sustainable urban development and following dialogue between developers and the owners of the estate, a contract was entered into by all parties, which included aims and objectives in relation to energy efficient buildings, smart energy installations and renewable energy.

The development featured the following initiatives to support smart and sustainable buildings, energy and renewable energy:

- A sustainable vacuum waste system, which automatically empties litterboxes through underground ducts and optically sorts material and waste for efficient recycling and disposal.
- Renewable energy systems where electrical and heating energy are supplied by wind turbines, an array of solar photovoltaic panels, solar collectors and geothermal energy systems.
- Energy efficient/passive houses. These were modelled using ICT to demonstrate how different efficiency measures could affect the energy performance of houses.
- Encouraged and protected biodiversity ecosystems.
- Smart meters for both new and older homes and participating commercial buildings, which provide immediate feedback on energy usage to customers, enabling them to understand, manage and optimize their personal usage.
- Electric vehicles to make the transport sector less carbon-intensive and the city quieter and cleaner.

Each individual project in the development was led by the relevant companies, including building developers and the utility company, E.ON Energy. Many of the projects did not have a reasonable payback period (and sometimes did not yield any repayment at all). Fullriggaren is a new sustainable district in the Western Harbor, which received grants from the Swedish Delegation for Sustainable Cities to enable it to pilot and provide examples of sustainability initiatives, including:

- passive and low energy houses,
- renewable energy,
- carpools,
- automatic food waste disposal systems,

- green walls and green roofs, and
- certification of green buildings.

The Western Harbor is also set to include eight smart energy villas, commissioned by E.ON. These will incorporate smart technology solutions to reduce energy use and allow for generation of electricity and hot water. Integrated electric and biogas vehicles will be used and all waste heat and household waste will be recycled. Residents will be able to monitor all electricity, heating and water consumption.

These smart and sustainable city initiatives give a good idea of where such technologies are heading and the advantages and applications of integrating technology, especially sensing networks with multiple nodes across cities. There are thousands of examples of cities, rural areas, households and private organizations implementing smart technologies to provide services that were not imagined feasible not too long ago. Advances in technology and miniaturization of microelectronic equipment are leading the way in distributing large networks of cost-effective nodes that gather and analyze data in real-time and provide citizens with immediate feedback on their environment. In the initiatives mentioned above, large sums of money have been invested in their respective applications and endeavors. These projects can only survive and reap benefits if the communities and city authorities, governments and municipalities support and back them.

New research and innovative equipment and devices are reported daily and information about these devices can be found from multiple sources on the internet. It is not an easy task to keep track of all new devices and applications, appearing almost daily, but the following section provides some examples of technologies that aim to improve smart cities and sustainable practices. Kickstarter/crowdfunding initiatives are another large driving force for new and exciting smart products and have been successful in many cases. Such products include Nest®, Pebble® Smartwatch, and the Tzoa®, which is a jewel-like sensor that can be clipped to clothing and measures air quality, temperature, humidity, atmospheric pressure and UV exposure; it is not yet mass-produced, but is likely to be a success. The following section provides examples of noteworthy reports of progress in smart technologies aimed at particularly sustainable cities.

The following section lists noteworthy reports of advancements in technology and new initiatives concerning the IoT, smart cities and sustainable development globally.

10.6 Noteworthy Reports of Smart and Sustainable Technologies

10.6.1 Wireless Sensor Networks to Monitor Food Sustainability (Libelium, 28 February 2012)

Food security refers to the ability of food systems to ensure that everyone has enough food to live a healthy life. To prevent food insecurity, reliable food systems at each stage of the food cycle are required, in production and harvesting, during transport and distribution, at points of sale and in social settings wherever food is consumed, as well as in the management of the resulting bio-waste outputs. Libelium's® Waspmote® sensors can be used to monitor and control the whole food cycle.

Food insecurity remains a major development problem across the globe, undermining people's health, productivity and often their survival. This problem usually affects developing countries and is especially severe in the Horn of Africa (Northeast Africa).

Food production by traditional means is already incorporating new technologies. In Italy, for example, residents can 'rent' a plot of land in a rural area and direct their crop cultivation over the internet. It allows them to share costs with other users. This is done by updating their internet website, which directs a rural farmer to plant and cultivate crops as requested by each individual crop 'renter'. Waspmote® takes this innovative business model to a more practical (and less resource-intensive) level by using sensor networks to maintain the monitoring capacity of crop cultivation throughout the production cycle.

By combining sensors detecting among others humidity, temperature and light, the risk of frost can be identified. Monitoring can ensure prevention of possible plant diseases or manage watering requirements based on soil humidity. This helps to control conditions in nurseries and to closely monitor high performance or delicate crops, such as vineyards and tropical fruit, where small changes in climate can affect the outcome. All this information also helps to determine the optimum conditions for each crop, by comparing the figures obtained during the best harvests. Given their ease of use and scalability, WSNs can also be used to monitor isolated areas where access is difficult, for example where mushrooms and truffles grow.

Waspmote® can detect and store environmental samples during the product's transport and therefore know whether it has been exposed to high temperatures or rain, whether it has been contaminated during the journey, or whether the container was opened in an unauthorized fashion or even whether it has been dropped or suffered an impact.

10.6.2 Google's® Pollution Sensing Gear in Street View Cars (Engadget, July 30th 2015)

A firm that builds environmental sensors teamed up with Google® to turn Street View cars into mobile pollution locators. Three of Google's® mapping vehicles have been equipped with hardware to measure harmful compounds in the atmosphere, including CO, methane and VOCs.

The pairing has already tested the system in Denver, USA, with the assistance of the EPA. In one month of driving around, the three vehicles managed to collect 150 million data points over 750 h of driving. The experiments will move to San Francisco, where local scientists and communities will be invited to design ways in which the information can be used in future.

10.6.3 International Space Station Photos Map Light Pollution

The Cities at Night project organizers began compiling hundreds of thousands of nighttime photos of metropolises taken from the international space station (ISS) in order to study light pollution in 2014. NASA® requested the public for help to sort through millions of images to be used for the initiative. Diffuse lighting is separate from lights emanating from buildings and vehicles, but scientists could not determine its exact origin from low-resolution satellite photos alone.

10.6.4 Samsung's Earth-Wide Internet Through 4600 Satellites (Engadget, 17 August 2015)

Facebook®, Google® and SpaceX® are not the only big corporations envisioning a world blanketed with satellite-based internet access. Samsung® has published research proposing earth-wide internet through a cloud of roughly 4600 micro-satellites. The vehicles would switch between multiple RF bands to deliver over a terabyte per second each, and a total capacity of 1 zettabyte (1 billion terabytes) per month—enough to handle the world's estimated mobile data traffic in 2028. The project aims to have ubiquitous internet access and expand the potential audience for its [Samsung's®] products, with worldwide internet access an (obvious) benefit.

10.6.5 Google Brings Android® One® Devices to Africa (Engadget, August 18th 2015)

Google's program to bring inexpensive Android® devices to developing countries has arrived in Africa with the launch of Infinix's® HOT® 2. The device has been available from August 2015 in six different countries, including Ghana, Ivory Coast and Kenya. It costs less than US $90. The HOT 2 has adequate hardware specifications with a quad-core CPU, 1 GB RAM and 16 GB storage, which is expandable with an additional microSD card slot.

10.6.6 Malaysia Will Use RFID to Track Vehicles (Engadget, 26 August 2015)

Malaysia plans to implement RFID-equipped road tax stickers that will help authorities track all vehicles, whether local or foreign. It will start with a pilot program at a border checkpoint in October 2015, but should expand to cover the whole of Malaysia by 2018.

The RFID system is officially intended to deter thieves and other criminals, and will also help pinpoint congestion through real-time traffic monitoring. Authorities could theoretically use the tracking device to follow the cars of political activists. This product has, however, been criticized in terms of the local community's privacy concerns.

10.6.7 Phone Data Can Predict Infectious Disease Outbreaks (Engadget, 30 August 2015)

A team of Princeton and Harvard researchers analyzed the phone call records of 15 million users in Kenya to study the spread of seasonal disease. The team wanted to see if the movement of phone users could predict outbreaks of seasonal diseases such as rubella (German measles). They used anonymous records for the period between June 2008 and June 2009, including the locations from where people made phone calls. They then compared the data to the areas where cases of rubella were reported in the country. It turned out the subjects' movement patterns matched the locations with the highest risk of rubella within that year-long period.

In future, the researchers plan to test if the same method is as effective in forecasting malaria and cholera outbreaks and whether it could eventually be used by doctors and other medical professionals.

The United Nations Sustainable Development Summit, held in September 2015, attempted to associate all smart and sustainable technologies with global efforts to improve the quality of life of all citizens on earth. Highlights of the summit are presented in the following section.

10.7 The United Nations Sustainable Development Summit 2015

The 193 member states of the United Nations, following negotiations that lasted from July 2012 until August 2015, have agreed upon the text of a new document entitled, "Transforming Our World: The 2030 Agenda for Sustainable

Development”. This agenda contains 17 goals and 169 targets (which have been criticized as being too many and too hard to keep track of). These goals and targets were officially adopted at the start of the UN Sustainable Development Summit in New York on 25 September 2015. The goals are to be achieved by all member countries within the next 15 years and were consequently named “Agenda for 2030”.

The post-2015 agenda will provide a new global framework for all countries to focus, coordinate and integrate their efforts better as they work towards sustainable development, while eradicating poverty in all its forms. The new 17 sustainable development goals (SDGs) are a universal set of goals, targets and indicators that UN member states are expected to use to frame their national development plans and policies over the next 15 years. The SDGs follow, and expand on, the MDGs, which were accepted by governments in 2000, and are due to expire at the end of 2015.

The 17 goals, which each contains a number of goals to achieve by 2030, are:

1. no poverty,
2. zero hunger,
3. good health and wellbeing,
4. quality education,
5. gender equality,
6. clean water and sanitation,
7. affordable and clean energy,
8. decent work and economic growth,
9. industry, innovation and infrastructure,
10. reduced inequalities,
11. sustainable cities and communities,
12. responsible consumption and production,
13. climate action,
14. life below water,
15. life on land,
16. strong peace and justice institutions, and
17. partnership for the goals

Two target goals that are specifically relevant to this book are goal 11 and goal 13; focusing on sustainable cities and communities and action against climate change. The targets for each of these goals are listed below.

The targets for sustainable cities and communities (goal 11) are that by 2030 members states should:

- ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums,
- provide access to safe, affordable, accessible and sustainable transport systems for all and improve road safety, notably by expanding public transport, with special attention to the needs of those in vulnerable situations such as women, children, persons with disabilities and older persons,

- enhance inclusive and sustainable urbanization and the capacity for participatory, integrated and sustainable human settlement planning and management in all countries,
- strengthen efforts to protect and safeguard the world's cultural and natural heritage,
- significantly reduce the number of deaths, the number of people affected and the direct economic losses relative to the global gross domestic product caused by disasters, including water-related disasters, placing the focus on protecting the poor and people in vulnerable situations,
- reduce the adverse per capita environmental impact of cities, including paying special attention to air quality and municipal and other waste management,
- provide universal access to safe, inclusive and accessible, green and public spaces, in particular for women and children, older persons and persons with disabilities,
- support positive economic, social and environmental links among urban, peri-urban and rural areas by strengthening national and regional development planning,
- by 2020, substantially increase the number of cities and human settlements adopting and implementing integrated policies and plans to effect inclusion, resource efficiency, mitigation and adaptation to climate change and resilience to disasters, and develop and implement, in line with the Sendai Framework for Disaster Risk Reduction 2015–2030, holistic disaster risk management at all levels, and
- support least developed countries, among others through financial and technical assistance, in building sustainable and resilient buildings using local materials.

The targets for climate action (goal 13) are to:

- strengthen resilience and adaptive capacity to climate-related hazards and natural disasters in all countries,
- integrate climate change measures into national policies, strategies and planning,
- improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning,
- implement the commitment undertaken by developed-country parties to the United Nations Framework Convention on Climate Change to a goal of mobilizing jointly US $100 billion annually by 2020 from all sources to address the needs of developing countries in the context of meaningful mitigation actions and transparency on implementation and fully operationalize the Green Climate Fund through its capitalization as soon as possible, and
- promote mechanisms for raising capacity for effective climate change-related planning and management in least developed countries and small island developing states, including focusing on women, youth and local and marginalized communities.

From the SDGs, which are extensions of the MDGs, it is evident that there is increased focus on changing current trends worldwide in terms of cities' sustainability and on addressing climate change issues. Targets include raising awareness and generating capital to implement processes to achieve set goals, which can be considered as important as implementing the strategies physically. Developing countries require larger contributions to be able to implement technologies that complement sustainable practices, decrease the demand on non-renewable resources and decrease their carbon footprint in a growing economy.

10.8 Conclusion

This chapter listed the categories of sustainable development, as well as indicators to measure the success of sustainability, especially for developing countries. Developing countries' cities are notably absent in the SCI's list of the 50 most sustainable cities worldwide. Funding and local backing, as well as instability in these countries, are slowing progress towards sustainability and preventing these cities from becoming future smart cites. Initiatives to improve these practices are being developed globally, but many developed countries are understandably first addressing their own needs and requirements to achieve smart status and ensure a sustainable future for their citizens. The internet is a facilitator of smart initiatives through technology such as the IoT, WSNs, and RFIDs and also serves as a database of information about new developments in technology. These new developments are often adopted relatively quickly by developed countries, leading to older technologies becoming less expensive and more affordable for developing countries. Developing countries are encouraged to take advantage of this in order to grow their technological profile and smart and sustainable footprints. Recently, the UN made its SDGs available, which aim to improve the quality of life of all people globally through investment initiatives and awareness campaigns. The goals of sustainable development and dealing with climate change are mentioned in this chapter, since they overlap thematically with other issues in this book.

References

Arcadis (2015). Sustainable Cities Index: Balancing the Economic, Social and Environmental Needs of the World's Leading Cities. Retrieved September 12, 2015 from http://www.arcadis-us.com

CIDA (2012). Indicators for Sustainability. How Cities Are Monitoring and Evaluating their Success. Retrieved September 6, 2015 from http://accredit-id.org

Hirst, P., Hummerstone, E., Webb, S., Karlsson, A., Blin, A., Duff, M., Jordanou, M., Deakin, M. (2012). JESSICA for Smart and Sustainable Cities.

IBM (2012). A Vision of a Smarter City. How Nairobi Can Lead the Way into a Prosperous and Sustainable Future.

ICLEI (2008). Sustainable Urban Energy Planning. A Handbook for Cities and Towns in Developing Countries. UN HABITAT, 2008.

Saha, D., Paterson, R. G. (2008). Local Government Efforts to Promote the “Three Es” of Sustainable Development. Journal of Planning Education and Research, 28(1), 21–37, 2008.

UNWATER (2008). Wastewater Management: A UN-Water Analytical Brief. Retrieved August 22, 2015 from http://www.unwater.org