

Erik Bründermann
Heinz-Wilhelm Hübers
Maurice FitzGerald Kimmitt

Terahertz Techniques



Springer

founded by H.K.V. Lotsch

Editor-in-Chief: W. T. Rhodes, Atlanta

Editorial Board: A. Adibi, Atlanta

T. Asakura, Sapporo

T. W. Hänsch, Garching

T. Kamiya, Tokyo

F. Krausz, Garching

B. Monemar, Linköping

H. Venghaus, Berlin

H. Weber, Berlin

H. Weinfurter, München

Springer Series in OPTICAL SCIENCES

The Springer Series in Optical Sciences, under the leadership of Editor-in-Chief *William T. Rhodes*, Georgia Institute of Technology, USA, provides an expanding selection of research monographs in all major areas of optics: lasers and quantum optics, ultrafast phenomena, optical spectroscopy techniques, optoelectronics, quantum information, information optics, applied laser technology, industrial applications, and other topics of contemporary interest.

With this broad coverage of topics, the series is of use to all research scientists and engineers who need up-to-date reference books.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors. See also www.springer.com/series/624

Editor-in-Chief

William T. Rhodes

Georgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, GA 30332-0250, USA
E-mail: bill.rhodes@ece.gatech.edu

Editorial Board

Ali Adibi

Georgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, GA 30332-0250, USA
E-mail: adibi@ee.gatech.edu

Toshimitsu Asakura

Hokkai-Gakuen University
Faculty of Engineering
1-1, Minami-26, Nishi 11, Chuo-ku
Sapporo, Hokkaido 064-0926, Japan
E-mail: asakura@eli.hokkai-s-u.ac.jp

Theodor W. Hänsch

Max-Planck-Institut für Quantenoptik
Hans-Kopfermann-Straße 1
85748 Garching, Germany
E-mail: t.w.haensch@physik.uni-muenchen.de

Takeshi Kamiya

Ministry of Education, Culture, Sports
Science and Technology
National Institution for Academic Degrees
3-29-1 Otsuka, Bunkyo-ku
Tokyo 112-0012, Japan
E-mail: kamiyatk@niad.ac.jp

Ferenc Krausz

Ludwig-Maximilians-Universität München
Lehrstuhl für Experimentelle Physik
Am Coulombwall 1
85748 Garching, Germany and
Max-Planck-Institut für Quantenoptik

Hans-Kopfermann-Straße 1

85748 Garching, Germany
E-mail: ferenc.krausz@mpq.mpg.de

Bo Monemar

Department of Physics
and Measurement Technology
Materials Science Division
Linköping University
58183 Linköping, Sweden
E-mail: bom@ifm.liu.se

Herbert Venghaus

Fraunhofer Institut für Nachrichtentechnik
Heinrich-Hertz-Institut
Einsteinufer 37
10587 Berlin, Germany
E-mail: venghaus@hhi.de

Horst Weber

Technische Universität Berlin
Optisches Institut
Straße des 17. Juni 135
10623 Berlin, Germany
E-mail: weber@physik.tu-berlin.de

Harald Weinfurter

Ludwig-Maximilians-Universität München
Sektion Physik
Schellingstraße 4/III
80799 München, Germany
E-mail: harald.weinfurter@physik.uni-muenchen.de

For further volumes:

<http://www.springer.com/series/624>

Erik Bründermann
Heinz-Wilhelm Hübers
Maurice FitzGerald Kimmitt

Terahertz Techniques

With 191 Figures

 Springer

Dr. Erik Bründermann
Honorable Guest Professor
Shizuoka University, Hamamatsu, Japan
Permanent Address:
Ruhr-Universität Bochum
NC 7 Rm 68
Universitätsstraße 150
44801 Bochum
Germany
erik.bruendermann@rub.de

Professor Dr. Heinz-Wilhelm Huebers
Technische Universität Berlin
Institut für Optik und Atomare Physik
und
Deutsches Zentrum für Luft- und
Raumfahrt e.V.
Institut für Planetenforschung
Rutherfordstr. 2
12489 Berlin
Germany
heinz-wilhelm.huebers@dlr.de

Professor Dr. Maurice F. Kimmitt
Physics Centre
University of Essex
Wivenhoe Park
Colchester Co4 3SQ
United Kingdom
mfk@kimmitt.org.uk

Springer Series in Optical Sciences ISSN 0342-4111 e-ISSN 1556-1534
ISBN 978-3-642-02591-4 e-ISBN 978-3-642-02592-1
DOI 10.1007/978-3-642-02592-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011939485

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: Integra Software Services Pvt. Ltd., Pondicherry

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Irene, Simona, and Mhairi

Preface

The terahertz (THz) spectral range is possibly the least explored part of the electromagnetic spectrum. This is despite the fact that THz research made a very significant early contribution to the development of physics when in 1900 H. Rubens and his colleagues measured the emission spectrum of a blackbody down to about 6 THz. This finally led to the discovery of the blackbody radiation law by M. Planck. For several decades it remained a rather exotic field, although there were interesting applications particularly in condensed matter research. However, in the 1970s it was realized that the THz portion of the electromagnetic spectrum – at that time still called the far-infrared or submillimeter region – offers unique opportunities in astronomical research. Subsequently, many dedicated THz telescopes were built, among them very successful air- and space-borne observatories. In the past few years, THz research and development has attracted ever-increasing attention. This has been triggered by tremendous progress in the development of sources, detectors, optics, and systems. THz technology is now on the verge of commercial applications for example in security, biomedicine, broadband communication, nondestructive testing, and process control.

Correspondingly, the number of researchers working in THz science and technology is rapidly growing. They might be developing new techniques and instruments for the exploration of the THz region or using THz instruments for specific research purposes. However, there are relatively few books, and they only cover some parts of the field. From our daily work with students we feel that there is a real need for a general introduction to all aspects of THz research, including methods, instrumentation, and physical principles. This book is in the spirit of an earlier, widely distributed textbook written by one of us (M. F. Kimmitt, *Far-Infrared Techniques*, Pion Ltd, London, 1970). It is a modern, comprehensive text aimed at covering the entire research field. The focus is on established THz techniques and instrumentation that are widely used over the frequency range 0.3–10 THz (1 mm to 30 μm). Recent developments are included in the belief that they are likely to become the standard techniques of the future. Wherever possible, references are given to excellent reviews on particular aspects. Figures and tables are based on published data and the reader is encouraged to read these in order to obtain

more detailed knowledge and insight. The book is primarily aimed at graduate students and researchers who are new to the field. But we also hope that active THz researchers may find helpful information and that the book will serve both as a useful introduction and as a reference for all involved in THz research.

As with our own THz research, the book has benefited from cooperation and discussion with many colleagues over a long period of time. They are far too many to be named here but to all of them we are greatly indebted. We also express our gratitude to the institutions where we are currently working, where we have worked, and which have supported our THz research. We especially thank our wives and children for their continuous support and never-ending patience during the time of writing, and, in particular, we thank Mhairi Kimmitt for her careful proof-reading of the text.

Bochum,
Berlin,
Brightlingsea

Erik Bründermann
Heinz-Wilhelm Hübers
Maurice FitzGerald Kimmitt

Contents

- 1 Introduction** 1
 - 1.1 The Terahertz Spectral Region 1
 - 1.2 History of Terahertz Research 5
 - 1.2.1 Early Twentieth Century Research 8
 - 1.2.2 The Years 1920–1940 8
 - 1.2.3 The Years 1940–1950 10
 - 1.2.4 The Years 1950–1960 12
 - 1.2.5 The Years 1960–1970 14
 - 1.2.6 The Years 1970–1980 15
 - 1.2.7 The Years 1980–1990 16
 - 1.3 Reasons for Increased Interest 19
- 2 Optical Principles at Terahertz Frequencies** 23
 - 2.1 Fundamentals of Gaussian Beams 24
 - 2.2 Gaussian Beam Propagation 29
 - 2.3 Coupling of Gaussian Beams 34
 - 2.4 Absorption of Radiation by Materials
and Resonant Effects 38
 - 2.5 Nonlinear Interactions with Matter and Terahertz Anisotropy 43
 - 2.6 Polarized Light 44
 - 2.7 Stokes Parameter 45
 - 2.8 Chirp 46
- 3 Optical Components** 51
 - 3.1 Introduction 51
 - 3.2 Reflection and Absorption 52
 - 3.3 Materials for Windows, Filters, and Lenses 53
 - 3.3.1 Polyethylene (PE) 53
 - 3.3.2 Polypropylene (PP) 55
 - 3.3.3 Polystyrene (PS) 55
 - 3.3.4 Polytetrafluorethylene (PTFE, Teflon) 55
 - 3.3.5 Poly-4-methylpentene-1 (PMT, TPX) 55

3.3.6	Tsurupica (Picarin).....	56
3.3.7	Polyethylene Terephthalate (PET, Mylar)	56
3.3.8	Parylene	56
3.3.9	Fluorogold.....	57
3.3.10	Zitex	57
3.3.11	Crystalline Quartz.....	57
3.3.12	Sapphire	58
3.3.13	Silicon and Germanium.....	59
3.3.14	Diamond.....	59
3.3.15	Alkali Halides	60
3.3.16	KRS-5	61
3.3.17	Material Considerations	62
3.4	Windows	62
3.5	Filters.....	63
3.5.1	Black Polyethylene (PE) Absorption Filters.....	64
3.5.2	Alkali Halide Filters (Yoshinaga Filters)	65
3.5.3	Scattering Filters	65
3.5.4	Reststrahlen Filters.....	67
3.5.5	Resonant Mesh Filters	67
3.5.6	Thick Metal Plate Filters	70
3.5.7	Cold Filters	71
3.6	Antireflection Coating	71
3.7	Lenses	72
3.8	Mirrors	74
3.9	Light Pipes	76
3.10	Light Concentrators and Winston Cone	78
3.11	Polarizers and Polarization Transformers	79
3.12	Beam Splitters	83
3.13	Attenuators	84
3.14	Absorbers and Absorbing Materials	85
3.15	Lens-Antenna Systems	86
3.16	Waveguides and Horn Antennas	91
3.17	Photonic Bandgap Materials and Metamaterials	92
4	Sources	103
4.1	Introduction	103
4.2	Thermal Emission	104
4.3	Practical Thermal Sources.....	106
4.3.1	The Globar	106
4.3.2	Plasma Sources.....	107
4.4	Gas Lasers.....	109
4.4.1	Electrically Excited Gas Lasers	109
4.4.2	Optically Excited Gas Lasers.....	112

4.5	Bulk Semiconductor Lasers	120
4.5.1	Germanium Lasers	121
4.5.2	Lasing from Optically Excited Donors in Silicon	126
4.6	Quantum Cascade Laser	128
4.6.1	Principle of Operation	128
4.6.2	Performance	131
4.7	Photomixing for the Generation of Terahertz Radiation	136
4.7.1	Low-temperature-grown Gallium Arsenide	137
4.7.2	Schottky Diode	138
4.7.3	Metal–Insulator–Metal Diodes	139
4.7.4	Photodiodes	140
4.8	Harmonic Generators	141
4.9	Backward Wave Oscillator	144
4.10	Smith–Purcell Sources	148
4.11	Gyrotrons	151
4.12	Terahertz Sources Based on Relativistic Electrons	154
4.12.1	Coherent Synchrotron Radiation from Electron Storage Rings	156
4.12.2	Coherent Synchrotron Radiation from Linear Accelerators	159
4.12.3	Free-Electron Lasers	160
4.13	Parametric Oscillators	164
5	Detectors	169
5.1	Introduction	169
5.1.1	Detector Cooling	171
5.2	Detector Theory	172
5.2.1	Detector Parameters	173
5.2.2	Relationship Between Detector Parameters	174
5.2.3	Sources of Excess Noise	178
5.3	Thermal Detectors	180
5.3.1	The Golay Detector	183
5.3.2	Pyroelectric Detectors	185
5.3.3	Thermopiles	190
5.3.4	Power Meters	191
5.3.5	Semiconducting Bolometers	195
5.3.6	Superconducting Bolometers	198
5.3.7	Room Temperature Microbolometers	202
5.4	Photoconductive Detectors	203
5.4.1	Extrinsic Germanium Detectors	209
5.4.2	Indium Antimonide Detectors	212
5.4.3	Gallium Arsenide Detectors	217
5.4.4	Blocked Impurity Band Detectors	218

5.5	Heterodyne Detection	221
5.5.1	Heterodyne Detection Theory	222
5.5.2	Schottky Diode Mixer	228
5.5.3	Superconductor–Insulator–Superconductor Mixer	234
5.5.4	Hot Electron Bolometer Mixer	239
6	Spectroscopic Methods	247
6.1	Introduction	247
6.2	Grating Spectrometer	251
6.3	Fourier-Transform Spectrometer	257
6.3.1	Dispersive Fourier-Transform Spectrometer	262
6.4	Polarizing (Martin–Puplett) Interferometer	263
6.5	Fabry–Pérot Interferometer	267
6.6	Time-Domain Spectrometer	269
6.6.1	Simplified Models of THz Pulse Generation and Detection	276
6.6.2	Applications	281
6.7	Coherent Source Spectrometers	282
6.7.1	Spectrometers with Multiplier-Based Sources	284
6.7.2	Spectrometers with Backward Wave Oscillators	286
6.7.3	Spectroscopy by Tunable Terahertz Sideband Generation	288
6.7.4	Spectroscopy by Difference Frequency Generation	291
6.8	Heterodyne Spectroscopy	293
6.8.1	System Aspects	294
6.8.2	Back-End Spectrometers	297
6.8.3	Examples of Heterodyne Spectrometers	298
7	Terahertz Imaging	301
7.1	Introduction	301
7.2	Some Fundamentals of Imaging	304
7.3	Camera-Like Imaging	307
7.3.1	Contrast Formation	308
7.3.2	Imaging with Direct Detectors	311
7.3.3	Heterodyne Imaging	314
7.3.4	Three-Dimensional Heterodyne Imaging	316
7.4	Short-Range Imaging	317
7.4.1	Imaging with cw and Long-Pulse Sources	318
7.4.2	Photomixer-Based Imaging	319
7.4.3	Imaging with Quantum Cascade Lasers	321
7.4.4	Real-Time Imaging with cw Sources	321
7.4.5	Spectroscopic Imaging	323
7.4.6	Three-Dimensional Imaging with a Quantum Cascade Laser	323
7.4.7	Imaging with Time-Domain Spectrometers	325

- 7.4.8 Spectroscopic Imaging with Time-Domain Spectrometers 327
 - 7.4.9 Three-Dimensional Imaging with Time-Domain Spectrometers 328
 - 7.4.10 Real-Time Imaging with Time-Domain Spectrometers 330
 - 7.5 Near-Field Microscopy 331
 - 7.5.1 Aperture-Based Scanning Near-Field Microscopy 331
 - 7.5.2 Aperture-Less Scanning Near-Field Microscopy 335
- References** 341
- Index** 375
- About the Authors** 381

Acronyms

AC	Alternating current
ACF	Autocorrelation function
ADC	Analog-to-digital converter
AFM	Atomic force microscopy
ALMA	Atacama Large Millimeter Array
ANSOM	Aperture-less near-field scanning optical microscope
AOS	Acousto-optical spectrometer
APEX	Atacama Pathfinder Experiment
ASE	Amplified spontaneous emission
aSNOM	Aperture-less scanning near-field optical microscope
ATGS	Alanine triglycine sulfate
BIB	Blocked-impurity-band
BS	Beam splitter
BTC	Bound-to-continuum
BWO	Backward wave oscillator
CCD	Charge coupled device
CPM	Colliding-pulse passively mode-locked
CR	Cyclotron resonance
CSL	Chirped superlattice
CSR	Coherent synchrotron radiation
CT	Computed tomography
CTS	Chirp transform spectrometer
cw	Continuous wave
DAQ	Data acquisition
DC	Direct current
DFG	Difference frequency generation
DFTS	Dispersive Fourier-transform spectroscopy
DLATGS	Deuterated L- α -alanine triglycine sulfate
DNA	Deoxyribonucleic acid
DSB	Double-sideband
EO	Electro-optic

EOS	Electro-optic sampling
EOS-Chem 2	Earth Observing System Chemistry 2
ESA	European Space Agency
ESF	Edge spread function
FEL	Free electron laser
FIR	Far-infrared
FMCW	Frequency modulated continuous wave
FPI	Fabry–Pérot interferometer
FTS	Fourier-transform spectrometer
FWHM	Full width at half maximum
GaAs	Gallium arsenide
Ge	Germanium
GREAT	German Receiver for Astronomy at THz Frequencies
HBV	Heterostructure barrier varactor
HEB	Hot electron bolometer
HIFI	Heterodyne Instrument for the Far-Infrared
ICR	Impurity cyclotron resonance
IF	Intermediate frequency
IMPATT	Impact ionization avalanche and transit time
InSb	Indium antimonide
IR	Infrared
IRAS	Infrared Astronomical Satellite
ISO	Infrared Space Observatory
I-V	Current-voltage
IVB	Intervalence band
JFET	Junction gate field-effect transistor
KAO	Kuiper Airborne Observatory
LATGS	L- α -alanine triglycine sulfate
LED	Light emitting diode
LHM	Left-handed material
LIGA	Lithographie, Galvanik, Abformung (lithography, electroplating, molding)
LO	Local oscillator
LSB	Lower sideband
LSF	Line spread function
LTG	Low-temperature-grown
MAX	Maximum
MBE	Molecular beam epitaxy
MEMS	Micro-Electro-Mechanical Systems
MIM	Metal–insulator–metal
MIN	Minimum
MIX	Mixer
MM	Metal–metal
MMIC	Monolithic microwave integrated circuit
mmW	Millimeter wave

MOM	Metal-oxide-metal
MOMED	Monolithic membrane diode
MPI	Martin–Puplett interferometer
MTF	Modulation transfer function
MW	Microwave
ND	Neutral density
NDT	Nondestructive testing
NEDT	Noise equivalent delta temperature
NEP	Noise equivalent power
NIM	Negative index metamaterial
NTD	Neutron transmutation doping
OAM	Off-axis mirror
OAP	Off-axis paraboloidal mirror
OFHC	Oxygen-free high conductivity
PACS	Photoconductor array camera and spectrometer
PBG	Photonic band gap
PD	Photodiode
PEATCAM	Passive European-American THz camera
PLL	Phase-lock loop
PSF	Point spread function
PZT	Lead zirconate
QPM	Quasi-phase-matching
RDX	1,3,5-Trinitro-1,3,5-triacetylcyclohexane explosive
REC	Receiver
RF	Radio frequency
RLT	Receiver Lab Telescope
RP	Resonant-phonon
SABOCA	Submillimeter APEX bolometer camera
SCUBA	Submillimeter common user bolometer array
SEM	Scanning electron microscope
Si	Silicon
SI	Semi-insulating
SIS	Superconductor–insulator–superconductor
SL	Superlattice
SMSMR	Short millimeter and sub-millimeter-wavelength range
SOFIA	Stratospheric Observatory for Infrared Astronomy
SOS	Sapphire-on-silicon
SP	Surface-plasmon or single-plasmon
SPIRE	Spectral and Photometric Imaging Receiver
SRR	Split-ring resonator
S-P	Smith-Purcell
SQUID	Superconducting quantum interference device
SSB	Single-sideband
s-SNOM	Scattering scanning near-field optical microscope
submm	Sub-millimeter

SWAS	Submillimeter Wave Astronomy Satellite
TDS	Time-domain spectroscopy
TE	Transverse electric
TEA	Transversely excited atmospheric
TEM	Transverse electric and magnetic mode
TEM	Transmission electron microscope
TES	Transition edge sensor
TGS	Triglycine sulfate
THz	Terahertz
Ti:Sa	Titanium-sapphire
TM	Transverse magnetic
TO	Transverse optical
TPO	THz-wave parametric oscillator
TPX	Transparent polymer X, trade name of poly-4-methylpentene-1 by Mitsui Chemical Inc.
TuFIR	Tunable far-infrared
UBI	Undulating beam interaction
USB	Upper sideband
UTC	Uni-traveling-carrier
VSF	Voltage-biased superconducting bolometer
YBCO	Yttrium barium copper oxide
YIG	Yttrium iron garnet

Symbols

α	Momentum compaction factor
α_{eff}	Effective polarizability
α_T	Polarizability of a nano-tip
A	Area
$\mathcal{A}$	Absorptivity
b	Blocking layer thickness
B	Bandwidth
C_E	Electrical capacitance
C_T	Thermal capacitance
d	Distance
Δ_{SSB}	Single sideband path length difference
Δ	Path length difference
E_0	Amplitude of electric field
E_{LO}	Electric field amplitude of the local oscillator
E_S	Electric field amplitude of the signal
E	Electric field
ϵ_0	Dielectric constant
ϵ	Relative (material) dielectric constant
ϵ_M	Permittivity of a medium
η	(Quantum) efficiency
$f^\#$	f-number (also F -number or $f/\#$)
γ	Damping constant, also: relativistic factor
g	Grating constant
G_{PC}	Photoconductive gain
$\Im$	Imaginary part
k_B	Boltzmann constant
L_0	Loss
L_{MIX}	Mixer loss
l	Resonator length
L	Length
L_c	Coherence length

M^2	M-square-value
m	m . order
μ	Mobility
N_A	Acceptor concentration
N_C	Carrier concentration
N_D	Donor concentration
N_v	Number of photons
n	Refractive index
N	Modulation depth
ν_I	Image frequency
ν_{IF}	Intermediate frequency
ν_{LO}	Local oscillator frequency
ν_S	Signal frequency
ν	Frequency
ω_P	Plasma frequency
P_0	Power amplitude
P_C	Constant power
P_E	Electrical power
P_t	Total power
Φ_R	Phase shift for one round trip
R_D	Detector resistance
r	Radius, also: Fresnel coefficient
R_L	Load resistance
R	Reflectivity
$\mathcal{R}$	Responsivity
$\Re$	Real part
S	Poynting vector, also: photon flux
Sinc	$\sin(x)/x$
T_0	Noise temperature
T_C	Cold load temperature
T_{DSB}	Double sideband temperature
T_H	Hot load temperature
T_{IF2}	Noise temperature at the 2nd intermediate frequency
T_{MIN}	Minimum noise temperature
T_{MIX}^{DSB}	Double sideband mixer noise temperature
T_{REC}	Receiver noise temperature
T_{REC}^{DSB}	Double sideband receiver noise temperature
T_S	Signal temperature
T_{SSB}	Single sideband temperature
t_t	Transit time
τ_0	Intrinsic time constant
τ_c	Carrier lifetime
τ_E	Electrical time constant
τ_e	Electron scattering time, also: electron-cooling time

τ_{ep}	Electron–phonon energy relaxation time
τ_{esc}	Escape time of phonons from a film into a substrate
τ_{T}	Thermal time constant
τ	Lifetime, also: integration time
Θ	Temperature change
v_{G}	Group velocity
V_{B}	Bias voltage
V_{b}	Built-in voltage
V_{J}	Johnson noise voltage
V_{LO}	Local oscillator amplitude
v_{P}	Phase velocity
V_{S}	Signal amplitude
w_0	Minimum waist radius (of Gaussian beams)
Y	Y -factor

Chapter 1

Introduction

1.1 The Terahertz Spectral Region

The terahertz (THz) spectral region, in this book considered as the range from 0.3 to 10 THz (1 mm–30 μm), is one of the most exciting and, at the same time, most frustrating sectors of the electromagnetic spectrum. Exciting because of the wealth of knowledge that can be gained from research at these frequencies, frustrating because, until recently, techniques have lagged behind those at lower and higher frequencies. But before introducing these real difficulties there is a more trivial dilemma – what is the best name for this spectral range?

For many years, the term “far-infrared” (FIR) was used to cover all wavelengths between $\sim 20\text{ }\mu\text{m}$ up to the shortest millimeter region. Later suggestions were “sub-millimeter” but as this, used strictly, provides a cutoff at 1 mm this was not entirely satisfactory. In 1978, Blaney suggested an accurate but rather long title “short millimeter and sub-millimeter-wavelength range” which could be abbreviated to SMSMR [1], but this did not become widely accepted. “Terahertz” is a relatively recent introduction and was initially closely associated with THz time-domain spectroscopy (THz TDS) but is now generally accepted as the convenient description for this spectral range regardless of whether it is approached from the IR or microwave direction. There is some doubt as to who first suggested the term terahertz, but it seems to have become increasingly accepted following its use by J. W. Fleming in a 1974 paper [2].

A major problem for research in this spectral region is the very high absorption of the Earth’s atmosphere, over most of the frequency range, mainly due to the vibrational–rotational levels of water vapor. This limited THz astronomy, and research in the upper atmosphere, to high altitude observatories and balloon flights in the very few “windows” available. Figure 1.1 shows the atmospheric transmission, between 0.2 to 2.2 THz, from the site of the Receiver Lab Telescope (RLT) on Cerro Sairecabur in the Andes range of northern Chile. The site is at 5525 m and appears to have some of the best THz transmission available. The spectrum was obtained in January 2005 with a Fourier-transform spectrometer

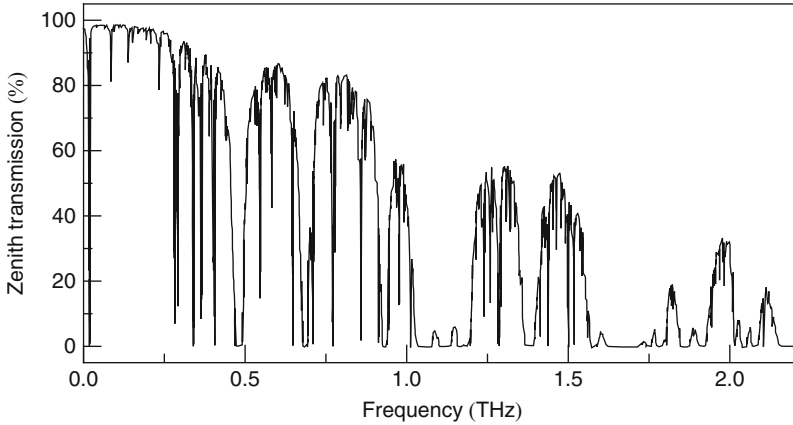


Fig. 1.1 Zenith atmospheric transmission at Cerro Sairecabur on 24 January 2005 (adapted from [5])

(FTS) with a resolution of 3 GHz [3]. Apart from other sites in the Andes, the only other locations for comparable transmission are on the Antarctic plateau. In what can only be described as a heroic experiment, led by the Polar Research Institute of China and the Chinese Academy of Science, exceptional THz transparency was found at Dome A, which is the highest point on the plateau at 4100 m [4]. The Antarctic atmosphere is very stable and it seems probable that a permanent observatory at Dome A would provide long periods of good THz transmission. However, logistically, the site presents severe problems, especially as the lowest water vapor content is inevitably in the Antarctic winter.

Above the frequency range shown in Fig. 1.1 the absorption increases very rapidly and for many years this restricted the majority of THz research to the laboratory. However, high-flying laboratories such as the Kuiper Airborne Observatory (KAO), and an ever-increasing number of orbital platforms, such as ESA's Herschel Space Observatory, have revolutionized astronomy and space science. While ground-based THz astronomy is with a few exceptions impossible, the strong atmospheric absorption also constrains laboratory experiments seriously, as one can see from Fig. 1.2.

The intensity of atmospheric absorption is remarkable, reaching a maximum of over 2×10^5 dB/km at around 8 THz [6]. To put that number in perspective, less than 1% of light at these frequencies would travel from a source to a detector over a path length of 10 cm. As an aside, one of the authors of this book failed to locate a powerful line at 9.1 THz from the first high-power THz laser [7] because there was a short open-air section in the optical path. As a further aside, when the distinguished cosmologist Sir Fred Hoyle turned to science fiction he wrote a book entitled "October the First is Too Late." In it a rocket is launched with a payload to measure the sun's output at a "wavelength roughly a hundred times less than the shortest radio waves," a wavelength at which no radiation reached the earth's

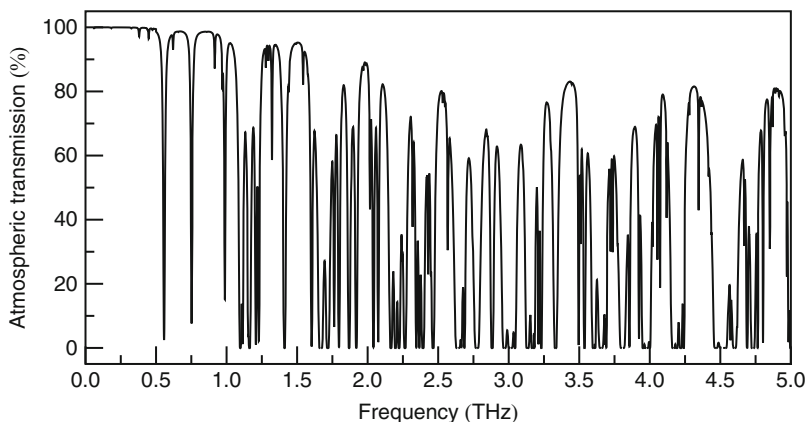


Fig. 1.2 Atmospheric transmission of a 1-m path at 1013 hPa and 40% relative humidity

surface. A powerful coherent signal is observed from a source with an apparent dimension of 10 times the radius of the sun. The scientists suggest that this would produce such a non-divergent beam that they must have located an inter-galactic relay station!

To achieve negligible absorption throughout the THz region, optical paths need to be evacuated to below 100 Pa, a pressure readily achievable with rotary vacuum pumps. However, if tunable sources are available, it is possible to find a frequency over most of the THz range, where short to medium path-length experiments can be performed at atmospheric pressure. Figure 1.2 shows the transmission under typical laboratory conditions.

Depending on the application, there are alternative units that may be more convenient for use in this spectral region. For example, the energy of the photon in electron-volts is very relevant for studying semiconductors, and the reciprocal centimeter (cm^{-1}), originally given the name “Kayser” after J. H. G. Kayser, who compiled a giant catalog of chemical spectra in the early 1900s, is convenient for describing spectra. The relationship between the various units is shown in Fig. 1.3. Before describing the history of the THz region, which began in the 1890s, it is relevant to consider the implications of the physical dimensions and electron energies shown in Fig. 1.3. In wavelength terms, 30–300 μm bridges the gap between the region where open path optical techniques are still convenient to use, up to the shortest wavelengths where waveguides can be employed. At 30 μm (10 THz), systems can be designed that are virtually identical to those used in the visible or near-IR regions with lenses, prisms, mirrors, etc., and spectroscopy can be performed with diffraction grating or Michelson interferometer instruments. However, as the wavelength is lengthened the size of optical components must be increased to avoid diffraction losses and interference effects. In the microwave region, these problems are avoided by restricting the radiation to a single mode in a waveguide. Although waveguides are manufactured for use up to ~ 2 THz, there

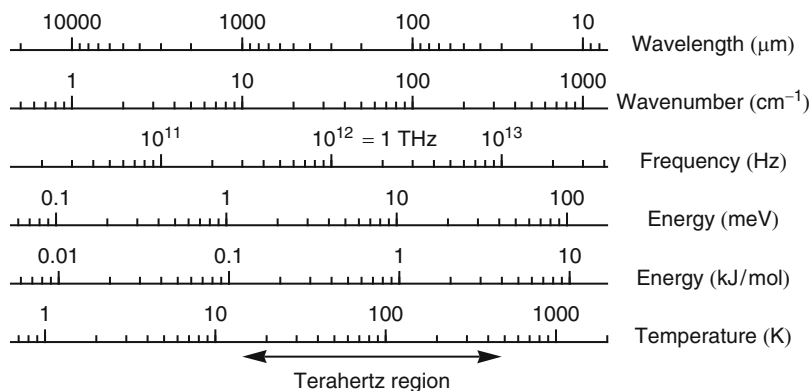


Fig. 1.3 Unit conversion in the THz frequency range

are major difficulties when the dimensions are reduced to match short millimeter, or sub-millimeter wavelengths. A typical rectangular waveguide for single-mode transmission at a wavelength λ has dimensions of approximately $\lambda/2$ by $\lambda/4$. Manufacture of very small cross-section guides of sufficient uniformity is difficult, but this is not the major problem. As a wave passes down a waveguide it sets up currents in the surface. These currents are required if the walls are to reflect the waves and, as the frequency of the electromagnetic wave is increased, only electrons near the surface of the waveguide have time to respond. In a resistive surface, this means that the effective resistance, through which the wall current flows, increases with frequency and the waveguide becomes increasingly lossy. Despite these difficulties, both optical and microwave-type arrangements are being employed over most of the THz region, and this has led to the development of what is called “Gaussian” optics. These are explained in more detail in Chap. 2.

While the physical dimensions of the wavelength range govern the types of spectroscopic systems used, it is the electron energy spread shown in Fig. 1.3 that makes this a fascinating region for research and application. The value of THz frequencies for a wide variety of studies can be readily understood by reference to Fig. 1.3. Room temperature corresponds to approximately 6 THz according to $h\nu = kT$. Hence, by varying the temperature from 14 K to 210°C, $h\nu$ can be made equal to kT over the range from 0.3 to 10 THz with comparative ease. This is the unique feature of this part of the electromagnetic spectrum. In energy terms, the frequency range covers 1.2 to 37 meV, corresponding, for example, to the energy levels of many phonon bands, shallow impurities in semiconductors, and rotational interactions in gases. It should be stressed that the relationship between $h\nu$ and kT is extremely important to a broad range of phenomena that can be studied in the THz region. For example, one noteworthy application is the investigation of the complex processes that lead to the formation and destruction of the ozone layer in the upper atmosphere. At THz frequencies, absorption and emission of radiation by gases are due to rotational transitions of molecules and the strength and the frequency of

maximum interaction is dependent on the molecular mass. The great advantage of the THz region is that the strength of the interactions increases, as the square to cube of the frequency, up to a mass-dependent maximum, but then falls exponentially at higher frequencies. For the great majority of molecules of interest to those studying the atmosphere, the peak absorption is in the THz range.

As mentioned earlier, now that there are a number of ways of reaching heights where atmospheric absorption is greatly reduced, or avoided altogether, the way has been opened for astronomers and astrophysicists to make a wide range of studies in the THz region. Because the vast majority of the Universe is either cold or very cold, the peak of the blackbody curves for most emissions falls into the THz region. For example, this frequency band contains information on the cosmic background at 2.7 K, which has a peak emission at 270 GHz while the interstellar medium, of great interest for star-formation studies, is as low as 10 K in the cooler regions to above 100 K in the denser parts. Thus, the peak emission from these gases varies from 1 THz to above 10 THz.

1.2 History of Terahertz Research

Research in the “gap” between the IR and microwave spectral regions took a giant step forward with the advent of THz pulse emission from photoconductive antennas (T-rays) but, reading some of the 1990s papers, one could be forgiven for thinking that these were the first experiments in the THz spectral region. In fact, the earliest studies were a hundred years earlier, almost entirely due to the work of Heinrich Rubens (Fig. 1.4), who was professor at the Technische Hochschule Berlin-Charlottenburg (now: Technische Universität Berlin) and at the Universität Berlin. The large majority of papers published up to 1920 were by Rubens, or researchers who had collaborated with him, but from then on there was a steady flow of papers, eventually leading to a rapid expansion from the 1950s onward. Rubens’ research concentrated on the extension of the IR spectral region to longer wavelengths into what soon became called the far infrared (FIR). In the late nineteenth century, there was also research on short “electric” waves, following the experiments of Heinrich Hertz, and Rubens participated in this, studying the polarization and reflection of electric waves, before beginning his IR work. There was a remarkable amount of microwave research in the last decades of the nineteenth century, some of it reaching to frequencies close to 0.1 THz. Distinguished scientists were involved, including Lodge, Fleming, Righi, Popov, and Lebedev, with theoretical contributions from Lord Rayleigh. The most remarkable experiments were those of the Indian physicist, J. C. Bose. Bose studied at Cambridge University, where he came under the influence of Rayleigh. On returning to India he took up a post at the Presidency College in Calcutta, and there he carried out a variety of microwave experiments, later returning to England to give demonstration lectures. Working at frequencies up to 60 GHz, he constructed prisms, lenses, grid polarizers, reflection gratings, a horn antenna and a double prism attenuator. For detection he



Fig. 1.4 Heinrich Rubens at work in the laboratory

used primitive point contact diodes. Bose measured the current–voltage curve of his junctions, noting their nonlinear properties and plotting curves very similar to those of modern diodes [8]. Sir Neville Mott, Nobel Laureate for his own contributions to solid-state electronics, is quoted as saying that “Bose was at least 60 years ahead of his time” and that “he had anticipated the existence of p- and n-type semiconductors.”

By 1900, the experiments of Marconi and others had shown that lower frequencies were much more useful for communication and interest in very high frequencies died. It was to be nearly forty years before its revival and it was to take another decade before there was significant encroachment into the THz region from the microwave direction. Progress in reducing the “gap” in the electromagnetic spectrum was to come initially by extending the IR to longer wavelengths.

During his thirty years of research into long-wavelength IR, Rubens had many collaborators, including several young Americans, and it was with one of them, B. W. Snow, that he performed his first significant experiments using prisms of rock salt (NaCl), sylvine (KCl), and fluorite (CaF₂) [9], eventually extending the spectrum to nearly 20 μm . IR prism spectroscopy was to have a long life-span, as it was widely employed into the 1950s. But it was another of the young Americans, E. F. Nichols, who made the discovery that opened the way to the first significant experiments in the FIR region. Working with crystalline quartz, he discovered that its reflectivity over a narrow wavelength range near 9 μm rose from a few percent to nearly that of a silvered surface [10]. This phenomenon, which Rubens later called the “reststrahlen” (residual ray) effect, is due to lattice vibrations and, with other crystals, Nichols and Rubens were to find similar narrow bands at longer wavelengths. With multiple reststrahlen reflectors, they were able to produce nearly monochromatic light at specific wavelengths. The elegant spectroscopic system used

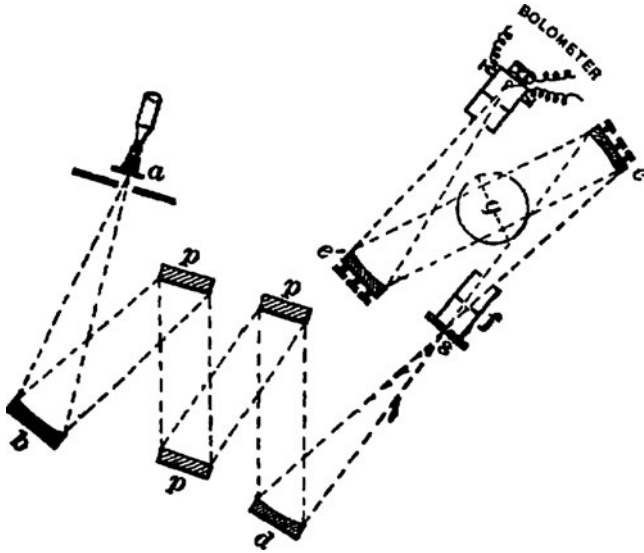


Fig. 1.5 Original reststrahlen filter arrangement. (a) source; (b) and (d) front-silvered concave mirrors; (p) reststrahlen plates; (s) adjustable slit; (f), (e) and (e'), with a wire grid grating (g) form a monochromator. Reprinted with permission from: [11]. Copyright (2012) by the American Physical Society

by Rubens and Nichols [11] is shown in Fig. 1.5. After Nichols left Berlin, Rubens and his colleagues extended the techniques to wavelengths beyond $50\text{ }\mu\text{m}$ (6 THz), and this research produced a dramatic outcome.

Toward the end of the nineteenth century one of the major problems engaging the attention of physicists and mathematicians was that of radiation from hot-bodies and, specifically, its variation with wavelength. A brief roll call of some of those involved is more than impressive: Wien, Angstrom, Rayleigh, Lummer, Pringsheim, Paschen, Larmor, and Planck. Various formulae for the radiation from an ideal source, the blackbody, were derived but none fitted all of the experimentally obtained results. A particular problem was to obtain really accurate data at very long wavelengths. In 1900, Rubens and Kurlbaum [12], using the reststrahlen spectrometer, obtained the necessary data and Rubens immediately visited Max Planck to give him the results and he, that same day, wrote down the equation which is now called Planck's Radiation Law.

After several weeks of what he later described as the most strenuous work of his life, Planck revolutionized physics by inventing quantum theory to explain the radiation formula, which he had derived empirically by combining the already well-known short wavelength radiation results with the long wavelength ones of Rubens. And after the death of Rubens in 1922 Planck was to write: "Ohne das Eingreifen von Rubens wäre die Formulierung des Strahlungsgesetzes und damit die Begründung der Quantentheorie vielleicht in ganz anderer Weise, vielleicht gar nicht einmal in Deutschland zustande gekommen." ("Without the intervention of

Rubens the formulation of the radiation law, and for this reason the foundation of quantum theory, would have been accomplished in an entirely different way, and maybe not even in Germany.”) [13].

1.2.1 Early Twentieth Century Research

As mentioned earlier, Rubens continued to dominate the FIR region until the time of his death. Between 1892 and 1922 just over 150 papers were published which included FIR research. Only 18 of them do not have Rubens as an author or joint author. Much interesting work was done, and three discoveries made at that time are still relevant today.

In 1911, Rubens and Baeyer showed that the mercury arc lamp in a quartz envelope was an excellent long-wavelength IR source [14]. In this region, the arc plasma is opaque and has an equivalent blackbody temperature of over 4000 K and, in modern grating spectrometers and interferometers, it is still the preferred source. One year earlier R. W. Wood, of Johns Hopkins University (USA), had produced the first “blazed” diffraction grating, which allowed the concentration of the majority of light from a source into a single order [15]. Wood achieved efficiencies of over 75% for the first order and showed that gratings of this type could be used out to beyond 100 μm (3 THz). Although Wood produced the first blazed gratings, the initial conception was by Rayleigh in 1874 [16]. Apart from their use in spectrometers, blazed gratings, with efficiencies close to 100% in one polarization, are employed in lasers as tuning elements. In carbon dioxide lasers, for example, replacing one of the mirrors with a blazed grating allows tuning over many wavelengths between 9 to 11 μm . These lasers are widely used as the “pump” for the optically excited THz lasers described in Sect. 4.4.2.

The third discovery, by Hagen and Rubens, concerned the reflectivity of metals [17]. They showed that the reflectivity depended on their electrical conductivity. The formula that they found empirically, which was later explained theoretically, shows that in the THz region any pure metal has a reflectivity of well over 99% so that even complex reflecting systems are virtually loss-free.

The “Rubens’ era” ended with an experiment where he used all his spectroscopic skills to measure the absorption of water vapor out to 400 μm [18] (Fig. 1.6). This was a remarkable achievement, as it had to be done point-by-point without any electronic aids.

1.2.2 The Years 1920–1940

Between 1920 and 1930 some 40 FIR papers were published and from 1930 to 1940 another 80. In 1920, after a relatively short but glittering career after leaving Berlin, E. F. Nichols became Director of Pure Science at the National Electric Light Association in Cleveland, Ohio. Here, before his untimely death in 1923, he returned to long-wavelength research, working with J. D. Tear, and in 1923 they succeeded in

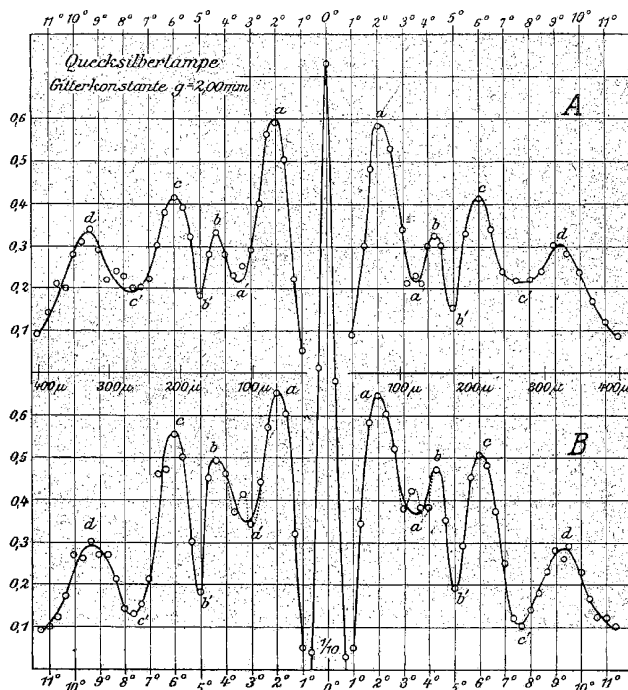


Fig. 1.6 Absorption of water vapor measured by Rubens in 1921 with a wire grid grating spectrometer. Traces A and B were measured with different illumination of the spectrometer slit by the mercury arc lamp. The *dashed line* around the minimum *a'* was measured with higher resolution. Rubens attributed the two minima in trace B at the position *a'* to water absorption. The origin of the minima *b'* and *c'* could not unambiguously be identified. Comparison with the spectrum in Fig. 1.2 shows that Rubens' assignment was correct (from [18])

joining the electric wave spectrum to the IR. Their experimental setup is shown in Fig. 1.7. Using a Hertzian oscillator, wavelengths down to $220\text{ }\mu\text{m}$ were produced, which overlapped with the extension of wavelengths from a mercury arc source to $420\text{ }\mu\text{m}$ [19]. In 1924, a Russian physicist, Glagolewa Arkediwa, showed that radiation of about $90\text{ }\mu\text{m}$ could be generated by exciting small Hertzian oscillators in the form of brass filings immersed in oil [20].

Apart from Berlin, the main laboratories involved in long-wavelength IR studies were at the University of Michigan, where H.M. Randall was to be a dominant figure in IR research for many years, and the California Institute of Technology, where the usefulness of the FIR region for the study of the rotational levels of gases was realized by R. M. Badger [21], who was later joined by C. H. Cartwright and J. Strong. Both Cartwright and Strong made significant contributions to improving FIR techniques between 1930 and 1940, with Cartwright following in the footsteps of earlier American researchers by spending time in Berlin. He made detailed studies of the absorption of a variety of materials [22] and designed a liquid oxygen-cooled thermopile with more than ten times the responsivity of its room-temperature



Fig. 1.7 Nichols and Tear in a laboratory of the National Electric Light Association in Cleveland, Ohio, where Nichols was the Director. The Hertzian type oscillator is within the metal container on the left. The output is collimated by a paraffin wax lens onto a Boltzmann interferometer which Nichols is adjusting. The reflected beam is then focused by the second lens onto a radiometer detector (for details of the experimental setup see [19], photograph courtesy of J. D. Tear’s relatives)

counterpart [23]. Strong collaborated with Randall on a “Self-Recording Spectrometer” [24], went on to study the absorption of 14 gases between 20 and 200 μm [25] and later, at Johns Hopkins University, made outstanding contributions in the optical field. In his latter years, he was one of the first to pursue the transition from grating to interference spectroscopy at THz frequencies [26].

Although there were major improvements in spectroscopic systems throughout these decades, there were only marginal improvements in sources and detectors. Thermocouples, thermopiles, or bolometers, feeding into very sophisticated galvanometers, were the detectors, usually with a mercury arc or some heated material such as the Welsbach mantle as a source [27].

Probably, the most significant experiment of the 1930s was the observation of the molecular absorption of ammonia gas over the wavelength range 1–4 cm using a coherent source, a split-anode magnetron [28]. This was the prelude to the field of microwave spectroscopy, which expanded dramatically from the late 1940s onward, when new coherent sources, such as the klystron, became available for research applications.

1.2.3 The Years 1940–1950

Not surprisingly, 1940–1946 was a barren time for research publications but, of course, there were many developments during this period. Of particular note is the research in Germany to extend the range of prism spectroscopy. Up to 1940,

25 μm was the maximum wavelength using KBr prisms but, with the mixed crystal thallium–bromide–iodide, KRS-5 (KRS standing for “Kristalle aus dem Schmelzfluss”, crystals from the melt), coverage was extended to 37 μm [29]. The introduction of cesium bromide [30] and cesium iodide in the early 1950s allowed prism systems to operate out to nearly 60 μm [31].

A very significant post-war event was the introduction of the pneumatic detector, now almost always referred to as the “Golay” [32]. This elegant device came within a factor of three of the theoretical limit for a room temperature detector. Although the detector is named after Golay, it is interesting to note that the first published paper was by H. A. Zahl and M. J. E. Golay [33] and Zahl filed a patent for a “pneumatic cell detector” in 1938. In 1939, Zahl also patented a “System for detecting Sources of Radiant Energy” with co-inventor M. J. E. Golay. At that time, there was considerable effort toward methods of detecting hot objects by their radiant emission. The first cooled bolometer was also a 1940s invention. This used the superconducting transition point of tantalum at 4.4 K, thus requiring the use of liquid helium [34]. A later version was made with niobium nitride (NbN), which has a transition at 14.3 K, a temperature which could be obtained by pumping on the then more readily available liquid hydrogen [35].

There were huge advances in electronics during the Second World War and this produced a dramatic change in spectroscopic systems. Virtually, all experiments up to 1940 used DC recording systems but from 1945 onward detectors were designed with shorter time constants, the incoming radiation was “chopped” and AC amplification employed. One earlier exception was a recording spectrograph for the FIR [36] covering the range 18–200 μm which used a very low frequency amplifier that had been designed as early as 1932 [37].

The development of high-frequency radar systems had led to new magnetron and klystron sources, and these were employed for a number of microwave spectroscopy experiments following on from Cleeton’s earlier measurement of the ammonia inversion spectrum. With a klystron, frequency doubled in a crystal harmonic generator, the absorption of oxygen between 50 and 60 GHz was measured in 1944 at the Massachusetts Institute of Technology (MIT) with the results published in 1946 [38]. Among those involved were W. Gordy, who later wrote a review covering this early research [39], and C. H. Townes and his colleagues [40] who, by 1952, had reached frequencies of 270 GHz [41]. In 1954, microwave spectroscopy was extended into the submm region by Gordy’s research group [42], thus providing a clear overlap with the longest wavelength IR spectroscopic systems. Gordy had a remarkable career, beginning his research with studies of hydrogen bonding. During World War II, he was at the MIT Radiation Laboratory, where he participated in the development of microwave radar, immediately realizing the potential of microwave technology for molecular spectroscopy. Joining the Physics Department at Duke University in North Carolina in 1946, he was instrumental in setting up a microwave laboratory to pursue research in what he described as the “gap in the electromagnetic spectrum.” This laboratory, which moved to Ohio State University in 1990, remains prominent in submm research after six decades. Notable among Gordy’s 75 PhD students is Frank de Lucia, who has been at the forefront of THz research since 1970.

1.2.4 The Years 1950–1960

By the end of the 1940s, the stage was set for a major expansion of IR spectroscopy and for the extension of microwave techniques into the THz region. The distinguished IR and FIR researcher, Armand Hadni, has described the period as the start of the “explosive” years [43], but the number of scientists involved in what was still considered to be a very difficult spectral region was relatively small, with less than 200 papers published in this decade. The majority of these came from the relatively few laboratories which had significant FIR groups. The important centers could almost be counted on the fingers of one hand: J. D. Strong at Johns Hopkins University, USA, later to be joined by G. A. Vanasse; E. E. Bell, and R. A. Oetgen at Ohio State University, USA; Earl Plyler at the National Bureau of Standards, USA; A. Hadni at the University of Nancy, France; L. Genzel at the University of Frankfurt, Germany, who was later to move to Freiburg; H. Yoshinaga, at Osaka University, Japan, and H. A. Gebbie at the National Physical Laboratory, England. There were also other more isolated researchers who were to have great significance in the opening up on the FIR region, including P. B. Fellgett at Cambridge in England and P. Jacquinot at Orsay, France. Toward the end of the 1950s new names appeared and these, with others, were to launch the FIR into the mainstream of optics research. Notable figures were P. L. Richards (USA), N. G. Yaroslavski (USSR), E. H. Putley (UK), and D. H. Martin (UK).

Grating spectroscopy still dominated, with the mercury arc as source and the Golay as detector, but improved electronics and recording systems greatly speeded up research. However, it was in this decade that the foundations were laid for the transition to FTS, which was to revolutionize FIR spectroscopy. Although the transition from gratings to FTS progressed comparatively slowly, mainly due to the limitations of early computers, the outcome marked a development that was certainly as important as the more dramatic arrival of THz TDS some three decades later. It was well known that interferometers such as the Michelson were excellent for the accurate study of monochromatic light but when there were more than three or four wavelengths from a source the interferogram produced was very difficult to analyze. Even so Michelson, with infinite patience and his own remarkable invention of a “harmonic analyzer” [44] (which was in essence a mechanical analogue computer) was able to derive spectra by observing fringe intensity. Both Michelson and Rayleigh were aware that the full spectrum could be obtained by Fourier transformation but had no way of performing this. The first person to derive an actual spectrum was Peter Fellgett, while studying stars in the IR. In the visible and very near-IR, a spread of wavelengths could be observed by placing a photographic plate at the exit of a diffraction grating monochromator but, at longer wavelengths, until the comparatively recent availability of detector arrays, only one wavelength could be studied at a time. After various efforts at coding different wavelengths with an ingenious chopping system – a method also employed by Golay [45] Fellgett realized that if a two-beam interferometer were to be placed in the path of undispersed radiation and its path difference scanned at a uniform rate, the

resultant signal from a single detector would contain information at all wavelengths simultaneously. Fellgett had to transform his interferogram by hand, using devices called Lipson–Beaver’s strips, a very tedious task [46].

The ability to study a wide range of wavelengths simultaneously became known as the multiplex or Fellgett advantage. What he had not immediately realized was that, while for high-resolution grating spectroscopy very narrow slits were required, thus reducing the amount of light passing through the instrument, interferometers had the further advantage of having large apertures. This was first pointed out by Pierre Jacquinot and came to be known as the Jacquinot (or throughput) advantage. In the 1950s, a number of experimenters, including Jacquinot, Connes, Strong, Vanasse, and Gebbie, began research with interferometers. The throughput, multiplex and resolving power advantages of interferometers were soon recognized but the limitations of early computers meant that it might take days to analyze the interferograms. When this limitation was overcome in the following two decades, FTS became the dominant force in spectroscopy in the FIR.

Jacquinot [47], Strong [48], Fellgett [49], and Gebbie [50] have written short review papers concerning the development of FTS in this period. Connes, in a detailed paper [51], has described the contribution of Michelson and even earlier researchers in the field. He points out that the modern form of FTS was almost discovered by Rubens and his collaborators in 1910 [52]. On the detector side, the Golay was virtually the only FIR detector of the 1950s but in 1959 the first convenient-to-use cooled device, the carbon bolometer, was invented [53] and the same year saw the first photoconductive detector to reach wavelengths longer than $100\text{ }\mu\text{m}$ [54].

Although no fundamental electronic source was produced for frequencies above 300 GHz in the 1950s, the foundations were laid with the first backward wave oscillator (BWO) or carcinotron [55, 56]. Extension to higher frequencies followed in the 1960s and BWOs are now available to above 1 THz, with a tuning range of $\pm 10\%$ of the center frequency.

Of the papers published in this decade, 80% were on applications, which included the first studies on semiconductors, including cyclotron resonance measurements [57], many papers on the absorption spectra of gases and, at 0.2 THz, observation of solar and lunar radiation [58]. Early work on the transmission of superconductors began, which was to lead to the direct measurement of the energy gap predicted by the BCS theory [59]. This decade also saw the first FIR diagnostics on the high-temperature plasmas produced in the attempts to achieve controlled thermonuclear fusion. These early experiments succeeded in measuring both the average electron density and the electron temperature in several different plasmas [60].

The late 1950s also saw the first steps on the long journey that would lead to the development of high-power gyrotrons. The earliest suggestions appeared in theoretical papers published in Australia [61] and the USSR [62], with the first experimental studies being made at microwave frequencies in the USA [63].

1.2.5 *The Years 1960–1970*

This was a flourishing decade for the THz region, with well over 1000 published papers and several new groups appearing. One of these, that of Fritz Kneubühl in Switzerland, was to have far-reaching consequences because, apart from his group doing outstanding FIR research, he, from the mid-1970s, organized a series of conferences in Zürich devoted to the IR and FIR spectral regions. Another series of conferences was to have longer-term importance. These began with a meeting in the USA at Atlanta in 1974, entitled “The First International Conference on Submillimeter Waves and their Applications.” The guiding hand was that of K. J. Button, then at MIT, who predicted then that it “would be the first of a long series because the submillimeter wave specialty has become well-defined and moderately well-populated.” Rome, in 2010, saw the 35th Conference in the series. But returning to the 1960s, the significant feature of this period is that it produced much of the instrumentation that is widely used today. If one looks back to the time before 1960 the only inventions which have not been superseded, apart from the spectrometers themselves, are the mercury arc source and the Golay detector. And for many applications the Golay detector is now replaced by the cheaper, and more convenient pyroelectric detector. Pyroelectric detectors date from this era although the use of the pyroelectric effect for detection had been suggested as early as 1938 [64]. The 1960s onward saw great progress in detector development with the n-InSb electron bolometer [65], the Ge bolometer [66], and a tunable FIR detector [67]. One of the most widely used detectors, the extrinsic photoconductor Ge:Ga, also dates from this decade [68]. When stressed this responds out to 200 μm [69]. While the “Putley” detector is still very widely used the Ge bolometer has been largely superseded by the more versatile Si composite bolometer [70]. A further helium-cooled detector based on the Josephson effect was also introduced [71]. One of the authors of this paper was Paul Richards who, in 1966, set up his research group at the Berkeley campus of the University of California and has been an outstanding figure in FIR/THz research over more than fifty years. Another important step was the invention of the so-called “honeycomb” diode chip design for Schottky diode mixers by D. T. Young and J. C. Irvin [72]. Originally applied to mmW frequencies this design became widely used in the 1970s and 1980s.

The major discovery at the start of this decade was the laser. The first THz laser, the water vapor laser, was invented in 1964 [7]. This was followed by several other long-wavelength gas lasers, including HCN, which provides several continuous wave (cw) lines close to 1 THz [73]. Lasers were to become a major tool for THz research in the proceeding years.

What was perhaps the most far-reaching paper published in this decade was a two-page Letter in the journal “Nature,” which began “A Michelson two-beam interferometer can be used for both refractive index determination and absorption by measuring the shift of the achromatic fringe when a specimen of known thickness is placed in one arm of the interferometer.” This was the birth of “dispersive Fourier-transform spectroscopy” (DFTS), which has proved to be a powerful

technique for studying the optical constants of solids, liquids, and gases in the IR and THz regions [74]. Even today, when much spectroscopic work is performed with TDS systems, considerable routine and innovative research is still performed with DFTS instruments. Credit must also be given to E. E. Bell, who reported independently on the advantages of using both arms of a Michelson interferometer to provide absorption and refractive index information [75]. In 1969, D. H. Martin and E. Puplett introduced a polarizing version of the Michelson interferometer that has become widely used in the THz region [76].

Grating spectroscopy was still widely used in this decade because obtaining spectra from interferograms was so time-consuming, although the use of a fast Fourier-transform algorithm introduced by J. W. Cooley and J. Tukey had reduced the time required by some two orders of magnitude [77]. A specific problem of grating systems is the need to suppress higher order radiation which, with thermal sources, is at a much stronger level than the first order. This difficulty was to a large extent overcome by the use of capacitive grid filters by R. Ulrich in 1967 [78]. These are described in Sect. 3.5. Another very helpful discovery was that the polymer TPX is transparent over much of the FIR [79]. This is a hard plastic, particularly useful for windows and lenses, with the great advantage of having virtually the same refractive index in both the visible and THz spectral regions.

The year 1968 saw the birth of time-domain spectroscopy (TDS), although in the microwave region, with the publication of a paper entitled “BroadBand Microwave Transmission Characteristics from a Single Measurement Transient Response” [80].

1.2.6 The Years 1970–1980

The year 1970 saw the arrival of one of the most useful laser sources, the optically excited THz gas laser [81]. These now produce hundreds of cw and many thousands of pulsed wavelengths throughout the THz region. On the strongest cw lines output powers of above 100 mW are available from commercial instruments with pulsed powers of several kW.

Also in 1970 carbon monoxide (CO) was discovered for the first time in the interstellar medium with a Schottky-diode heterodyne receiver operating at 115 GHz [82]. Seven years later, T. G. Phillips et al. made the first real THz detection of an interstellar molecule [83]. They detected CO at 0.345 THz with a heterodyne receiver based on an InSb hot electron bolometric mixer. The 1970s were the pioneering years for astronomy at THz frequencies, mainly because of significant progress in receivers and because high altitude observatories became available. Most notably, the Kuiper Airborne Observatory (KAO) started its operation in 1974.

The first THz imaging experiments for other than astronomical applications were reported by D. H. Barker et al. in 1975. Already at that time the authors pointed out that “an FIR imaging system could be developed for industrial, military, law enforcement, and medical applications in the next few years” [84].

Early in this decade the first THz TDS results were obtained by the Paul Richards group, using a mode-locked Nd:glass laser and lithium niobate as an electro-optic (EO) crystal [85]. Due to the comparatively long pulse length of about 2 ps the shortest wavelengths produced were about 700 μm . Similar results were obtained independently by a Japanese group [86].

There was further significant progress toward improved detectors in this decade. Detection using metal–semiconductor contacts, so-called cat’s whiskers, dated back to Bose’s research in the 1890s. These are also the devices that Gordy used to provide harmonics of the fundamental frequency. Now described as Schottky diodes, these were already familiar devices for both direct detection and as mixers in microwave and mmW heterodyne systems. It was realized that if these diodes could be designed to operate at higher frequencies, they would have the great advantage, compared with other fast THz detectors, of operating at room temperature. Because various relatively high-power sources were also becoming available, there was the opportunity to extend microwave-type heterodyne systems to higher frequencies. A very useful review of the remarkable progress in Schottky diode optimization was written in 1980 [87]. As Gordy had shown, Schottky-type diodes are not only useful for heterodyne detection systems. Due to their nonlinear characteristic they are valuable as harmonic generators. Gunn and IMPATT sources were producing significant power levels in the mmW region and the use of these with Schottky structures began to provide significant harmonic power deep into the THz range.

A main reason for developing room-temperature systems for the THz region was that various platforms were becoming available for both astronomy and upper atmospheric research. However, video detection remained important and it was well known that the background noise that limited the ultimate detectivity of extrinsic photoconductive detectors, such as Ge:Ga, when used in a 300 K environment, would be greatly reduced in space applications. Intensive research on improving Ge:Ga detectors led to an improvement of four orders of magnitude in their performance [88]. Another He cooled detector that dates from this era, and was to become very important, particularly for THz heterodyne systems, is the superconductor–insulator–superconductor mixer (SIS) [89, 90].

1.2.7 The Years 1980–1990

The improvements in THz technology over the previous thirty years had led to a widespread recognition of the usefulness of this spectral region in many disciplines. Developments in computer technology had made FTS systems the choice for spectroscopy throughout the IR and THz regions. But, with an increasing interest in astronomical and upper atmosphere research, it was realized that heterodyne systems were required for these and other applications to achieve the necessary resolving power. This can be understood by looking at a paper published in 1984, which reported a balloon-borne FT interferometer study of stratospheric emission at ~ 40 km altitude [91]. At the shortest wavelength observed of 110 μm the resolving

power was $\sim 27,000$, falling to just over 2,000 at the longest wavelength of 1.4 mm. This instrument had a path length difference of up to 1.5 m between the fixed and moving mirror and was approaching a realistic limit in the resolving power of an FTS instrument. Despite this impressive achievement, the resolving power is not sufficient to determine the true lineshape of a molecular line.

The need for higher resolving powers was pointed out in a review paper outlining the state-of-the-art of diagnostics on interstellar plasmas [92]. What was required was the extension of the high resolution of heterodyne systems familiar in the microwave region up to THz frequencies. Until that time the only heterodyne studies had been with the “Putley” InSb electron bolometer, which has a very limited bandwidth. In the short-mm region rectifying diodes, usually Schottky devices were normally used as the mixer with electronic sources providing the LO power. The main problem in reaching higher frequencies is that losses increase, resulting in a requirement for more local oscillator (LO) power, typically in the mW range. At that time, the only practical cw sources available for frequencies above ~ 0.5 THz were optically excited gas lasers. Although these only produce specific wavelengths, a successful heterodyne detection of an interstellar CO line at $434\text{ }\mu\text{m}$ was achieved in 1981 [93] by mixing the CO line with a laser line of similar wavelength. Although optically excited gas lasers are normally bulky, a sufficiently compact version for use in the Kuiper Airborne Observatory (KAO) was designed and used for astronomical studies [94, 95]. The KAO was one of a number of platforms that were becoming available for both astronomy and upper atmospheric research, and for these and other applications the search for better detectors and mixing devices, as well as for improved sources to act as LOs, was intensified. The 1980s also saw the first IR/THz space-borne observatory. The Infrared Astronomical Satellite (IRAS), launched in 1983, was the first observatory to perform an all-sky survey at IR wavelengths, including two THz bands around 60 and $100\text{ }\mu\text{m}$. This was the first of a very successful series of space-borne IR/THz observatories.

As already mentioned, conventional extrinsic Ge detectors had been hugely improved and were becoming widely used in astronomy and atmospheric studies. However, there was a particular problem with these detectors when used in space applications. This was because excess noise was produced when high-energy photons struck the detectors. This disadvantage was to a large extent overcome in the 7–10 THz region by the invention of blocked-impurity band detectors [96].

A fundamental source covering the frequency range from 1 THz to above 4 THz, invented in Russia and Japan in this decade, was the p-type Ge laser [97, 98]. Operating at or near liquid helium temperature, this is a useful tunable pulsed source that can also be operated in near cw mode [99]. Ge lasers are electrically excited and, as the power requirement, especially for small versions, is quite low, these can be conveniently packaged with a mechanical cooling device [100].

The first free electron laser (FEL) operated in the early 1970s and by 1984 Elias and his colleagues [101] had constructed a laser operating between 0.3 and 0.77 THz, with a peak power of 10 kW, as a user facility at the University of California at Santa Barbara. One of its main application areas was in biology, including photobiology. Since then a number of THz laser user facilities have been

installed in France, Germany, Italy, Russia, The Netherlands, and the USA, which are providing very valuable results in a wide range of disciplines. The tunable high-power and short-pulse output of FELs means that they are particularly useful for pump-probe experiments and time-resolved spectroscopy.

Alternative facilities for THz research are the IR ports of certain synchrotrons [102]. These provide a significantly higher average source brightness than an arc lamp in the 0.3–30 THz region and the ps pulse structure is useful for a number of applications. Stable coherent synchrotron radiation (CSR), first observed in 2002, has produced a power increase of 10^5 over the 0.3–1 THz range and is an exciting development for spectroscopy [103].

The dramatic event of this decade can almost be described as the “rediscovery” of the FIR region of the electromagnetic spectrum. With the advent of high-power, very short-pulse length near-IR lasers, the pioneering research of Nicholson in the microwave region, and Richards and Yajima, with their colleagues, in the FIR, came to fruition. The essential features of this system of THz research are fairly simple. When a very short pulse, typically 50 fs or less, of near-IR radiation falls on a photoconductor, THz pulses are generated within the photoconductor, which then radiates. The highest frequency produced depends on both the length of the laser pulse and the mobility of the electrons in the photoconductor. An alternative arrangement is to use an EO crystal where the ultra-short laser pulse changes the electrical polarization and causes THz radiation to be emitted. In this case, the upper frequency produced only depends on the length of the laser pulse, as no charge carriers are involved. In practice, the essential difference between the two systems is that photoconductive devices produce higher power but are less efficient above 3 THz. EO crystals provide less power, in comparison with photoconductive devices with the same excitation power, but can reach to above 100 THz. In both cases, the spectrum is obtained from the Fourier-transform of the emitted pulse.

In early experiments, the power levels produced were quite low. However, in 1984 Auston and his colleagues [104] introduced what became known as the “Auston switch,” to produce higher powers, and used these to study the transmission and absorption of various materials. This paper created widespread interest and, as mentioned previously, some of the early accounts seemed to suggest that this was an entirely new field of research.

Just before the close of this decade the first results of a new detector that was to become widely used as a THz mixer were published by G. N. Goltsman and coworkers. This was the hot electron bolometer (HEB), remarkable for the fact that it is a thermal detector [105]. Essentially, this relies on the same mechanism as the first helium-cooled detectors, reported in 1942 [34], as it operates in the narrow temperature range that marks the transition from the normal to the superconducting state of a superconductor. The particular advantage of this bolometer made of a thin NbN film is its combination of a small heat capacity with high heat conductivity, giving a time constant of approximately 40 ps for small devices. This is equivalent to a bandwidth of several GHz, thus making it very useful as a mixer. Furthermore, as it is a fast thermal detector it is useful throughout the IR and THz regions.

1.3 Reasons for Increased Interest

From being a rather specialist sector of the electromagnetic spectrum for research, the THz region has become of great importance in a wide range of disciplines. These include many aspects of space research, astronomy, atmospheric studies, biology, medicine, plasma physics, chemistry, nondestructive testing, and process control. There are also a number of important defense and security applications. For all of these applications, imaging as well as spectroscopic systems have been developed, and there are continuous, intense efforts around the world to improve their performance and capabilities.

A major reason for the increased interest in THz research comes from astronomy. The THz part of the electromagnetic spectrum possesses an amazing scientific potential for spectroscopy and imaging. Many absorption and emission lines of the important astrophysical and astrochemical molecules and atoms occur in the THz frequency region. By 2011 approximately 150 molecules and atoms have been detected in space. There are fairly simple molecules, such as CO, but molecules consisting of ten or more atoms have also been detected. Besides their existence and abundance, valuable information about the physical conditions such as temperature, density, and dynamics of the observed astronomical object can be obtained from spectroscopy. Many galaxies or interstellar clouds emit most of their energy at THz frequencies, either as broadband blackbody-like emission or as line emission from molecules and atoms. Imaging and photometry allow the study of the structure and morphology of these objects. Ground-based telescopes at many locations around the world, as well as the space-borne THz observatories IRAS, ISO, SWAS, Odin, Akari, and Spitzer, have contributed significantly to the exploration of the THz universe. At the time of writing this book ESA's Herschel Space Observatory is in operation and will revolutionize our view of the THz universe.

It is the demand from these observatories that has driven forward the development of detectors, radiation sources, and spectrometers. The requirement for very sensitive detectors and large format detector arrays has led, for example, to the development of Ge:Ga photoconductive detector arrays with hundreds of elements, and superconducting composite bolometer arrays with thousands of elements. Amazingly, the sensitivity of these detectors is background noise limited. For heterodyne spectrometers SIS and HEB mixers approach the quantum noise limit, and small heterodyne arrays are now operational at several telescopes. The development of local oscillator (LO) sources has made all-solid-state multiplier-based THz sources with μW to mW output power available up to 2 THz. New observatories such as the Stratospheric Observatory for Infrared Astronomy (SOFIA), or the Atacama Large Millimeter Array (ALMA), which comprises an array of 66 giant 12-meter and 7-meter diameter telescopes for mmW and submm/THz observations, will continue to provide a strong incentive for the development of THz receivers and spectrometers.

Imaging has become an important application in the THz region. Some of the initial excitement of the early TDS research was the "discovery" that a variety

of materials, including clothing and packaging materials, which are opaque in the visible and near-IR regions, are reasonably transparent over much of the THz range. Imaging had already been performed using conventional THz techniques [106] but TDS imaging has found wide application over diverse disciplines, since the first paper was published in 1995 [107]. This also has the advantage of providing three-dimensional imaging. Because THz TDS does not require cooling, while providing large spectral coverage and high sensitivity at the same time, this technique rapidly spread and found new applications in research as well as paving the way for commercial uses. Along with THz TDS, breakthroughs in component development have contributed to the rapid development of the field since 1990. Most notable has been the appearance of THz quantum-cascade lasers (QCLs) in 2002 [108], which now cover a large part of the THz region with cw and frequency-tunable radiation. In addition, these only require modest cooling and provide mWs of output power. Such systems have been employed for a large variety of proof-of-principle experiments [109] for applications in biomedicine, nondestructive testing, security, and other fields such as art conservation [110].

Fourier-transform spectroscopy is commercially available and is the workhorse in chemistry for spectroscopy of gases, liquids, and solids. THz spectroscopy with these spectrometers is widely used in chemistry not only with pyroelectric detectors, but also with liquid helium-cooled bolometers. Early research on astrochemistry activated advances in gas spectroscopy of simple molecules in the laboratory, but current increased interest in THz spectroscopy is related to the extensive computer power now available. This allows calculation and simulation of an expanding number of molecules with many atoms, weak interaction forces related to the structure of biomolecules like DNA and proteins, and interaction of molecules with a large number of solvent molecules, for example, for the investigation of chemical reactions in solvents such as water.

Both large mass and weak interaction forces lead, independently, to low vibration frequencies located in the THz frequency range. Also, collective modes in crystal lattices consisting of many, and even assorted, molecules, as well as liquids, show THz spectra and complex THz responses, respectively. Polymorphs are very difficult to separate by IR spectroscopy, but show significantly different spectral signatures in the THz frequency range. High signal-to-noise ratios, now available with room temperature THz spectrometers, in contrast to THz Fourier-transform spectrometers, may lead to applications in the pharmaceutical industry to distinguish generic drugs, and to support patent applications on newly developed pharmaceuticals. Molecular systems, and reactions with surfaces and interfaces, are of major interest for the development of new catalytic and energy efficient processes, as well as in energy storage, such as batteries, which are relevant for future electromobility. THz spectroscopy can investigate the interaction of molecules with surfaces because the heavy mass of the surface lowers vibrational frequencies into the THz range.

While common chemical reactions are performed in solvents, the reaction formula is typically written down neglecting solvent molecules. It has recently been recognized that the solvent can play a more active role in many chemical reactions, beyond just controlling the diffusion of molecules for the reaction. Especially for

biomolecules, water, the solvent of life, actively participates in interactions, and these can be probed on ps and sub-ps timescales corresponding to THz frequencies. A current large new class of solvents, ionic liquids, is also being considered for industrial and research applications. Since many macroscopic properties, for example viscosity, may be related to THz responses, it is conceivable that THz spectroscopy may also help to develop designer fluids.

THz TDS will, in general, have a strong impact in chemistry, medicine, and biology. With the development of sub-ps mode-locked dye lasers in 1972 by Ippen and Shank [111] and subsequently, in 1986, the development of titanium-sapphire (Ti:Sa) lasers by Moulton [112], it did not take long for chemists to use such lasers to investigate chemical reactions. Short sub-ps pulses and pump-probe techniques allow the study of kinetics and dynamics of chemical reactions, which range from femtoseconds via milliseconds up to seconds. Since misfolding of proteins triggers a variety of illnesses, unraveling the process of protein folding, which takes place in many subsequent stages from femtoseconds to seconds, will be of prime medical interest.

THz pulses also permit the study of processes which are far from equilibrium, because the time scales involved are much shorter than thermal time scales. Plasma reactions for research and industrial applications can be studied in detail, with the added advantage that the thermal background of the typically hot plasma is negligible when coherent detection systems are used. Many processes, for example near equidistant rotational lines of molecules, are coherently locked for short time scales after excitation, and THz short pulses can provide coherence and dephasing times relevant to the understanding of molecular processes. Short pulses are also very relevant to condensed-matter research.

Although optical and electronic components to manipulate amplitude and phase of short pulses are readily available in the optical range, such components are in their infancy in the THz range. This, in part, has fuelled interest in systems based on plasmons, metamaterials, photonic band gap materials, and fast semiconductor switches, which may contribute to new components for THz pulse manipulation and shaping. Nuclear magnetic resonance is a standard technique for identifying chemicals, whilst in medicine the technique can be used to generate images in magnetic resonance imaging. Special pulses, pulse trains, and sequences of pulses allow a variety of applications and methods related to multi-dimensional spectroscopy. Similar techniques permeate into the THz field and may provide new insights into materials. Medical imaging is still in its infancy but, for example, the high absorption of water and low absorption of fatty tissue contribute to THz contrasts, and may lead to medical THz imaging augmenting the presently available infrared and ultrasound techniques.

The increasing need for extra bandwidth for communication may also impact heavily on THz technology and component prices, especially in the low THz frequency range where research applications are approaching 0.3 THz. However, the strong absorption of THz waves in air will restrict their use to local networks. This is beneficial because it allows an increased occupancy of frequency bands for spatially separated networks. Therefore, in-house and, in general, short-range (and

therefore locally secure) broadband communication is possible, without interference between neighboring networks in the same frequency band. Proximity initiated communication links are in service in some countries in the 0.1 THz range.

Many research centers, such as laser facilities, synchrotrons, and free electron lasers are increasing the number of THz sources and beam lines for external users, but much of the increased interest stems from the close connection between electrons and THz incoherent and coherent radiation. THz spectroscopy is an excellent tool for the diagnostics of charged particle beams [113, 114] and may even revive interest in electron systems and tubes, although, most probably in micro- and nanoscale devices [115], currently being investigated for new types of plasma displays.

This close relationship of electrons and THz radiation is also reflected in the use of THz spectroscopy to investigate charges in semiconductors, and to measure conduction without the need to apply physical contacts and electrodes. The latter is especially relevant, because an increasing number of nanostructured and nanoscale materials are being applied to devices, ranging from fast electronics to systems converting, for example, solar energy into electrons, which are key fields of interest.

The increased interest in THz techniques is also related to their broader availability and, in many situations room temperature operation, which allows users from different disciplines to incorporate THz technology into their research or applications. This is in part a consequence of enabling technologies such as ultra-fast optics, nanopositioning, fast electronics for regulation and stabilization, advanced material processing, increased speed in simulations, and computer power. The proliferation of THz techniques from research into everyday life is benefiting from declining prices of commercially available components such as femtosecond fiber lasers, liquid-cryogen free cooling systems requiring little maintenance, sub-nanometer precise piezo-positioning systems, and off-the-shelf GHz electronics.

To summarize, there has been a push resulting from major THz technological breakthroughs, leading to significantly improved performance, greater ease of use and, last but not least, more affordable systems. This is being accompanied by a pull from both research and “real-world” applications. The combination of both push and pull has stimulated the dramatic increase in THz research and development that has occurred since about 1990, and will continue in the foreseeable future.

Chapter 2

Optical Principles at Terahertz Frequencies

Waves at THz frequencies are part of the electromagnetic spectrum between the IR and the mmW frequency range. As would be expected, some properties are similar to visible and infrared waves, while others resemble propagation in the mmW region. This chapter prepares the ground for the various definitions, figure-of-merits and equations used to describe interactions of electromagnetic waves with materials and components.

THz components mostly respond linearly to the electric field amplitude, leading to regular absorption, dispersion, or reflection which is used, for example, to guide light through optical systems (Chap. 3), in detection (Chap. 5), or for the investigation of samples, in spectroscopy (Chap. 6). The description of the electric field distribution in space requires Gaussian beams and beam transformation. Gaussian optics are also required for the design of lasers and analysis of the properties of their beams (Chap. 4). The introduction of high-power and short fs-pulse lasers in the 1970s, followed by the development of THz time domain spectroscopy and many other pulse detection techniques, have the potential to drive materials into nonlinear regimes. The dielectric function then depends on higher orders of the electric field to describe the nonlinearity (Sect. 2.5). Near-field microscopy also relies on nonlinear interactions (Chap. 7). Near-field properties support the function of metamaterials (Sect. 3.17), while both theory and experiments on metamaterials also stress the importance of interactions between the magnetic component of electromagnetic waves with materials and devices. For these materials, the relative magnetic permeability μ not only deviates from the vacuum value of one, but also can be negative. Most equations used within this book treat the effect of THz components on the electric field in space and time in a similar fashion to common text books on electromagnetic theory, for which the assumption is made that μ is approximately one, which is true for most regular materials.

Free space THz wave propagation is best described by using Gaussian wave optics or, in short, Gaussian optics. Furthermore, the term *quasi-optics* has been coined for the mmW to THz frequency region. This term acknowledges the hybrid role of THz radiation, propagated in free space using optical elements or in waveguides as used in radio technology. The textbook by P.F. Goldsmith

provides an in-depth treatment of quasi-optics with Gaussian wave optics [116]. The wavelengths for a frequency range from 0.3 to 10 THz of 1–0.03 mm are often significant compared to the physical dimensions of common optics. For example, a typical tolerance for machining parts in a workshop during the manufacturing process is 0.1 mm, corresponding to 3 THz. As an illustration, a vacuum window for a detector cryostat requires that the thickness provides sufficient physical strength to seal the vacuum against normal pressure, and has maximum transmission with minimal etalon effects. Therefore, the size and optical path length of each optical component are significant and deviations from geometric optics have to be considered.

Approximations, typically valid in the visible range in which the wavelength λ is very small relative to a physical dimension d ($\lambda/d \ll 1$), need to be reinvestigated to avoid aberrations, ripple formation in the wavefront, and the effects of apertures. In addition, a size of more than 0.03 mm is challenging for micromachining, and rather difficult when the growth of active layers for quantum cascade lasers and quantum wells is considered. In such devices the active layer should approach the size of the wavelength in order to reduce beam divergence and avoid asymmetric mode profiles, which complicate beam coupling to detectors or waveguides.

In electronics, electromagnetic waves are guided by wires, but with increasing frequency waves tend to radiate into free space. In microwave electronics and mmW propagation, waveguides, strip lines, coaxial cables, and other structures are used. With the decreasing wavelength corresponding to the THz region it becomes increasingly difficult to fabricate suitable structures. Consequently, the wave is often emitted to and propagated into free space. Antenna structures and horns couple guided waves to free space, relying on solutions in radio technology. A further difficulty is the increasing impact of the skin effect and decreasing skin depth with increasing frequency, which shifts the field energy into the surface layer of structures. Antennas, fabricated by lithography, can be considered as two-dimensional structures for radio waves because λ is very large compared to the thickness of the structure d ($\lambda/d \gg 1$). For shorter THz wavelengths, these structures need to be viewed as three-dimensional objects or three-dimensional surface objects, if the skin depth decreases with increasing frequency.

2.1 Fundamentals of Gaussian Beams

Gaussian beams best describe radiation propagation at THz frequencies. A Gaussian intensity distribution determines a Gaussian beam profile. In its simplest form, it is radially symmetric. If the phase over the radius r is uniform, the Gaussian intensity reproduces itself under Fourier transformation, a unique property of such beams. The electric field of the Gaussian beam follows as

$$E = E_0 \exp\left(-\frac{r^2}{w^2}\right). \quad (2.1)$$

With the full solution of the Fresnel integral, the Gaussian distribution is present at *every* point along the propagation direction in an optical system [117]. Focusing or defocusing does not change the Gaussian character, only its size. The intensity is given by

$$I(r) = |E|^2 = E^* E = E_0^* E_0 \exp\left(-2\frac{r^2}{w^2}\right) = I_0 \exp\left(-2\frac{r^2}{w^2}\right). \quad (2.2)$$

Lasers often oscillate naturally in a Gaussian mode, since the electromagnetic field remains self-reproducing after repeated oscillation in a resonator. Multiplication of the fundamental Gaussian mode with certain polynomials also leads to self-reproducing field distributions called higher order transverse modes. The Gaussian beam waist radius w is defined to assign a physical dimension to the infinite field distribution. The intensity drops to $(1/e)^2$ or 0.135 of the intensity at the waist radius w from the optical axis. The power within a circle of radius r can be calculated by integration from 0 to r giving

$$P(r) = P_\infty \left(1 - \exp\left(-2\frac{r^2}{w^2}\right)\right). \quad (2.3)$$

The intensity curve is reproduced by introducing the total power of the beam P_∞ :

$$P_\infty = \frac{\pi w^2}{2} I_0. \quad (2.4)$$

Figure 2.1 illustrates (2.1) and (2.2). Inserting an aperture can severely affect a Gaussian beam. An aperture should be chosen with a diameter exceeding a value of $4.6w$ to minimize disturbances and far-field ripple formation. Small apertures will result in intensity structures in the far-field. Hence, the design of an optical system requires the choice of, for example, mirrors that are large enough, compared to the propagating waist diameter, to prevent such far-field artefacts and beam distortion.

The propagation of the intensity profile for Gaussian beams is easy to follow. The transverse distribution is Gaussian at every point, but the radius of the Gaussian beam and the radius of the wavefront curvature can change as displayed in Fig. 2.2. The curvature is infinite at the position of the beam waist minimum w_0 . This position, at which a flat phase front is found, can correspond, for example, to a flat outcoupling mirror of a laser in a confocal laser design. The following equations describe the Gaussian beam radius $w(z)$ and the radius of curvature $R(z)$ of the wavefront (Fig. 2.3). The beam can be described by these two parameters alone, using the single parameter z_R , the Rayleigh range:

$$z_R = \frac{\pi w_0^2}{\lambda}, \quad (2.5)$$

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R}\right)^2}, \quad (2.6)$$

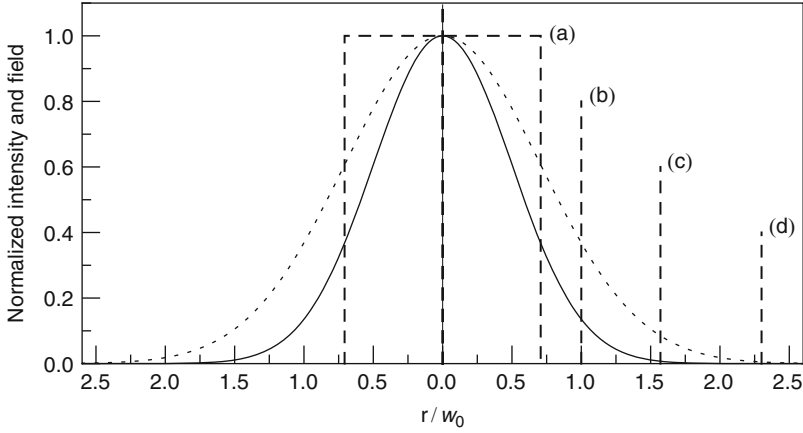


Fig. 2.1 Intensity I (solid line) and field E (dotted line) of a Gaussian beam as a function of radius r and minimum waist radius w_0 . Dashed lines indicate characteristic widths. (a) The equivalent top hat profile of a cylindrical beam with the same peak and total power transmission of $\sim 63\%$ has a diameter of $\sqrt{2}w_0$. (b) The field and intensity (transmitted power $\sim 87\%$) drop to a level of $1/e$ and $(1/e)^2$, respectively. (c) For an aperture of radius $\pi w_0/2$ the power transmission exceeds 99% with a far-field ripple of $\pm 17\%$. (d) At a radius of $2.3w_0$, the ripple reduces to $\pm 1\%$ and the power transmission exceeds 99.99%

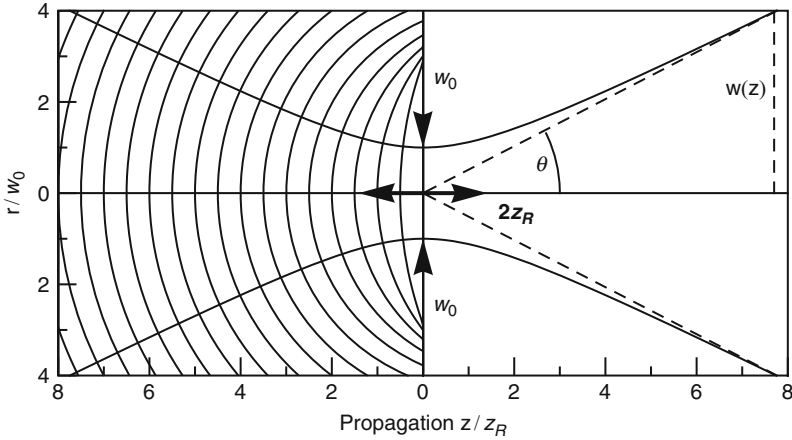


Fig. 2.2 Gaussian $(1/e)^2$ intensity distribution as a function of the dimensionless ratio between the beam propagation z and the Rayleigh range z_R . The minimum waist radius w_0 is indicated at $z = 0$, the solid line curves manifold corresponds to phase fronts and the solid line to the propagating waist diameter. The angle θ describes the opening angle of the beam (dashed lines) as though it is emerging from a point source

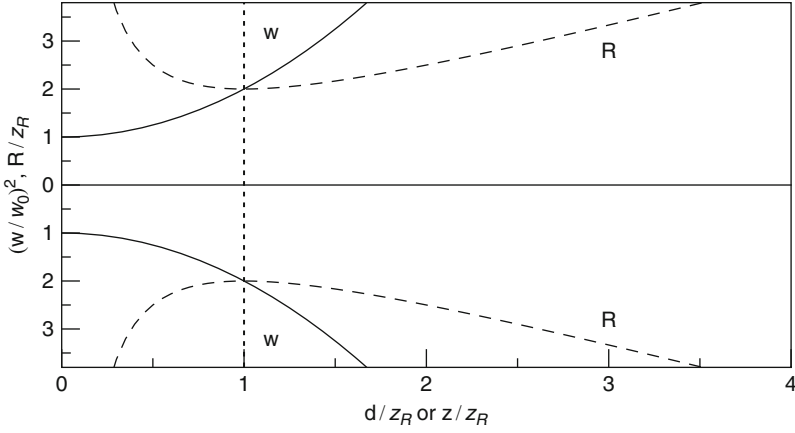


Fig. 2.3 Curvature radius R (dashed line) and waist w (solid line) in dimensionless units extracted from the real and imaginary part of the complex beam parameter q ((2.6) and (2.7)) as a function of the ratio z and z_R . Note that z can also be replaced by the propagation distance d according to (2.22) and (2.23). The dotted line indicates the depth of focus and the special role of the Rayleigh range z_R (2.14)

$$R(z) = z \left(1 + \left(\frac{z_R}{z} \right)^2 \right). \quad (2.7)$$

The wavefront is flat at the position of the minimum waist radius: $R(z \rightarrow 0) \rightarrow \infty$. The complex radius of curvature q is defined by the waist radius w and the radius of curvature R as

$$\frac{1}{q(z)} = \frac{1}{R(z)} - i \frac{\lambda}{\pi w^2(z)}. \quad (2.8)$$

It follows that q at the position of the minimum waist $w(z = 0) = w_0$ for infinite radius of curvature R has only an imaginary part equal to the Rayleigh range z_R .

At a distance, several times the Rayleigh range, the change of the Gaussian beam intensity can be approximated by a linear relationship, because the propagating waist radius $w(z)$ grows nearly linearly with z (Fig. 2.2). From a far distance the Gaussian beam can be viewed as a point source originating from the center of the minimum waist. The opening angle θ determines the beam divergence defined in a sphere due to the beam curvature. This angle θ is the ratio $w(z)/z$ at which the intensity drops to $(1/e)^2$ and is given by

$$\tan(\theta) = \frac{\lambda}{\pi w_0} = \frac{w_0}{z_R}. \quad (2.9)$$

For small opening angles θ , the tangent can be approximated by the angle itself. The beam parameter product, often used to describe laser beam profiles, then follows as

$$\theta_{w_0} \approx \frac{\lambda}{\pi}. \quad (2.10)$$

This product depends only upon the wavelength and, importantly, it does not change if the beam passes through linear optical systems. For real lasers, the desired Gaussian beam with minimal divergence is difficult to obtain, because of the presence of higher order modes, or an inhomogeneous gain medium. Higher transverse modes lead simultaneously to an increase of the waist radius and divergence, which is accounted for by a factor M increasing the fundamental Gaussian mode, the transverse electric and magnetic TEM₀₀ mode. Equation (2.10) is modified for real lasers to

$$\theta_{w_0} = M^2 \frac{\lambda}{\pi}. \quad (2.11)$$

As a measure of the beam quality, the beam propagation factor $k = 1/M^2$ is used. The values, k and M^2 , are both equal to one for ideal Gaussian beams, but k is smaller than one for larger beam divergence and lower beam quality. These parameters are determined experimentally by using a focusing lens, followed by the measurement of the beam caustic. A beam with a smaller M^2 leads to a smaller focus spot and smaller divergence after the lens.

In the case of Gaussian beam propagation over long distances (several times the Rayleigh range), for example, because of focusing elements with large focal lengths, a relation between Gaussian beam parameters, opening angle, and focal length similar to geometric optics can be made (2.9). In geometric optics, the full viewing angle $\theta' = 2\theta$ corresponding to the lens diameter is used. For a focusing lens, the angle is calculated from the focal length f and the lens diameter D as

$$\theta' \approx \frac{D}{f} = \frac{1}{f^\#}. \quad (2.12)$$

The photographic aperture value of a lens, also called relative aperture or speed of the lens, is given by f/D which is abbreviated to the f -number $f^\#$. It follows that the minimum radius of the beam waist is given by

$$w_0 = \frac{2\lambda}{\pi} \frac{f}{D} = \frac{2\lambda}{\pi} f^\#. \quad (2.13)$$

The depth of focus (DOF) defines the region for which the beam expands to $\sqrt{2}w_0$ and can be calculated by

$$\text{DOF} = \frac{8\lambda}{\pi} (f^\#)^2 = \frac{\pi w_0^2}{\lambda/2} = 2z_R. \quad (2.14)$$

The DOF is twice as large as the value of the so-called confocal parameter, confocal distance, or Rayleigh range. The Rayleigh range z_R is a very useful quantity, since it describes a length scale at which the Gaussian beam propagation changes from a near-field to a far-field dominated region, or from Fresnel to Fraunhofer diffraction. It also gives a length scale over which a beam is quite well collimated. Beyond this

distance, the beam divergence becomes significant. A further advantage is that many equations related to Gaussian beam propagation can be written in a compact form by using the Rayleigh range as a parameter.

2.2 Gaussian Beam Propagation

The previous section introduced Gaussian beams in a static form describing the intensity profile as a function of position in space. This section considers the propagation of Gaussian beams, and the transformation of beam parameters entering and leaving an optical system. Light ray propagation in geometric optics, for example through a lens, can be calculated by a matrix formalism. In this formalism, matrices (with 2×2 elements) represent different optical elements. In analogy with a ray in geometric optics, which can be described by ray parameters, Gaussian beams can be represented by beam parameters. Gaussian beams are described by the paraxial wave equation, a simplified version of the Helmholtz wave equation, using the following approximations leading to Gaussian beams as solutions of the wave equation. The approximations are that the variation of the field amplitude for a distance comparable to the wavelength along the direction of propagation is small (variation in time is neglected), and the axial variation is small in comparison with the variation perpendicular to this direction. These conditions imply that Gaussian beams originate from an extended source and that Gaussian fundamental beams represent wave propagation well for limited divergence. If focusing would result in a beam waist comparable to the wavelength or less corresponding to an angle of $\theta < 18^\circ$ as described by (2.9), then corrections are required beyond the paraxial limit. These corrections lead to a field amplitude distribution deviating from a Gaussian beam distribution. In the following paragraphs Gaussian beams are considered. Details on corrections to the paraxial approximation can be found in [116] and references within.

The propagation of a ray or beam through an optical element is mathematically equivalent to a linear matrix transformation of its parameters. Parameter transformation through each element is calculated by left-multiplication of each matrix, or by matrix-multiplication of all individual matrices, giving a total 2×2 matrix. The so-called ray transfer matrix of the system simulates a single propagation through the whole system. The incoming ray or beam is incident from the left according to convention. The parameters of such a matrix A , B , C , and D form a complex bi-linear transformation using left-side multiplication of the matrix:

$$\begin{pmatrix} r_{\text{out}} \\ \phi_{\text{out}} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_{\text{in}} \\ \phi_{\text{in}} \end{pmatrix} = \begin{pmatrix} Ar_{\text{in}} + B\phi_{\text{in}} \\ Cr_{\text{in}} + D\phi_{\text{in}} \end{pmatrix}. \quad (2.15)$$

In geometrical optics, the incoming ray is defined by its distance r_{in} to the axis of propagation along z with the ray slope ϕ_{in} (Fig. 2.4). The angle between the ray and the z -axis should be small so that the sine and the tangent of this angle can be

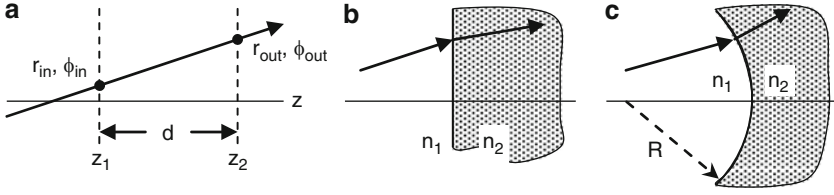


Fig. 2.4 (a) Propagation along a straight section d represented by matrix (2.20), (b) refraction at a dielectric interface separating two media with different refractive indices n_1 and n_2 , and (c) refraction at a dielectric spherical interface with a radius of curvature R . (b) and (c) are represented by matrix (2.24) with $R \rightarrow \infty$ and $R \rightarrow -R$ to account for the concave interface, respectively

approximated by the angle value itself. The resulting ray is defined by r_{out} and ϕ_{out} , respectively. The radius of curvature is then given as $R = r/\phi$, which leads to the expression

$$R_{out} = \frac{r_{out}}{\phi_{out}} = \frac{Ar_{in} + B\phi_{in}}{Cr_{in} + D\phi_{in}} = \frac{AR_{in} + B}{CR_{in} + D}. \quad (2.16)$$

For Gaussian beams, the radius of curvature is expressed by the complex radius of curvature q (2.8), which transforms (2.16) into

$$q_{out} = \frac{Aq_{in} + B}{Cq_{in} + D}. \quad (2.17)$$

The elements A , B , C , and D of the matrix are the same as in geometrical optics ray theory, allowing calculations with Gaussian beams using the same formalism. In most cases, the propagation of the waist is of interest, so that the input curvature q_{in} is taken to be equal to the beam waist expressed by the Rayleigh range as $q_{in} = iz_R$ (2.5). Evaluation of the real and imaginary part of (2.8) of the output q_{out} gives the respective curvature and waist radius as

$$R_{out} = \left(\Re \left(\frac{1}{q_{out}} \right) \right)^{-1}, \quad (2.18)$$

$$w_{out} = \sqrt{\frac{\lambda}{\pi} \left(\Im \left(-\frac{1}{q_{out}} \right) \right)^{-1}}. \quad (2.19)$$

In the following paragraphs, a few examples are discussed, concentrating on fundamental Gaussian beam optics [117–119].

Propagation through a homogeneous medium from a point P_1 to P_2 by a distance $d_{1 \rightarrow 2} = d_{12} = d$ transforms with the matrix

$$\mathbf{M} = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}. \quad (2.20)$$

In geometrical optics, the input ray $\mathbf{r}_{\text{in}} = (r_{\text{in}}, \phi_{\text{in}})$ defines the output ray $\mathbf{r}_{\text{out}} = (r_{\text{out}}, \phi_{\text{out}})$, with the distance from the propagation axis $r_{\text{out}} = r_{\text{in}} + d\phi_{\text{in}}$ and unchanged slope $\phi_{\text{out}} = \phi_{\text{in}}$. The curvature is $R_{\text{out}} = R_{\text{in}} + d$, respectively. The determinant $\det(\mathbf{M}) = AD - BC$ of the 2×2 matrix $\mathbf{M}$ is equal to one when there is no change in the refractive index. In analogy, the input waist $q_{\text{in}} = iz_{\text{R}}$ of a Gaussian beam inserted in (2.17) gives $q_{\text{out}} = q_{\text{in}} + d = iz_{\text{R}} + d$. Calculating the complex radius of curvature of the exiting Gaussian beam

$$\frac{1}{q_{\text{out}}} = \frac{d}{z_{\text{R}}^2 + d^2} - i \frac{z_{\text{R}}}{z_{\text{R}}^2 + d^2}, \quad (2.21)$$

and using (2.18) and (2.19) leads to

$$R_{\text{out}} = \left(\Re \left(\frac{1}{q_{\text{out}}} \right) \right)^{-1} = \frac{z_{\text{R}}^2 + d^2}{d} = d \left(1 + \left(\frac{z_{\text{R}}}{d} \right)^2 \right), \quad (2.22)$$

$$w_{\text{out}} = \sqrt{\frac{\lambda}{\pi} \left(\Im \left(-\frac{1}{q_{\text{out}}} \right) \right)^{-1}} = \sqrt{\frac{\lambda}{\pi} \frac{z_{\text{R}}^2 + d^2}{z_{\text{R}}}} = w_0 \sqrt{1 + \left(\frac{d}{z_{\text{R}}} \right)^2}. \quad (2.23)$$

These equations reproduce (2.6) and (2.7) if z and d are exchanged. Therefore, the matrix formalism can be used for Gaussian beam propagation, if the ray parameters are replaced by the complex radius of curvature of the incident Gaussian beam. A complex quasi-optical system for Gaussian beams can thus be represented by a simple sequence of matrices.

Refraction at any interface between two materials with different refractive indices n_1, n_2 and a curved surface with the radius of curvature R leads to

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ \frac{n_1 - n_2}{R n_2} & \frac{n_1}{n_2} \end{pmatrix}. \quad (2.24)$$

Here, the determinant of the matrix is $\det(\mathbf{M}) = n_1/n_2$, which is larger than 0 due to the refractive indices of normal materials. At most, two elements of $\mathbf{M}$ can be zero. If two elements are actually zero, they must be either diagonal terms or off-diagonal terms. For plane surfaces $R \rightarrow \infty$, (2.24) gives Snell's law for small angles.

The correspondence between physical properties and the matrix elements can be illustrated by setting different matrix elements sequentially to zero. If A is set to zero propagation parallel to the optical axis $\phi_{\text{in}} = 0$ is obtained, in which case r_{out} is always independent of the value of r_{in} and is equal to zero. This is the definition of a focal point in P_2 . If B is set to zero, the result is $r_{\text{out}} = Ar_{\text{in}}$. In this case, the two points P_1 and P_2 have an object–image relationship with the magnification being $m = r_{\text{out}}/r_{\text{in}} = A$. If C is set to zero, the equation simplifies to $\phi_{\text{out}} = D\phi_{\text{in}}$. This relation means that any parallel beam incident on the optical system leaves the system as a parallel beam, which is in essence a telescope focused at infinity. The angular magnification or power of a telescope can then be defined as

$p = \phi_{\text{out}}/\phi_{\text{in}} = D$. If D is set to zero, a focal point in P_1 is found, in analogy with the first case.

Lens-like systems can be studied by investigating the matrix of a thick lens. A thick lens is characterized by combining two curved surfaces R_1 and $-R_2$ (the negative sign introduces the opposite orientation of the curved surface) at a distance d separating three materials with refractive indices n_1 , n_2 , and n_3 by left-multiplying the three corresponding matrices:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ \frac{n_2 - n_3}{-R_2 n_3} & \frac{n_2}{n_3} \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{n_1 - n_2}{R_1 n_2} & \frac{n_1}{n_2} \end{pmatrix}. \quad (2.25)$$

The total matrix is rather complex

$$\mathbf{M} = \begin{pmatrix} 1 + \frac{d}{R_1} \frac{(n_1 - n_2)}{n_2} & d \frac{n_1}{n_2} \\ \frac{1}{R_1} (n_1 - n_2) \left(\frac{1}{n_3} - \frac{d}{R_2} \frac{n_2 - n_3}{n_2 n_3} \right) - \frac{1}{R_2} \frac{n_2 - n_3}{n_3} & \frac{n_1}{n_3} - \frac{d}{R_2} \frac{n_1 (n_2 - n_3)}{n_2 n_3} \end{pmatrix}, \quad (2.26)$$

but can be simplified for different types of lens. If the input and output beam are in air, the matrix simplifies due to the refractive indices $n_1 = n_3 \approx 1$. For a thin lens, the distance d is small and $d \ll R_1, R_2$. If the assumption of equal curvature is made ($R = R_1 = R_2$), the matrix is simplified to

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ \frac{2}{R} (1 - n_2) & 1 \end{pmatrix} \quad (2.27)$$

with the matrix element C always negative, because the refractive index of the lens is $n_2 > n_1 = 1$. Using the focal length f of a thin lens, the matrix element C can be identified with $-1/f$ giving $f = R/(2(n_2 - 1))$.

Various versions of the matrix in (2.26) are found for the plano-convex lens ($R_1 \rightarrow \infty$), convex-plano lens ($R_2 \rightarrow \infty$), and free-standing lens ($n_1 = n_3$). For the free-standing plane-parallel plate or etalon ($R_1 \rightarrow \infty$ and $R_2 \rightarrow \infty$) in air ($n_1 = n_3 \approx 1$), the following matrix results

$$\mathbf{M} = \begin{pmatrix} 1 & \frac{d}{n_2} \\ 0 & 1 \end{pmatrix}. \quad (2.28)$$

Note that the lens formulas are immediately transferable to spherical mirrors, which are curved surfaces determined by a radius R . For a mirror, the ray on the optical axis is reflected back onto itself. For a concave mirror, the focal length is $f = R/2$ while a convex mirror is described by a focal length of $f = -R/2$.

Aberrations, i.e., deviations from first-order Gaussian optics, are introduced by higher order powers of r and ϕ . First-order Gaussian optics prevails as long as the curvature R of a refracting interface is large compared to the ray parameters, i.e., $(r/R)^2 \ll 1$ and $\phi^2 \ll 1$. This relation already suggests the correct use of a

plano-convex lens for image formation with the object far away from the lens. If a beam, that is nearly parallel to the optical axis, is incident on the convex side of the lens, the aberrations are smaller since it is refracted twice, which introduces smaller changes in ϕ at each interface resulting in smaller aberrations. Chromatic aberrations are neglected here. They are normally compensated for by using achromats.

Telescopes are not only used in astronomy, but are also useful instruments for matching beam sizes to spectrometers and other instruments. As shown earlier, the telescope is described by setting the matrix element C to zero using $A = 1/p$, the angular magnification or power of the telescope. For any system in which the input beam and the output beam are in a medium with the same refractive index ($n_{\text{out}} = n_{\text{in}}$), the condition $\det(\mathbf{M}) = 1$ holds and the matrix simplifies to

$$\mathbf{M} = \begin{pmatrix} \frac{1}{p} & B \\ 0 & p \end{pmatrix}. \quad (2.29)$$

Points with object-image relationships are derived by propagating an object point at P_1 via a distance $d_{1 \rightarrow 2} = d_{12}$ to the telescope at P_2 , and further out via $d_{2 \rightarrow 3} = d_{23}$ to the image point at P_3 . Combination of the three matrices leads to

$$\mathbf{M} = \begin{pmatrix} 1 & d_{23} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{p} & B \\ 0 & p \end{pmatrix} \begin{pmatrix} 1 & d_{12} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{p} & B + \frac{d_{12}}{p} + d_{23}p \\ 0 & p \end{pmatrix}. \quad (2.30)$$

The object-image relation is fulfilled for $B^* = B + d_{12}/p + d_{23}p = 0$ or $d_{23} = -d_{12}/p^2 - B/p$. The following conclusions are derived from this equation:

- The lateral magnification m is the same for all points, which fulfill an object-image relationship, and is equal to the reciprocal of the telescope power $1/p$.
- If the object is moved by a small amount Δd_{12} , the image moves by $d_{12}/p^2 = m^2 d_{12}$ in the same direction. This is equivalent to a constant axial magnification $m_{\text{ax}} = m^2 = 1/p^2$.

The first result gives a simple derivation of the telescope power by determining the ratio of the objective aperture (entrance pupil) and its image (exit pupil).

One very practical and useful example is the ‘‘Gaussian beam telescope’’ [116]. Setting two focusing elements at a distance equal to the sum of their focal lengths $d_{12} = f_1 + f_2$ renders the waist transformation wavelength independent. The magnification simplifies to $m = f_2/f_1$, transforming the input waist radius $w_{0,\text{in}}$ and location d_{in} to the output waist radius $w_{0,\text{out}}$ and location d_{out} :

$$w_{0,\text{out}} = mw_{0,\text{in}}, \quad (2.31)$$

$$d_{\text{out}} = \frac{f_2}{f_1} \left(f_1 + f_2 - \frac{f_2}{f_1} d_{\text{in}} \right) = m(d_{12} - md_{\text{in}}) = md_{12} - m^2 d_{\text{in}}. \quad (2.32)$$

Gaussian beam telescopes with focusing mirrors are preferred because they are usable over a wide frequency range with good control of the beam parameters, while lenses normally have frequency-dependent reflection and transmission losses.

The magnification in a focusing system depends on the position of the object on the optical axis. A ray originating from a focal point is described by $f_1 = r_2/\phi_1$ and for a focal point on the exit side of the optical system by $f_2 = -r_1/\phi_2$. The sign has been introduced to provide a positive focal length. The resulting matrix is then

$$\mathbf{M} = \begin{pmatrix} 0 & f_1 \\ -\frac{1}{f_2} & 0 \end{pmatrix}. \quad (2.33)$$

Using further the notation z for the distances to the focal points, z_1 and z_2 , the matrix result is

$$\mathbf{M} = -\frac{1}{f_2} \begin{pmatrix} z_2 & z_1 z_2 - f_1 f_2 \\ 1 & z_1 \end{pmatrix} \quad (2.34)$$

with $\det(\mathbf{M}) = f_1/f_2 = n_1/n_2$. Setting B^* as above to zero gives a condition for points in an object-image relationship $z_1 z_2 = f_1 f_2$. For these points, the lateral magnification is $m = -z_2/f_2 = -f_1/z_1$.

2.3 Coupling of Gaussian Beams

A common challenge in THz experiments and applications is the efficient transport of the electromagnetic wave from the source to the detection system, even when neglecting absorbers along the optical path such as water vapor. Each component, source and detector, augmented with horns, waveguides, antennas or resonator structures will define a certain mode pattern, and only with an appropriate design can an azimuthally symmetric fundamental Gaussian mode be obtained. Efficient coupling or mode matching of source and detector or in spectroscopy, between source, sample, and detector, is often disregarded. This can lead to severe power loss as, for example, tilt and off-axis misalignment errors between two Gaussian beams reduce the power transfer exponentially. This is especially important if the source or detector is much smaller than the minimum beam waist radius. Gaussian beam coupling is described in detail in the textbook by Goldsmith [116]. It covers a variety of coupling schemes with components used in the microwave region, which are readily transferable to the THz region. For practical purposes, a few common causes for coupling losses due to misalignment of an experiment are described (Fig. 2.5):

- *Beams with waist mismatch:* coupling loss due to different beam radii and/or radii of curvature along the propagation axis due to a minimum waist position shift Δz and axial offset.
- *Tilted beams:* coupling loss due to misaligned beam axes by a tilt angle $\Delta\theta$.
- *Beams with lateral offset:* coupling loss due to axes misaligned in position by Δx , perpendicular to the propagation axis.

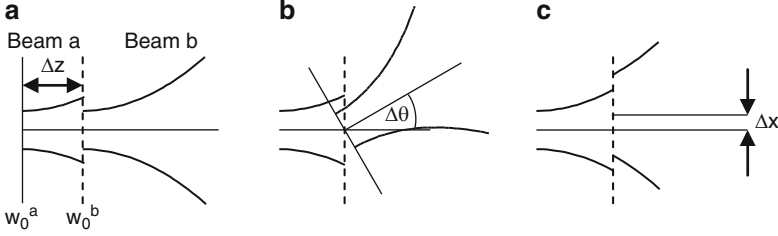


Fig. 2.5 Coupling loss as a result of misalignment: (a) beam radii and waist size mismatch by Δz , (b) angle tilt $\Delta\theta$, or (c) off-axis shift Δx for two misaligned beams a and b with minimum waist radius w_0^a and w_0^b , respectively. The *dashed lines* indicate the arbitrary reference plane for which the power coupling affected by misalignment is evaluated

The coupling loss between two beams can be quantified by calculating the overlap integral f_{ab} in two dimensions defined by the complex conjugate of the infinite electric field distribution E_a^* of the first beam and the infinite field distribution of the second beam E_b :

$$f_{ab} = \iint E_a^* E_b dA. \quad (2.35)$$

The integral is evaluated perpendicular to the propagation direction in a reference plane A . In the case of coinciding minimum waist locations, the reference plane can simply be taken in the plane of the minimum waist. For azimuthally symmetric fundamental Gaussian modes, the coupling is described by the following three equations for the corresponding cases in Fig. 2.5a–c.

$$f_{\Delta z} = \frac{4}{(x + 1/x)^2 + \chi^2 x^2}, \quad (2.36)$$

$$f_{\Delta\theta} = f_{\Delta z} \exp\left(-2\left(\frac{\Delta\theta}{\theta_0}\right)^2\right), \quad (2.37)$$

$$f_{\Delta x} = f_{\Delta z} \exp\left(-2\left(\frac{\Delta x}{x_0}\right)^2\right). \quad (2.38)$$

Misalignment errors in angle or position transfer into relative phase shifts of both beams a and b , resulting in power coupling loss. The mode mismatch factor $f_{\Delta z}$ in (2.36) is related to two quantities, $x = w_0^a/w_0^b = w^b/w^a$ and $\chi = \Delta z/z_R^a$. The value x is the ratio of the beam waist minima, which is the same as the inverse ratio of the propagating waist of both beams in a reference plane. The position misalignment for the minimum waist relative to the Rayleigh range z_R of the initial beam a along the propagation axis is given by χ . The power coupling is unity, i.e. without loss if the axial shift is zero, which is equal to $\chi = 0$, and if

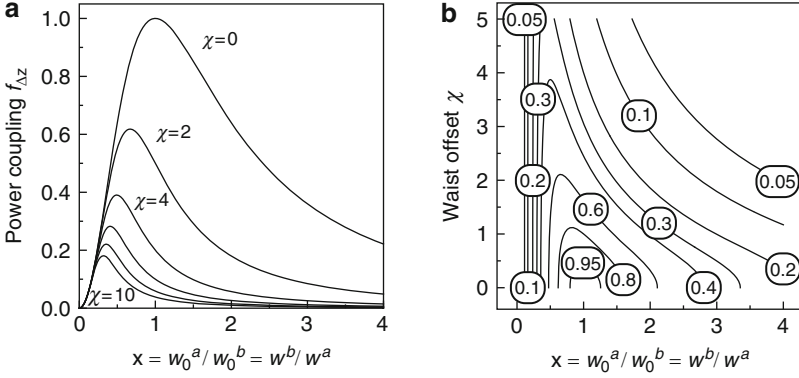


Fig. 2.6 (a) Power coupling $f_{\Delta z}$ of two beams a and b for a misalignment along the optical axis expressed by χ with values 0, 2, 4, 6, 8, and 10, in units of the Rayleigh range of beam a , as a function of the beam waist ratio x at an arbitrary reference plane. (b) Plot of contours for the power coupling $f_{\Delta z} = 0.05, \dots, 0.95$

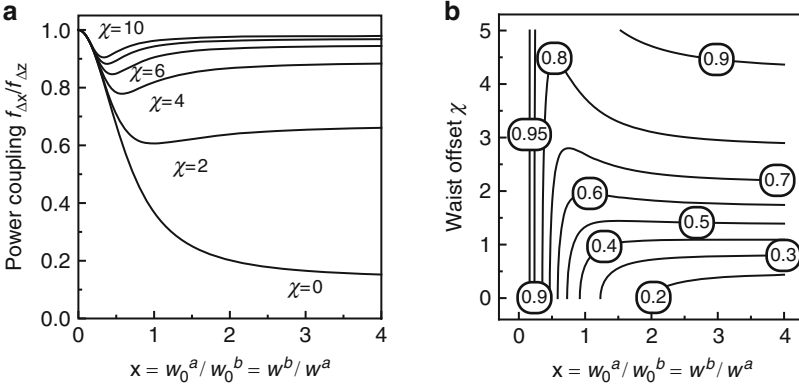


Fig. 2.7 (a) Power coupling factor $f_{\Delta x}/f_{\Delta z}$ of two beams a and b for an off-axis misalignment of $\Delta x = w_0^a$ excluding the additional factor $f_{\Delta z}$ due to axial misalignment. (b) Plot of contours for the power coupling $f_{\Delta x}/f_{\Delta z} = 0.2, \dots, 0.95$

both beam waist minima are the same or $x = 1$ (Fig. 2.6). Angle tilt and off-axis displacement errors reduce the power transfer further and need to be avoided. Off-axis misalignments and beam tilts lead to additional exponential reduction factors of the power coupling efficiency (Figs. 2.7 and 2.8). The characteristic parameters for the exponential reduction are

$$x_0 = w_0^a \sqrt{1 + \frac{1}{x^2} + \frac{\chi^2}{1 + 1/x^2}}, \quad (2.39)$$

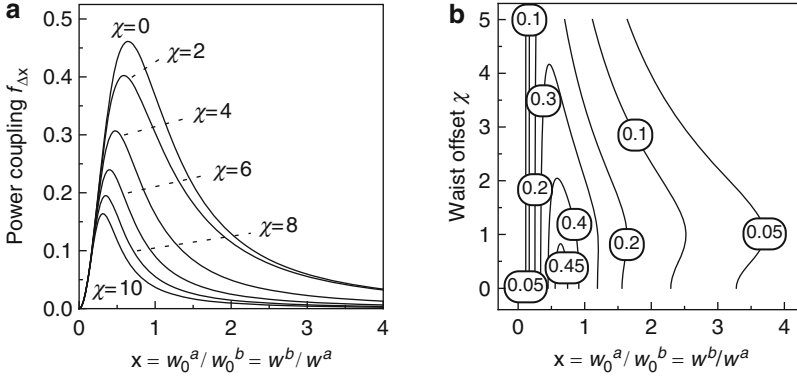


Fig. 2.8 (a) Total power coupling including axial and off-axis misalignment $f_{\Delta x}$ of two beams a and b for an off-axis misalignment of $\Delta x = w_0$. (b) Plot of contours for the power coupling $f_{\Delta x}$ as a function of the parameter x and χ

$$\theta_0 = \frac{\lambda}{\pi w^a} \sqrt{1 + \frac{1}{x^2} + \frac{\chi^2}{1 + 1/x^2}}. \quad (2.40)$$

The parameter x_0 scales with the minimum waist radius w_0^a and θ_0 with the inverse of the propagated waist radius w^a at the reference plane. For the special case of a tilt at the waist minimum, $\Delta z = 0$, the characteristic angle is a function of the far-field angles (2.9) and for small tilt the even simpler square sum of the far-field angles. This angle is also proportional to the inverse of the minimum waist radius:

$$\theta_0 = \frac{\lambda}{\pi} \sqrt{\frac{1}{(w_0^a)^2} + \frac{1}{(w_0^b)^2}} = \sqrt{\tan(\theta^a)^2 + \tan(\theta^b)^2}. \quad (2.41)$$

The off-axis shift Δx at the waist minimum, $\Delta z = 0$, is given by the square root of the sum of the waist radii:

$$x_0 = \sqrt{(w_0^a)^2 + (w_0^b)^2}. \quad (2.42)$$

The similarity of (2.39) and (2.40), and their special cases (2.41) and (2.42), indicates that x_0 and θ_0 are not independent. Similarly, both tilt and off-axis misalignment errors reduce the power coupling exponentially. Considering only the plane for the minimum waist radii and symmetric beams with equal waist radius, the tilt tolerance angle $\theta_0 = \sqrt{2}\lambda/(\pi w_0)$ and the lateral offset distance $x_0 = \sqrt{2}w_0$ multiply to

$$\theta_0 x_0 = 2\lambda/\pi. \quad (2.43)$$

This equation is related to the antenna theorem [120] and illustrates the inverse proportionality of angular divergence and beam size, and sensitivity to tilts and off-axis errors, respectively. A similar result can be obtained by rewriting (2.9), the Gaussian beam angle and spot size $2w_0$, as $\tan(\theta)2w_0 = 2\lambda/\pi$. Larger beam radii are less sensitive to off-axis misalignments, and smaller beam radii are less sensitive to tilts, due to their large far-field divergence.

From relation (2.43), an important conclusion follows: in a spectroscopic THz system efficient power coupling between source and detector is not sufficient on its own. If a sample is inserted into the spectroscopic system, it may be misaligned due to tilt and off-axis errors. Hence, the power coupling is further lowered. Power coupling loss may introduce systematic errors independent of the sample properties if several samples need to be measured and positioned in the sample area. Optimizing the waist diameter at the sample position, which affects both parameters x_0 and θ_0 , can minimize sample placement errors. However, the inverse dependence on the waist diameter of x_0 and θ_0 sets a limit to this optimization, because improving the insensitivity to tilts will increase the sensitivity to off-axis misalignment errors, and vice versa. Beam propagation in the THz frequency range is so closely connected to Gaussian mode propagation and quasi-optics that even small optical systems require a Gaussian beam propagation analysis. This is in contrast to optical designs in the visible where the radiation propagation is easily sketched by geometrical optics.

2.4 Absorption of Radiation by Materials and Resonant Effects

All materials absorb radiation to a greater or lesser degree. Clearly, windows should be designed for minimum absorption but, very often, other samples being used, or studied, may have very high absorption. The Lambert–Beer absorption law characterizes losses

$$I = I_0 \exp(-\alpha d), \quad (2.44)$$

where (assuming no reflection losses) I_0 is the initial intensity, α is the linear absorption coefficient, and d the material thickness. The coefficient can be described as $\alpha = NBh\nu/c = N\sigma$ with the Einstein coefficient B , the particle density N , and the absorption cross-section σ . For materials with absorbing components, for example particles mixed in a matrix material for the fabrication of THz windows and filters (Chap. 3), the coefficient can also be expressed with the extinction coefficient ε_{EXT} as $\alpha = \varepsilon_{\text{EXT}}c$ by inserting the concentration c of the components or particles. With Fourier-transform spectrometers (Chap. 6), the absorbance $A = -\log_{10}(I/I_0)$ is measured for a transmitted intensity I and an initial light source intensity I_0 . This absorbance is related to the absorption coefficient $\alpha = \ln(10)A/d$. The penetration depth $d_p = 1/\alpha$ of a wave into a material can be defined by determining the depth at which the intensity is reduced to $1/e$ of the initial value.

In general, an electromagnetic wave is not described by its measured intensity but by the propagating electric or electromagnetic field. Nowadays, electro-optical sampling techniques allow the measurement of the amplitude and phase of an electric field. A wave propagating along the $+z$ direction through a medium is described by its amplitude E_0 and complex phase $\Phi = \hat{k}z - \omega t$:

$$E = E_0 \exp(i\Phi) = E_0 \exp\left(i\left(\hat{k}z - \omega t\right)\right). \quad (2.45)$$

The complex wavevector $\hat{k} = k + i\alpha/2$ connects the real wavevector and the absorption coefficient, while the angular frequency $\omega = 2\pi\nu$ depends on the wave's frequency. The definition of $\hat{k}$ and the intensity of the wave $I = |E|^2 = E^*E$, after propagation over a distance d , gives Lambert–Beer's law (2.44).

The real part of the wavevector $k = \omega/v_p = n\omega/c = 2\pi/\lambda$ in the medium is related to the phase velocity v_p and is given by the wavevector $k_0 = \omega/c$ in vacuum and the frequency-dependent refractive index n . The refractive index in materials is not constant. The index varies as a function of frequency leading to the phenomena of dispersion, the reason why prisms disperse light into its individual colors. The complex refractive index $\hat{n}$ can be extracted by rearranging $\hat{k} = k_0n + i\alpha/2 = k_0(n + i\alpha/(2k_0))$ and defining $\kappa = \alpha/(2k_0)$, the imaginary part of $\hat{n} = n + i\kappa$.

In general, the interaction of an electromagnetic wave with matter is not only described by the relative dielectric permittivity ε but also by the relative magnetic permeability μ of the medium in comparison with a vacuum. The refractive index is described by the Maxwell relation

$$\hat{n}^2 = \varepsilon\mu. \quad (2.46)$$

The wave's impedance Z connects the two values by

$$Z = \sqrt{\frac{\mu\mu_0}{\varepsilon\varepsilon_0}}, \quad (2.47)$$

which is in vacuum $Z_0 \approx 377 \Omega$. Normal optical materials have positive ε and μ (and typically $\mu \approx 1$) and then $\hat{n}$ is simply the positive square root of the dielectric function. As shown above, absorption corresponds to a complex quantity and anisotropic materials are described by tensors. Negative values of n can be achieved with so-called metamaterials (Sect. 3.17), with simultaneous negative permittivity and negative permeability [121]. In the following paragraphs μ is set to one.

The refractive index can be determined experimentally by using methods such as THz time domain spectroscopy (Sect. 6.6), dispersive Fourier-transform spectroscopy (Sect. 6.3.1), and ellipsometry. These techniques either measure the electric field directly, or they are sensitive to the phase of the wave by measuring interference or polarization. The refractive index and the absorption are closely related. This follows from the Kramers–Kronig relations, which connect real and imaginary parts of the dielectric function. They are often used to calculate the refractive index

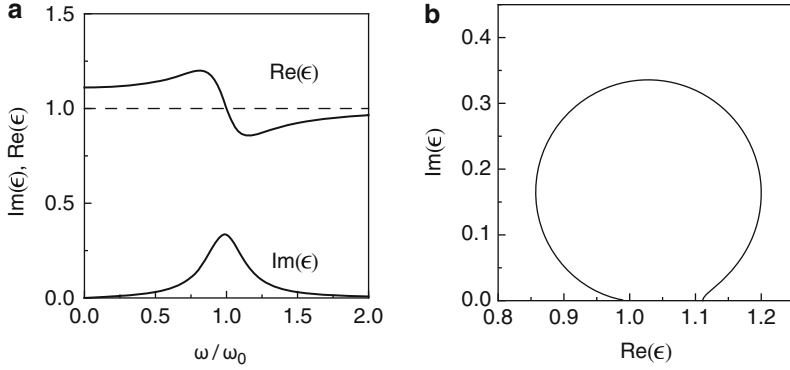


Fig. 2.9 Lorentz oscillator for a plasma frequency $\omega_p/\omega_0 = 1$, damping parameter $\gamma/\omega_0 = 1$, and resonance frequency ω_0 (2.54)

from measured absorption. Although the relation is exact it is defined over infinite frequency space, while experimental data spans a limited and discrete frequency range.

The complex dielectric constant of the medium ϵ is often expressed by the complex refractive index, and further separated into real and imaginary parts, which are used to describe material properties such as the polarizability:

$$\epsilon = \hat{n}^2 = (n + i\kappa)^2 = \epsilon' + i\epsilon''. \quad (2.48)$$

The real and imaginary parts are separated in $\epsilon' = n^2 - \kappa^2$ and $\epsilon'' = 2n\kappa$ and can be plotted in the complex plane (Fig. 2.9).

A simple description of the dielectric constant is possible with the damped harmonic oscillator model, in which an external electric field E with frequency ω drives, harmonically, a single electron of charge e and mass m with a resonance frequency $\omega_0 = \sqrt{f/m}$, depending on the restoring force constant f . With the phenomenological damping factor γ , which can be often related to relaxation times due to scattering events or bond breaking, the equation of motion follows as

$$\frac{\partial^2 x}{\partial t^2} + \gamma \frac{\partial x}{\partial t} + \omega_0^2 x = \frac{e}{m} E. \quad (2.49)$$

The equation leads to the solution

$$x = \frac{e}{m} \frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega} E. \quad (2.50)$$

The polarization of a medium is described by the susceptibility χ , generally a tensor. For the linear response in E described here, the first susceptibility coefficient $\chi^{(1)}$ is sufficient, and this is a scalar. The coefficient is labeled by the superscript (1).

Nonlinear effects using further susceptibility coefficients and anisotropic effects are discussed in the following section. The polarization is written as

$$P = \varepsilon_0 \chi^{(1)} E = \varepsilon_0 (\varepsilon - 1) E. \quad (2.51)$$

The dipole moment of a single electron is obtained from $p = ex$. For N molecules per unit volume with Z electrons, and with individual oscillator strengths f_i for each oscillator i obeying the sum-rule $\sum_i f_i = Z$, the dielectric constant follows as

$$\varepsilon = 1 + \frac{Ne^2}{\varepsilon_0 m} \sum_i \frac{f_i}{\omega_i^2 - \omega^2 - i\gamma_i \omega}. \quad (2.52)$$

Here, the assumption is made that all oscillators i have potentially different damping constants γ_i , and resonance or binding frequencies ω_i . Assuming certain relaxation processes, and distributions of frequencies and damping, the dielectric function is often fitted empirically and displayed in the complex plane (Fig. 2.9) referred to as Cole–Cole plots [122], Davidson–Cole plots [123], etc.

To illustrate the importance of the resonant Lorentz dielectric permittivity (2.52), it can be simplified, assuming the same oscillator resonance frequency ω_0 and strength $f_0 \approx 1$. The introduction of the total number of electrons $N_e = NZ$ and definition of the plasma frequency

$$\omega_p = \sqrt{\frac{N_e e^2}{\varepsilon_0 m}} \quad (2.53)$$

leads to a simple expression of the dielectric permittivity

$$\varepsilon = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega}, \quad (2.54)$$

which is illustrated in Fig. 2.9.

Equation (2.54) is of fundamental importance for describing a variety of phenomena. These include reflection from metals (Sect. 3.2), resonances in dielectrics and crystals (Sect. 3.3), and the dielectric response of metamaterials (Sect. 3.17), in which the damping and resonance frequencies are mainly determined by the design geometry. The Drude form for a plasma consisting only of free electrons N_e is obtained if ω_0 is set to zero:

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)}. \quad (2.55)$$

For high frequencies beyond resonance ($\gamma \ll \omega$), this equation simplifies to

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega^2}, \quad (2.56)$$

and shows that for $\omega < \omega_p$ the dielectric constant is always negative. Therefore, waves with a frequency $\omega \ll \omega_p$ cannot penetrate and are reflected from a plasma, a metal, or heavily doped semiconductor. In semiconductors, the mass m in (2.53) is replaced by the effective mass m^* and the free electrons are related to the free electron density as a function of the doping concentration.

Macroscopic properties of materials can be derived from the Lorentz oscillator model, for example, the reflectance of metals in the THz frequency range. Free electrons in metals lead to a refractive index and dielectric constant, which is in turn connected to the metal DC conductance by

$$\sigma_{DC} = \varepsilon_0 \frac{\omega_p^2}{\gamma} = N_e e \frac{e}{m} \frac{1}{\gamma}. \quad (2.57)$$

The right hand side of (2.57) is defined as the mobility. In semiconductors, the phenomenological damping γ is related to the inverse scattering time, due to various scattering events in the semiconductor material. From (2.55) it follows, for the conductance-dependent dielectric function, that

$$\varepsilon = 1 - \frac{\sigma_{DC} \gamma}{\varepsilon_0 \omega (\omega + i\gamma)} \rightarrow_{\omega \ll \gamma} 1 + i \frac{\sigma_{DC}}{\varepsilon_0 \omega}, \quad (2.58)$$

which includes the assumption that the frequencies investigated are much smaller than the damping frequencies $\omega \ll \gamma$. By further neglecting the real part of the dielectric function, the complex refractive index, according to Drude, is obtained:

$$\hat{n} = (1 + i) \sqrt{\frac{\sigma_{DC}}{2\varepsilon_0 \omega}}. \quad (2.59)$$

The reflectance is calculated from the complex refractive index as

$$R = \left| \frac{\hat{n} - 1}{\hat{n} + 1} \right|^2. \quad (2.60)$$

Inserting (2.59) into (2.60) leads, after a further simplification, to the same result as the equation, derived empirically by Hagen and Rubens in 1904 [17], for the reflection of metals in the THz region (Chap. 3).

The ability to connect microscopic electronic phenomena to macroscopic values such as the conductance is very useful. The mathematically equivalent treatment of an electric oscillator consisting of an inductance L , capacitance C , and resistance R allows the formalism of equivalent circuit diagrams. The Lorentz oscillator model can then be used to investigate THz sources, devices, spectroscopic techniques, and optical components such as metal meshes. The drawback of dropping the physical microscopic meaning is balanced by the easier calculation of circuit diagrams. Typically, the damping γ is related to the resistance R/L , and the resonance frequency is defined by $\omega_0 = 1/\sqrt{LC}$. In general, the inductance L is related

to the permeability and the capacitance C to the permittivity. The performance of mesh structures and also metamaterials can be calculated from equivalent circuit diagrams, where the values L , C , and R have been evaluated from the geometry of the resonant structures, which are predominantly made from metals (Sect. 3.17).

2.5 Nonlinear Interactions with Matter and Terahertz Anisotropy

Nonlinear effects are increasingly important in the THz frequency range due to the availability of high-power and short pulse lasers. THz TDS is a broadly applied technique and THz opto-parametric oscillators are efficient THz sources relying on nonlinear processes. Both are room temperature devices, an added benefit which is in part related to the initial high intensity of the visible lasers. This high intensity allows rather low conversion factors to generate THz photons. In nonlinear processes even three or four photons can be involved, giving rise to three-wave and four-wave mixing processes.

Optical rectification uses a nonlinear material to generate THz radiation (Sect. 6.6). The process down-converts a visible photon into two photons in the IR and THz range. This is comparable to difference frequency generation (DFG), and, importantly, the efficiency increases with intensity. Second-order rectification and generation of THz radiation strongly depends on the material anisotropy. For example, oriented $\langle 110 \rangle$ crystals in the zinc-blend structure like ZnTe are efficient emitters, and are also used as detecting elements for electro-optical sampling (EOS). This section outlines the basic equations for nonlinear processes in materials, while devices and techniques using these processes are discussed in the following chapters.

For the investigation of nonlinear responses, the linear response of the polarization (2.51) represents only one term of

$$\mathbf{P}(\mathbf{E}) = \varepsilon_0 \chi^{(1)} \mathbf{E} + \mathbf{P}^{\text{nl}} \quad (2.61)$$

with $\mathbf{P}^{\text{nl}}$, which describes the nonlinear polarization. The polarization can also be developed into a Taylor series as a function of the electric field, and then the nonlinear polarization is described by higher orders of the electric field.

$$P(\mathbf{E}) = \varepsilon_0 (\chi^{(1)} \mathbf{E} + \chi^{(2)} \mathbf{E}^2 + \dots). \quad (2.62)$$

The index (1), (2), ..., labels the susceptibility coefficients $\chi^{(1)}$, $\chi^{(2)}$, First-order terms describe normal polarization, second-order terms frequency conversion processes, while third and higher order terms are responsible for Raman and Brillouin scattering, and the Kerr effect. The latter is used in mode-locked femtosecond lasers (Sect. 6.6) generating frequencies with a bandwidth from 10 to 100 THz.

The second susceptibility coefficient $\chi^{(2)}$ describes three-wave mixing. It is a tensor for most materials. Crystals with anisotropy interacting with high-power laser pulses are described by such tensors with specific nonzero values. The three-component tensor in 2nd order $\chi_{ijk}^{(2)}$ leads to an equivalent formulation of (2.62):

$$P_i = \varepsilon_0 \left(\chi_{ij}^{(1)} E_j + \chi_{ijk}^{(2)} E_j E_k + \dots \right). \quad (2.63)$$

For materials with inversion symmetry, the equation $\mathbf{P}(\mathbf{E}) = -\mathbf{P}(-\mathbf{E})$ is valid and the coefficient $\chi^{(2)}$ is zero. Therefore, one condition for a suitable nonlinear material, for frequency conversion and THz generation, is the absence of inversion symmetry in the crystal. Exchanging indices does not change the physical properties, and using the general convention of replacing $\chi^{(2)}$ by $d = \chi^{(2)}/2$ further simplifies the tensor to a few independent parameters

$$\mathbf{P}^{\text{nl}} = \begin{pmatrix} P_x^{\text{nl}} \\ P_y^{\text{nl}} \\ P_z^{\text{nl}} \end{pmatrix} = 2\varepsilon_0 \begin{pmatrix} d_{11} & \cdots & d_{16} \\ \vdots & \ddots & \vdots \\ d_{31} & \cdots & d_{36} \end{pmatrix} \begin{pmatrix} E_x^2 \\ E_y^2 \\ E_z^2 \\ 2E_y E_z \\ 2E_z E_x \\ 2E_x E_y \end{pmatrix}. \quad (2.64)$$

Some parameters of d can be zero due to crystal symmetries. However, for a nonlinear crystal, one nonlinear coefficient has a maximum value with respect to all other coefficients for a certain crystallographic orientation of the crystal and incident wave directions.

2.6 Polarized Light

In the previous sections, the assumption was made that the polarization is preserved in the optical system. In practice, polarization is very important, not only for coupling radiation, but also in the study of various materials, e.g., in magnetic fields as well as for balanced detection schemes and spectroscopy. Various antenna structures employed in THz detectors and emitters determine the polarization and form of the spectral response, e.g., dipole antennas select linear polarization preferentially, while systems in magnetic fields can either respond, emit, or transmit circularly to elliptically or linearly polarized light. Clockwise and anticlockwise helical antennas respond to the appropriate circularly polarized light.

Common window materials such as quartz exhibit birefringence in the THz region. The different propagation characteristics, and different refractive indices along the crystallographic axes, can be minimized by choosing, e.g., z-cut (or c-cut) quartz. This material is typically employed as a cryostat vacuum window. Polarization rotation due to impinging THz fields on nonlinear crystals, such

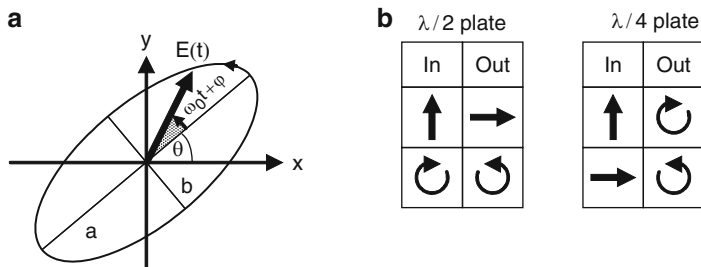


Fig. 2.10 (a) A monochromatic wave is specified by its propagation vector and complex amplitude of the electric field vector $E(t)$ in time. In general, the tip of the electric field vector traces out an ellipse. The polarization can be described by the semi-major a and semi-minor b axes of the polarization ellipse, its orientation determined by θ , and sense of rotation. (b) change of polarization states due to $\lambda/2$ and $\lambda/4$ wave plates inserted in the beam path: linear polarization can be rotated by 90° or transformed into left-hand (counter-clockwise rotation) or right-hand circular polarization, respectively

as ZnTe, is an integral part of the detection scheme in electro-optical sampling (EOS, Sect. 6.6). THz near-field microscopy (Sect. 7.5) uses scattering nano-tips, comparable to long wire antennas with whisker contacts, for coupling the radiation preferentially along the wire axis, while linear polarized light, perpendicular to the antenna, does not couple well.

Current research concentrates on the development of suitable components for manipulating the polarization, or more generally the phase of the wavefront, in order to obtain similar facilities to those routinely available in the visible region. Devices based on electronically controlled liquid crystals, photonic bandgaps, metamaterials, and the latter combined are considered ([124, 125], Sect. 3.17). A few examples are THz lambda-quarter ($\lambda/4$) [126] and lambda-half ($\lambda/2$) plates to convert linear to circular polarization and vice versa, or to attenuate power and rotate polarization (Fig. 2.10). Other requirements are for electronic pulse retarders, phase shifters [126–128], filters [129], THz modulators [124, 130], and hybrid antennas to emit both perpendicular and linear polarized wavefronts to study circular dichroism [131], Faraday rotators, and further polarization sensitive devices and measurement techniques.

2.7 Stokes Parameter

Light polarization changes while passing through an optical system. The outgoing polarization can be calculated by applying Müller calculus on the Stokes vector of the incident light. The Stokes vector is a convenient and compact way to describe polarization of waves. The Stokes vector consists of four parameters and values designated as I , Q , U , and V , or S_0 , S_1 , S_2 , and S_3 . George Gabriel Stokes introduced the parameters in 1852 to describe the polarization state [132]. They

Table 2.1 Elements of the Stokes vector for different polarizations normalized to the electric field amplitude E

Polarization	Linear (horizontal)	Linear (vertical)	Circular (left)	Circular (right)	Unpolarized
S_0	1	1	1	1	1
S_1	1	-1	0	0	0
S_2	0	0	0	0	0
S_3	0	0	-1	1	0

were chosen because they are easily measured by determining the intensity of light after passage through different polarizers. The Stokes vector describes unpolarized, partially polarized, and fully polarized light (Table 2.1), defined as

$$\mathbf{S} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix}. \quad (2.65)$$

For polarized light, the four parameters are related

$$S_0^2 = S_1^2 + S_2^2 + S_3^2. \quad (2.66)$$

For fully unpolarized light, most parameters are zero. The power P is measured with polarizers at 0° , 90° , 45° , 135° , and for right- and left-polarized light, P_R and P_L , respectively, and allows the calculation of the Stokes parameters as follows

$$\begin{aligned} S_0 = I &= P_{0^\circ} + P_{90^\circ} = \langle E_x^2 + E_y^2 \rangle, \\ S_1 = Q &= P_{0^\circ} - P_{90^\circ} = \langle E_x^2 - E_y^2 \rangle, \\ S_2 = U &= P_{45^\circ} - P_{135^\circ} = \langle 2E_x E_y \cos(\delta) \rangle, \\ S_3 = V &= P_R - P_L = \langle 2E_x E_y \sin(\delta) \rangle. \end{aligned} \quad (2.67)$$

Alternatively, the time averages indicated by the brackets in (2.67) of the electric field amplitudes E_x and E_y in an (x, y) -orthogonal coordinate system can be used. The relative phase between the field amplitudes is denoted as $\delta = \phi_y - \phi_x$.

2.8 Chirp

Ultrashort laser pulses used in THz time-domain spectrometers have a broad frequency spectrum. Optical elements such as mirrors and lenses may not reflect, transmit, or focus all frequencies in precisely the same way, often introducing pulse broadening in time or, in general, pulse shaping. Two effects are typically distinguishable, amplitude filters if, for instance, a mirror reflects waves at some frequencies with lesser amplitude, or dispersion if the phase changes at different

frequencies. In a laser cavity, these effects become amplified. In an experiment, attempts are made to avoid this outcome, or to design the optical system in order to manipulate the amplitudes and phases in a controlled way.

The effects can be evaluated by considering a pulse, which has a Gaussian field distribution in time and a width τ

$$E(t) = E_0 \frac{1}{\tau} \exp\left(-\frac{t^2}{2\tau^2}\right). \quad (2.68)$$

The frequency spectrum has also a Gaussian distribution with its maximum field at ω_0

$$E(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(t) \exp(-i\omega t) dt = \frac{E_0}{\sqrt{2\pi}\tau} \exp\left(-\frac{1}{2}((\omega - \omega_0)\tau)^2\right) \quad (2.69)$$

with a full width at half maximum (FWHM) of $\Delta\omega_{\text{FWHM}} = 2\sqrt{\ln(2)}/\tau$. The product of pulse time width and bandwidth at FWHM is approximately 0.44.

If such a pulse experiences dispersion, due to reflection at a surface or through a medium, then its spectral components will be modified by amplitude and phase factors, $A(\omega)$ and $\Phi(\omega)$, respectively. The resulting pulse can be determined by inverse Fourier transformation

$$E(t) = \int_{-\infty}^{\infty} E(\omega) A(\omega) \exp(-i\Phi(\omega)) \exp(i\omega t) d\omega. \quad (2.70)$$

Considering a pure amplitude filter with a constant phase factor $\Phi(\omega) = \Phi$ and an amplitude function of

$$A(\omega) = \exp\left(-\frac{1}{2} \frac{(\omega - \omega_0)^2}{\omega_F^2}\right) \quad (2.71)$$

leads to a modified frequency bandwidth of

$$\Delta\nu_{\text{FWHM}} = \frac{\omega_F}{2\pi\sqrt{\ln 2}} \quad (2.72)$$

with the frequency bandwidth ω_F of the filter. The modified field

$$E^m(t) = E_0 \frac{1}{\tau^m} \exp\left(-\frac{t^2}{2(\tau^m)^2}\right) \quad (2.73)$$

is determined by a modified time

$$\tau^m = \tau \sqrt{1 + \frac{1}{\omega_F^2 \tau^2}}. \quad (2.74)$$

The pulse broadens due to the limited bandwidth of the filter or mirror unless the bandwidth is very large, i.e. $\omega_F \gg 1/\tau$, and $\tau^m \rightarrow \tau$.

With dispersion, the phase at each frequency of an amplitude filter varies. For simplicity, the phase can be expanded in a Taylor series around a center frequency ω_0

$$\Phi(\omega) = \Phi_0 + \frac{\partial \Phi}{\partial \omega}(\omega - \omega_0) + \frac{1}{2} \frac{\partial^2 \Phi}{\partial \omega^2}(\omega - \omega_0)^2 + \dots \quad (2.75)$$

The first term adds only a constant phase shift. The second term does not change the initial pulse shape $E(t)$ either, however, $E^m(t)$ is shifted in time by $\Delta t = \partial \Phi / \partial \omega$. Furthermore, the second derivate $\delta = \partial^2 \Phi / \partial \omega^2$ contributes to pulse broadening or pulse compression. For a pulse with a width of $\tau_b = \sqrt{|\delta|}$ and using (2.70) the modified pulse is

$$E^m(t) = E_0 \frac{1}{\tau^2 + i\delta} \exp\left(-\frac{t^2}{2\tau^2}\right) \exp\left(i \frac{\delta}{\tau^2} \frac{t^2}{2\tau^2}\right) \quad (2.76)$$

with a pulse width of

$$\tau^m = \tau \sqrt{1 + \frac{\tau_b^4}{\tau^4}}. \quad (2.77)$$

The pulse broadens, and for large broadening the pulse length approaches τ_b^2/τ . The broadening effect by second-order dispersion affects shorter pulses more severely, in contrast to pure amplitude filters, as shown by (2.74). The second-order dispersion δ relates to the transfer of frequencies at different velocities through a medium. It leads to chirp modulation of the initial pulse (Fig. 2.11). THz time domain spectroscopy uses lasers with pulses of a few to a few tens of femtoseconds. Typical values of δ for optical materials used at 800 nm range from 360 fs²/cm for fused silica to 580 fs²/cm for sapphire. The phase $\Phi(\omega)$ can be related to the refractive index $n(\omega)$ (Sect. 2.4) if the radiation passes through a material of length L

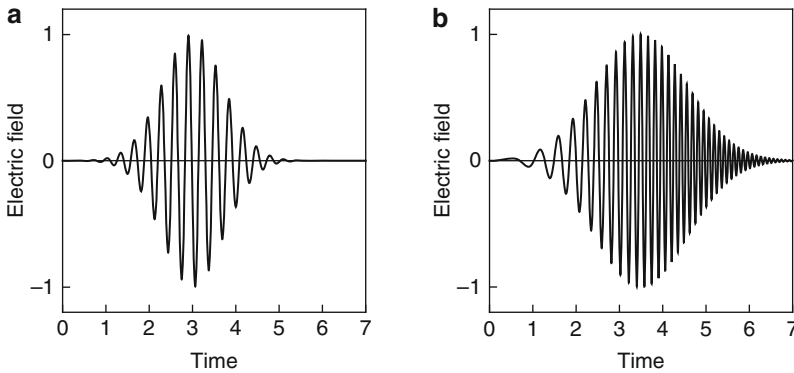


Fig. 2.11 (a) Illustration of an electric field pulse as a function of time and (b) a chirped and broadened pulse, both with a Gaussian intensity profile

$$\Phi(\omega) = k(\omega)L = \frac{\omega}{c}n(\omega)L. \quad (2.78)$$

The first-order dispersion can be evaluated as

$$\frac{\partial \Phi}{\partial \omega} = \frac{\partial k}{\partial \omega}L = \frac{n}{c} \left(1 + \frac{\omega}{n} \frac{\partial n}{\partial \omega} \right) L = \frac{L}{v_G} \quad (2.79)$$

with the group velocity v_G in the medium, which leads to a pulse retardation or delay of L/v_G . The second-order dispersion follows as

$$\frac{\partial^2 \Phi}{\partial \omega^2} = \frac{\partial^2 k}{\partial \omega^2}L = \left(\frac{2}{c} \frac{\partial n}{\partial \omega} + \frac{\omega}{c} \frac{\partial^2 n}{\partial \omega^2} \right) L. \quad (2.80)$$

The combination of dispersive elements within a laser cavity, such as prism pairs and grating pairs, can compensate for second-order dispersion caused by optical elements such as laser gain media or windows. Additional prism or grating pairs can compress pulses further in time or lead to pulse shaping.

Chapter 3

Optical Components

3.1 Introduction

As the THz region is between the IR and mmW parts of the electromagnetic spectrum the types of components are, on the one hand, similar to optical equipment, for example mirrors and lenses, while, on the other hand, mmW devices, such as antennas or waveguides, are used. The optical components are much the same as in infrared or visible systems. However, there are differences in detail. As there are relatively few suitable lens materials, mirrors assume greater importance than in the visible region. As the most sensitive detectors at THz frequencies are cryogenically cooled in dedicated cryostats, windows are required, which hold the vacuum while allowing the THz radiation to enter with as little loss as possible. Due to the absorption of THz radiation by atmospheric water vapor, the optical path often needs to be evacuated and, therefore, windows are also needed for this. The subsidiary components of an optical system such as filters, polarizers are nowadays as well developed in the THz region as for the visible and infrared regions, and methods for providing such components with photonic bandgap structures or metamaterials are under way.

The optical components that have attracted great attention are filters. The most important requirement for filters is to remove all radiation up to a predetermined wavelength and then transmit at longer wavelengths. This need arises because of the very low emission of thermal sources in the THz region compared with those at shorter wavelengths. Besides lenses and mirrors, there is another condensing component that is often employed. This is the tapered light pipe which has the same property as a mirror or lens, namely that of changing the area to solid-angle ratio of radiation. A particularly widely used condensing light pipe is the Winston cone. In addition to tapered light pipes for condensing radiation, parallel pipes are used to transfer radiation from one part of a THz system to another. Light pipes are very efficient in the THz region because metals have a very high reflectance. In some sense, light pipes can be viewed as waveguides. With state-of-the-art machining, it is now possible to fabricate waveguides, with a similar performance to that in

the mmW region, up to about 2 THz. Antennas are either used in the form of horn antennas, which couple radiation into a waveguide, or as planar antennas, which are lithographically produced on a substrate. To overcome substrate losses, the substrate is glued to a lens of the same material as the substrate. This is called a quasi-optical antenna and is possibly the best example of the merging of optical with radio technology.

3.2 Reflection and Absorption

Before discussing optical components in detail, it is useful to give a brief review of some relevant features of reflection and absorption processes, which are pertinent at THz frequencies. The reflectance R is given by

$$R = \frac{(n - 1)^2 + \kappa^2}{(n + 1)^2 + \kappa^2}, \quad (3.1)$$

where n is the refractive index and $\kappa = \alpha\lambda/(4\pi)$ is the extinction coefficient (λ : vacuum wavelength, α : absorption coefficient). The derivation of (3.1) is given in many books on optics (see also Sect. 2.4). When κ is small (3.1) reduces to

$$R = \left(\frac{n - 1}{n + 1} \right)^2, \quad (3.2)$$

which is the equation used to calculate the reflection from a surface of a low absorption material.

The reflectance of metals is high at low frequencies and can be described with reasonable accuracy from an equation obtained empirically by Hagen and Rubens in 1904 [17]:

$$R = 1 - 2\sqrt{\frac{4\pi\epsilon_0\nu}{\sigma_{\text{DC}}}} \quad (3.3)$$

with reflectance R , DC conductance σ_{DC} , and frequency ν . However, the equation can be derived from the Drude equations, assuming that n and κ of metals are approximately equal (Sect. 2.4).

The Hagen–Rubens equation does not take into account any frequency dependence of the conductivity. It holds in the limiting case of low frequencies, i.e. $\sigma_{\text{DC}}/\nu \gg \epsilon_0$ where Drude-type absorption by free electrons is negligible. With increasing frequency, the conductivity of a metal becomes complex and (3.3) no longer describes the reflectance accurately. The assumption $\sigma(\nu) = \sigma_{\text{DC}}$ basically means that $2\pi\nu\tau \ll 1$ where τ is the mean free collision time of the electrons in the metal. However, taking the frequency-dependent conductivity into account changes the reflectance by <0.02% at 0.6 THz and <0.15% at 2.5 THz [133]. For most applications, this is negligible.

Nichols observed a high reflectance of certain crystalline materials in the infrared as long ago as 1897 [10]. The term “reststrahlen” (residual rays) is used for this phenomenon. It occurs at wavelengths where the absorption coefficient is so high that $R \rightarrow 1$ (cf. (3.1)). Reststrahlen reflection occurs in certain crystals, where the ion vibrations in the crystal lattice interact directly with light. The absorption band caused by this interaction leads to a high reflectance. Alkali halides, quartz, sapphire, and the two-atom ionic mixed crystal KRS-5 have strong reststrahlen bands. Houghton and Smith [134] have given a detailed account of the process involved and they also include a fuller list of reststrahlen materials.

Covalent crystals such as Si or Ge have no reststrahlen reflection. Si is quite transparent throughout the THz region. But this only holds good for crystals of high purity, i.e. the concentration of impurities with absorption bands at THz frequencies has to be as small as possible. Shallow donors or acceptors in particular contribute to the THz absorption. Due to their large refractive index of 3 to 4, the reflection loss of these crystals is high. Many polymers have a smaller refractive index (~ 1.5) and consequently less reflection loss. In addition, their absorption is quite low. It is governed by phonon absorption but there may be resonant absorption features caused by the structure of the polymers. As mentioned below, the absorption of polyethylene around 2.2 THz is a good example.

3.3 Materials for Windows, Filters, and Lenses

In this section, the optical and other important properties of materials used for windows, filters, and lenses in the THz region will be described. Many nonmetallic materials transmit over some part of the THz range, but comparatively few of them are suitable for application in a THz instrument or experiment.

3.3.1 Polyethylene (PE)

Polyethylene (PE) is the most generally useful window material for the THz region. PE is available as a high-density (HDPE) with a density above, and low-density (LDPE) material with a density below 0.95 g/cm^3 . The refractive index and the absorption coefficient are shown in Figs. 3.1 and 3.2. There is also a rather elusive absorption band at 2.2 THz, which is much more marked in HDPE. This is unfortunate because it is the more useful material as it has a higher transmission and is stronger and more rigid than the low-density form because HDPE is $\sim 20\%$ more crystalline than LDPE. In many ways, PE comes close to being an ideal window material. It is very cheap and has a low refractive index of 1.54 ± 0.01 , which is reasonably constant below 2 THz [135]. It is quite inert, surprisingly strong mechanically, and can be used as a vacuum seal at liquid nitrogen temperatures. At lower temperatures, it becomes very brittle and tends to crack if held tightly in place.

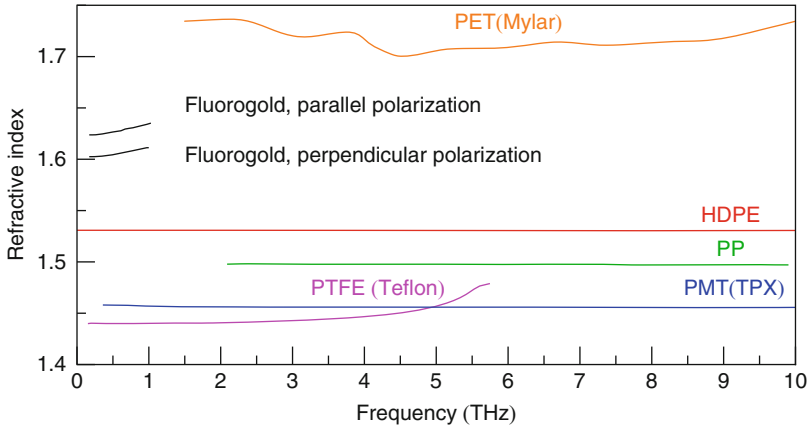


Fig. 3.1 Refractive indices of a number of polymer materials at 290–300 K (adapted from various sources)

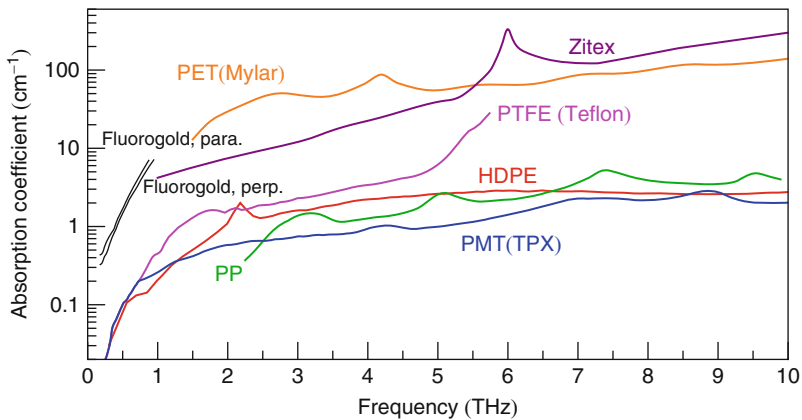


Fig. 3.2 Absorption coefficients of a number of polymer materials at 290–300 K (adapted from various sources)

LDPE softens rapidly above 70°C and melts between 110°C and 120°C, while HDPE melts around 130°C. PE can be heated without additives for extruding and hot embossing to fabricate various components. It is also used as a substrate material for components such as metal mesh filters and polarizers. Other materials can be dispersed into it to provide selective filters, and ruled PE is used as a transmission filter grating. It can also be machined to make lenses and other optical components.

3.3.2 Polypropylene (PP)

Polypropylene (PP) is similar to PE. Its refractive index is about 1.5 and varies by less than 0.5% throughout the THz region [136, 137]. Above 2 THz, the spectrum has several strong absorption bands, which are superimposed on a background that rises with increasing frequency (Fig. 3.2). These are caused by internal motion of the PP molecules. It is worth noting that the values shown in Fig. 3.2 are for pure isotactic PP. Commercial PP often contains additives which increase the absorption coefficient by more than a factor of 2 [136]. Because of its relatively large absorption, in comparison with other polymers, PP is rarely used as a lens material. Due to its mechanical strength PP makes good vacuum windows [138]. It can also be used as a beam splitter, but in a Fourier-transform spectrometer (FTS) it is somewhat less efficient than Mylar because of its lower refractive index.

3.3.3 Polystyrene (PS)

Polystyrene (PS) is a polymer developed in the 1930s by BASF and is the first industrial, fully synthetic thermoplastic used for mass production. It is especially known in its form as foam (Styrofoam), which is produced by adding chemicals which form a gas at the temperatures used during production, for example methyl chloride. It is very cheap and has a reasonably constant and low refractive index of 2.08 ± 0.01 , while the absorption coefficient gradually increases to 3 cm^{-1} from 0 to 2.4 THz [135]. The refractive index value is very close to that of crystalline quartz.

3.3.4 Polytetrafluorethylene (PTFE, Teflon)

Polytetrafluorethylene (PTFE), fabricated by DuPont under the trade name Teflon, has a refractive index of about 1.43, which increases with frequency (Fig. 3.1). Its absorption coefficient is slightly larger than that of PE but increases strongly with frequency (Fig. 3.2). At 8 THz, it has a strong absorption feature. Teflon is a good choice for lenses at low THz frequencies. In contrast to HDPE, it does not make a good vacuum window because it is permeable to helium.

3.3.5 Poly-4-methylpentene-1 (PMT, TPX)

Poly-4-methylpentene-1 is manufactured by Mitsui Chemical Inc. under the trade name TPX. It is a hard material which can be molded under pressure, is resistant to deformation by heat, and can be polished quite well. Its transmission is excellent

below ~ 6 THz. TPX is transparent in the visible region and has virtually the same refractive index (~ 1.45) in both the visible and THz regions (Fig. 3.1 [139]). This is a very considerable advantage as it allows complex optical systems to be set up using visible light. The absorption is slightly lower than that of HDPE (Fig. 3.2 [139]).

3.3.6 *Tsurupica (Picarin)*

Tsurupica (sometimes referred to as Picarin) is similar to TPX and highly transparent in the THz as well as in the visible region. The refractive index is almost the same for THz ($n = 1.56$) and visible light ($n = 1.55$). Like TPX it is very rigid and can be polished mechanically. Its major use is for lenses in systems where optical pre-alignment is important or as windows in cryostats, gas cells, or in similar situations.

3.3.7 *Polyethylene Terephthalate (PET, Mylar)*

Mylar is the trade name of DuPont for Polyethylene terephthalate (PET)-film. Other trade names are Melinex and Hostaphan. It is chemically inert, mechanically stable, and transparent to visible light. The refractive index is about 1.72, and the absorption coefficient increases from about 15 to 150 cm^{-1} in the frequency range from 1.5 to 10 THz [140], with a distinct resonance at 4.2 THz (Fig. 3.2). Mylar films are available in thicknesses between 1 and $500\text{ }\mu\text{m}$. Due to the large absorption coefficient its main application is as a beam splitter where only thin, 3–12 μm , foils are required. Since Mylar is used in thin films the influence of its biaxial structure can be significant, caused by the elongated PET molecules affecting the refractive index (up to 2%) and the absorption coefficient (up to 10%).

3.3.8 *Parylene*

Parylene is the generic name for a family of thermoplastic polymers made from Para-xylylene. It can be deposited from the gas phase at room temperature in virtually any thickness up to several hundreds of micrometers. The vacuum deposition technique allows the thickness of the film to be controlled with an accuracy of about $1\text{ }\mu\text{m}$. In addition, vacuum-deposited Parylene films are conformal, uniform, free of pinholes or defects, and chemically inert. Since vacuum deposition requires no catalyst or solvent the films are very pure and outgassing is negligible. Three forms of Parylene are known: Parylene C, D, and N. Parylene C is less transmissive for moisture and less permeable for gases than the other two and is therefore most widely used. At THz frequencies Parylene C has a refractive index of about 1.62 at

300 K, which drops to ~ 1.5 at 4.5 K [141]. Its absorption coefficient is in the order of several 10 cm^{-1} . These properties make Parylene C a good antireflection coating for Si and Ge lenses and windows.

3.3.9 Fluorogold

Fluorogold is a registered trademark of Fluorocarbon Inc. It consists of PTFE filled with glass grains. The grains are to a certain extent aligned in the PTFE and Fluorogold is birefringent. The refractive index increases from 1.602 at 0.3 THz to 1.610 at 1 THz for radiation polarized perpendicular to the direction of alignment and from 1.623 to 1.630 for radiation polarized parallel to it [142]. The birefringency makes Fluorogold unique compared with all other polymer materials [143]. The absorption coefficient increases from 1 to 10 cm^{-1} around 1 THz. This makes Fluorogold a good low-pass filter, although the cutoff is not very sharp [142].

3.3.10 Zitex

Zitex, a sintered Teflon material from Saint-Gobain Performance Plastics, is available in different varieties. Most important is Zitex G, which is made from sintered Teflon spheres with pore sizes ranging from 1.5 to $5.5\text{ }\mu\text{m}$. It comes in thicknesses from 0.1 to 0.4 mm. The density of Zitex G relative to Teflon is 0.45 to 0.60, with smaller density for larger pores. Like glass bead filters [144], Zitex relies on scattering. A dielectric sphere of radius r embedded in a material with a different refractive index n will scatter strongly for wavelengths $\lambda < \pi r(n - 1)$, i.e. for Teflon a $5\text{ }\mu\text{m}$ radius sphere will produce a shadow for wavelengths below $7\text{ }\mu\text{m}$. At these wavelengths, scattering redistributes the power in an incident beam in all directions. The transmittance of Zitex at frequencies below 3 THz is about 90%. Above this frequency it drops steeply to less than 10% at 6 THz. With a refractive index of 1.2 the reflection loss is less than that of other polymers [145]. Since most losses are due to scattering, heating of the filter is not a major issue even though the heat conductivity is low. For this reason, Zitex is often used as a cryogenic IR blocking filter, where very high transmission is required and some out-of-band transmission is acceptable.

3.3.11 Crystalline Quartz

Until the advent of polymers and high-purity semiconductors, this was the only useful material for THz windows and it is still frequently employed. Thin windows are usable up to about $\sim 7\text{ THz}$ ($\sim 40\text{ }\mu\text{m}$). Quartz is birefringent. It has a refractive index of 2.1 to 2.6 throughout the THz region for the ordinary ray as well as for the extraordinary ray (Figs. 3.3 and 3.5, [146]), with the one for the extraordinary ray about 2% larger. The reflection loss of quartz is therefore high compared with

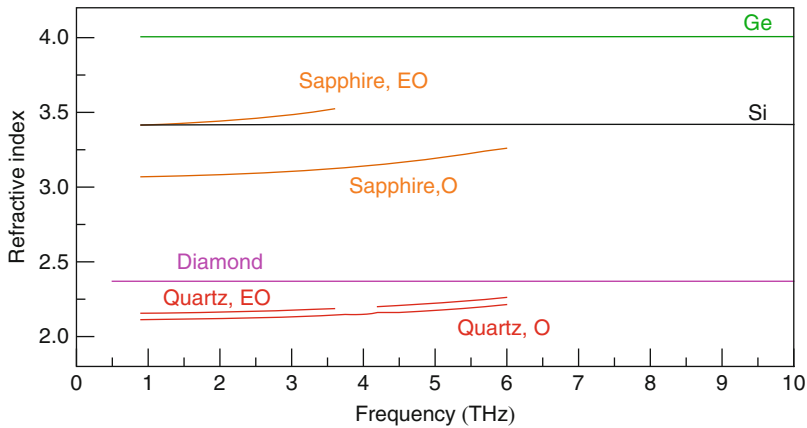


Fig. 3.3 Refractive indices of various crystalline materials at 290–300 K (EO: extraordinary ray, O: ordinary ray; adapted from various sources)

polymer materials but less than that of Si and Ge. The absorption coefficient for the extraordinary ray is significantly lower (typically 50%) than for the ordinary ray, and it decreases above 4 THz upon cooling. In addition, the ordinary ray has a strong narrow absorption at ~ 4 THz. Quartz is opaque between 4 and $40\ \mu\text{m}$ due to two reststrahlen bands at 9 and $21\ \mu\text{m}$, which means that it is a useful filter for removing these wavelengths, and it is often used for this purpose even when it is not required as a window. It can be cooled to 4.2 K, when its cut-on at ~ 7 THz becomes much sharper. It will, however, usually fracture if clamped to form a low-temperature vacuum seal as it has a large thermal expansion coefficient. Due to its opaqueness between 4 and $40\ \mu\text{m}$, it is a very good cold filter for removing background radiation from detectors. For optimum performance as a filter or window, z-cut Quartz is used because of its low birefringence. Fused quartz is not such a good transmitter of THz radiation but is useful below 1.5 THz.

3.3.12 Sapphire

Sapphire is monocrystalline Al_2O_3 and like quartz it is birefringent. Both sets of optical constants are displayed in Figs. 3.3 to 3.6 [146]. The refractive index increases from ~ 3.4 to ~ 4.3 (extraordinary ray) and ~ 3.1 to ~ 3.5 (ordinary ray) in the range from 1 to 10 THz. The refractive index decreases by $\sim 1\%$ upon cooling. Sapphire is transparent in the visible range. At 300 K, it is virtually opaque above 4 THz due to an absorption coefficient $> 50\ \text{cm}^{-1}$. The absorption coefficient decreases dramatically upon cooling. However, in the reststrahlen bands from 11 to $18\ \mu\text{m}$ and 21 to $23\ \mu\text{m}$ the absorption remains high. Because of the high absorption between 10 and $30\ \mu\text{m}$, it is often used to block high intensity radiation from CO_2 lasers or IR free electron lasers in pump-probe experiments.

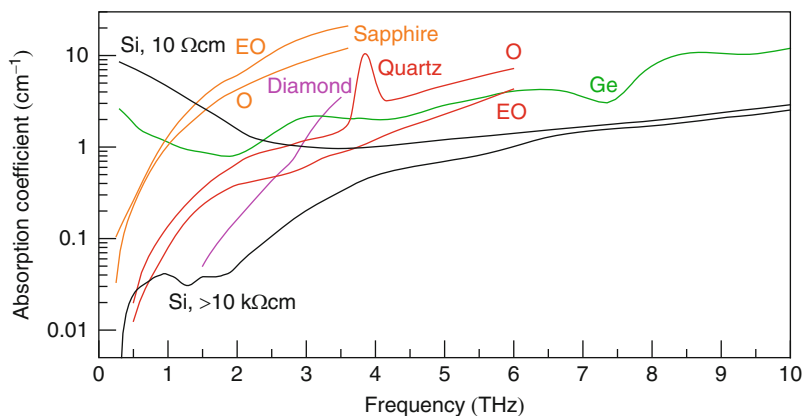


Fig. 3.4 Absorption coefficients of some crystalline materials at 290–300 K (EO: extraordinary ray, O: ordinary ray; adapted from various sources)

3.3.13 Silicon and Germanium

Semiconductors absorb at THz frequencies because of both lattice and free-carrier absorption. The former is fundamental to the crystal structure but the latter can be overcome to a large extent in Si by growing very pure crystals with resistivity of 5–10 k Ω cm (Fig. 3.4). However, due to the relatively narrow energy gap between its valence and conduction bands, the maximum room-temperature resistivity of Ge is approximately 50 Ω cm, and this leads to very significant absorption in the THz region [147] (Fig. 3.4). Si and Ge have no reststrahlen reflection due to the covalent nature of their bonds, but they are lossy in the mid-infrared because of two-phonon absorption. In Si, the absorbing region is between 7 and 30 μ m and in Ge from 16 to 50 μ m [148]. The covalent bonds have a high polarizability, leading to rather large refractive indices of 3.4 for Si and 4 for Ge [146, 149] (Figs. 3.3 and 3.4). Thus, optical components from these materials suffer from significant reflection losses. However, these can be overcome by an appropriate antireflection coating [150, 151].

3.3.14 Diamond

Diamond is the only material which is transparent over the entire region from the visible to mmW except for an absorption feature at ~ 5 μ m due to two-phonon absorption. Its index of refraction of 2.37 is almost frequency independent (Fig. 3.3). Another very useful feature is the large heat conductivity of diamond. Because of its obvious expense, it is only used to a very limited extent. One place where diamond windows are installed is at the output of IR and THz beamlines at synchrotron facilities or linear accelerators, where its high transmittance from infrared to THz frequencies, and excellent heat conductivity are important. Another useful application is as a window for the Golay detector, allowing wide frequency coverage. However,

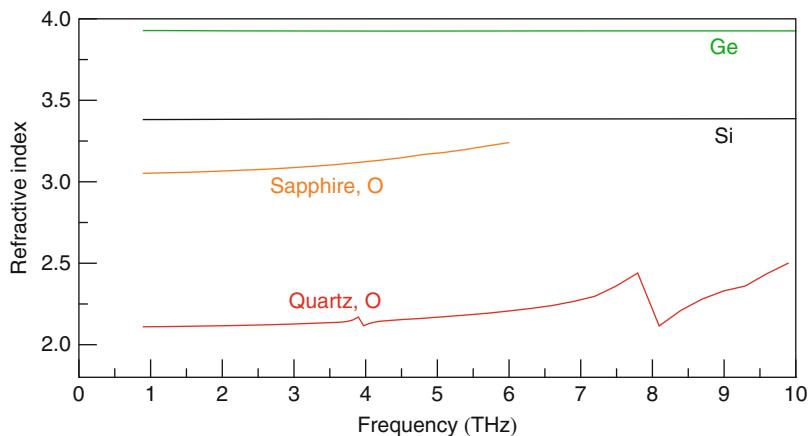


Fig. 3.5 Refractive indices of some crystalline materials at approximately 1.5 K (O: ordinary ray; adapted from various sources)

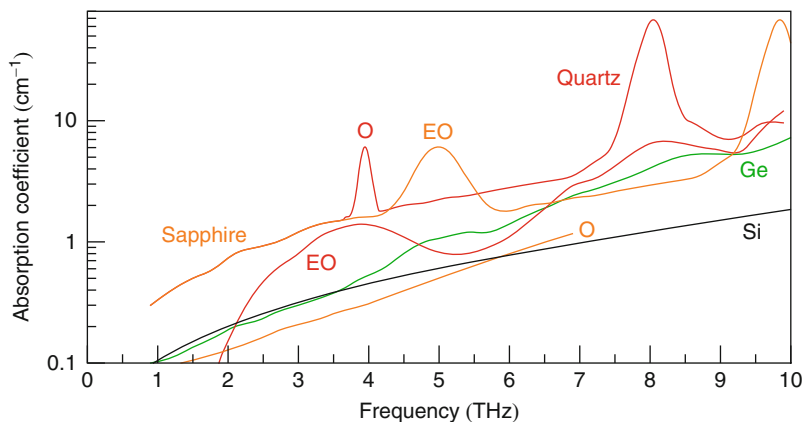


Fig. 3.6 Absorption coefficients of some crystalline materials at approximately 1.5 K (EO: extraordinary ray, O: ordinary ray; adapted from various sources)

for wavelengths longer than $50\text{ }\mu\text{m}$ crystalline quartz is preferred because of its advantage in removing the $4\text{--}40\text{ }\mu\text{m}$ radiation (Figs. 3.5 and 3.6).

3.3.15 Alkali Halides

Alkali halides are more frequently used in the IR but in some cases they are also useful at THz frequencies. These ionic crystals serve three purposes in the THz region. For high THz frequencies, they are windows with the lowest frequency transmission, that of CsI, reaching to 5 THz. They then have a region of opaqueness, over part of which they become very highly reflecting. In this region, they are employed as reststrahlen reflection filters. The high-frequency transmittance is

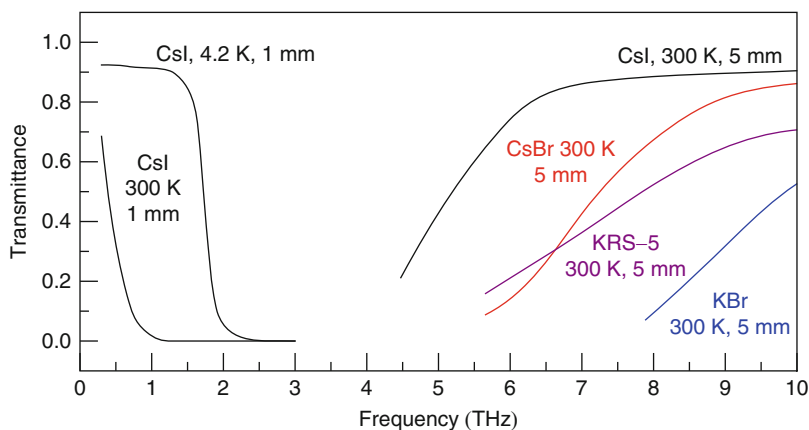


Fig. 3.7 Transmittance of some alkali halides at approximately 4.2 and 300 K (adapted from various sources)

shown in Fig. 3.7. These materials are useful as prism materials above ~ 5 THz, as well as for windows. Their disadvantage is that they are all hygroscopic but this has not prevented their widespread use. The absorption coefficient at short wavelengths is very low for all the materials and, with a comparatively low refractive index of ~ 1.5 , they have been widely used as laser windows, particularly at $10.6\text{ }\mu\text{m}$ for the CO_2 laser. NaCl and KBr are cheap materials but CsBr and CsI are more expensive. Other alkali halides such as KCl also have good transmittance at short wavelengths. Hadni et al. [152] have shown that alkali halides have much sharper transmission characteristics at 4.2 K, and they refer to this effect as “supertransmission.” The sharpening of the absorption characteristic arises from the change in phonon population when $k_B T \ll h\nu$. Figure 3.7 shows the transmission of CsI at 4.2 and 300 K and it can be seen that at the low temperature it is close to being an ideal filter. The sharpening of the characteristic also occurs on the shorter wavelength side of the absorption, and this means that the actual absorbing region may be quite narrow.

3.3.16 KRS-5

KRS-5 is one of a series of crystals developed in Germany during the 1939–1945 war and appears to be the only one that has had any significant use since then. KRS stands for “Kristalle aus dem Schmelzfluss” (crystals from the melting pot) and it consists of 42% TlBr and 58% TlI. Its high frequency transmission reaches to about 8 THz (Fig. 3.7), but it has a fairly high refractive index of about 2.3 so there is considerable reflection loss. It is, however, much more stable than the single crystals of alkali halides with only slight water solubility. There is considerable difficulty in polishing the material, but this can be done, and KRS-5 is available in both window and prism form. It shows a more pronounced reststrahlen reflection at a longer wavelength than any other material reported in the literature. It should be

noted that thallium compounds are very poisonous and care is required when KRS-5 materials are used.

3.3.17 *Material Considerations*

There is a wealth of data available on the dielectric properties of many different materials employed at THz frequencies and only the most important are described above. More details and references can be found in [116, 153]. The available data show considerable differences between the optical constants measured on different specimens of the same material. There are at least two reasons for this. First, the material properties might be different although the specimens are nominally from the same material. This could be due to different fabrication procedures or ageing and is especially relevant for polymer materials. The biaxial structure, which may arise in some of the polymers (e.g., Mylar), contributes if thin films are considered. In the case of crystalline materials, the residual doping and other impurities which may give rise to optically active centers might not be well characterized. Secondly, measurements of optical constants are susceptible to systematic errors. This topic has been addressed in a comparison study involving eleven institutions [154]. Seven specimens from different materials were exchanged between the institutions, and their refractive index and absorption coefficient were measured in the frequency region from 0.03 to 0.9 THz with monochromatic and broadband techniques. Systematic errors were found to be present in all measurements and were the dominant source of error. In the case of the refractive index, the error was less than 1%. With increasing absorption coefficient, the error increased by up to an order of magnitude. This should be kept in mind when considering the optical data in the previous section or from any other source. From this study, it becomes clear that inter-comparison of measurement techniques, and the development of material standards of refraction and absorption for the whole THz region remain a major issue.

3.4 Windows

An ideal window – such as glass in the visible – needs to have a large number of desirable features. It should be cheap and strong, insoluble in water, and stable over a wide range of temperatures; it should also be reasonably inert, have a low reflection loss, and high transmission. Similar requirements apply for lenses. No material fulfills all these requirements at THz frequencies but there are several which perform quite adequately. At temperatures around 300 K TPX, HDPE, Tsurupica, and PP are often used, because they combine low absorption and reflection loss over a wide range of frequencies, with sufficient mechanical strength for use as vacuum windows. Also, high resistivity Si with surface grooves as antireflection coating is a good choice. It is difficult to make a sharp distinction between windows and some of the transmission filters because very often a single material performs

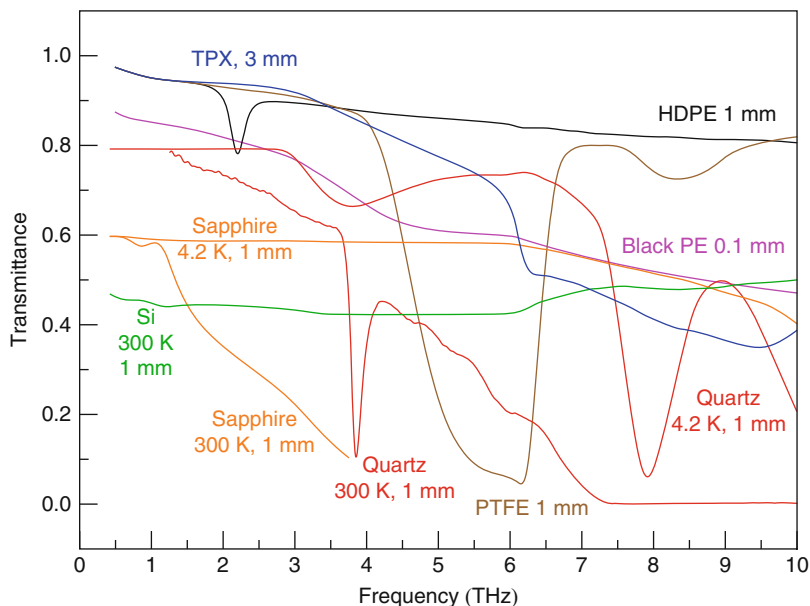


Fig. 3.8 Transmittance of some commonly used materials with a typical thickness (adapted from various sources)

both functions over part of the THz range. In Fig. 3.8, the transmittance of the most important window materials is shown for some typical thicknesses.

3.5 Filters

The primary function of filters in the THz region is the removal of unwanted shorter-wavelength radiation. It is not only visible and near-IR energy that must be removed. The grating spectrometers used at long wavelengths employ “blazed” echelette gratings (Sect. 6.2), and these have the property of being very efficient at the sub-multiples of the chosen wavelength. A grating blazed for wavelength λ will also be efficient at $\lambda/2$, $\lambda/3$, ..., and λ/m . It is therefore necessary when using this type of instrument to remove all wavelengths up to and including half the chosen wavelength. This is difficult to achieve and has led to a considerable amount of effort to improve filters. Other spectroscopic systems such as Michelson and lamellar grating interferometers also need filters of this type but the requirement is not so severe. They are also required for some detectors but in this case the filters need to be usable at liquid helium temperature. There are three types of filter which are important:

- Short-pass filters transmit wavelengths shorter than a chosen wavelength. They are described by the wavelength at 50% of peak transmission. This is called the cutoff wavelength.

- Long-pass filters transmit wavelengths longer than a chosen wavelength. They are described by the wavelength at 50% of peak transmission. This is called the cut-on wavelength.
- Band-pass filters isolate a small wavelength interval, normally by transmission and ideally block all wavelengths outside this interval.

It should be noted that the nomenclature for filters is derived from the IR, where it is more common to refer to wavelength instead of frequency. Short-pass filters are used in the near-IR where, for example, a grating is deliberately used in a higher order. Long-pass filters are vital for THz spectroscopy as they are needed, for example, to block IR background radiation. Band-pass filters are very important for photometry, e.g., in astronomy. In an ideal filter, the transition between zero and maximum transmission is infinitely sharp. In practice, even the best filters do not approach the ideal. There are various effects which produce filtering. The most important of these are selective absorption and selective reflection, but other methods include refraction, interference, and diffraction. Filters can also be divided into those that rely on a naturally occurring effect, and those that are deliberately constructed with some periodic structure.

Absorption filters remove the unwanted radiation, usually by converting it into heat. A second advantage of an absorption filter, in common with all transmission filters, is that its orientation is not very important. They are usually mounted at right angles or with a small tilt angle in order to avoid standing waves to the axis of the radiation beam to reduce reflection losses and to make full use of their area, but a variation of $\pm 10^\circ$ is not important. A reflection filter, which takes the place of a mirror, must be much more accurately aligned and also, of course, it needs to have a uniform reflecting surface. When this type of material is used as a transmission filter, it is very important that the reflected radiation should not find an alternative path to the detector.

3.5.1 Black Polyethylene (PE) Absorption Filters

PE has been used as the basis of a range of absorption filters. The simplest of these consists of a suspension of carbon particles in PE. Carbon in one form or other has been used to filter out short wavelengths since the beginning of THz spectroscopy; sooted quartz or black paper were used for many years. Black PE – of the heavier weight type used by gardening enthusiasts – is quite suitable. However, it has been pointed out that the suspended particles scatter as well as absorb and that the black PE should be placed some distance from the detector to avoid any of the scattered radiation reaching it [155]. The transmission of a sheet of black PE is shown in Fig. 3.8 but because the density of carbon varies from one piece of PE to another, these figures are only an indication of the actual performance. A good preliminary test is to look at the sun through the material. If any sunlight is transmitted, it is not suitable. Black PE is often used with crystalline quartz to provide an opaque filter at wavelengths shorter than $40\text{ }\mu\text{m}$, but very often a small amount of radiation below $4\text{ }\mu\text{m}$ is not removed by this combination.

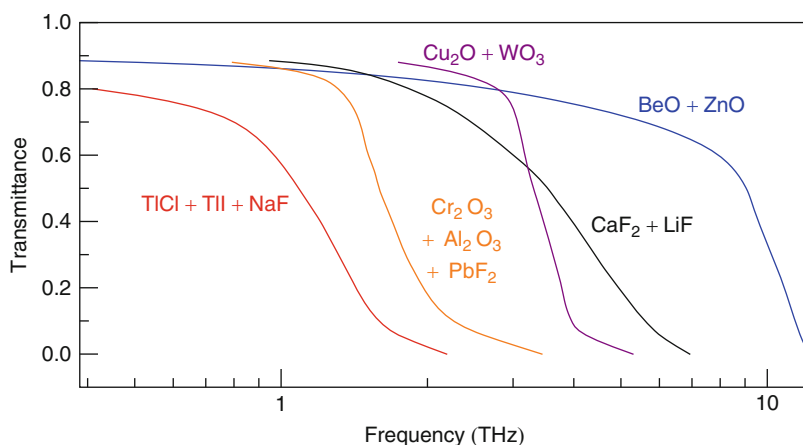


Fig. 3.9 Transmittance of several typical alkali halide and composite filters with cutoff wavelengths in the region from 25 to 300 μm (adapted from [156, 157])

3.5.2 Alkali Halide Filters (Yoshinaga Filters)

Yoshinaga and his colleagues [156] developed a very useful range of filters in which reststrahlen crystals, originally thallium bromide, were suspended in black PE sheets. By careful control of particle size and by suitable mixing of materials, filters with cut-on wavelengths from 20 to 200 μm were produced. A useful selection of these filters is shown in Fig. 3.9. Manley and Williams [157] followed up this work by using various other materials to produce a further helpful group, which is also shown in Fig. 3.9. The Japanese researchers relied on black PE to remove wavelengths below 15 μm but Manley and Williams included various organic dyes in their mixtures and produced totally opaque filters out to the cut-on point. Their longest-wavelength filter containing $\text{TiCl} + \text{TlI} + \text{Cu}_2\text{O} + \text{WO}_3$ is particularly useful when working beyond 200 μm . This type of filter is not easy to make. The particle size, the thickness, and the amount of each constituent are important, as well as the distribution within the sheet. Because many of the crystals used in these filters are ionic, there is once again a reflection problem. Another difficulty connected with all PE filters is their low melting point, which precludes their use near a thermal source in a spectrometer.

3.5.3 Scattering Filters

If a powder is suspended in another medium, the radiation passing through will be scattered, except when the refractive index of the powder and the medium are the same. Filters based on this principle are called Christiansen [158] filters.

The refractive index of ionic crystals passes through unity at a higher frequency than the resonance frequency of the crystal (which causes reststrahlen reflection) and so it is possible to make small area free standing filters. Filters depending on this principle have a limited use as their transmission rarely exceeds 50% and they are not very practical. A second type of scattering filter can be provided by using a roughened metal plate in reflection. By grinding a metal plate with suitable powder (for example carborundum) a surface is produced which will scatter short wavelengths but specularly reflect long ones. By varying the coarseness of the grinding powder reflecting plates for different wavelengths can be produced. A third way is to use a powder layer on a transmissive substrate, for example sapphire, or on a material which already has some filter functionality. The latter is especially useful for blocking near-IR leaks. The transmission characteristics of a layer of dielectric powder are determined by the size, distribution, index of refraction, and thickness of the particles, therefore allowing tailoring of the filter properties.

Ruled reflection gratings of the type used for diffraction grating spectroscopy (Sect. 6.2) can also be used as scattering filters. They are, in effect, a more sophisticated version of the roughened plate filter. With a ruled grating wavelengths shorter or comparable with the grating spacing are diffracted into various orders, but long wavelengths see the grating as a plane mirror and are specularly reflected. Reflection filters are, however, somewhat inconvenient to use as they need to be set up very accurately in an optical system. A more useful filter can be produced by employing a transmission grating ruled into a window material. PE grating filters of this type were first described in 1965 [159]. Details of the manufacture of a more sophisticated version, with crossed rulings on either side, are available [160]. The advantage of crossed rulings is that any polarization effects are avoided. The transmittance of this type of filter is shown in Fig. 3.10. With appropriate rulings, these filters are useful over the entire THz range. However, as the gratings diffract

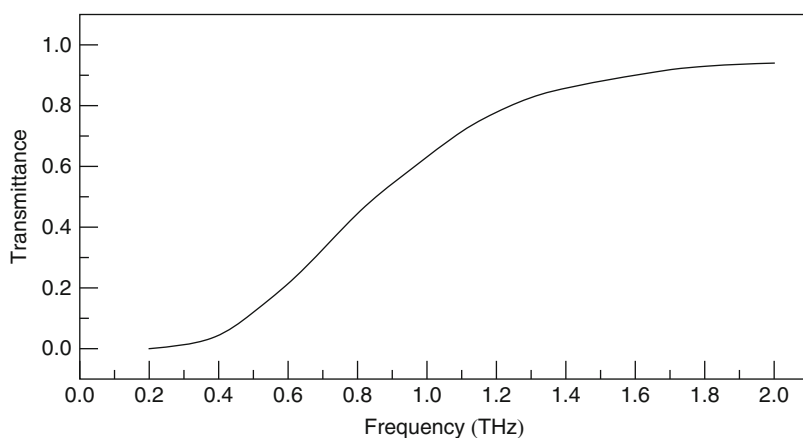


Fig. 3.10 Transmittance of transmission grating ruled into a window material such as PE

rather than absorb the unwanted radiation, care must be taken to avoid the diffracted frequencies reaching the detector.

3.5.4 Reststrahlen Filters

The reststrahlen reflectance of crystals finds a useful application in reflection filters. In fact, Rubens developed a spectrometer which provided narrow-band radiation with a wavelength as long as $51\text{ }\mu\text{m}$ by filtering the emission from a blackbody with a series of reststrahlen filters. As mentioned in Sect. 1.2, this eventually led to the discovery of the Planck radiation law. The reststrahlen reflection of several materials is shown in Fig. 3.11. Some other materials which also have a pronounced reststrahlen reflection and which have been discussed in previous sections are quartz, KRS-5, and the alkali halides.

3.5.5 Resonant Mesh Filters

Meshes are two-dimensional arrays of holes in a thin sheet of metal, or an array of metal islands on a thin dielectric substrate. Generally, one can distinguish two types: the inductive mesh and the capacitive mesh (Fig. 3.12). The name comes from the equivalent circuits, which represent the mesh. In its simplest form, the inductive mesh consists of a grid of wires or metallic stripes with diameter or breadth $2r$ and thickness d , which are separated by a distance g . The ideal inductive mesh can be modeled by an inductance, which shunts a transmission line. However,

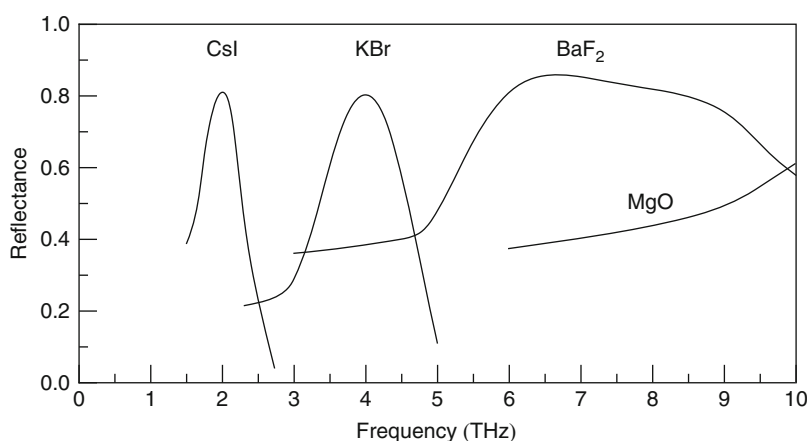


Fig. 3.11 Reststrahlen reflectance of some alkali halides (adapted from various sources)

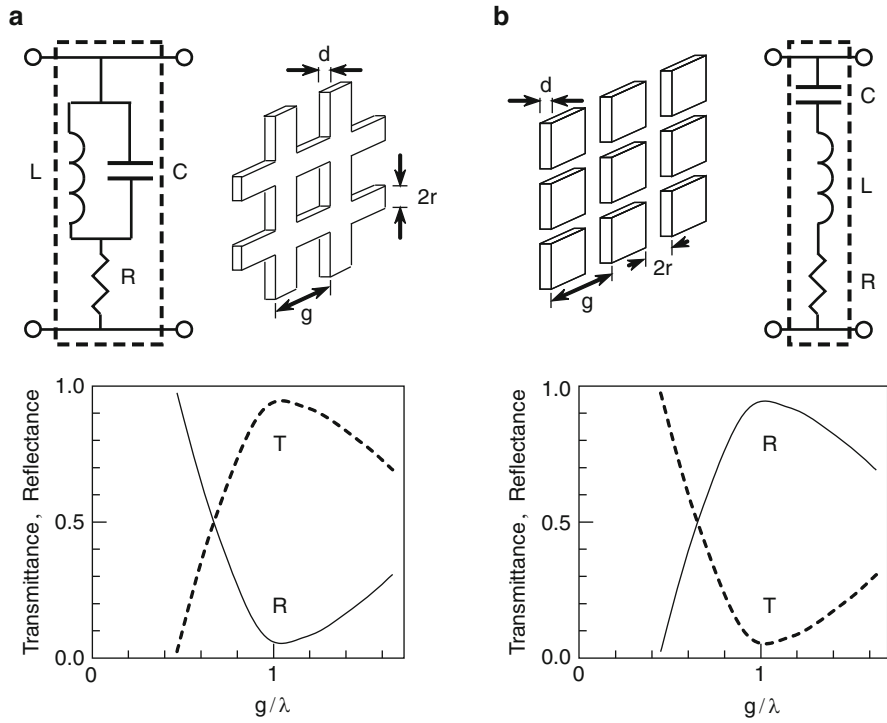


Fig. 3.12 (a) Inductive and (b) capacitive mesh filters. At the top are the structures and the corresponding equivalent circuits. The transmittance T and reflectance R are qualitatively shown as a function of the normalized frequency g/λ in the lower part of the figure

better agreement with real meshes is found when additional circuit elements are added [78]. A capacitance in parallel with the inductance is required to form a parallel resonant circuit, and a resistor is required to simulate optical losses. The inductive mesh can be either free-standing or on a supporting thin ($\sim 3 \mu\text{m}$) transparent dielectric film made, for example, from Mylar. The transmission through an inductive mesh reaches a maximum when $\lambda \approx g$. At longer wavelengths, the transmission drops steeply, while for shorter wavelengths the decrease is less pronounced. It should be noted that for an ideal inductive mesh the transmission should approach unity for $\lambda < g$. However, diffraction leads to a decrease in transmission. As the ratio r/g increases, the maximum transmission decreases and the bandwidth of the mesh becomes narrower. Approaching from the long wavelength direction the edge steepness increases with rising ratio r/g . It turns out that $2r/g \approx 1$ is the best value when using the mesh as a bandpass filter. For other values, the bandpass performance is not as good because either the bandwidth is too large or a double peak structure appears (Fig. 3.13). Sakai and Yoshida [161] have characterized single mesh narrow bandpass filters, which were optimized in this way in the IR and throughout the THz region. They found Q -values around

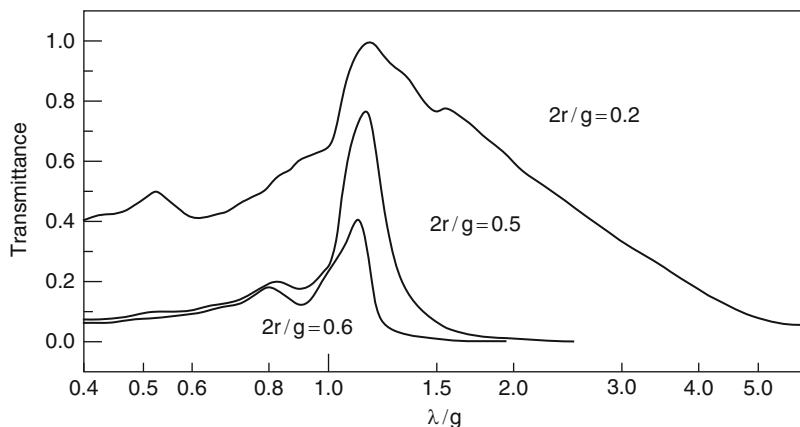


Fig. 3.13 Transmittances of several inductive mesh filters with different characteristic ratios $2r/g$ as a function of the normalized wavelength λ/g (adapted from [161])

four and a peak transmission of about 70%. This can be compared with reststrahlen filters from alkali halides, which have Q -values of about three and a transmission of around 80%. The performance of mesh filters can be improved by placing the meshes in series. If the distance is not too small, the transmittances simply multiply. By accurate spacing of the grids, an interference pattern can be produced which provides a further improvement and an even steeper transmission characteristic.

The capacitive mesh is complementary to the inductive one (Fig. 3.12b). R. Ulrich [78] has studied the properties of this type of mesh. It consists of metallic islands of thickness d , which are separated by a distance $2r$. The distance between two similar edges of the island is g . Because of this structure, the capacitive mesh is always on a dielectric film or substrate. The equivalent circuit of an ideal capacitive mesh is a capacitance shunting a transmission line. As with the inductive mesh, better agreement with measured transmission curves is obtained if additional circuit elements are taken into account. In this case, it is an inductance and a capacitance in series. At $\lambda \approx g$, its reflection has a maximum while the transmission is almost zero. At longer wavelengths, the transmission increases steeply. At shorter wavelengths, there is an increase of transmission, although it is less pronounced. Somewhat below $\lambda < g$, interaction of evanescent waves of the mesh leads to an increase of transmission. At even shorter wavelengths, the transmission increases due to diffraction [162]. This cannot be described by the equivalent circuit model.

Meshes have a sharp reflection characteristic, but care must be taken with their mounting in a spectrometer, as they also have to act as mirrors. When used at or near normal incidence a mesh produces negligible polarization and there are only very small reflection maxima below the cut-on wavelength. When used in transmission, the mesh has a fairly low transmission until just before it becomes reflecting, when it has a sharp transmission maximum.

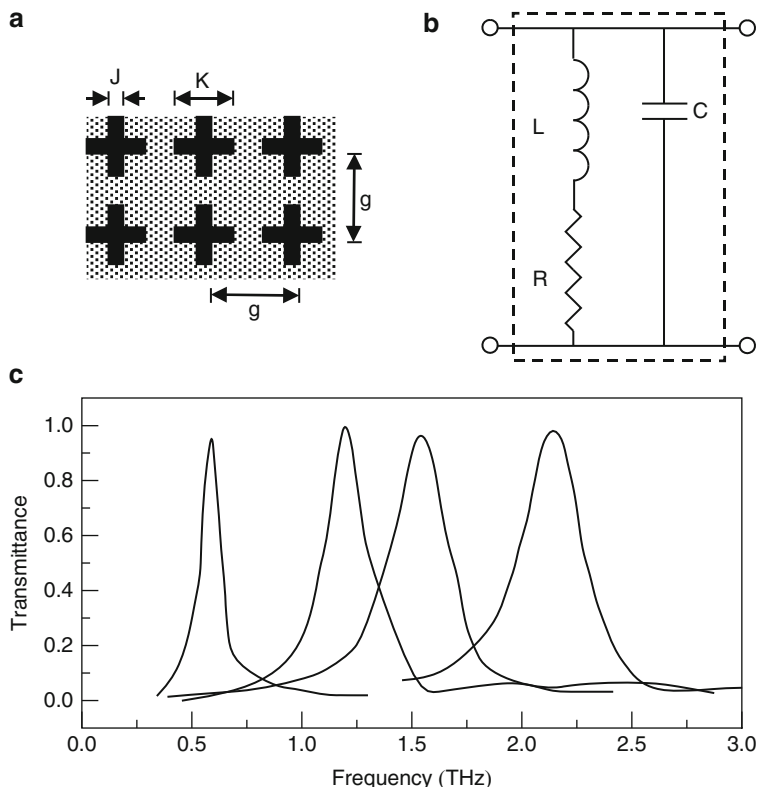


Fig. 3.14 Inductive mesh filter geometry with cross-shaped holes (a) and the corresponding (b) transmission line model of the bandpass filter. (c) Transmittance of four bandpass inductive mesh filters. The size of the holes (given in μm) varies from $g/K/J = 402/251/56$ to $113/71/19$ for center frequencies from 0.6 THz to 2.1 THz (adapted from [163])

Besides meshes with rectangular structures, a number of other designs have been reported in the literature. As an example, very good bandpass performance has been demonstrated with freestanding meshes with cross-shaped apertures, which are fabricated from a $12\text{-}\mu\text{m}$ thick electroplated copper film as shown in Fig. 3.14. Essentially, this is a combination of an inductive and a capacitive mesh. The Q value is as large as 20, the power transmission is in the range of 97–100% at the center frequency, and the out-of-band transmission is below 2% [163].

3.5.6 Thick Metal Plate Filters

The thick variant of the metal meshes described in the previous paragraphs is a metal plate with a closely spaced array of holes. This acts as an efficient short-pass filter. Such a plate can be viewed as an array of waveguides with a particular

cutoff frequency ν_c . For frequencies well below this limit, the perforated plate acts as a reflector, while well above ν_c power is largely transmitted. The damping below the cutoff frequency grows exponentially with the thickness of the plate. The cutoff frequency depends on the shape of the cross-section of the holes. For a circular waveguide with a diameter a , which is filled with a lossless dielectric with a refractive index of one, it is given by $\nu_c = 0.586(c/a)$ [164, 165]. Thus, for THz frequencies the hole diameter is about 60% of the wavelength. At longer wavelengths, these filters can be produced by drilling holes in a metal plate or foil. Other ways of manufacturing are to solder bundles of steel tubes [166], or by high-precision diamond machining of copper sheets [167]. The latter method allows rectangular and other noncircular cross-sections to be created.

3.5.7 Cold Filters

Although cold filters are physically not different from filters used at room temperature, some peculiarities are described here because of their importance for cryogenic detectors and experiments. When using helium-cooled detectors it is often necessary to remove background radiation from room-temperature sources. This can only be done by using cooled filters, which are normally placed close to the detector, where they will be at or near the detector temperature. Alkali halides, with their sharp transmission characteristic at cryogenic temperature, are often used. However, a disadvantage is the reflection losses of 20–50% which are found with these materials. These losses can be reduced with single or multilayer antireflection coatings of PE, yielding a transmittance of 90% or better over selected, and relatively narrow, parts of the THz range. In the case of hygroscopic crystals the antireflection layer also protects the crystal. Near-IR leaks can be blocked by using scattering powder layers of transparent particles. Depending on the design, transmittances from <0.1% to 80% over a wide frequency range can be achieved. Multiple filters are combined in order to achieve specific band-pass characteristics or other special features.

3.6 Antireflection Coating

It is essential to minimize the reflection loss of lenses, filters, or windows for many THz applications and systems. For crystalline materials, the refractive index is between 1.4 and 4 and the reflection loss may easily exceed 30%. Most polymer materials have a refractive index around 1.5 and the reflection loss is about 4% per surface. A single layer antireflection coating with an optical thickness of a quarter of a wavelength is often sufficient. The ideal index of refraction of the coating material for a transition from vacuum is equal to $\sqrt{n}$ with the refractive index n of the material to be coated. The reflection loss can be reduced to a few percent, and the bandwidth of a $\lambda/4$ antireflection coating is about $\pm 20\%$.

Alumina-loaded epoxy that is diamond machined to the correct thickness has been used up to 1 THz [168]. It has an effective refractive index of about 2 and is well suited to Si. This yields smooth, very rugged, cryogenically recyclable coatings, with excellent performance. At low THz frequencies Cirlex ($n = 1.84$ at 0.3 THz), a laminate of a Kapton polyimide film makes a very good antireflection coating for Si lenses [169]. Cirlex is glued to the Si lens with the epoxy adhesive Stycast. Another suitable material is Parylene C. It has a refractive index of ~ 1.62 and ~ 1.5 at 300 and 4.2 K, respectively, and can be deposited from the gas phase. Using this method, a film of almost any thickness, with very high homogeneity, can be produced. The Si lens of a quasi-optical hot electron bolometer mixer with a Parylene antireflection coating has close to 30% less optical loss than a mixer with an uncoated lens [141]. The coating survives many cryogenic cycles and is used for the lenses of the 1.4–1.9 THz quasi-optical mixers for the Herschel Space Observatory [170]. LDPE is also a suitable, although not optimal, antireflection coating. Its refractive index is ~ 1.52 at 2.5 THz. A major problem with LDPE is achieving the appropriate thickness and obtaining a good adhesion, because LDPE is available only in films of certain thicknesses. The coating needs to be applied by stretching the LDPE film around the Si sample. By this means a transmittance of 90% for a plane parallel 1.5 mm thick sample with an 18.5- μm thick antireflection coating on both sides can be achieved [171]. SiO_2 has been used for antireflection coating of Ge detectors [150, 151] and later for the facets of a QCL operating at 4.7 THz. It has good transmission with a refractive index between 1.9 and 2.1, depending on the deposition method and parameters [172].

An alternative method is to design the required dielectric by changing the geometrical characteristics at the surface of the material. This can be done in various ways but the most common is a grooved layer at the surface. When the groove spacing is much smaller than the incident wavelength (with a ratio less than 0.1), the grooves and the corresponding ridges can be considered as capacitors, with the impedance depending on the geometry of the grooves and the material's dielectric constant. Proper design leads to a simulated dielectric constant, which fulfills the $\sqrt{n}$ requirement of a $\lambda/4$ antireflection coating [173, 174].

3.7 Lenses

Lenses were used in one of the first THz experiments by Rubens and Wood in 1911 [14] and they are still a major prerequisite of THz experiments. They can be made from polymer materials such as HDPE, Teflon, TPX, or Tsurupica. These materials have in common a refractive index of about 1.4 to 1.7, which changes little across the THz range. This corresponds to reflection losses of between 3% and 7% per surface and normally antireflection coatings are not necessary. Due to the longer wavelengths in the THz region than in the visible, lenses can be made relatively easily by machining on a lathe or a milling cutter. Injection molding is also possible. Regarding the surface accuracy, the same considerations as for mirrors hold,

meaning that polishing is not required in many applications. This makes polymer lenses fairly cheap. While HDPE and Teflon are not transparent to visible light, TPX has a good transparency, and a 3 mm slab of Tsurupica has a transmittance above 90% throughout the visible range. As mentioned earlier, they also have almost the same refractive index in the THz range as in the visible. This favors the latter two materials for applications where optical pre-alignment is required. Besides polymers, crystalline materials are frequently employed. Those used most often are quartz, where some care has to be taken because it is birefringent, and Si or Ge. Their absorption decreases significantly upon cooling. With refractive indices of 2.1, 3.4, and 4, respectively, their reflection loss is rather high and antireflection coatings are required. In addition, they are expensive compared to polymer lenses and more demanding regarding fabrication. Therefore, they are mainly used in specialized systems such as quasi-optical lens–antenna systems (Sect. 3.15). Absorption is another important factor when choosing a lens for a particular experiment. For a lens of focal length f , diameter D , and refractive index n with $f^\# = f/D \gg 1$ the thickness t_c at the center of the lens can be approximated by [116]

$$t_c \approx \frac{D^2}{8f(n-1)}. \quad (3.4)$$

The fractional power absorbed on-axis is determined by αt_c where α is the absorption coefficient of the lens material. For example, a TPX lens ($n \approx 1.45$, $\alpha \approx 1 \text{ cm}^{-1}$) with 50 mm diameter and a focal length $f = 200 \text{ mm}$ has a center thickness of $\sim 3.5 \text{ mm}$ and an absorption loss at the center of $\sim 30\%$. The same lens from HDPE ($n \approx 1.52$, $\alpha \approx 2 \text{ cm}^{-1}$) has a center thickness of $\sim 3.0 \text{ mm}$ and an absorption loss of $\sim 45\%$.

A particular lens design is the zoned lens or Fresnel lens. In ordinary lenses, the phase shift is a function of the distance from the optical axis. For lenses with $f > 0$, the phase shift, as well as the lens thickness, increases toward the optical axis. This may result in quite a thick and lossy lens. A Fresnel lens is based on the independence of the focusing properties with respect to a phase change of 2π radians. The Fresnel counterpart of an ordinary lens is generated by reducing the thickness of the ordinary lens by an amount required to produce a 2π phase change, which results in a series of jumps in the lens thickness each time a 2π phase change occurs. The profile of the refracting surface of the Fresnel lens is parallel to the surface of the original lens. But even using only those surfaces that are perpendicular to the optical axis may result in a Fresnel lens with good performance. Figure 3.15 illustrates the differences between these lens types. Disadvantages of Fresnel lenses

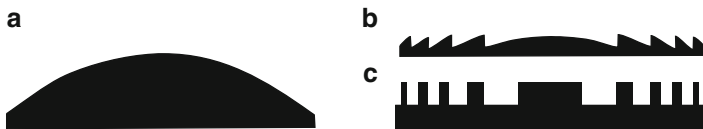


Fig. 3.15 (a) Regular lens, (b) corresponding Fresnel lens, and (c) approximated Fresnel lens

are limited bandwidth, because the phase change is strictly obeyed only at one frequency, and losses due to shadowing caused by the jumps in the lens profile.

3.8 Mirrors

Mirrors are usually of solid metal or glass with the front surface metallized. In order to obtain good reflectance, the actual thickness of the metal must be substantially larger than the skin depth d_{skin} , i.e. the distance perpendicular to the interface in which the field in the conductor drops by a factor of e . This is given by

$$d_{\text{skin}} = \sqrt{\frac{\rho}{\pi \nu \mu_0 \mu}} = \frac{1}{\sqrt{\pi \nu \sigma \mu_0 \mu}}. \quad (3.5)$$

A typical range of values for metals is 50–100 nm at 1 THz. Normal commercial mirror coatings appear to be quite satisfactory down to at least 0.3 THz. Among the metals that have very high conductivity are silver, copper, gold, and aluminum. There is, however, very little direct information available about their THz reflectance and one has to rely on calculations. Silver is not often used for mirrors because it tarnishes quickly and this is also true of copper. Gold, however, is excellent and its very high reflectance has led to its use for laser mirrors, where a 1% difference in reflectance can be very important. Mirrors of aluminum evaporated onto glass are employed as they have a high reflectance, are relatively cheap, and deteriorate very little during long periods of use. Aluminum is also a very good reflector in the visible region which assists in the “setting-up” of optical systems using visible light sources. They can also be machined with high (sub- μm) accuracy from the bulk, with the oxide, which rapidly forms on fresh aluminum, serving as a protective layer. Other metals and alloys have higher resistivities and, therefore, lower reflectance; while this is usually not important for a single reflecting surface, it becomes significant in multiple-reflection devices such as light pipes (Sect. 3.9) or laser resonators.

Another important consideration regarding mirror quality is the surface accuracy, which determines the correctness of the phase transformation produced by the mirror. There are only a few calculations or measurements available on this subject. J. Ruze [175] has performed an analysis of a reflecting antenna, which in fact resembles the problem of a mirror surface. He obtained the following expression for the gain reduction

$$\frac{G}{G_0} = 1 - \left(\frac{4\pi\sigma_{\text{rms}}}{\lambda} \right)^2. \quad (3.6)$$

Here, G and G_0 are the true antenna gain and the gain of a lossless antenna, respectively, and σ_{rms} is the “rms” surface roughness. I. Anderson has analyzed the effect of small phase errors upon transmission between confocal apertures and arrived at a similar expression [176]. For a single reflecting surface, $\sigma_{\text{rms}} \leq \lambda/16$

yields a gain reduction $G/G_0 = 0.5$. Even for one reflecting surface this is hardly acceptable and $\sigma_{\text{rms}} = \lambda/100$ corresponding to $G/G_0 = 0.98$ seems to be a better criterion.

The scattering problem can be also approached in a similar way to classical radar theory [177]. In this model, the surface roughness is described as a collection of identical flat elemental reflectors. Each has a height h from the average surface. If the surface roughness is much smaller than the wavelengths, i.e. $z = |h| \ll \lambda$ the reflected wave can be calculated as the coherent sum of the wavefronts coming from all elemental reflectors with different optical delays. In the case of a Gaussian height distribution $P(z)$ with standard deviation σ_{rms} , the specular reflection $R_{\text{spec}}(\phi, \lambda)$ can be described by

$$R_{\text{spec}}(\phi, \lambda) = R(\phi, \lambda) \exp \left(- \left(\frac{4\pi\sigma_{\text{rms}}}{\lambda} \cos(\phi) \right)^2 \right). \quad (3.7)$$

Here, ϕ is the angle of incidence and $R(\phi, \lambda)$ is the ideal specular reflectance in the absence of any surface roughness. It should be noted that this reflectance depends only on the surface roughness, and that scattering and diffraction from mirror edges is neglected. $R_{\text{spec}}(\phi, \lambda)$ can also be calculated for different forms of $P(z)$, as this is related to its Fourier-transform. Measurements of samples with different surface roughness have shown good agreement with this model [178].

Because mirrors are widely employed in place of lenses, considerable ingenuity has been used in the design of systems to overcome the disadvantages of the “change of direction” associated with a reflecting device. Frequently, ellipsoidal or paraboloidal mirrors are used. The focal length f_e of an ellipsoidal mirror is given by

$$f_e = \frac{R_1 R_2}{R_1 + R_2}, \quad (3.8)$$

where R_1 and R_2 are the distances from a point on the surface of the ellipse to the focal points. The paraboloidal mirror is a special case of the ellipsoidal one with $R_2 = \infty$. For a symmetric paraboloid, any incident ray parallel to the rotation axis passes through the focal point after it is reflected. However, a symmetric paraboloid is not very useful in many systems because of the inevitably high blockage caused, for example, by a detector placed at the focal point of the paraboloid. Therefore, an off-axis segment, called an off-axis paraboloid, is more frequently employed. Provided that $r \sin(\theta_1)/\rho \ll 1$, where r is the radius of the off-axis paraboloid, θ_1 is the angle of incidence, and ρ is the distance between the center of the parabolical section, its focal length is equal to ρ . Note that this is not simply the focal length of the paraboloid from which the off-axis paraboloid is a section. Ellipsoidal or paraboloidal mirrors are normally used for off-axis beams. Upon reflection, distortion of the amplitude and phase, as well as cross-polarization, occurs. This is particularly troublesome because systems are normally designed for fundamental Gaussian modes and such distortion degrades the beam coupling, especially when many subsequent mirrors are used. For a fundamental Gaussian

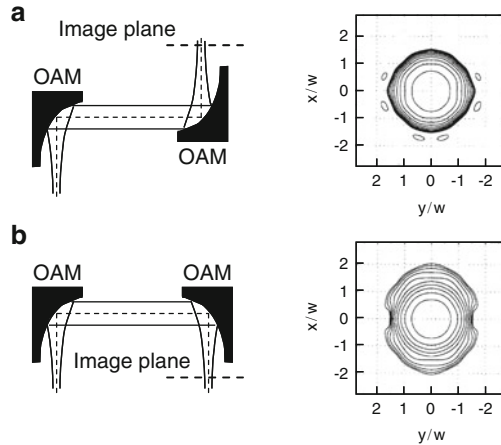


Fig. 3.16 Two configurations of a Gaussian beam telescope with off-axis ellipsoidal mirrors (OAM) illuminated with a fundamental Gaussian beam and cuts in the image planes of configuration (a) and (b). Note that in configuration (a) the beam profile in the image plane is more symmetric, because aberrations are corrected by the differently oriented mirrors (adapted from [180])

beam reflected at an ellipsoidal or paraboloidal mirror, the fraction of power that remains in the fundamental mode is

$$K_f = 1 - \frac{\tan^2(\theta_i)}{8} \left(\frac{w_m}{f} \right)^2, \quad (3.9)$$

where f is the focal length, w_m is the beam waist radius at the mirror, and θ_i is the angle of incidence. From this, it follows that a small angle of incidence and a large focal length are desirable. For example, to achieve $K_f \geq 0.99$ and to have negligible spillover losses, the diameter of the mirror needs to be four times the beam radius at the mirror, and the mirror has to have f -numbers $f^\# \geq 1.0$ and $f^\# \geq 0.5$ for $\theta_i \leq 45^\circ$ and $\theta_i \leq 30^\circ$, respectively. It should be noted that distortions can be minimized by using certain arrangements of mirrors [179, 180]. An example is shown in Fig. 3.16. In case (a), the mirrors are oriented to minimize the distortion, while in case (b) the orientation leads to much larger distortion.

3.9 Light Pipes

The light pipe is a specific THz device. It simply relies on the high reflectance of metals and is a cheap and convenient means of transferring and collecting radiation. Both condensing and straight light pipes are used in THz instrumentation and some consideration has been given to straight pipes by Ohlmann et al. [181]. They find that the transmittance T of a light pipe is

$$T = \frac{1}{2} \exp(-2q) + \left(1 - \exp\left(-\frac{1}{2} \frac{q}{f^{\#2}}\right)\right) \frac{f^{\#2}}{q} \quad (3.10)$$

with $q = (L/d)\sqrt{4\pi\epsilon_0 c \rho/\lambda}$, L is the pipe's length, d is the diameter, λ is the wavelength, $f^{\#}$ is the f -number of the incoming radiation (Sect. 2.1), and ρ is the resistivity of the light pipe material in Ω in Wm. Using (3.10) Ohlmann et al. as well as Harris et al. [182] find reasonable agreement between theory and experiment. Another approach by Fu for calculating the transmittance of a light pipe yields [183]

$$T \approx \exp(-q) = \exp\left(-\frac{L}{d} \sqrt{\frac{4\pi\epsilon_0 v}{\sigma}}\right). \quad (3.11)$$

The significant difference between the Ohlmann result and that of Fu is that the latter does not include any term for the angular spread of the beam entering the light pipe. Fu obtains good agreement with (3.11) at wavelengths between $70 \mu\text{m}$ to $570 \mu\text{m}$ but as the source was an optically excited THz gas laser this, presumably, provided a beam with small divergence.

The best light pipes are made from high-conductivity metal, or materials plated with high conductivity metal. This presents a problem when used with low-temperature systems where the light pipe is immersed in liquid helium, because metals with high electrical conductivity also have a high thermal conductivity. Therefore thin-walled stainless steel tubes are normally used, although the optical losses are higher than for high-conductivity tubes. When light pipes are required to change the direction of radiation, it is not sufficient just to bend the pipe, even on a large radius. The best method [181, 184] of producing a 90° change of direction is to construct a coupler containing a plane mirror set at 45° . This effectively receives the light from one pipe and reflects it down another one, set perpendicularly to the first pipe. Such an arrangement can be 90% efficient, while using a bend reduces the efficiency to 50% or less.

The approximations in (3.10) and (3.11) neglect skew rays and the distribution of the light at the entrance of the pipe. Consequently, they overestimate the transmittance. Hawthorn and Timusk [185] have performed a ray-tracing analysis including skew rays and different cross-sections of the light pipe, as well as 90° bends and gaps. They found that the transmittance is not exponential with respect to length because of the greater attenuation of the p component of the electric-field vector. In agreement with (3.10) and (3.11), the transmittance was larger for materials with higher conductivities and at longer wavelengths. The transmittance also depends on the initial distribution of the incident light. Light pipes with circular cross-section have a better transmittance than elliptical, square, and semicircular light pipes. However, despite some shortcomings, (3.10) and (3.11) are useful because of their simplicity. The larger the conductivity of the light pipe material and the longer the wavelength the better is the approximation.

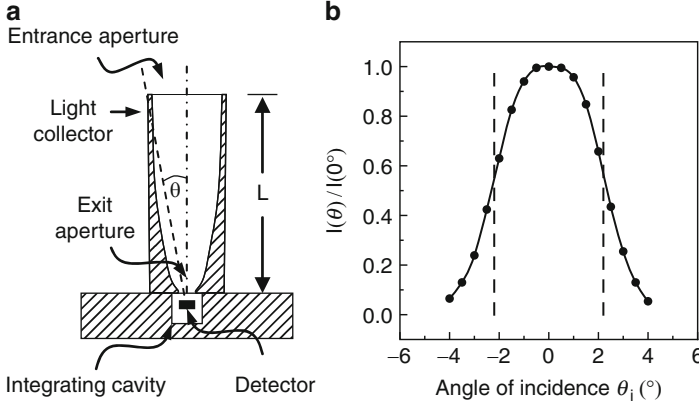


Fig. 3.17 (a) Cross section and (b) beam profile of a Winston cone with an integrating cavity. The beam pattern was measured at a wavelength of $280\ \mu\text{m}$ ($\sim 1.1\ \text{THz}$). The cutoff angle θ is 2.2° (adapted from [188])

3.10 Light Concentrators and Winston Cone

The principle of a light concentrator is shown in Fig. 3.17. It is a hollow, axially symmetric reflector. Rays coming into the concentrator at the entrance aperture are reflected and channeled to the exit aperture. In effect, it is a replacement for a mirror or a lens, because it has the same property of altering the area-solid-angle relationship of the beam. Radiation of large area A_1 and small solid angle Ω_1 enters the pipe and leaves it at a small area A_2 and large solid angle Ω_2 . In a 100% efficient light concentrator, the general condition is that the integral $\int A d\Omega$ of the projected area A over the solid angle Ω is conserved when light passes through the concentrator. The relation between the diameters of the entrance and exit apertures d_1 and d_2 and the maximum angle at which the entrance aperture accepts radiation follows from a generalization of Abbe's sine law [186]. According to this, the maximum concentration of a light beam with angular divergence θ of the concentrator is given by

$$\frac{A_1}{A_2} = \left(\frac{d_1}{d_2} \right)^2 = \left(\frac{1}{\sin(\theta)} \right)^2. \quad (3.12)$$

Diffraction effects within the concentrator can lead to losses. However, for wavelengths $\lambda \ll d_2$ the light concentrator can achieve almost the maximum concentration of radiation. Based on Abbe's sine law it can be shown that a system of lenses or mirrors that act as ideal light concentrators should have an f -number of $f^\# = 0.5$ [187]. In a practical system, the lowest achievable f -number is approximately one. Hence, the concentration efficiency of a practical lens or mirror system is about four times less than that of an ideal light concentrator. For detectors which are signal-to-noise limited, the (S/N) -ratio that can be achieved from extended sources

with an ideal light concentrator is about four times larger than with a lens or mirror system. The angular acceptance characteristic of an ideal light concentrator has a sharp cutoff at θ for meridional rays. The cutoff averaged over all rays occurs over a finite angular interval $\Delta\theta \ll \theta$. Other effects which contribute to the angular interval are diffraction and reflection losses within the concentrator [188].

A widely used design is the parabolic concentrator, which is an off-axis parabola of revolution. It is often referred to as a Winston cone after its inventor R. Winston [186]. Its length is given by

$$L = \frac{d_1(2F + 1)\sqrt{4F^2 - 1}}{4F} \approx d_1 F \quad (3.13)$$

with $F = 1/(2 \sin(\theta))$. The approximation holds for $F \gg 1$. A useful equation describing the shape of the parabola, especially for fabrication on a computer-controlled lathe is given in [189]. A beam profile of a Winston cone is shown in Fig. 3.17. As d_1 and F increase, the length of the concentrator becomes inconveniently large. A more compact design can be achieved by placing a lens in front of the entrance aperture [190]. The simplest concentrator is the linear light cone [191, 192]. It is important that it should taper slowly. A good rule is that the angle of taper should be a few times less than the acceptance cone angle at the wide end of the pipe. For example, if the cone's f -number is $f^\# = 3$ corresponding to a 10° cone angle, the taper of the pipe should be 2° to 3° . Nevertheless, the angular cutoff interval is typically an order of magnitude greater, the average number of reflections is larger, and the light concentration is smaller when compared to an ideal light concentrator.

One particular use for light concentrators is in transferring radiation onto cold detectors in cryostats. But the light concentrator has one serious disadvantage compared with mirrors or lenses: there is no focus at the exit. To overcome this, the detector must be extremely close to the cone exit or be placed in an integrating cavity at the exit aperture of the concentrator (Fig. 3.17). The cavity is designed to contain the strongly divergent beam emerging from the collector and to optimize its absorption by the detector through multiple reflections inside the cavity. It should be noted that the angular acceptance characteristic is influenced by the integrating cavity. The whole system of concentrator, integrating cavity, and detector is called a heat trap. Heat traps are used in many THz systems, starting from single-pixel detectors for laboratory use and ending in multi-pixel arrays for astronomical observations, both with ground and space-based telescopes.

3.11 Polarizers and Polarization Transformers

Polarized radiation is required for measurements of the optical constants of semiconductors and anisotropic crystals, to study Faraday rotation, and for a variety of other purposes. Several instruments such as ellipsometers or certain types of

diplexers and Fourier-transform spectrometers require a linearly polarized input and output or the rotation of polarization.

The most common way to prepare a linearly polarized beam is by using a wire grid. Metal grids have widespread application in THz optics as filters, beam splitters, polarizers, and reflectors in interference filters, lasers or Fabry–Pérot interferometers. Metal grids are made from wires with a small radius r (typically 3–25 μm) spaced by a distance g , which is called the grid constant (Fig. 3.18). The electromagnetic properties of metal grids have been thoroughly studied both experimentally and theoretically with the first work dating back to H. Lamb in 1897 [193]. Grids are transparent to radiation, which is polarized perpendicularly to the wires and for wavelengths much larger than the grid constant and the wire diameter. Radiation which is polarized parallel to the wires is strongly reflected if the wavelength is much larger than the diameter and the grid constant. Casey and Lewis [194] derived a formula for the absorbance A of a wire grid. If the wire radius is large compared to the skin depth of the grid material, this is given by

$$A = R \frac{g}{\pi r} \sqrt{\frac{4\pi\epsilon_0 c}{\sigma \lambda}}, \quad (3.14)$$

where R is the power reflectance, and σ is the conductivity of the wire. In the limiting case of $\lambda \gg g$ and negligible absorption loss, the transmittance T of a wire grid is given by [194, 195]

$$T = 1 - R = \left(\frac{2g}{\lambda} \ln \left(\frac{g}{2\pi r} \right) \right)^2 = 1 - (\tan(\Phi))^2, \quad (3.15)$$

where R is the reflectance of the grid and Φ is the phase of the reflected wave. An alternative way of creating a wire grid is by deposition of metal stripes on a thin film, for example from Mylar or PE [196]. Provided that the supporting film is lossless, which is a good assumption for a typical film thickness of a few μm , the transmittance for $\lambda \gg g$ is given by

$$T = 1 - R = \left(\frac{2g}{\lambda} \ln \left(\sin^{-1} \left(\frac{\pi r}{g} \right) \right) \right)^2 = 1 - (\tan(\Phi))^2, \quad (3.16)$$

where r is the half width of the stripes.

Another rather simple type of polarizer consists of transmission or reflection from Brewster angle plates. At the Brewster angle only that polarization component with the electric field vector parallel to the plane of incidence is transmitted. The transmitted beam still contains some contribution from the perpendicular component. To minimize this, Mitsuishi et al. [197] have used a stack of twenty sheets of PE. In principle, any material with low absorbance such as Si might be used. However, such a polarizer is a rather large device, with the additional disadvantages that it requires nearly parallel radiation and produces a lateral offset of the outgoing beam.

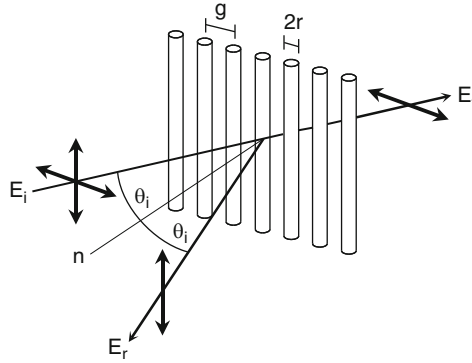


Fig. 3.18 Schematic of a wire grid and its effect on radiation of different linear polarization states. The grid consists of round wires of diameter $2r$ and spacing g . This configuration can be used to separate a beam into two orthogonal components. It can also be used to combine two beams with orthogonal polarizations, which need not to have the same frequency (adapted from [116])

The polarization of a beam can be changed by generating a differential phase shift between the two orthogonal polarization directions (Sect. 2.6). The polarization can be rotated by introducing a differential shift of π between the two directions (half-wave plate). Linearly polarized radiation can be transformed into circular polarization by a quarter-wave plate, which produces a differential shift of $\pi/4$. One way of making a wave plate is to pass the beam through a birefringent crystal, which has different refractive indices for two orthogonal polarizations. Examples are quartz or sapphire, which have to be cut perpendicular to their optical axis. The phase difference $\delta\phi$ between two polarizations after one passage through the crystal is $\delta\phi = (2\pi d \Delta n)/\lambda$, where d is the thickness of the crystal, Δn is the difference between the two refractive indices, and λ is the wavelength of the incoming beam. For example, a quarter-wave plate requires a thickness of $d = \lambda/(4\Delta n)$. For quartz at 1 THz, this corresponds to 1.7 mm, which is convenient from the point of view of fabrication and handling as well as from the point of view of absorption loss. With a quarter-wave plate the polarization state of the incoming beam can be changed continuously from linear to left- and right-handed circular. If φ is the angle between the plane of the incoming linearly polarized beam and the optical axis, the helicity of the outgoing beam, i.e. the ratio between the amplitudes of the right- and left-handed components, varies according to $P_{\text{circ}} = \sin(2\varphi)$. The outgoing beam is linearly polarized ($P_{\text{circ}} = 0$) for $\varphi = m\pi/2$ (integer m) and circularly polarized ($P_{\text{circ}} = \pm 1$) for $\varphi = (2m + 1)\pi/4$ ($m = 0$ or even: right-handed, m odd: left-handed).

Another effective method of making a low loss wave plate is the combination of a wire grid and a mirror separated by a distance d as shown in Fig. 3.19. The component of the incoming beam, which is polarized parallel to the wires, is reflected by the wires while the perpendicular component passes through the grid and is reflected by the mirror. The latter component experiences a phase delay

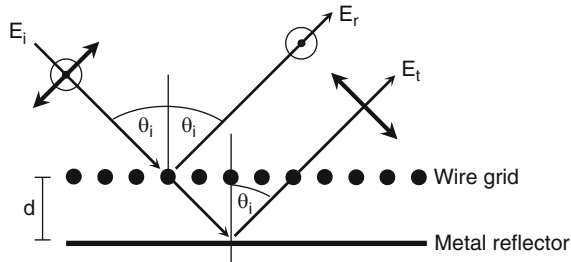


Fig. 3.19 Cross-section through a wave plate consisting of a metal reflector and wire grid separated by a distance d (adapted from [116])

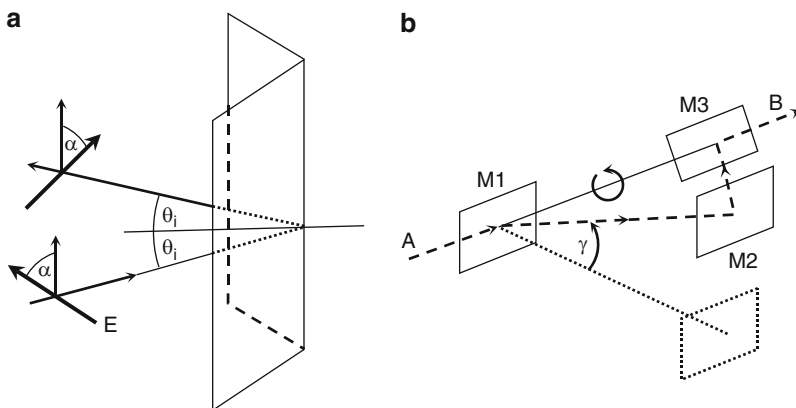


Fig. 3.20 Two polarization rotators. (a) Rooftop mirror used at non-normal incidence. The incident field direction is at angle α relative to the mirror axis and the rooftop mirror rotates the field direction by an angle 2α . (b) *K*-mirror consisting of three mirrors. The incident field enters from A along axis AB is successively reflected from mirrors M1, M2, and M3, and exits toward B. As the assembly is rotated about axis AB, mirror M2 moves in a circle. As shown, the assembly has rotated by an angle γ and thus has rotated the field direction by an angle 2γ (adapted from [116])

of $\delta\phi = (4\pi d \cos(\theta_i))/\lambda$ (θ_i : angle of incidence). For a quarter-wave plate, the distance is given by $d = (2m+1)\lambda/(8 \cos(\theta_i))$ ($m = 0, 1, 2, 3, \dots$). The absorption loss of this device is quite low. The major limitation is the lateral offset between the two polarization components, which results in an incomplete overlap of the outgoing beams.

The plane of polarization can be rotated by a rooftop mirror (Fig. 3.20). This is an example of a broadband polarization rotator and it rotates the polarization by a fixed angle. If the rooftop mirror is used at non-normal incidence, i.e. if the plane of polarization of the incident beam is tilted by an angle ϕ with respect to the edge of the rooftop, the polarization of the outgoing beam is rotated by 2ϕ . The rooftop mirror is especially important in the Martin–Puplett type polarizing interferometer (Sect. 6.4).

While the rooftop changes the direction of the incoming beam and allows only a fixed rotation, the K-mirror overcomes these limitations. It is an assembly of three mirrors that can be rotated about an axis, which is defined by the first and third mirrors (Fig. 3.20). Rotating the whole device about this axis by an angle ϕ leads to the rotation of the polarization of the outgoing beam by 2ϕ . As with the rooftop mirror, the K-mirror has basically no bandwidth restriction, but its size is larger and beam truncation could be a limitation in some systems. Besides polarization rotation, the K-mirror is useful for rotating the field of view of an imaging system, especially in astronomical instruments.

3.12 Beam Splitters

Beam splitters are key components in Fourier-transform spectrometers (FTS). They are also used for diplexing, which is the combination of beams from two sources, for example signal and local oscillator radiation in a heterodyne receiver. The most important properties that a beam splitter requires are low absorption loss, sufficient mechanical strength, and a refractive index that allows the optimum amount of splitting to be obtained. For an FTS, the optimum is $R = T = 0.5$ with R the reflectance and T the transmittance of the beam splitter. Diplexing in a heterodyne receiver has quite the opposite requirement: the transmittance of the signal radiation should be as high as possible, with a limit set by the required amount of local oscillator power that needs to be reflected from the beam splitter. The function of dielectric beam splitters is based on the interference of multiple internally-reflected beams. The wavelength-dependent reflectance R and transmittance T of a free-standing, nonabsorbing, parallel-sided dielectric material is

$$R = \frac{2(R_{p,s})^2(1 - \cos(\delta))}{1 + (R_{p,s})^2 - 2R_{p,s} \cos(\delta)}, \quad (3.17)$$

$$T = \frac{1 - (R_{p,s})^2}{1 + (R_{p,s})^2 - 2R_{p,s} \cos(\delta)}. \quad (3.18)$$

Here, p denotes parallel and s denotes vertical polarization of the electric field relative to the plane of incidence, $\delta = (4\pi nd) \cos(\theta_t)$ is the phase difference between two adjacent transmitted rays, n is the refractive index, d is the thickness of the beam splitter, and θ_t is the angle of transmission, i.e. the angle of the beam inside the film relative to the surface normal, and R is the single-surface reflectance. R and T oscillate with a period δ with R at a maximum when T is at a minimum and vice versa. For any nonzero angle of incidence, the reflectance of the beam splitter depends on the polarization of the incident beam according to

$$R_p = \left(\frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \right)^2, \quad (3.19)$$

$$R_s = \left(\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \right)^2. \quad (3.20)$$

Here, θ_i and θ_t are the angles of incidence and transmission.

The beam splitter must have a high refractive index to achieve the optimum performance of an FTS. Among the polymers Mylar is the best choice. PP and PE are less good because of their lower refractive index. But even a Mylar beam splitter is quite far from ideal. For example, the reflectance of p-polarized radiation at an angle of incidence of 45° is 0.05 and the transmittance is 0.95 (0.37 and 0.63 for s-polarized, 0.21 and 0.79 for unpolarized radiation, respectively). Another inconvenience arises from the pronounced frequency dependence of thin dielectric beam splitters (cf. (3.17) and (3.18)). Usually, several Mylar beam splitters with different thicknesses need to be employed if the whole THz range is to be covered. A Si beam splitter, several millimeters thick, is an alternative, at least for modest resolution, above about 1 cm^{-1} (30 GHz). It comes close to the optimum performance, has little polarization dependence, and the entire THz range can be covered with a single beam splitter [198]. Using a thin (several μm) Si beam splitter allows the avoidance of interference patterns, which prohibit the use of thick Si beam splitters for high-resolution spectroscopy [199, 200]. Other designs rely on Mylar films onto which thin Si or Ge films are deposited. This latter device is commercially available.

3.13 Attenuators

Many of the materials discussed in Sect. 3.3 can be used as attenuators. Sheets or discs of Teflon or PE are good attenuators but any other material with a reasonable absorbance, which does not depend strongly on the frequency, is also acceptable. An advantage of using polymer attenuators is that, due to their comparatively low refractive index, reflection and interference effects are less pronounced than for crystalline materials.

A neutral density (ND) filter is a gray filter. Ideally, it reduces radiation of all frequencies equally. They are typically used in systems where it is necessary to reduce the intensity of a transmitted beam without affecting its spectral characteristics. ND filters can be made from a crystal or polymer substrate that is coated with a semi-opaque metallic film.

For linearly polarized radiation, a rotatable wire grid can be used. The power transmission of the wire grid is $T = (\sin(\theta))^2$, where θ is the angle between the polarization of the incident beam and the wires of the grid. This allows a continuous attenuation. To avoid back reflection, the wire grid must be tilted with respect to the optical axis. In this case, the projection of the wires onto the plane orthogonal to the optical axis needs to be taken into account. The drawback of a single wire grid polarizer is a change of polarization of $(\pi/2) - \theta$. A wire grid attenuator which

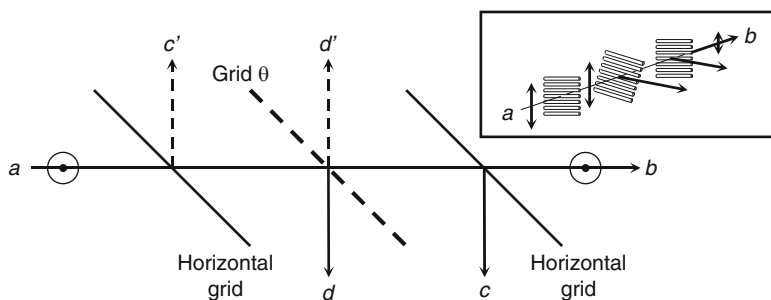


Fig. 3.21 Wire grid attenuator/power divider designed to be free of reflections back into the incident beam, which is polarized orthogonal to the plane of the paper. The three wire grids are all tilted by $\phi = 45^\circ$ to the initial beam propagation direction. The first and third grids have wires oriented horizontally, and the wires of the middle grid are rotated by a projected angle θ . With input power from port a power is coupled into ports b , c and d while if power is incident from b there can be output at ports a , c' and d' (adapted from [116])

avoids changing the polarization is shown in Fig. 3.21. This design could also be used as a power divider.

3.14 Absorbers and Absorbing Materials

Absorbers and absorbing materials have two important functions in THz systems. They are used to reduce unwanted power and stray light, and they are also employed as blackbody emission standards for system calibration. The absorbing material should have maximum absorption and minimum reflection throughout the frequency range of interest. Other common requirements are vacuum compatibility, machinability, and suitability for cryogenic temperatures. A lossy dielectric is an effective absorber provided that the reflection coefficient is sufficiently small. The reflection is predominantly due to the mismatch in the real part of the index of refraction. A rough surface on the absorber minimizes specular reflection and increases the absorption of incoming radiation. A commercially available coating is Herberts 1356H from DuPont, which is a two-component coating, which needs to be sprayed onto a surface. The coating is typically $400\text{--}440\text{ }\mu\text{m}$ thick and it forms a rough layer. Its reflection in a specular direction is below 1% at frequencies above 1 THz. Below that the paint becomes partially transmitting and interference fringes appear, which result in a modulated reflectivity of up to 3% (Fig. 3.22). Another coating can be made from a mixture of the epoxy adhesive Stycast as binder material and SiC grains of 1 mm average size as filler material [201]. The Stycast/SiC weight ratio is 1:4. While Stycast is a good absorber at THz frequencies the SiC grains reduce the specular reflection at longer wavelengths. The specular reflectance for incident angles $<50^\circ$ is below 0.1%. The reflectance of this coating increases with increasing wavelength, angle of incidence, and diminishing surface

roughness (Fig. 3.22). A material which is often used as an absorber is the CR-series of Eccosorb. It was originally developed for the microwave region and consists of small iron spheres (a few micrometers in diameter) uniformly dispersed in an epoxy base. The specular reflectance of CR110, the most useful material of the CR series, is below 0.1% in the range from 0.3 THz to 3 THz [202] (Fig. 3.22). Yet another approach to realizing absorbers is structured materials. An example is the tessellated THz radar absorbing material (RAM) from Thomas Keating Ltd. It consists of injection molded 25-mm squares with sharp pyramidal surfaces formed in conductive polymer. Its reflectance is below 1% (Fig. 3.22). Nanostructured coatings are currently being investigated to provide an even lower reflectance [203].

For measuring the reflectance, the sample under test is illuminated by radiation from a narrow band source such as a gas laser or BWO. Alternatively, it can be placed in an FTS. The reflected radiation is collected by suitable optics and focused onto a detector. Typically, the detector is placed in the direction for specular reflection. The design of the optics determines the solid angle over which reflected power is collected. The sample reflectance is determined by measuring its reflection relative to reflection from a metal mirror with close to 100% reflectance. A more sophisticated procedure is to measure the bidirectional reflectance distribution function, which describes the scattering properties of a surface. In this case, the reference sample needs to be a Lambertian surface, i.e. a nonabsorbing perfectly diffuse scatterer. This can be obtained, for example, with a gold-coated roughened surface. For any kind of reflectance measurement, care is required to eliminate any radiation leakage.

Blackbody calibration sources require very high emissivities (typically >99.9%). This can be achieved by an appropriate geometrical design, somewhat similar to a Winston cone, in combination with a highly absorbing coating of the calibrator. When designing a blackbody calibrator, several aspects must be considered. The cavity of the calibrator should have the same design as one that would trap all incoming radiation. A straight cone with an opening of the order of 20°–30° or a Winston cone are good choices. In order to avoid reflection from the tip of the cone, which is not perfectly sharp, an opening with an integrating cavity can be attached. The surface of the calibrator should be covered with a highly absorptive coating. A specularly reflecting surface is better than a diffusely reflecting one. Care should be taken to ensure that the coating is not thicker than necessary for good absorption. Thick coatings may have a temperature gradient which would diminish the accuracy of the blackbody.

3.15 Lens-Antenna Systems

The combination of a planar antenna with a substrate lens is a so-called lens-antenna system. Other terms are hybrid antenna or integrated antenna. It is a widely used approach for either coupling of THz radiation into a detecting element or for coupling THz radiation out of an emitting element [204, 205]. It is one of the best

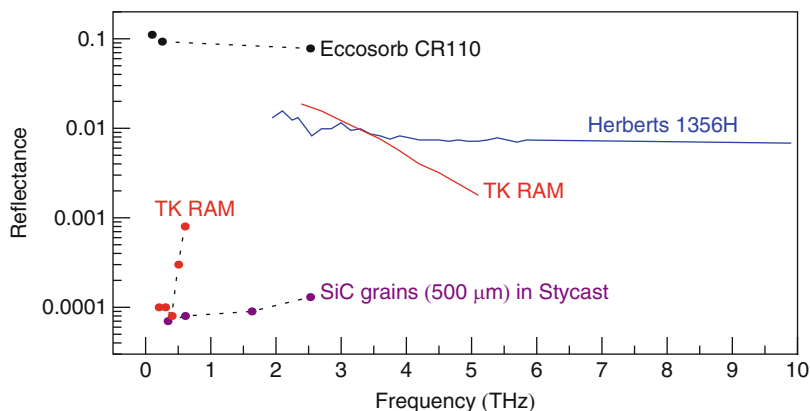


Fig. 3.22 Reflectance of some common absorbing coatings. The *points* indicate single frequency measurements with cw sources (*dashed lines* as guides to the eye). *Solid lines* indicate measurements with a Fourier-transform spectrometer (adapted from various sources)

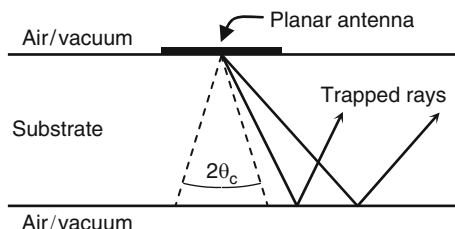


Fig. 3.23 Planar antenna on a dielectric substrate. At angles larger than θ_c the rays are trapped in the substrate due to total internal reflection ($\theta_c = 17^\circ$ for Si)

examples of how radio techniques, e.g. a planar antenna, and optical techniques, e.g. a lens, meet in the THz frequency range. In general, it consists of a planar antenna, which is attached to the flat side of a plano-convex lens with a spherical or elliptical surface. The detecting/emitting element is located at the feed of the planar antenna. The lens transforms the broad, highly diverging beam of the planar antenna into one which can be more easily coupled to other optical elements. Substrate and lens are made from the same low loss material, or from materials with similar dielectric constants, in order to minimize reflection loss at the substrate–lens interface and to avoid excitation of slab modes due to total reflection at the surface of the substrate (Fig. 3.23). Since the material of the lens and substrate usually has a high dielectric constant, an antireflection coating is applied to the lens–air interface, which minimizes the reflection loss. Lens–antenna systems are widely used in heterodyne receivers above ~ 1 THz, where their performance becomes comparable to horn antennas, which suffer from increasingly severe fabrication inaccuracies at high frequencies. In addition, they are cheaper to make. Furthermore, semiconducting emitters and detectors in time-domain spectrometers are coupled

with lens-antenna systems. This is quite natural because the generation and detection process involves a planar antenna.

The lens designs which are most often used are elliptical or hemispherical lenses with a cylindrical extension. There are several design considerations. One is how near to Gaussian is the beam that is launched or accepted by the lens-antenna system. This is defined as the coupling efficiency of the far-field pattern of an antenna compared to the far-field beam pattern of a Gaussian beam. Another issue is the directivity of the beam, i.e. how divergent it is. Often the directivity is a requirement set by the optical system of which the lens-antenna is part. In principle, there are many combinations of surface shape and extension length possible. The most common ones are described here.

An elliptical lens is characterized by its major and minor axis a , b and the foci at $\pm c$. If the eccentricity is chosen in such a way that the geometric focus becomes the optical focus, these parameters are connected via the refractive index n of the lens and substrate material by $b/a = n/\sqrt{n^2 - 1}$ and $c/a = 1/\sqrt{n^2 - 1}$. In ray optic terms, this lens produces an aberration-free focus of a collimated on-axis beam. In Gaussian optics terms, the beam is a fundamental Gaussian with a waist that is proportional to the lens diameter ($2a$), which is located at the back of the elliptical lens. The main beam is diffraction limited by the aperture of the elliptical lens and is therefore very narrow. The elliptical lens is compatible with large f -number systems because it produces a very narrow beam pattern, and it couples well to a Gaussian beam at its minimum waist, where the equiphase front is planar. An elliptical lens can be synthesized from a hemispherical lens and a spacer (Fig. 3.24). Since a hemispherical lens is cheaper than an elliptical one this is often done. There are several ways to synthesize an ellipse [206, 207]. At the tip of the lens, the curvature of the hemisphere should be the same as for the elliptical lens. The larger the radius of the lens, the larger its refractive index, and the narrower

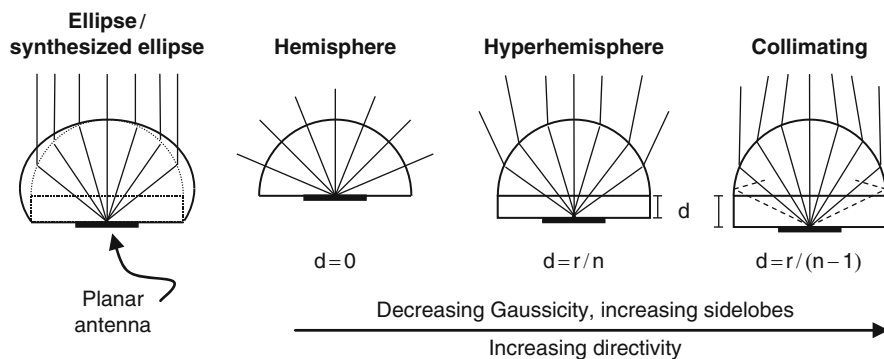


Fig. 3.24 Scheme of some common lens-antenna systems. The diagram is for an Si lens and substrate ($n = 3.42$) with a lens radius of 3 mm. The rays indicate the basic properties of the antenna. Solid lines represent emerging rays while dashed lines represent trapped rays. It should be noted that the description by ray optics only yields a qualitative result. For accurate results, the patterns of the planar antenna and physical optics need to be taken into account. The qualitative properties in terms of Gaussian optics are indicated by the arrow

the beam of the planar antenna, the better is this approximation. For a particular design, it has been shown that the Gaussian coupling efficiency decreases by less than 6% [206].

A hemispherical lens with the planar antenna at the center does not refract the rays but magnifies them by n . The somewhat divergent beam, which essentially resembles the pattern of the planar antenna, is not well adapted to most optical systems. Adding an extension to the hemisphere improves this. The properties of such lens designs depend on the length of the extension. The hyperhemisphere is a specific extended hemisphere with an extension $d = r/n$, where r is the radius of the lens. In terms of ray optics, the hyperhemispherical lens is free of circular coma and spherical aberration. It magnifies by n^2 (in antenna terms it increases the gain of the antenna by n^2), which is useful for detectors. To a first order it produces a fundamental Gaussian beam with low side lobes, i.e. it is highly Gaussian, with a small waist located behind the lens. The small waist results in relatively large far-field divergence. The hyperhemisphere couples well to a converging Gaussian beam but not to a planar equiphase front. Increasing the extension length decreases the Gaussianity of the beam but increases its directivity, i.e. the divergence angle becomes smaller. If the extension is further increased, another particular position is $d = r/(n-1)$. This is called a collimating lens because rays emitted near the optical axis are collimated, while at larger angles the rays emerge at angles which lead to an astigmatic beam. At even larger angles rays are internally reflected at the lens-air boundary. In order to overcome the reflection loss at the lens-air/vacuum interface, the lens needs to be covered with an appropriate antireflection coating e.g. Parylene for an Si lens (Sect. 3.6).

To summarize, in terms of ray optics, increasing the extension decreases the divergence of the beam and eventually leads to a collimated beam with astigmatism. Also, the effective lens aperture decreases and more radiation is lost, due to internal reflection at the lens-air interface. In terms of Gaussian optics, increasing the extension length leads to a less Gaussian beam with higher sidelobes, a larger beam waist and an increased directivity. It has been shown that the hemispherical lens yields the best coupling to an input Gaussian beam. In order to reduce the divergence of the beam an additional lens in front of the lens-antenna system is required. The elliptical lens couples less well to a Gaussian beam but, because of better collimation, no additional lens is required. In time-domain spectrometers it was found that the hyperhemispherical lens restricts the bandwidth of the system, because it exhibits large interference fringes. The collimating lens approach is superior with respect to bandwidth [208]. Which approach is preferable depends on many details, especially the design of the optical system of which the lens-antenna is a part.

There is a wide range of planar antennas, which are used in lens-antenna systems. These may be roughly divided into two groups: frequency independent broadband antennas, such as the logarithmic-spiral and the logarithmic-periodic antenna, and resonant antennas such as the double-dipole and double-slot antenna. It is beyond the scope of this book to discuss the details of these antennas. Instead, their main features will be briefly summarized and the reader is referred to a specific textbook



Fig. 3.25 Layout of three planar antennas. From left to right: logarithmic-spiral, logarithmic-periodic, and double-slot antenna. The metallized parts are shown in black. The *arrows* indicate the positions of the detecting or emitting element. Some typical dimensions are given (for explanation see text)

on antenna theory and design for details [120]. The arms of a spiral antenna are described by $R = R_0 \exp(a\phi)$, with ϕ being the azimuth angle and R being the distance from the geometric center of the spiral. The antenna receives elliptically polarized radiation with the degree of ellipticity given by the parameter a . As a rule of thumb, the outer diameter D of the maximum circle that encompasses the spiral determines the longest wavelength, and the inner diameter d determines the shortest wavelength at which this antenna still performs well. A structure of a logarithmic-periodic antenna is shown in Fig. 3.25. The antenna receives linearly polarized radiation with a polarization angle, which varies with frequency.

Dual antenna elements have a more symmetric beam pattern than the single antenna designs. The double-dipole antenna is used in semiconducting emitters/detectors of time-domain spectrometers, where the optoelectronic switch is embedded between the two arms of the double-dipole. A double-slot antenna is the inverse of a double-dipole antenna. It consists of two thin slots in a metallic ground plane. The slots are approximately λ/n long and separated by about $\lambda/(2n)$. The E -plane pattern is broad because it is determined by the coherent addition of the fields generated by the two slots, while the H -plane pattern is defined by the radiation pattern of a single slot. If the slots are connected in anti-phase the resulting radiation pattern is relatively symmetrical, having linear polarization orthogonal to the slot. It should be noted that for small antennas there may be a parasitic contribution to the beam pattern due to the wiring of the detector element located between the two slots.

The choice of planar antenna depends on the application. Obviously for broadband detectors, one of the broadband antennas will be chosen. In narrow-band applications, for example as a mixer in a heterodyne receiver, the double-slot antenna has proved to be an excellent choice. It provides a symmetric beam pattern and its impedance can be very well matched to that of an SIS or HEB mixer. However, it should be kept in mind that changing the dimensions of a double-slot antenna, in order to improve the impedance matching to the mixer, could alter the radiation pattern, which in turn needs to be taken into account when designing the lens and extension. In general, this problem can only be solved accurately with a complete electromagnetic modeling of the mixer and its radiation pattern in

conjunction with a physical optic calculation (i.e., solving the Fresnel–Kirchhoff diffraction integral) of the radiation pattern of the lens-antenna system.

3.16 Waveguides and Horn Antennas

Besides quasi-optical techniques, THz radiation can also be manipulated with waveguide components of the type used in the mmW range. The most important components are rectangular waveguides and horn antennas. Essentially, these components can be described by classical transmission line theory and the reader is referred to specialized textbooks on these topics [209, 210]. The highest frequency band for which rectangular waveguides are specified extends from 220 to 325 GHz (WR-3 band according to EIA, Electronic Industries Alliance, band designation). Thus, it just touches the lower end of the THz range. The inner dimensions of this waveguide are $0.864 \times 0.432 \text{ mm}^2$ and the tolerance is $\pm 1.27 \mu\text{m}$. The loss for a silver waveguide at the high-frequency end of this band is 11.6 dB/m. These numbers illustrate the main challenges occurring for THz waveguides, namely considerable fabrication difficulties due to the ever-smaller dimensions (typical dimensions are $0.254 \times 0.127 \text{ mm}^2$ for a band from 0.75 to 1.1 THz and 0.130 to 0.065 mm^2 for a band from 1.4 to 2.2 THz) and increasing loss due to surface inaccuracies and surface roughness. Also connecting THz waveguides repeatedly and in a reproducible manner is not a trivial task. Several fabrication technologies for THz waveguide components have been developed. Direct machining with high precision lathes and milling machines in combination with high precision metrology enables the fabrication of THz waveguide components up to 2.5 THz. These components are commercially available. Alternatively, micromachining techniques such as X-ray or ultraviolet LIGA (Lithographie, Galvanik, Abformung, in English: lithography, electroplating, molding) can be employed, especially in combination with the photoresist SU-8, which can be patterned into complicated three-dimensional structures several mm in height. Si micromachining using anisotropic etching, DRIE (deep reactive ion etching), or laser ablation is yet another possibility. Besides the considerable progress in fabrication of THz waveguides, the availability of software tools that allow the modeling of such three-dimensional components is a prerequisite for the implementation of THz waveguide components.

Waveguides and horn antennas are mostly used in radiation sources based on frequency multiplication (Sect. 4.8) up to ~ 2 THz and in heterodyne mixers up to ~ 1 THz. In the case of multiplier sources, the multipliers are integrated into waveguides and the outcoupling is done with a horn antenna. Mixers used in heterodyne receivers at frequencies below about 1 THz are integrated into waveguide structures, and horn antennas are used for coupling the radiation into the mixer. When comparing quasi-optical with waveguide-based radiation handling, it can be said that, in general, for integrated systems, which are used in a turn-key manner, waveguide structures are superior. But for experimental setups, which are

often changed and rearranged, the cheapness and the ease of handling favors quasi-optical components.

3.17 Photonic Bandgap Materials and Metamaterials

In the previous sections, optical components, such as lenses and THz filters, were described which are made from bulk homogeneous materials, for example, from Si or PE. These materials display macroscopic properties such as absorption, reflection, and transmission bands. In semiconductors, band gaps exist for which wave propagation is forbidden. However, a homogeneous material is, on an atomic level, a highly inhomogeneous structured assembly of individual atoms. Photonic bandgap (PBG) materials and metamaterials belong to the group of artificially structured quasi-bulk materials, which have properties that are basically unattainable in an unstructured bulk homogeneous material [211] (Fig. 3.26). Structuring technology introduces a second structural level where the “atoms” on that level are themselves artificial “quasi-atoms” put together from true natural atoms. These materials promise the development of unique devices such as the “perfect” lens, which is a lens without aberrations and with unlimited resolution, and even invisibility cloaks à la Harry Potter. Due to the advances in micro- and nanostructuring, PBG and metamaterials are being intensively studied to design new THz materials with previously unobtainable physical properties. Many terms used in this very active

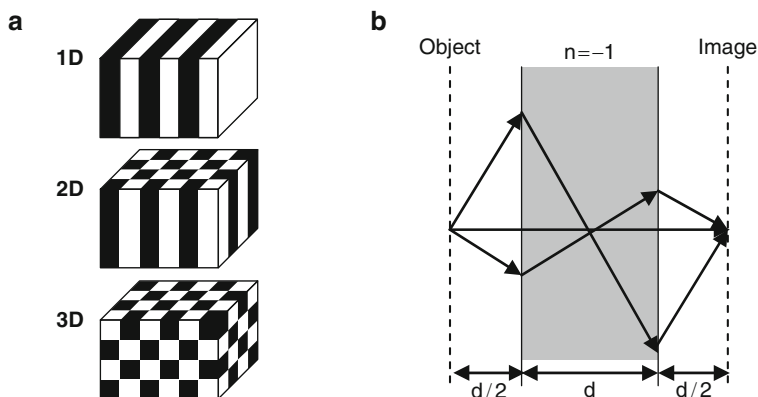


Fig. 3.26 (a) PBG materials structured in one, two, and three dimensions (1D, 2D, and 3D materials). 1D materials are, for example, multilayer mirrors or Bragg mirrors. 3D materials approximate crystal structures and can be made, for example, from stacked microscale spheres, (b) perfect lens of thickness d made from a lossless metamaterial with $n = -1$ which images all near-fields from the object plane in a distance $d/2$ in front of the lens to the image plane at $d/2$ behind the lens

research field distinguish subtle but important differences, which are described in the following paragraphs.

PBG materials exhibit, in their simplest form, a dielectric permittivity, which varies periodically with position on the order of the wavelength of the radiation inside the medium (Fig. 3.26a). A simple example for a one-dimensional (1D) PBG material in the THz frequency range is a Si multilayer mirror using a layer thickness of $\lambda/(4n)$ with n the refractive index of Si. For high reflectivity in a frequency band centered at ~ 2.3 THz ($128\ \mu\text{m}$), a stack of Si wafers with a thickness of $\sim 10\ \mu\text{m}$ was spaced by gaps close to $\lambda/4$, achieving a reflectance beyond 99.9% [212]. Multilayer mirrors are employed in laser cavities, especially Bragg mirrors are used in vertical cavity surface emitting lasers.

The name photonic crystals refers to materials which are structured in all three spatial dimensions [213]. In analogy with a semiconductor band structure or a reststrahlen crystal, artificially structured periodic media display allowed and forbidden bands of wave propagation independent of the nature of the waves. In photonic crystals stop bands or band gaps arise from multiple Bragg scattering in a periodic array of dielectric scatterers, with periodicities on a length scale of the wavelength. Due to the need for high reflectivity mirrors, for example for laser cavities, most work on photonic crystals has concentrated on achieving frequency bands in which wave propagation is forbidden. However, unusual effects also take place inside the allowed propagation bands. Just as the allowed bands in an electronic band structure contain regions where the effective mass of electrons becomes negative (at least in certain crystallographic directions), a photonic band structure contains regions where the refractive index can become negative [214].

The interest in what are now called *metamaterials* surged when Pendry [215] showed in a theoretical paper that the construction of a “perfect” lens is possible if it is built from a flat piece of metamaterial with a negative refractive index of $n = -1$ (Fig. 3.26). A perfect lens creates an exact image of an object by transferring all electric and magnetic field components, including near-field components, without loss from the object plane to the image plane (Fig. 3.26b). Such a lens can provide in principle unlimited resolution, in contrast to a regular lens, which is governed by diffraction-limited resolution (Sect. 7.2).

In textbooks on optics, the refractive index n is typically introduced as the optical density of a material with $n = v/c$ for a speed v of an electromagnetic wave in the medium relative to the speed of light in vacuum. Hence, a negative index value is, at first, counterintuitive. However, investigating Maxwell’s equations, the refractive index is introduced by the relative dielectric permittivity ε and relative magnetic permeability μ :

$$\hat{n}^2 = \varepsilon\mu, \quad (3.21)$$

where $\hat{n}$ denotes the complex index of refraction. For normal optical materials, ε and μ are positive and the positive square root can be used as $\hat{n} = \sqrt{\varepsilon\mu}$. Most materials have a low response to the magnetic component of the electromagnetic field so that the equation further simplifies to $\hat{n} = \sqrt{\varepsilon}$ for $\mu = 1$. To account for absorption, the

refractive index is a complex number and, in anisotropic materials, the index turns into a tensor.

The negative sign of the square root ($\hat{n} = -\sqrt{\epsilon\mu}$), the negative refractive index, was considered by Veselago, who showed that a material with such a property requires a negative permittivity and permeability, simultaneously [216]. The dielectric permittivity ϵ and the magnetic permeability μ describe the macroscopic response of a homogeneous medium to applied electric and magnetic fields. Atomic and molecular oscillators are not discernible in the wave's response because the fields are sufficiently averaged, both in time and on spatial dimensions of the wave's wavelength. The wave responds to the macroscopic resonances of the material and the microscopic properties are not resolved. In metamaterials, these oscillators are resonant structures, but typically much smaller than the wavelengths for which they have been designed. This is in contrast to photonic crystals or band-gap materials, which are typically periodic on the length scale of the wavelength (Fig. 3.26). In metamaterials, the periodicity is not important since the response is defined mainly by the single scatterer resonances and the structure can be treated in the context of a so-called *effective medium*.

Figure 3.27 illustrates the wave propagation in materials and metamaterials by distinguishing four quadrants depending on the sign of the two parameters ϵ and μ .

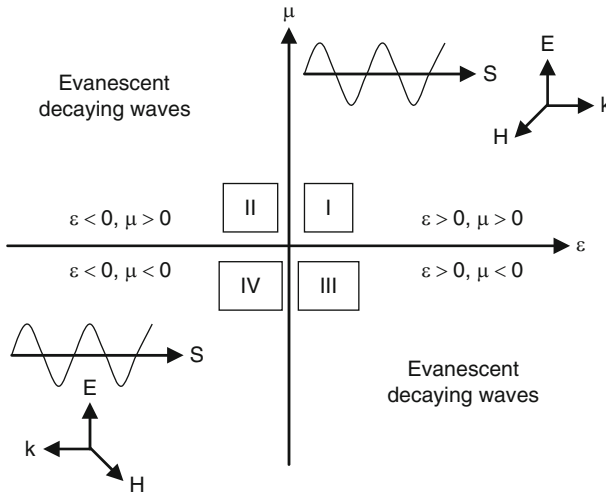


Fig. 3.27 Permittivity ϵ and permeability μ define four quadrants I to IV by their sign. Quadrant I describes normal optical materials and right-handed propagating waves for which the wavevector k , defined relative to the electric and magnetic fields, E and H , is oriented in the same direction as the Poynting vector S corresponding to the energy flow. Quadrant II describes electrical plasma and evanescent decaying waves inside materials, for example in metals. Quadrant III, in a similar way, describes evanescent decaying waves in magnetic materials, while quadrant IV corresponds to negative refractive index materials, left-handed propagating waves and artificial metamaterials. Here, the wave vector, and with it the phase velocity, has the opposite direction to the Poynting vector

Negative index metamaterials (NIMs) are also called left-handed materials (LHMs), which are related to the reversal of the wavevector with reference to the electric and magnetic field of the wave (Fig. 3.27). The concept of left-handed materials, in which the wavevector is reversed, can be traced back to so-called “backward” waves discussed as early as at the beginning of the twentieth century [217]. In 1904, Lamb [218] considered linear mechanical structures which approximate PBG materials, and Poklington [219] studied a linear mechanical model that provided backward waves.

Oscillators operating with backward waves are well known in the THz range and are discussed in Sect. 4.9. These devices contain a slow-wave structure, a particular set of periodic structures used in microwave tubes. Slow-wave structures slow down the electromagnetic wave so that they can interact with electron beams drifting at the same velocity.

Since the 1950s, the driving force for artificial dielectrics has been radar technology with its need for high permittivity, low loss materials for shielding, guiding, or reflecting waves. Periodic resonant structures are easily fabricated in the THz frequency range and resonant filters have been developed since the 1960s, introduced by Vogel and Genzel [220], and Ulrich [78], to tailor the response to electromagnetic waves. Metamaterials made of periodically placed resonant units of metal wires or squares (although periodicity is not required for metamaterials) resemble resonant filters (Sect. 3.5.5). However, metamaterials avoid the grating modes of a resonant filter because the resonant units are much smaller in size or more densely packed than the wavelength. If the characteristic lengths of the metamaterial structure are much smaller than the wavelength, it can be treated mathematically as an effective medium with an effective refractive index n , which can be negative for properly designed resonant units.

Experimentally, it is necessary to build a resonance into a material to obtain negative index values (Fig. 3.32). Such a resonance is typically connected to losses, for example, an absorption line or a plasma resonance edge in the spectrum. The theoretical concept of a perfect lens is best approximated by the so-called “super-lens” made from an etalon-shaped slab containing a resonant metamaterial. Such a lens permits the transfer of sub-wavelength sized near-fields from object to image plane, thus allowing sub-diffraction imaging, but often only in a limited frequency band close to the resonance frequency of the metamaterial, which also introduces losses. Since the year 2000 several concepts have been derived and proved experimentally, such as a near-field magnifying glass, and a cloak which renders an object invisible, at least over a small frequency band.

The need to introduce a loss mechanism for a metamaterial follows from causality. According to

$$\mathcal{E} = \frac{1}{2}\varepsilon_0\varepsilon\mathbf{E}^2 + \frac{1}{2}\mu_0\mu\mathbf{H}^2 \quad (3.22)$$

the static energy density $\mathcal{E}$ of the wave, in a material without dispersion, would be negative if both ε and μ were negative, which contradicts causality. However, if a medium with dispersion is considered, the parameters $\varepsilon(\omega)$ and $\mu(\omega)$ are functions

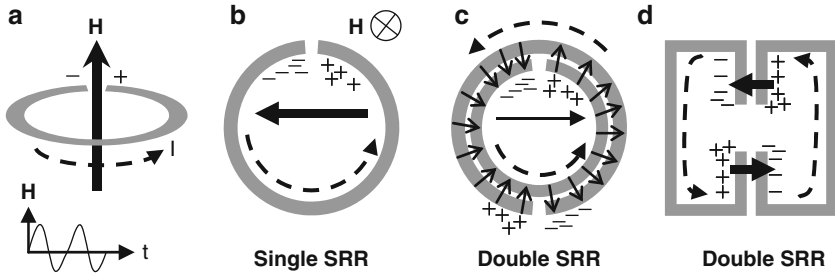


Fig. 3.28 (a) An oscillating magnetic field $H(t)$ induces in a split-ring resonator (SRR) an oscillating current I (dashed arrow) with negative and positive charge accumulation at the gap, (b) top view of (a) illustrating the electrical dipole moment formed due to the accumulated charge, (c) two SRRs with opposite slits reduce the net electrical dipole moment, (d) reshaped SRRs reduce the net electrical dipole moment further, due to the symmetric geometry, to achieve mainly a magnetic response

of frequency. For narrow band radiation, the energy density can be expanded in a Taylor series. Retaining only the linear term gives the energy density in the transparency regions at frequencies far away from resonance as:

$$\mathcal{E} = \frac{1}{2} \epsilon_0 \frac{\partial(\omega \epsilon)}{\partial \omega} \mathbf{E}^2 + \frac{1}{2} \mu_0 \frac{\partial(\omega \mu)}{\partial \omega} \mathbf{H}^2. \quad (3.23)$$

This energy density is positive even if both parameters are, simultaneously, negative in the region of dispersion. Dispersion is typically connected to dissipation and losses. Large changes in dispersion are often limited to a small frequency band, for example, close to an absorption or resonance line.

The simplest approach to achieve negative permittivity is the use of very thin wires. Negative permeability can be realized by so-called “split-ring resonators” (SRRs), which are short pieces of wire bent into a circle, so that the wire ends form a small slit or capacitor (Fig. 3.28). A resonant semiconductor split-ring is obtained if an SRR is deposited on a semiconductor. This structure is very attractive in the THz frequency range because it can be manipulated by modifying the semiconductor properties by light [221], current [222, 223], temperature [224], or with an external magnetic field [225]. An optical laser pulse can create an electron–hole plasma in the gap of the SRR by which the ring appears fully conducting and the charges cannot accumulate at the gap, i.e. the capacitance is shorted out [221] and the THz metamaterial resonance is lost. Using highly doped n-GaAs as a substrate, each metal structure of the metamaterial forms a Schottky contact. If all the metal structures are connected by a very thin wire which does not alter the resonance of the individual resonators and if, in addition, an ohmic ground contact is formed on the substrate, it is possible to apply an electric field to the structure. This field modifies the charges at the resonant gaps and it is possible to modulate the THz metamaterial resonance by the external electric field [222, 223].

The deposition of two SRRs with slits positioned opposite to each other (double SRR) allows current flow via capacitive coupling between the two rings (Fig. 3.28b). This structure is often used because the net electrical dipole moment is lower compared to a single SRR, so that in the main a magnetic response is observed. Metal layers in resonant rings need at least a thickness comparable to the skin depth to interact strongly with the incident radiation ((3.5) in Sect. 3.8).

Metamaterials mainly influence near-field properties at which far-field descriptions, such as the refractive index of a plane wave, may deviate from common experience resulting in, for instance, negative refractive index. Electric fields in the near-field region, or evanescent waves, are damped quickly on a length scale corresponding to the radiation wavelength. They do not propagate unless converted by resonances in the surface of a material, or by means of surface-plasmon polariton generation. As a consequence, a metamaterial needs strong resonances to function, and typically anomalous dispersion. The main challenge is to produce a metamaterial, which can provide unusual optical properties over a large frequency band. Broadband metamaterial properties may be obtained by chiral metamaterials.

It has been shown that an array of thin wires could be described with an equivalent plasma frequency [226]. Plasma oscillations were first found experimentally by Tonks and Langmuir in 1929 [227]. Within the Drude model, it was shown that the plasma frequency ω_p (Sect. 2.4) as a macroscopic quantity can explain a variety of physical phenomena [228].

Alternatively, a plasma oscillation in a free electron gas can be described by a quasiparticle, the plasmon. If the plasmon is confined to a surface, for example in a metal, the term surface plasmon is used. In metamaterials, an electromagnetic wave couples to a resonant metal structure on a surface, which gives rise to surface-plasmon polaritons or surface-plasmon resonances (Fig. 3.29). The permittivity which describes a plasma can be negative below the plasma frequency, explaining the good reflectance of metals far below the plasma frequency. Above the plasma frequency, the transparency of bulk metals increases, for example, at ultraviolet and shorter wavelengths. Different particle sizes in metal colloids or glasses lead to the well-known appearance of different colors, for example, in church windows. Localized surface-plasmon resonances appear due to the metal micro- and nanoparticles. Such resonances are relevant in the context of near-field microscopy (Sect. 7.5), Mie and Rayleigh scattering at surfaces.

Research based on the theoretical models of Pendry et al. [229] led to the choice of thin wires for producing negative permittivity and negative refraction experimentally [230, 231]. The thin metal wire medium is one of the main building blocks for negative permittivity (Fig. 3.30a). A popular alternative in the THz frequency range are metal structures, which contain a capacitor structure within a wire frame (Fig. 3.30b). Such structures can be illuminated in normal incidence and have a negative permittivity. The periodicity of typically fabricated THz metamaterial structures, such as SRRs, ranges from 36 to 54 μm with somewhat smaller outer dimensions of the resonant structure ranging from 26 to 40 μm . The gap at which the charges accumulate, and the wire widths of SRRs, are typically 2–4 μm wide. The resonances then occur in the range from 0.5 to 2 THz [232].

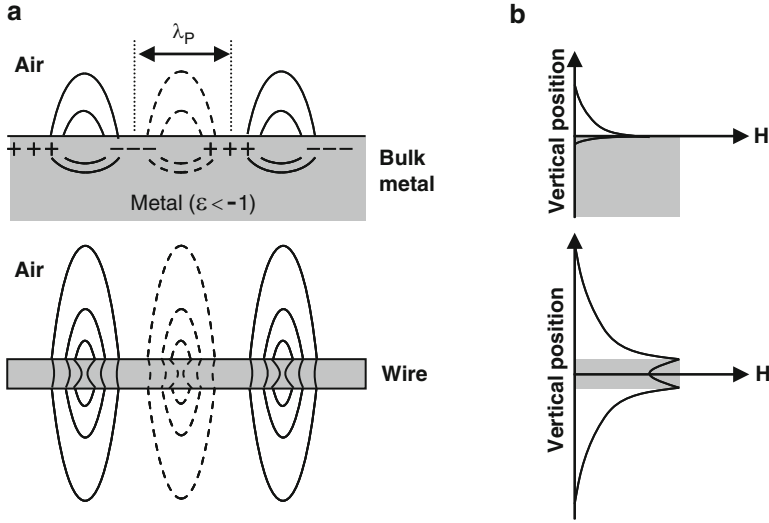


Fig. 3.29 (a) Surface plasmon of wavelength λ_p at an air/metal interface with a negative dielectric constant $\epsilon < -1$ in the metal, (b) corresponding magnetic field strength H in air, in bulk metal, and along the metal wire. Such a metal wire can transport a wave, for example, to a small detection device

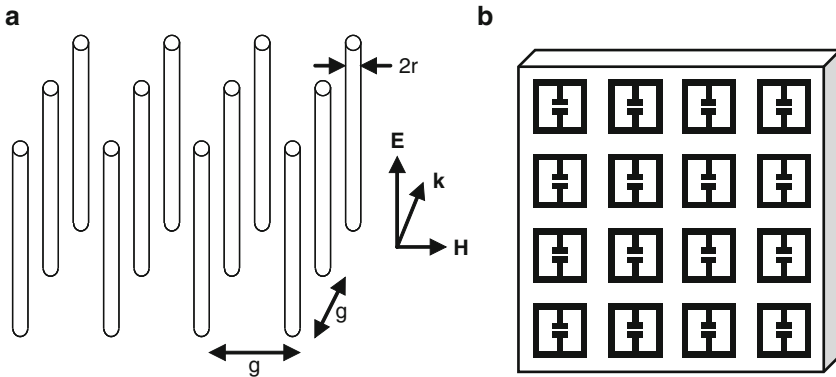


Fig. 3.30 (a) Thin wire array which approximates the physical properties of a low-frequency plasma resulting in negative permittivity for an electric field incident parallel to the metal wires of radius r and periodicity g . (b) Typical metamaterial structure of resonant units used in the THz frequency range. The units contain a capacitive gap in the center connected with a wire inductance leading to an oscillating current flow from one side of the capacitor to the other side

The wire structure is set up in such a way that the wire radius r is much smaller than the lattice constant g and the wavelength λ ($\lambda \gg g \gg r$). In the limit of long wavelengths, or small geometric dimensions, the radiation fails to resolve the wire structure and the medium can be considered as an effective medium independent

of microscopic quantities [211, 229]. The electrons are confined so as to move only in the metal wires due to the applied electric field, which is oriented parallel to the wires. The wave cannot sense the individual wires so that an effective electron density results as

$$N_{\text{eff}} = \frac{\pi r^2}{g^2} N_e = f N_e \quad (3.24)$$

with N_e , the actual density of conduction electrons, and the filling factor $f = \pi r^2/g^2$.

Thin wires have a large inductance. Considering induction and current flow in the individual wires an effective mass for the electron can be derived as

$$m_{\text{eff}} = \frac{\mu_0 r^2 N_e e^2}{2} \ln(g/r). \quad (3.25)$$

The plasmonic mode for this system described by the plasma frequency follows

$$\omega_p = \sqrt{\frac{N_{\text{eff}} e^2}{\varepsilon_0 m_{\text{eff}}}} = \frac{c}{g} \sqrt{\frac{2\pi}{\ln(g/r)}}. \quad (3.26)$$

Equation (3.26) depends only on the geometry and not on the electron density, or in general it does not depend on microscopic parameters. Therefore the thin wire metal model can be recast in terms of capacitance and inductance per unit length. The oscillating electric field of the incident wave induces a current in the wires

$$E = i\omega L I = i\omega L \pi r^2 N_e v \quad (3.27)$$

with the electron velocity $v = -i\omega r$. The polarization per unit volume is given as

$$P = -N_{\text{eff}} e r = -\frac{E}{\omega^2 g^2 L}. \quad (3.28)$$

The inductance L can be estimated by the flux per unit length Φ passing through a plane between the wires, assuming that at the mid-point between wires the field is zero:

$$\Phi = L I = \frac{\mu_0 I}{2\pi} \ln\left(\frac{g^2}{4r(g-r)}\right). \quad (3.29)$$

With the polarization $P = (\varepsilon - 1)\varepsilon_0 E$ in the limit $r \ll g$, the effective permittivity is obtained

$$\varepsilon(\omega) = 1 - \frac{2\pi c^2}{\omega^2 g^2 \ln(g/r)}, \quad (3.30)$$

which is the identical result for the microscopic derivation of the plasma frequency for such a medium (3.26). A finite conductivity σ of the wires modifies the electric field to

$$E = i\omega LI + I/(\sigma\pi r^2), \quad (3.31)$$

and the effective permittivity follows as

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega(\omega + i(\varepsilon_0 g^2 \omega_p^2 / (\pi r^2 \sigma)))}. \quad (3.32)$$

In similarity with (2.55), the damping constant γ is related to the conductivity and the plasma frequency (2.57) but includes the geometric filling factor f :

$$\gamma = R/L = \varepsilon_0 \omega_p^2 / (f\sigma). \quad (3.33)$$

The permittivity of a thin wire structure is displayed in Fig. 3.31. The anisotropy of such a structure, i.e. with response along the wires but without response perpendicular to the wires, can be circumvented by combining wires in three dimensions, resulting in an isotropic effective medium behaving like a low-frequency plasma.

Figure 3.32 shows the permeability constant for a resonance in a double SRR (Fig. 3.28c) with different electrical conductivity for the metals used. The ring radius r is assumed to be much larger than the spacing d between the two SRRs. In the gap of the SRR, the material is described by the dielectric constant ε . The effective permeability of a double SRR is given by

$$\mu_{\text{eff}} = 1 + \frac{f\omega^2}{\omega_0^2 - \omega^2 - i\gamma\omega} \quad (3.34)$$

with filling factor f , the resonant frequency is

$$\omega_0 = \sqrt{\frac{3d}{\mu_0 \varepsilon_0 \varepsilon \pi^2 r^3}} \quad (3.35)$$

and, due to the metal resistivity ρ , the damping parameter is

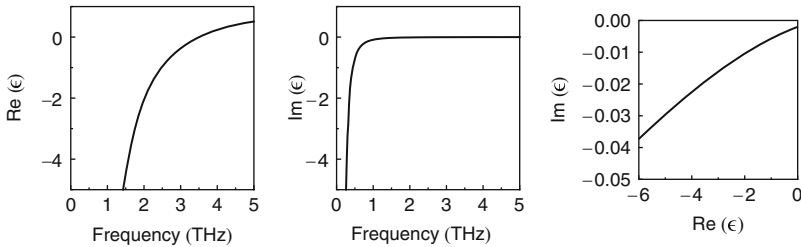


Fig. 3.31 Real and imaginary parts of the dielectric constant of an effective medium made of thin wires with a wire radius of $r = 1 \mu\text{m}$, a wire spacing of $g = 100 \mu\text{m}$, and electric conductivity of gold at 273 K: $\sigma = 4.9 \times 10^7 / (\Omega\text{m})$. These parameters lead to a plasma frequency of 3.5 THz

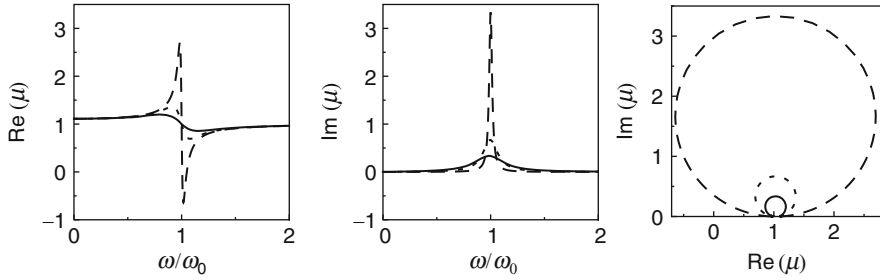


Fig. 3.32 Real and imaginary parts of the permeability constant near the resonance frequency ω_0 with high, medium, and low damping constants (*solid, dotted, and dashed lines*). This resonant behavior occurs for split-ring resonators with low, medium, and high electrical conductivity of the SRR, respectively

$$\gamma = \frac{2\rho}{\mu_0 r}. \quad (3.36)$$

If the damping and the losses are low, it is possible to achieve negative values of the permeability for $\omega > \omega_0$ similar to the regions of anomalous dispersion in permittivity. Then the response is out of phase with the driving magnetic field and μ_{eff} is negative up to the “magnetic plasma” frequency

$$\omega_M = \sqrt{\frac{3d}{(1-f)\mu_0\epsilon_0\epsilon\pi^2r^3}} = \frac{\omega_0}{\sqrt{1-f}}. \quad (3.37)$$

From the magnetic plasma frequency, it can be seen that the filling factor f determines the bandwidth for which μ_{eff} is negative and that the response can be tuned by the permittivity ϵ . A finite resistivity broadens the peak, while a very resistive material can prevent the formation of negative permeability (Fig. 3.32). A combination of SRRs and thin wires can result in negative permittivity and negative permeability, simultaneously [230]. Current research is concerned with minimizing losses, at least in the desired frequency band, by a proper design, and broadening the frequency band for which such unusual metamaterial properties can be obtained. The electrical or optical manipulation of metamaterial resonances, for example by using semiconductor substrates, opens a window to spatial and frequency modulators, phase shifters, or pulse shaping of THz signals.

Chapter 4

Sources

4.1 Introduction

Many THz experiments involve some type of source. This may be the object of the investigation, for example in astronomy or the study of molecular emission in the atmosphere, but more often it is an integral part of a spectroscopic system. Until about 1950, the only available sources were thermal sources. These were either certain heated solids that retain good emissivity from the IR into the higher frequency portion of the THz region or the hot plasma of discharge lamps. The latter have the advantages of operating at much higher temperatures than the solids and of behaving as almost perfect blackbodies over much of the THz range. The 1950s saw the first inroads by electronic beam sources similar to those already widely used in the mmW region, culminating in the very successful backward wave oscillator (BWO). Relatively soon after the invention of the laser in 1960 electrically excited gas lasers were providing significant power over much of the THz range. These were followed by the extremely useful optically excited gas lasers in 1970. By using different gases thousands of narrow-band lines, many in cw mode, can be produced, covering the entire THz region. The sixties also saw the birth of solid-state negative-resistance devices, for example, IMPATT and Gunn diodes. While primarily for lower frequencies their power level is such that at high harmonics, they provide sufficient power to act as local oscillators for room temperature heterodyne systems operating to above 1 THz.

Later in the 1900s, fundamental solid-state sources became available. These include p-Ge lasers tunable over a wide range of THz frequencies and quantum cascade lasers (QCLs), originally operating in the near-IR but now available at THz frequencies. There are various methods for mixing sources of two wavelengths to provide useful power. One example is a very compact tunable source, which can be made by mixing two short-wavelength lasers in a nonlinear element. Initially, this required two pump lasers but these can be replaced by a two-color semiconductor laser, which generates two closely spaced visible or near-IR lines simultaneously. An alternative system, that has been widely used, is the mixing of an optically

excited, fixed frequency THz gas laser with a tunable mmW source such as a Gunn diode. Because of the availability of many cw laser lines, this allows tuning over almost the entire THz region.

Electron storage rings, although primarily designed to provide radiation in the X-ray and vacuum UV spectral regions, also produce significant power throughout the THz range. This is typically one to two orders of magnitude greater than the power from thermal sources. The output is normally incoherent but at longer wavelengths it is possible to achieve coherent radiation with much higher power levels. Coherent power is also available from THz free-electron lasers (FELs), and there are a number of FEL facilities available where visitors can carry out experiments.

THz time-domain spectroscopy (THz TDS), which is discussed in Sect. 6.6, requires extremely short high-power pulses. Initially, these were from mode-locked Nd:glass or YAG lasers, with pulse widths of a few ps. They were followed by mode-locked dye lasers with less than 100 fs long pulses. Self-mode-locked Ti: sapphire oscillators and fiber lasers have reduced this to a few fs, allowing THz TDS to be extended into the IR spectral region.

4.2 Thermal Emission

The ideal thermal emitter is the blackbody radiator. Full derivations of the equations appropriate to THz radiation are readily available, e.g. [233], and only the important features are included here.

Planck's radiation formula states that the brightness radiation $B_\nu^P d\nu$ per unit source area between frequencies ν and $\nu + d\nu$ is given by

$$B_\nu^P d\nu = \frac{2h\nu^3}{c^2} \left(\exp\left(\frac{h\nu}{k_B T}\right) - 1 \right)^{-1} d\nu. \quad (4.1)$$

The photon energy $h\nu$ in relation to the thermal energy $k_B T$ is important to derive the Rayleigh–Jeans or Wien approximations $h\nu/(k_B T) \approx 48 \text{ K/THz } (\nu/T)$. For high temperature sources at low frequencies where $h\nu \ll k_B T$, the approximation $\exp(x) - 1 \approx x$ is used and (4.1) reduces to

$$B_\nu^{\text{RJ}} d\nu = \frac{2k_B T \nu^2}{c^2} d\nu, \quad (4.2)$$

which is the Rayleigh–Jeans approximation. For low temperatures at high frequencies, $\exp(x) - 1$ is approximated by $\exp(x)$, giving the Wien formula.

In laboratory spectroscopy, where source temperatures typically vary between 1,000 to 5,000 K, (4.2) is sufficiently accurate for most of the THz range. But in many of the more recent applications of THz techniques, particularly in astronomy

and atmospheric spectroscopy, source temperatures can be very low and Planck's formula must be used.

For detectors, it is common to integrate the brightness over one hemisphere which increases the photon flux S to the detector by 2π , i.e. $S = 2\pi B_\nu$. Many of the detectors used in THz research respond to the photon flux rather than the power falling on them. Because the number of “photons per watt” is inversely proportional to frequency and nearly all data graphs for photon detectors are plotted in terms of watts, this can cause confusion in assessing the quality of a detector for use over a range of frequencies. What looks like a rapid reduction in performance as the frequency is increased can be very misleading and it is convenient to rewrite (4.1) in terms of the photon flux ($N_\nu = B_\nu/(h\nu)$):

$$N_\nu^P d\nu = \frac{2\nu^2}{c^2} \left(\exp\left(\frac{h\nu}{k_B T}\right) - 1 \right)^{-1} d\nu \quad (4.3)$$

and (4.2) becomes

$$N_\nu^{RJ} d\nu = \frac{2k_B T \nu}{hc^2} d\nu. \quad (4.4)$$

In applications, the radiation within a small solid angle Ω_S is of most interest, and this can be obtained from the expression

$$\Omega_S = 2\pi(1 - \cos(\theta)), \quad (4.5)$$

where θ is the linear half-angle of the radiation and Ω_S is in steradians. The energy is derived from the brightness by integrating over the full solid angle of 4π sr. For very small angles θ the cosine term reduces to $\cos(\theta) \approx 1 - \theta^2/2$ and (4.5) becomes

$$\Omega_S \approx \pi\theta^2. \quad (4.6)$$

In situations where the Rayleigh–Jeans approximation is nearly valid, i.e., when $h\nu < k_B T$ rather than $h\nu \ll k_B T$, it is often convenient to apply a correction factor to (4.2) instead of using the Planck formula. Figure 4.1 is a plot of the required correction factor:

$$\frac{B_\nu^P}{B_\nu^{RJ}} = \frac{N_\nu^P}{N_\nu^{RJ}} = \frac{h\nu}{k_B T} \left(\exp\left(\frac{h\nu}{k_B T}\right) - 1 \right)^{-1}. \quad (4.7)$$

In practice, no source achieves the output of a blackbody, although some come very close, particularly in the visible and near-IR regions. The total output of a source with area A is calculated by $A\Omega_S \varepsilon B_\nu$. The emissivity ε of a source is defined as the ratio of its output compared to that of a blackbody. Kirchhoff's law of electromagnetic radiation states that in thermodynamic equilibrium the quantity of energy absorbed is equal to the quantity emitted. Thus, the requirement for a useful thermal source is that it should be a good absorber of radiation. Finding materials that are reasonably

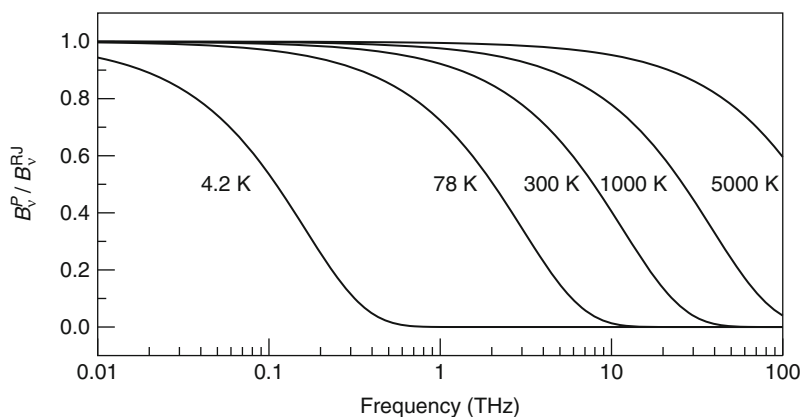


Fig. 4.1 Ratio of the blackbody emission predicted by the Planck (B_v^P) and Rayleigh-Jeans (B_v^{RJ}) equations at different temperatures

absorptive in the THz region has proved to be important, not only for providing good sources but also in producing detectors with uniform spectral absorbance.

The need for thermal sources of high emissivity is particularly important over the THz range. This can be seen by comparing (4.1) and (4.2). When $h\nu < k_B T$ the output of a source increases very rapidly with temperature but in the Rayleigh-Jeans region the increase is linear. Thus, the main requirement for a high-frequency source is its temperature, with emissivity of less significance, while in most of the THz region both parameters are of equal importance.

4.3 Practical Thermal Sources

Although a variety of solid-state thermal sources is available for the visible and IR regions, only the globar has sufficient emissivity to be useful at THz frequencies and then only above ~ 3 THz. The alternatives are plasma sources, where the emissivity increases toward lower frequencies, with the further advantage of operating at much higher temperatures than solid-state sources.

4.3.1 The Globar

This source typically consists of a cylindrical rod of silicon carbide 20–60 mm long and 5 mm in diameter, capped with metallic electrodes contained in a water-cooled housing with a slit to allow light out with the angular spread required for spectroscopy. The operating temperature is 1,350–1,500 K, and its emissivity in the visible and near-IR is ~ 0.85 .

Studies have shown that the heated envelopes of arc lamps (Sect. 4.3.2) provide equal or superior emission to the glowing bar or, in short globar, in the 5–10 THz region [234] but not at higher frequencies. For spectroscopy, the relatively uniform output of the globar, which is shown in Fig. 4.3, is advantageous.

4.3.2 Plasma Sources

The radiation from hot gas plasmas has been investigated in great detail, particularly in connection with the attempts to produce controlled thermonuclear fusion. Studies of the THz emission from the very hot plasmas required for fusion have yielded valuable information on their density and temperature. At more modest temperatures, plasmas can produce useful outputs in the THz region, as Rubens discovered in 1911 when using a mercury arc lamp in his research [14]. These lamps are still widely used sources for THz frequencies.

A typical design of a mercury arc source used for THz spectroscopy is the Philips HPK 125 lamp (since 2005 a product of Heraeus Noblelight GmbH, Hanau, Germany) shown in Fig. 4.2. This lamp has a cylindrical fused quartz envelope 30 mm long and 12 mm in diameter, with electrodes at either end of the cylinder. In operation, the pressure within the lamp is about one atmosphere. The discharge is pinched into a filament approximately 1–2 mm in diameter. This design is very convenient for the long narrow slits used for grating spectroscopy but is not ideal for interferometers, where sources with circular symmetry are preferable. As its number

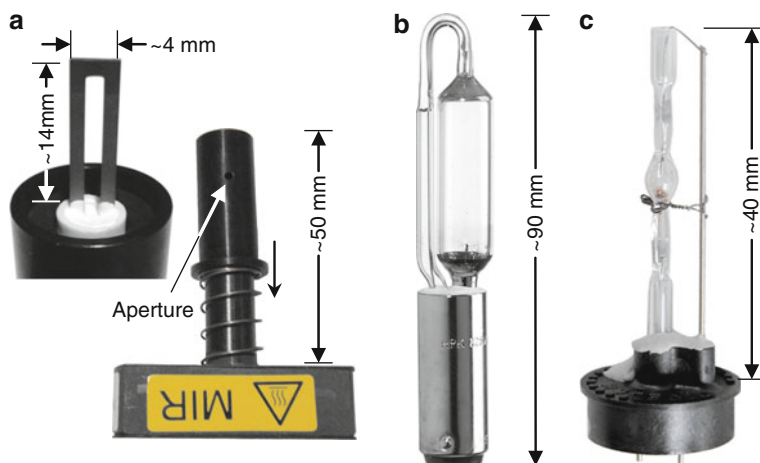


Fig. 4.2 Three THz sources: (a) globar (Bruker Optik GmbH, Ettlingen, Germany, photo copyright, courtesy of one of this book's authors), (b) HPK 125W mercury arc lamp (photo copyright, courtesy of Heraeus Noblelight GmbH, Hanau, Germany), (c) metal halide Welch Allyn M21N002, 21W Solarc lamp (photo copyright, courtesy of Welch Allyn Inc., Skaneateles, USA)

implies, the lamp consumes 125 W and requires water-cooling in the majority of applications. A particular advantage of this lamp is the stability of the arc. This is achieved by inserting a diaphragm, with a small hole at its center, just in front of one electrode. Should the arc position change on the electrode, the main arc is constrained by the diaphragm and the arc does not swirl around. Nevertheless, for stable operation a warm-up time of several minutes is required. A second advantage is that material sputtered and evaporated from the electrode is trapped in the end chamber and does not cause blackening of the central arc region. The instructions for using this lamp say that it should be used in an upright position but one of this book's authors has employed them horizontally over long periods and encountered no problems.

The output of mercury arc lamps is a combination of the emission of the arc itself, which is at $\sim 5,000$ K, and the quartz envelope that reaches $\sim 1,000$ K when the lamp is at its full operating temperature. Fused quartz becomes increasingly absorbing at frequencies above ~ 1 THz, and the 1–2-mm-thick envelope of the lamp is essentially opaque above 3 THz. Therefore, at high frequencies the lamp becomes, in effect, a thermal source with the area of the envelope. The relatively high refractive index of quartz in the THz region causes the emissivity to be approximately 0.88.

Below 1 THz, the effective temperature of the source rises because the plasma in the discharge is opaque due to electron-ion collisions, and it acts as a blackbody at the plasma temperature. The theory of plasmas is complex but a good description of that relevant to the mercury arc can be found in [236]. Below the plasma frequency ν_P , the discharge becomes highly reflective and therefore nonemitting, but for the electron densities in the medium-pressure arc lamps that are used for spectroscopy ν_P falls well below 0.3 THz. Figure 4.3 shows the effective temperature of the Philips HPK 125 lamp throughout the THz region. Similar results have been

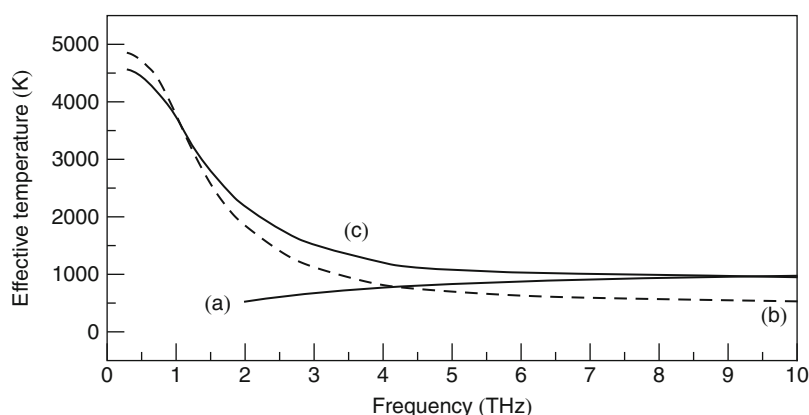


Fig. 4.3 Effective blackbody temperature of the three thermal sources discussed in the text: (a) global, (b) Philips HPK 125W mercury arc lamp, (c) Welch Allyn lamp no. M21N002, 21W Solarc lamp (adapted from [235])

obtained with other lamps but the output varies, due mainly to the difference in thickness of the quartz envelope.

In practice, it is important when using mercury arcs over the full THz range to remember that the effective area changes, dependent on whether the output is from the envelope, the plasma or a combination of both. The best arrangement is to use an optical system that provides an image of the plasma, which fills the entrance aperture of any spectroscopic system used, at a sufficient solid angle to exploit the full throughput of the instrument. Off-axis ellipsoidal mirrors are particularly useful for collecting light from small area sources at large solid angles and expanding the image to fill the entrance aperture.

A problem that can occur with arc lamps, in the spectral region where the emission is due to the plasma, is interference, because the inner and outer faces of the envelope can act as a Fabry–Pérot cavity, producing interference fringes. This type of interference can be clearly seen when the source is used, without a mirror or lens, to focus the light on the entrance aperture of a spectroscopic system [237]. Interference effects can also occur when, for example, an upright cylindrical mirror is used to provide a vertical image of the source at the entrance slit of a grating monochromator. However, in most situations, the optical system before the entrance aperture is designed to change the shape of the image and under these circumstances the light is traveling at varying angles through the envelope and any interference effects are very small.

The lamps used in the THz region are primarily manufactured as ultraviolet sources and exposure of the skin, and particularly the eyes, is hazardous. It is essential to take the necessary precautions. Fortunately, glass goggles are opaque to ultraviolet radiation and do not cause too much inconvenience when aligning equipment.

A more compact plasma source that has been shown to be a useful alternative to conventional mercury arc lamps is a metal halide lamp manufactured by Welch Allyn (Fig. 4.2) [235]. This has an approximately spherical diameter of 3 mm and consumes only 21 W and a surface temperature of nearly 1,400 K. In the region above 3 THz, it also has the advantages that its shape makes it suitable for circular symmetry systems, and that it does not normally require water-cooling. Below 3 THz, its brightness is very similar to the Philips HPK 125 lamp but the plasma area is small at $1.5 \times 0.3 \text{ mm}^2$ (Fig. 4.3). However, it is proving very useful in situations where a small source size is acceptable or required.

4.4 Gas Lasers

4.4.1 *Electrically Excited Gas Lasers*

The earliest lasers to provide significant power at THz frequencies were produced by a glow discharge in pure gases enclosed in a Fabry–Pérot cavity. H_2O was the first

to produce stimulated emission [7], and other examples are D_2O , NH_3 , OCS , H_2S , SO_2 , DCN , and HCN . To a large extent, this type of source has been superseded by the optically excited lasers, which are described in the following section. However, the relative simplicity of construction of electrically excited lasers makes them a convenient tool when one of the wavelengths produced fulfils a specific requirement. Two species that produce useful power levels, in both continuous wave and pulsed modes, are HCN and DCN . Producing a number of frequencies in the 1–1.5 THz range, these two lasers have been quite widely employed in the diagnostics of plasmas in controlled fusion research.

Laser action at THz frequencies in electrically excited HCN and DCN systems is due to population inversion between rotational levels of vibrational states with closely related energies. For the HCN laser, these are the (11^10) and (04^00) states and a simplified vibrational–rotational energy level diagram is shown in Fig. 4.4a. Two strong laser lines are produced due to the $J(11)$ and $J(10)$ rotational levels of the (11^10) vibrational state having near coincidence with the $J(10)$ and $J(9)$ levels of the (04^00) vibrational state. The line at 0.890 THz is of particular importance as it is within an atmospheric window, allowing propagation of the beam over relatively long distances.

The energy levels in DCN are similar to those in HCN and are shown in Fig. 4.4b, with the two strong lines being at 1.537 and 1.578 THz. Laser action can occur at other frequencies as a consequence of the two initial lasing processes. For example, the 0.890 THz transition reduces the population of the $J(10)$ level of the (11^10) state and over populates the $J(9)$ level of the (04^00) state, leading to oscillation at

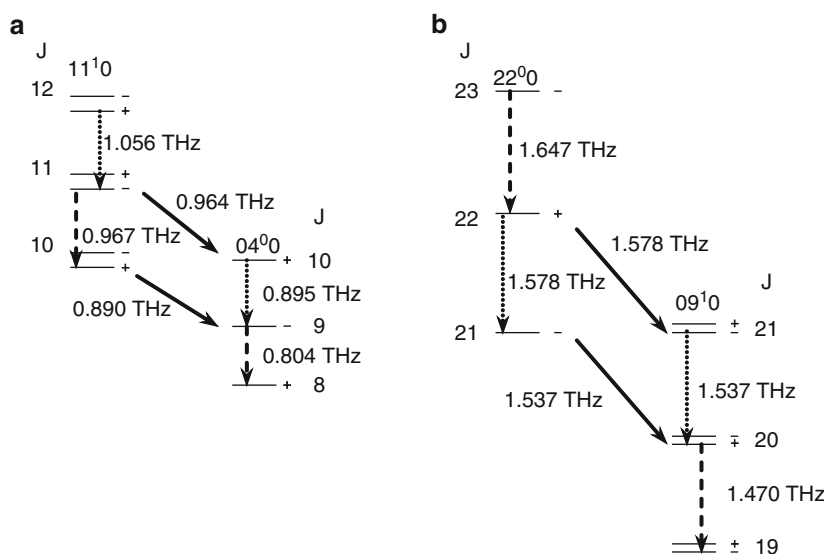


Fig. 4.4 The energy levels and to scale lasing transitions of (a) HCN and (b) DCN with the splitting energies exaggerated. The quantum number J indexes the laser levels. Dotted and dashed lines indicate transitions within a vibrational state

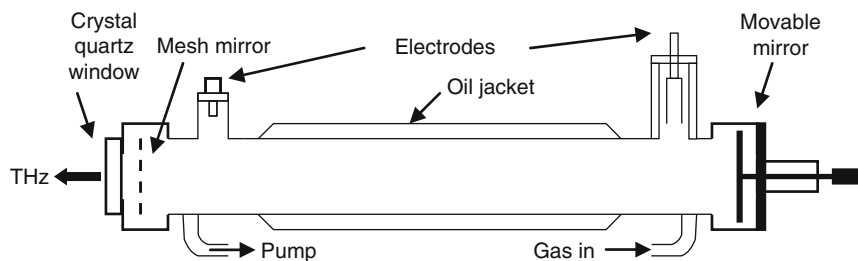


Fig. 4.5 Design of an HCN or DCN laser

0.804 THz and 0.967 THz. However, these lines are very weak compared with those due to transitions between the vibrational states.

A typical design for an HCN or DCN laser is shown schematically in Fig. 4.5. With resonator lengths of ~ 3.5 m and diameter ~ 50 mm, cw powers of up to 250 mW for HCN and about 220 mW for DCN have been obtained on the strongest lines. With longer lasers, higher powers can be achieved. For example, a 6.5-m long laser, 100 mm in diameter, produced power levels up to 600 mW in a TEM_{00} mode [238]. Operation in pulsed mode provides powers of ~ 1 kW in a 20- μ s long pulse with a similar size system [239]. In the cw mode, the voltage across the tube is a few kV and the discharge current is 1–2 A. The reason for the oil jacket surrounding the discharge is the need to keep the tube at a relatively high temperature. An early problem of HCN lasers was the building up of a brown deposit on the tube wall which steadily reduced, and eventually extinguished, laser emission. A gas mixture of nitrogen and methane is fed into the laser which, under electrical discharge conditions, forms HCN. It has been shown that by adding helium and operating at temperatures above 100°C lasers can be operated for long periods without the development of deposits.

The ratio of the three gases is dependent on a number of factors, including the laser tube diameter and the operating temperature. Typical ratios for $N_2:CH_4:He$ are 1:2:10 but as the gases flow continuously through the laser tube it is relatively easy to optimize the mixture for maximum output. The total pressure of gas is a few hPa. The role of helium is not entirely clear, but it seems likely that, despite the high wall temperatures, its high thermal conductivity leads to a reduced gas temperature. Addition of helium also produces a more stable discharge.

An alternative gas mixture replaces helium with hydrogen in HCN and DCN lasers. Although the power is $\sim 25\%$ lower, the maximum output is reached at 150°C wall temperature, when there no deposit is produced on the tube walls [240]. Increased output from electrically excited lasers can be obtained by using a transverse electrode structure, with higher current at greater gas pressure. With these modifications, powers approaching 10 kW have been achieved [241].

In some of the other electrically excited gas lasers, the energy levels are much more complex. In H_2O , the vibrational–rotational spectrum has so many near-coincident energy levels that over 80 lasing transitions occur at frequencies from

below 1 THz to above 40 THz. Many of these lines are only available with pulsed excitation but two cw lines at 10.73 and 2.53 THz have been used for a number of experiments. As with HCN the design of H₂O lasers can be very simple. The very rich spectrum of water vapor is the reason why atmospheric absorption is so intense over much of the THz region. G. Dodel has written a useful review of electrically excited gas lasers, and the optically excited gas lasers discussed in the following section [242].

4.4.2 *Optically Excited Gas Lasers*

Gas lasers, which are optically excited with a CO₂ pump laser, have been the major source of cw coherent THz radiation above 0.3 THz for many years. Although there are now alternatives such as harmonic generators, or quantum cascade lasers, optically excited lasers are still a workhorse in many laboratories around the world. The first laser emission from such a system was reported by Chang and Bridges in 1970 [81]. Since then, more than 1,000 laser lines with cw emission have been found. They cover the frequency range from 0.15 to 8 THz with an output power ranging from several μ W up to several 100 mW. Some useful reviews on optically excited THz gas lasers are available [242–246]. The last reference is especially helpful because it contains an extensive list of laser lines.

The optically excited THz gas laser is based on rotational transitions of molecules with a permanent electric dipole moment. The molecules are excited into a vibrational state by a pump laser, usually a CO₂ laser. Thus, population inversion is obtained between particular rotational levels of an excited vibrational state (Fig. 4.6). This constitutes a four-level laser scheme coupled by the IR pump and THz laser radiation. The simplicity of the laser scheme, the large number of laser-active molecules, the richness of rotational/vibrational spectra, and the availability of more than 200 pump lines from different CO₂ lasers (different isotopologues, sequence band transitions, hot band transitions, etc.) is the reason for the discovery of so many laser lines. To be a strong THz emitter, the molecule should have IR active vibrational modes and rich THz spectra. Efficient absorption of the IR pump radiation is possible for molecules with a large dipole moment and small rotational partition function. This guarantees large absorption cross sections for the IR radiation. In addition, the vibrational relaxation rate should be large in order to reduce excess excited state population. Fast vibrational relaxation allows operation at higher pressure because interaction with the laser walls is not required for de-excitation. In general, lightweight polyatomic molecules fulfill these requirements. Some examples are CH₃OH, CH₂F₂, HCOOH, and their isotopologues. An example is shown in Fig. 4.7, which displays the absorption of CO₂ laser radiation by CH₃OH molecules. The measurement is based on the photoacoustic effect. Chopped radiation from the CO₂ laser is absorbed by the molecules in the THz resonator. The absorbed energy induces a pressure wave or sound at the chopping frequency, which is detected with a microphone inside the laser resonator.

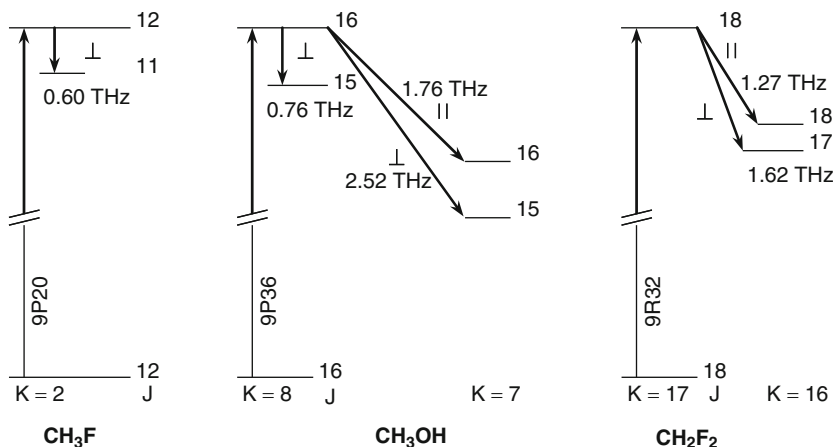


Fig. 4.6 Laser schemes of several emission lines. CH_3F is a prolate symmetric top molecule with a dipole moment solely along its a-axis of moment of inertia. The laser transitions obey the selection rule $\Delta K = 0$. CH_2F_2 is a near prolate asymmetric top molecule with a dipole moment solely along its b-axis and the laser transitions obey the selection rule $\Delta K = \pm 1$. CH_3OH is an intermediate case with a dipole moment along both axes of inertia. Therefore, laser transitions with $\Delta K = 0$ and $\Delta K = \pm 1$ occur. J is the total rotational quantum number and k is the quantum number associated with the projection of J on the rotational axis. Polarized radiation orthogonal to the pump radiation is labeled with $\perp$ and parallel polarization with $\parallel$ (adapted from [243])

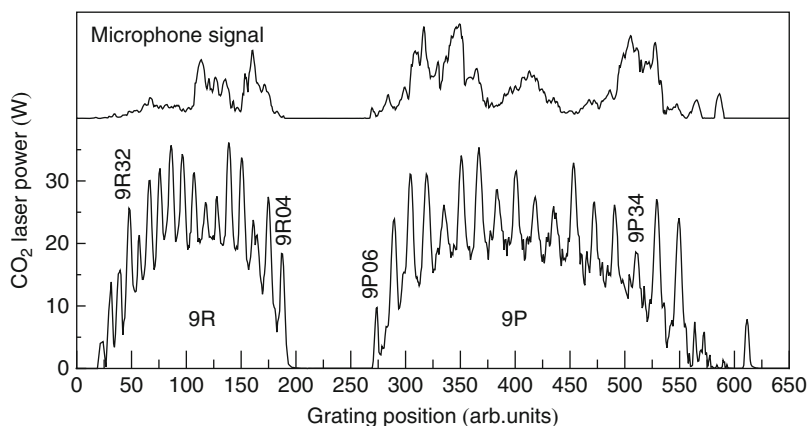


Fig. 4.7 Output power of a CO_2 laser as a function of the position of the laser grating, which selects the laser transition (lower trace). The upper trace is the photoacoustic signal of a CH_3OH THz laser, which is measured with a microphone inside the laser cavity. The peaks indicate strong absorption of CO_2 laser pump power. A strong absorption is a good indicator of a strong laser line. The absorption peaks of the 9P34, 9P36, and 9R10 lines, which correspond to strong laser lines (Table 4.1) are clearly visible

In order to understand the selection rules of the laser emission, three cases can be considered (Fig. 4.6): CH_3F is a symmetric top molecule with a permanent dipole moment solely along its a-axis of least moment of inertia. For transitions induced by this dipole moment (type a) the quantum number k has to remain unchanged. The quantum number k is the projection of the total rotational quantum number J on the rotational axis of the prolate top. The opposite situation holds for the near-prolate asymmetric top molecule CH_2F_2 , which has a large permanent dipole moment solely along the b-axis of the intermediate moment of inertia. In this case, only transitions with $\Delta k = \pm 1$ are allowed. The intermediate case is CH_3OH , which is a near-prolate asymmetric top with permanent dipole moments along the a- and b-axes. Therefore, CH_3OH has both types of rotational transition. For the relative polarization of the THz emission and the pump radiation, the change in the rotational quantum number J of the IR pump transition, ΔJ_1 , and the change in the laser transition, ΔJ_2 , need to be taken into account. If $\Delta J_1 + \Delta J_2$ is even, pump and laser radiation are parallel polarized while, when $\Delta J_1 + \Delta J_2$ is odd, they are orthogonally polarized with respect to each other. The molecular reaction rates are essential to achieve cw operation. Collisions between molecules tend to thermalize the population among the rotational states in each vibrational state. Therefore, relaxation from the excited vibrational state has to be sufficiently rapid. This process is dominated either by diffusion out of the active laser medium to the walls of the resonator ($\sim 10^{-8} \text{ s Pa}^{-1} \text{ cm}^{-2}$) or vibrational relaxation ($\sim 10^{-8} \text{ s Pa}^{-1}$). These relatively slow rates limit the typical operation pressure to tenths of a hPa. Such low pressures reduce pump absorption and limit the output power. The maximum efficiency for conversion of pump power into THz power is given by [246]

$$P = \frac{1}{1 + g_i/g_j} \left(\frac{L_M}{L + L_M} \right) \left(\frac{\alpha l}{\alpha l + \gamma} \right) \left[1 - \frac{h\nu}{k_B T} f_i \frac{\tau_v}{\tau_{\Delta J}} \right] \frac{\nu}{\nu_i} P_{\text{pump}} \leq \frac{1}{2} \frac{\nu}{\nu_i} P_{\text{pump}}, \quad (4.8)$$

where g_i and g_j are the degeneracy of the upper and lower laser states, P at the frequency ν and P_{pump} at the frequency ν_i are the power of the THz and CO_2 pump radiation, respectively. l is the resonator length and L_M is the loss at the THz frequency due to outcoupling. L is the THz cavity loss and α is the absorption coefficient of the pump radiation. γ is the cavity loss at the pump wavelength and f_i is the Boltzmann factor of the upper rotational laser state. τ_v is the vibrational relaxation time and $\tau_{\Delta J}$ is the rotational collision time. The fundamental limitation is given by the Manley–Rowe limit expressed by the ratio of the frequencies and multiplied by the factor containing the degeneracy of the laser states. The two factors in curved brackets account for cavity losses of the THz radiation and the pump radiation. The factor in square brackets is the condition for gain. Gain is possible if this factor is larger than zero. The inequality gives the upper limit for the conversion efficiency.

The gain profile of the lasing transition is determined by Doppler and pressure broadening. The first one yields an inhomogeneously broadened gain profile with a width ranging from 1 to 10 MHz. The actual value depends on the particular frequency and mass of the molecule. The Doppler width increases with laser

frequency and the mass of the laser molecule. Pressure broadening is independent of frequency and mass. Collisions and saturation result in a homogeneous broadening of $\gamma_{\text{hom}} = \gamma_{\text{coll}} / \sqrt{1 + I/I_{\text{sat}}}$, where I is the laser intensity, I_{sat} is the saturation intensity, and γ_{coll} is the collisional broadening. The latter is relatively high for laser molecules in the range 10 to 40 MHz/hPa. At typical pressures of 0.1–0.8 hPa, γ_{hom} is of the order of a few MHz. It is nearly independent of pressure, because the collisional broadening increases with pressure while the saturation intensity decreases. Above about 1.5 THz, Doppler broadening dominates for typical laser molecules. The width of the gain profile allows the laser frequency to be tuned over a few MHz by changing the length of the laser cavity. Very narrow linewidths of less than 1 kHz and fractional frequency instabilities as low as 2×10^{-12} can be obtained with appropriate phase-locking techniques. The frequency reproducibility for a free running laser is determined by the accuracy in obtaining the center of the gain curve, which is about 2×10^{-7} . A list with some useful, easy to operate gas laser lines and typical performance parameters is given in Table 4.1.

Many different resonators have been employed for THz gas lasers. Fundamentally, one can distinguish the open resonator, and the waveguide-type resonator. In the case of an open resonator, the walls of the laser cavity play a negligible role in

Table 4.1 Selected THz gas laser lines with CO₂ pump laser lines and output characteristics. The polarization is the relative polarization of THz emission and pump radiation. The output power is classified as follows: vvs (very very strong): >1 mW THz power for 1 W pump power, vs (very strong): 0.1–1 mW/W, s (strong): <0.1 mW/W. Note that the output power depends strongly on the design of the THz gas laser system, which is usually designed to operate on a large number of emission lines. The numbers represent typical values from the literature. Optimization for one particular emission line may increase efficiency and power output

Molecule	ν_{CO_2}	ν (THz)	λ (μm)	Polarization	THz power
HCOOH	9R28	0.5843882	513.002		s
HCOOH	9R38	0.6538222	458.523	⊥	s
HCOOH	9R20	0.6928950	432.667		vs
HCOOH	9R18	0.7616083	393.631		vs
¹⁵ NH ₃	10R42	0.8029860	373.347	⊥	vs
CH ₂ F ₂	9R34	1.0421504	287.667		vs
CH ₂ F ₂	9R32	1.2721714	235.654		vs
CH ₂ F ₂	9R34	1.3971186	214.579	⊥	vs
CH ₂ F ₂	9R32	1.6266026	184.306	⊥	vvs
CH ₃ OH	10R38	1.8388393	163.033		s
CH ₂ F ₂	9P10	1.8912743	158.513		s
CH ₂ F ₂	9R22	2.4479685	122.466	⊥	s
CH ₃ OH	9P36	2.5227816	118.834	⊥	vvs
CH ₂ F ₂	9R20	2.5464950	117.727	⊥	vvs
CH ₃ OH	9R10	3.1059368	96.522		vs
¹³ CH ₃ OH	9P22	3.5138530	85.317		s
CH ₃ OH	9P34	4.251674	70.512	⊥	vs
CH ₃ OD	9R08	5.2456129	57.151	⊥	s

determining the resonant frequencies. Usually, these lasers are constructed using large diameter dielectric tubes, for example, made of quartz glass, as walls for the resonator. With this assumption and the approximation of mirrors with infinite diameter, the resonant frequencies are given by

$$\frac{\nu}{\nu_0} = (q + 1) + \frac{1}{\pi}(2p + m + 1) \cos^{-1} \left(\sqrt{\left(1 - \frac{l}{R_1}\right) \left(1 - \frac{l}{R_2}\right)} \right), \quad (4.9)$$

where R_1 and R_2 are the radii of curvature of the two mirrors which form the resonator, and l is the resonator length. p , m , and q are the number of field variations in the azimuthal, radial, and axial directions and $\nu_0 = c/(2l)$. It should be noted that (4.9) assumes negligible diffraction loss, i.e. it does not apply for resonators with Fresnel numbers less than about 20. The emission frequency ν is determined by the lasing medium. In order to achieve laser emission from the active medium, the distance between the mirrors has to be tuned to fulfill (4.9). For a given mode (p, l) , the change in resonator length between successive axial modes $q, q + 1$ is very nearly equal to $\lambda/2$, with the wavelength λ corresponding to the emission frequency. This approximation is valid to about 0.05%. Thus, by varying the resonator length the emission frequency of the laser can be determined. This is called the laser resonator interferogram. An example is shown in Fig. 4.8.

It should be noted that the gain bandwidth is much narrower than the frequency separation of axial modes. The requirement is that the resonator length needs to be tuned accurately to the resonance frequency in order to obtain laser emission. THz gas lasers may also be constructed as a waveguide-type resonator with metal walls.

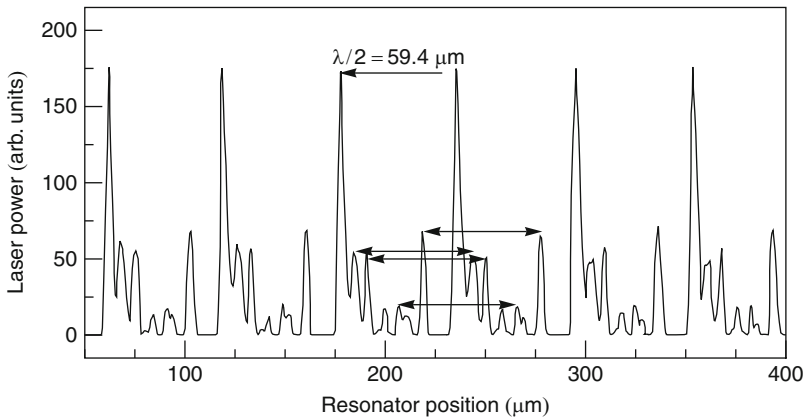


Fig. 4.8 Resonator interferogram of a THz gas laser with a nearly confocal resonator and with CH_3OH as the active medium. The output power is shown as a function of the position of one of the resonator mirrors. Each peak appears every $59.4 \mu\text{m}$ indicating that the laser emission is $118.8 \mu\text{m}$ ($\sim 2.5 \text{ THz}$). The highest peak corresponds to the fundamental mode while all other peaks correspond to higher order transverse modes. Higher order modes can be suppressed by introducing an intra-cavity iris diaphragm

Here, the transverse size and shape of the laser cavity determine the type of mode, and shifts the emission frequency to a certain extent. The wavelength λ_{wg} inside the waveguide depends on the cut off wavelength λ_{co} , the longest wavelength at which radiation can propagate in that mode, according to

$$\lambda_{wg} = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_{co})^2}}. \quad (4.10)$$

Close to cut off λ_{wg} can differ significantly from the free space wavelength. This effect needs to be taken into account when analyzing the resonator interferogram. Two more categories are standing wave resonators and ring resonators (Fig. 4.9). The specific advantage of a ring resonator is that, due to the formation of a traveling wave, no spatial hole burning occurs and the active medium is used more efficiently. However, the price to be paid is greater complexity because at least one more mirror needs to be used, in contrast to the standing wave resonator, which can be achieved with a minimum of two mirrors. Moreover the THz emission can occur in either clockwise or counter-clockwise direction. In the case of a homogeneously

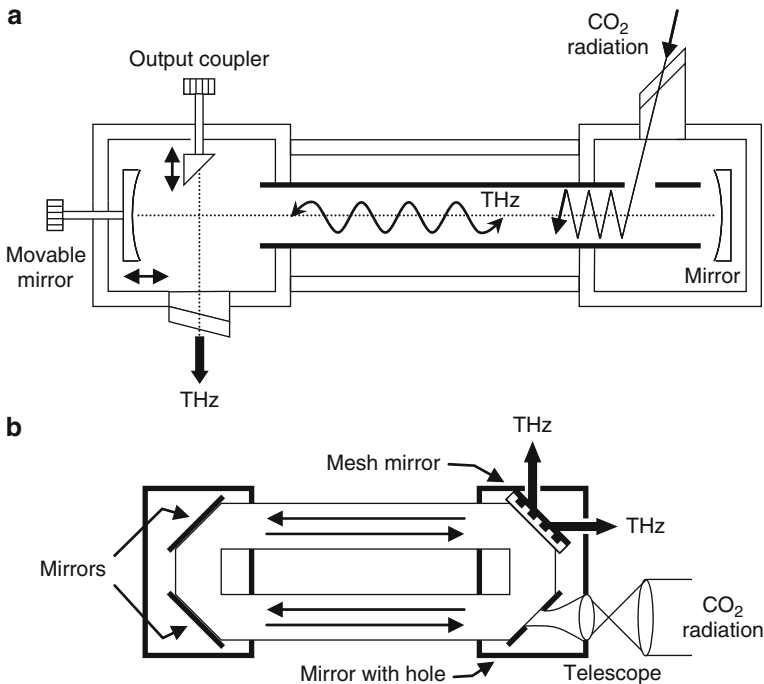


Fig. 4.9 Standing wave resonator (a) and ring resonator (b) with different excitation schemes. The standing wave resonator is transversely excited by radiation from a CO₂ laser and the ring laser is longitudinally excited

broadened gain profile, switching between both directions occurs close to the center of the gain profile.

THz radiation can be coupled out of the resonator in many ways. Some examples are a hole in one of the mirrors, a partially transmissive mesh filter which replaces one of the mirrors, a mirror which sticks into the resonator and reflects a small amount of radiation transversely to the optical axis of the resonator, and a Michelson interferometer. Frequency and power stabilization are mainly an issue for the CO₂ pump laser. Most types of stabilization scheme for CO₂ lasers can be employed. Examples are Lamb-dip stabilization to a fluorescence transition of CO₂, stabilization to a gas absorption line with an photoacoustic cell [247] or stabilization with respect to a Fabry–Pérot interferometer, which in turn is stabilized to a Zeeman-stabilized HeNe laser [248]. The THz laser resonator is usually passively stabilized by an appropriate thermal design.

As mentioned earlier, optically excited THz gas lasers have been one of the workhorses of THz research since the early 1970s. They have been used not only for fundamental research such as molecular spectroscopy (Sect. 6.7.4) but also in imaging (Chap. 7) and industrial inspection [249]. During the past ten years, the technology of optically excited THz gas lasers has advanced significantly, resulting in smaller, more reliable and easier to use THz gas lasers. Primarily, this new generation was developed as the LO for heterodyne receivers. A prominent example is the gas laser LO of the Microwave Limb Sounder on the Aura satellite, which monitors stratospheric OH at a frequency of 2.5 THz [247]. This laser needs a total electrical input power of only 120 W and delivers 20 mW at 2.5 THz. Although for a long time considered to be bulky, energy consuming and difficult to operate, THz gas laser technology is now mature and will continue to play a role in exploring the THz spectral range. Nevertheless, the lack of significant frequency tunability, which is due to the nature of the active medium and cannot be overcome, ultimately limits its use.

However, in pulsed THz gas lasers this lack of tuning can be overcome to some extent. In its simplest design, the laser consists of a CO₂ laser typically producing 0.1–0.5 μ s long pulses of very high power followed by a mirrorless tube approximately 1–4 m long containing the gas to be optically excited. As with the cw laser the CO₂ laser creates an inverted population in an appropriate gas in the long tube, but in the pulsed situation the gain is high and saturation can occur in a single pass. This method for providing THz radiation is described as amplified spontaneous emission (ASE).

The CO₂ laser employed is very different from the type used for cw optically excited lasers as the gas mixture is at atmospheric pressure or above. The required electric field to break down high pressure gas is very large and for this reason the laser is transversely excited. In one of the first experiments, a reasonably uniform discharge was obtained by using a metal plate as the anode. Opposite this was a long row of resistors, attached to a wire, acting as a distributed cathode [250]. In later versions, two identical solid aluminum electrodes were used. These had profiles of a specific cross-section designed to provide a homogeneous electric field [251]. To ensure a uniform discharge between the electrodes, the gas needs to be pre-ionized and a convenient method is to run a fine trigger (tungsten) wire parallel to

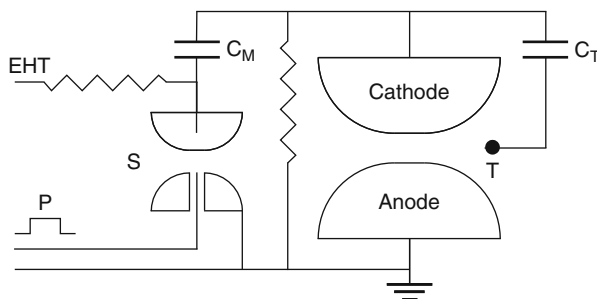


Fig. 4.10 Cross section of a TEA laser. Trigger pulse P, spark gap S, high voltage supply EHT, trigger capacitor C_T , trigger wire T, main capacitor C_M

the main electrodes but offset from their center line, as shown in Fig. 4.10. Each end of the wire is connected to the cathode by a small coupling capacitor. When a high-voltage pulse is applied to the main electrodes field emission from the wire induces a sheet discharge between the wire and the anode. This discharge, which is limited in energy by an appropriate choice of coupling capacitors, produces UV radiation that ionizes the CO_2 gas mixture and initiates a uniform discharge between the main electrodes. It was later found that an even more reliable discharge could be obtained by using wires on either side of the main electrodes. CO_2 lasers of this type are called transversely excited atmospheric (TEA) lasers and can produce many tens of megawatts peak power in pulse lengths of 0.1–0.5 μs .

As mentioned earlier, the gas to be inverted can be in a tube without mirrors, made of either glass, silica, or plastic. Typically, this has an entrance window that transmits $10\text{ }\mu\text{m}$ radiation, usually ZnSe, and an output window to pass the THz radiation, which is opaque at $10\text{ }\mu\text{m}$. Crystalline quartz is ideal, with TPX a cheaper alternative unless the output is in the 6–10 THz range, when CsI is the best choice. More sophisticated designs resemble those for cw excited lasers with mirrors at either end of the cavity. These are required for one of the early applications of pulsed systems, Thomson scattering from plasmas produced in controlled fusion research. Due to the small scattering cross-section, powers approaching 1 MW are required to achieve acceptable (S/N)-ratios. Although this power level is comparatively easy to obtain in multi-mode systems [252], this is only one of the requirements for the scattering source. In addition, it has to have a narrow linewidth ($<100\text{ MHz}$) and a sufficiently long ($>1\text{ }\mu\text{s}$) pulse length. In a remarkable experiment, the necessary parameters were obtained by using an energy of 600 J in a pulse length of 1.4 μs at 0.8 THz [253]. The cost and complexity of such a sophisticated system has meant that optically excited gas lasers have not been widely used for Thomson scattering. High-frequency gyrotrons (Sect. 4.11) are the preferred alternative.

In early experiments with high-pressure systems, it was discovered that, unlike low pressure cw lasers, exact matching to the absorption line of the pumped gas was normally not necessary. The reasons for this are complex, involving a Raman-type two-photon process with some additional frequency offset due to the AC Stark

effect. The Raman process is favored because of its higher quantum efficiency compared to the conventional inversion process in low pressure optically excited THz gas lasers [254]. The result is that a very wide range of THz frequencies is available from ASE systems.

Soon after the invention of TEA CO₂ lasers, it was realized that by increasing the pressure of the gas mixture to above 1 MPa the rotational levels begin to overlap, allowing quasi-continuous tuning over much of the 9–11 μm band [255]. By using this tunable radiation as a pump, it was suggested that it should be possible to produce tunable radiation over a significant portion of the THz spectral region [256] and, for example, tuning over 85% of the range from 0.75 to 1.4 THz using CH₃F has been reported [257]. As so many gases can be pumped by TEA CO₂ lasers it is, in principle, possible to produce pulsed tunable radiation over virtually all the THz region at very high output power. High pressure gas lasers, unlike low pressure ones, are not a very convenient source as they require systems operating to well above 1 MPa and voltages of up to 100 kV. However, particularly in the region 3 to 10 THz, they are sources which will provide very high power tunable radiation.

4.5 Bulk Semiconductor Lasers

Semiconductor lasers are extensively used in the visible region, e.g. for data storage in CD and DVD players, and in the near-IR range for telecommunication around 1550 nm, where glass fibers have minimal absorption. Erbium-doped fibers can amplify the data packages if pump lasers excite the transitions of erbium dopants at 807, 980, and 1485 nm simultaneously. Therefore, such lasers at the respective erbium wavelength can be obtained with very high output powers, due to the demand in the telecommunications industry. These semiconductor and solid state lasers are employed, for example, in THz TDS (Sect. 6.6) and THz OPO laser systems (Sect. 4.13). These very efficient optical and near-IR semiconductor lasers consist of a p-doped/n-doped junction and use direct band gap transitions in GaAs, InGaAs, etc. The photon emission is due to electron–hole pair recombination. In the THz frequency range, this mechanism cannot be used because the smallest direct band gaps in semiconductor materials are found in lead-salt lasers [258], but these exceed 18 THz (600 cm^{-1}). Furthermore, it is difficult to obtain an inversion at room temperature in semiconductors because they are vulnerable to thermal effects, since $k_{\text{B}}T$ at 300 K corresponds to 6 THz, a frequency just in the middle of the THz frequency range.

Nevertheless, special mechanisms can support laser action in bulk semiconductors, for example between Landau levels due to an external magnetic field, optical pumping of impurity energy levels, and external uniaxial stress, which may be extended to internal built-in strain in mismatched lattices. With structuring of bulk semiconductors into quantum wells, THz quantum cascade lasers can be fabricated as described in Sect. 4.6. Promising concepts for future THz emitters and lasers

include excitons within a semiconductor generated by an optical light field, quantum wires, and quantum dots.

4.5.1 Germanium Lasers

The first electrically excited semiconductor THz laser was made from Ge. The possibility of creating such an inversion in p-type Ge had already been suggested by Krömer in 1958 [259]. He proposed a so-called negative mass amplifier and generator utilizing nonelastic optical phonon scattering of carriers accelerated by electric fields in the kV/cm range. This leads to a so-called “streaming motion” along the electric field E direction and the crystal axis, along which the field is applied [260] (Fig. 4.11). Due to the nonparabolic bandstructure and warping of the equienergy surfaces of heavy holes, a negative effective mass can be derived which leads to amplification. Applying a magnetic field in the direction parallel to the electric field and to the streaming motion resulted in heavy hole cyclotron resonance masers [261] and lasers [262].

In 1972, Maeda and Kurosawa suggested applying, in addition to the electric field E , a magnetic induction B which is oriented in a perpendicular direction [263]. Often, this is referred to as crossed electric and magnetic fields. This configuration leads to a so-called “accumulation region” for which scattering with optical phonons is negligible (Fig. 4.11a,b). With this configuration, Andronov et al. reported amplification of THz radiation in Ge [264]. Subsequently, population inversion in Ge was reported by several other groups in Russia and in Japan [265, 266]. It is worth noting that in the same year, 1979, S. Komiyama et al. [267] showed, in a different material, silver halide, streaming motion, and population inversion of electrons [267], indicating that the mechanism can be transferred to n-type and other materials. For example, THz emission has been found for electrically excited Si [268].

In the 1980s, THz emission from p-type Ge [97] and THz Ge lasers were reported [98, 269]. The electric field applied can vary from 0.2 to 6 kV/cm with the magnetic induction applied from 0.2 to 6 T. The magnetic induction leads to a splitting of the valence bands into Landau levels. The laser transition typically occurs between a Landau level of a light hole and a lower level, which can be either a light, heavy, or hybridized light-heavy hole Landau level (Fig. 4.11c). At low magnetic induction, approximately less than 1.5 T, the splitting of light hole Landau levels is not sufficient to distinguish laser lines easily. Ge lasers with such a broadband emission were named interband or intervalence band lasers (IVB). With increasing magnetic induction the number of possible light hole Landau levels, which do not interact with the optical phonon energy, reduces rapidly until only one emission line is observed. The line is a result of a cyclotron resonance (CR) transition giving rise to the name CR Ge laser. The frequency is tunable by the magnetic induction with a rate of ~ 0.6 THz/T. At a magnetic induction just below 1.5 T, it is possible to more easily distinguish different types of CR transition including spin-flip CR transitions

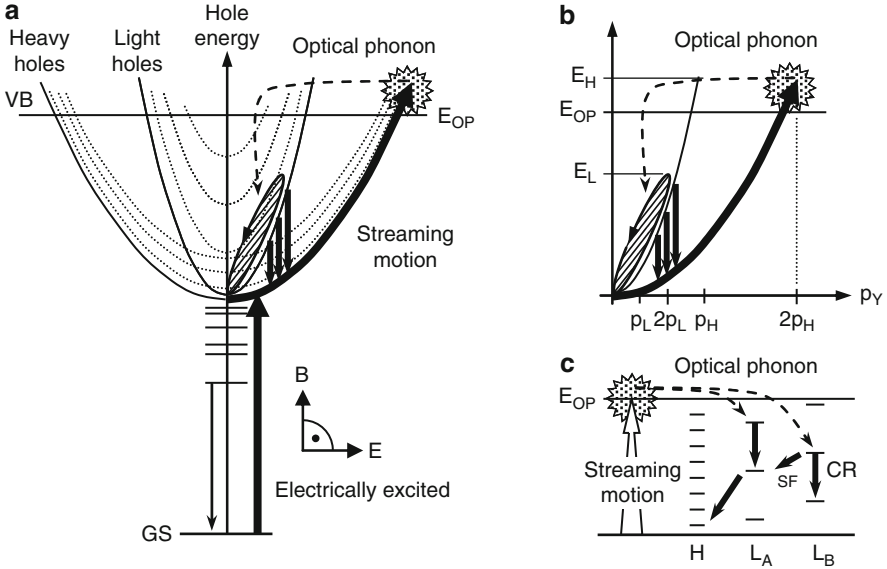


Fig. 4.11 (a) THz radiation over a wide frequency range due to transitions from the light hole to heavy hole valence band (VB) in Ge. The magnetic induction B splits the VB into Landau levels. Heavy holes are impact ionized by the electric field E from the acceptor ground state (GS) and accelerated in a streaming motion to reach the optical phonon (OP) energy E_{OP} . OP emission (dashed arrow) leads to an accumulation of holes in light hole Landau levels. Intracenter emission to the GS of shallow acceptors can occur (for further details see text). (b) Simplified scheme of laser operation. In crossed electric and magnetic fields, the heavy (H) and light (L) holes reach different maximum energies E_H and E_L . THz broadband emission (downward straight arrows) is achieved for $E_L < E_{OP} < E_H$ by tuning the E/B -ratio. (c) Simplified laser level diagram. The light hole levels are populated by excited heavy holes via optical phonon emission (dashed arrows). They are spaced by ~ 0.6 THz/T within two spin sets denoted A-set (L_A) and B-set (L_B). The spin sets are offset to each other by ~ 0.3 THz/T. THz transitions (solid downward arrows) occur due to CR, spin-flip (SF), combined CR and SF [270], and transitions from light to heavy hole levels. The heavy hole Landau level spacing is smaller by a factor of 8 due to the higher effective mass. The second heavy hole spin set and the nonradiative decay from lower levels to the GS is not shown for clarity (adapted from [271])

and second-order CR transitions at twice the frequency [270]. The latter are a result of the high applied and internal electric fields. The spin-flip transitions are offset by ~ 0.3 THz/T with respect to normal CR transitions, due to the offset of the two spin sets of Landau levels denoted A-set and B-set (Fig. 4.11c).

Single Ge crystals doped with acceptors were initially cut in rectangular parallelepipeds with ohmic electrical contacts on two opposite sample facets (Fig. 4.12a,c). Several different laser designs such as Bragg mirrors [273], ring resonators [272], and planar lasers have been demonstrated [274] including external resonators, which achieved a single longitudinal mode and tunable THz laser emission [275–279] (Fig. 4.12). External metal mirrors or outcouplers are attached

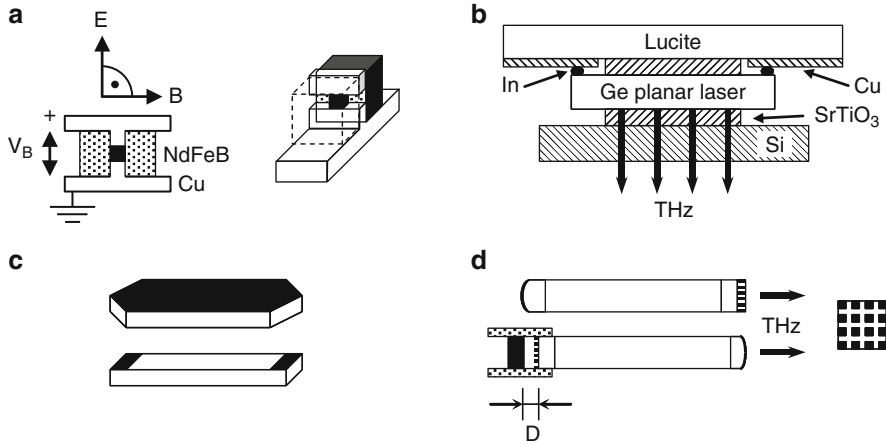


Fig. 4.12 Ge laser setups (adapted from [271, 272]). (a) Laser mounted between two NdFeB permanent magnets. The opposing electrical contacts are attached to a Cu heat sink biased at a voltage V_B . (b) Planar Ge laser operated in a vertical cavity arrangement. The one-sided contacts were attached via In “bumps” to Cu contacts. THz radiation was observed through an ultra-pure Si wafer while the laser was mounted between two SrTiO₃ crystal plates. The plates serve as high reflectivity mirrors due to their THz reststrahlen band. (c) Single crystal laser ring resonator in short-sample Hall geometry and planar laser in long-sample Hall geometry. (d) Confocal resonator with capacitive mesh outcoupler and lamellar grating with a tunable mirror at a distance D (electrical contact and metallized areas are shown in black)

to the Ge crystal. For electrical isolation, they are supported on ultra-pure Si or thin polymer foils. A tunable resonator consists of a lamellar, metal grating and a supported metal mirror which form a Fabry–Pérot resonator of length L . This resonator is coupled to the Fabry–Pérot resonator of the polished Ge crystal and selects the emitted THz frequency or wavelength λ according to $L = m\lambda/2$ at the order $m = 1, 2, \dots$. A mesh outcoupler supported on ultra-pure Si with a small curvature radius completes the confocal resonator (Fig. 4.12d).

The magnetic field is supplied by a superconducting solenoid, a conventional electromagnet or, more conveniently, by a pair of rare earth, high remanence, permanent magnets (Fig. 4.12a). The very compact design in Voigt configuration with permanent magnets [280] allowed the operation of an electrically excited THz semiconductor laser in a mechanical cryocooler, providing long-term liquid helium free operation [100].

Despite the low gain of the Ge laser of $\sim 0.14 \text{ cm}^{-1}$ [281], with respect to QCLs (Sect. 4.6), it can reach a quite high peak power of 40 W in pulsed operation [282] because the laser crystal can be made with a length of several tens of millimeters. Using mode-locking even higher peak power in ps pulses can be achieved [283–285]. The conversion efficiency can reach up to 10^{-4} . All crystal dimensions can be several millimeters in size, which favors the integration of quasi-optical, external tunable resonators (Fig. 4.12d). Ge lasers have developed into a tool for

THz spectroscopy utilizing, for example, their high spectral brightness to penetrate liquid water and aqueous protein solutions [286]. THz Ge lasers are described in detail in [271] and its references. In the following paragraphs the basic properties of Ge lasers and the laser mechanism are described.

The population inversion in Ge lasers can be understood by a simple semi-classical model [271, 272]. The E/B -ratio can be tuned so that heavy holes are accelerated by the electric field E in a streaming motion to the optical phonon energy, and then inelastically scattered by optical phonons into the light hole band (Fig. 4.11a). The magnetic induction B splits the valence band of the heavy and light holes into equally spaced Landau levels. Due to the lower effective mass of light holes in Ge, they experience negligible optical phonon scattering and are accumulated in light hole Landau levels. The ratio E/B is chosen in such a way that light holes have a small and heavy holes a large momentum p_L and p_H , respectively, so that the energy is below the optical phonon energy for light holes ($E_L < E_{OP}$) and above for heavy holes ($E_H \geq E_{OP}$). Subsequently, a population inversion between light and heavy holes builds up and amplification and laser action on direct optical transitions between the Landau levels of light-hole to heavy-hole subbands occur over a wide frequency range (Fig. 4.11b).

At low lattice temperatures, the scattering rate of holes, having a kinetic energy below the optical phonon energy ($E_{OP} = 37 \text{ meV}$ for Ge), is relatively low (10^{10} to 10^{11} s^{-1}) and controlled by the interaction with acoustic phonons and impurity scattering. The latter is limited by a low acceptor doping concentration in ultra-pure Ge in the range from 10^{12} to 10^{15} cm^{-3} and negligible compensation of 1% and less. Above the optical phonon energy holes emit optical phonons very rapidly with a rate exceeding 10^{12} s^{-1} . As a result, a highly anisotropic distribution of hot carriers in momentum space exists when an electric field accelerates holes up to the optical phonon energy.

The effective mass of carriers in semiconductors depends on the crystallographic direction and, during carrier transport, on the electric and magnetic field direction in which these are applied. The effective mass ratio of heavy and light holes is approximately eight if the magnetic induction is applied along the $\langle 110 \rangle$ crystal axis. The laser performance also depends strongly on the homogeneity of the electric field in the laser crystal, which is a function of the Hall fields, local doping imperfections, and crystal anisotropy.

Two laser geometries of parallelepipeds are favorable due to the Hall effect. The first, or short-sample Hall geometry, is planar, in which two opposite surfaces are closest and fully covered with an ohmic contact forming an oversized metal–metal waveguide [272] (Fig. 4.12c). A homogeneous electric field is obtained between the ohmic contacts, only marginally affected by the Hall effect. This design further allows efficient device cooling via the ohmic contacts attached to the Cu heat sinks. The second, or long-sample Hall geometry, is a long parallelepiped with ohmic contacts on the most distant opposite surfaces (Fig. 4.12c). The field distribution is strongly influenced by the Hall effect, leading to a homogeneous electric field distribution in which the potential is tilted by the Hall angle with respect to the long axis of the crystal. Figure 4.12b shows a design which combines the long-sample

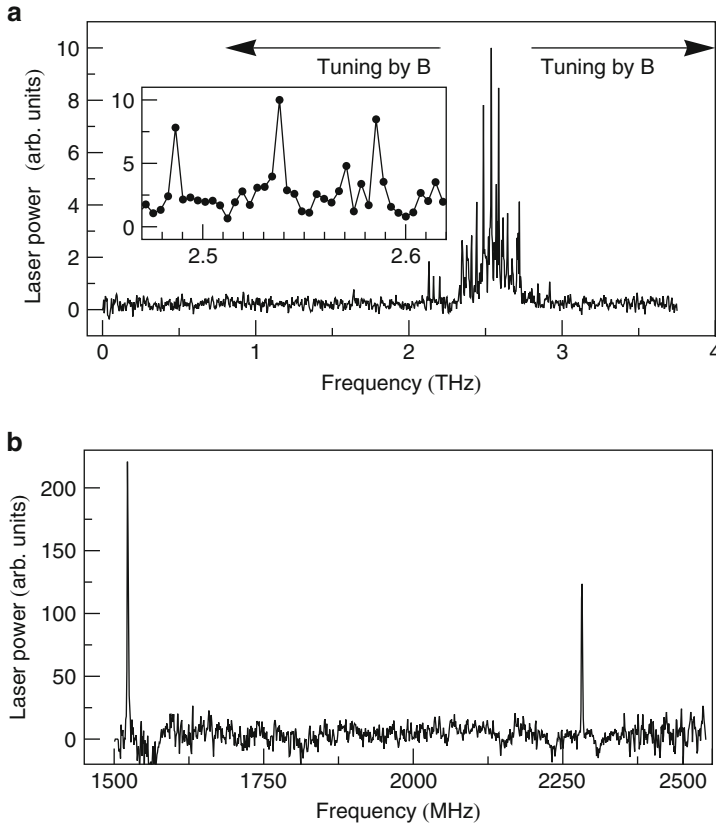


Fig. 4.13 (a) Typical spectra of a Ge:Be laser measured with an FTS and Ge bolometer for 12 μ s long laser pulses at 300 Hz repetition rate. The spectral resolution of the FTS was ~ 4 GHz. The main lines are spaced by ~ 0.5 THz. The magnetic induction B of 1 T is oriented along the $\langle 110 \rangle$ crystallographic axis and the electric field E of 1 kV/cm along the $\langle 100 \rangle$ axis. The THz emission can be tuned by the magnetic induction beyond 4 THz. The low THz range is achieved at low magnetic induction of ~ 0.2 T or with CR Ge lasers according to 0.6 THz/T. (b) Ge:Ga laser emission measured with heterodyne spectroscopy for a 4 μ s long laser pulse using an acousto-optical spectrometer. The linewidth of less than 1 MHz is Fourier-transform limited due to the emission pulse length (adapted from [271, 278])

Hall geometry with a planar design [274]. The latter is favorable for polishing a Ge laser crystal into a thin etalon forming a Fabry–Pérot resonator similar to a vertical cavity surface emitting laser.

Without any external cavity selection, Ge lasers have a broadband multi-mode emission, which is tunable in the range from 1 to beyond 4 THz by the applied fields (Fig. 4.13). The linewidth is less than 1 MHz. THz laser radiation, detected along the magnetic field axis, in Faraday configuration, is elliptically or circularly polarized. In the Voigt configuration (Fig. 4.12a,b), linearly polarized THz light is detected perpendicular to both applied fields.

The actual emission spectrum depends on the dopant. Originally, lasers were doped with Ga but group-III acceptors Tl, Al, and In also support laser emission. These shallow acceptors absorb on impurity-to-continuum transitions in the range from 11 to 13 meV, which overlaps with the emission causing frequency gaps in the laser spectrum. In order to eliminate this effect, Ge was doped with double acceptors (Be, Zn) and Cu instead [287, 288]. The dopant Cu is easily introduced by diffusion into ultra-pure Ge simplifying the production of Ge laser material. In contrast with shallow acceptors, the ionization energies of Be, Zn, and Cu are larger, 25, 33, and 43 meV, respectively, and well above the emission frequency of the Ge laser.

Very small Ge:Ga lasers have also been demonstrated under uniaxial stress of up to 0.4 GPa (4 kbar), showing laser emission with pressure but no laser emission without pressure [289]. Typical stress values exceeding 0.4 GPa are applied along a crystal axis such as $\langle 111 \rangle$ or $\langle 100 \rangle$. Uniaxial stress can lift the band degeneracy of heavy and light hole bands at the Γ -point. This suggests that uniaxial stress leads to a reduction of loss mechanisms and resulting in laser action.

Ge laser emission covers a broad spectral band, so it is possible to achieve short pulses by mode-locking techniques [290]. Active mode-locking modulates the gain through additional electric fields applied parallel to the magnetic field, thus breaking the condition of crossed fields. This decreases the lifetime as well as the population of the light-hole subband. If the parallel electric field is modulated at a radio frequency (RF) corresponding to twice the round-trip time of the laser radiation, the gain is at its maximum each time the RF field is zero. Therefore the gain is lowest each time the RF field is at its maximum value. The gain is periodically modulated and mode-locking is induced, which has been demonstrated with Ga-, doped [283, 284] as well as Cu-doped Ge laser crystals [285]. Pulses as short as ~ 50 ps were generated. The pulse length was limited by the number of equidistant laser modes. In addition, passive mode-locking has been obtained [291].

4.5.2 *Lasing from Optically Excited Donors in Silicon*

THz photoconductive detectors rely on the ionization energy of shallow or deep acceptors or donors to define a sensitive band of detection. Impurities have a hydrogen-like energy level diagram with a ground state, determined by impurity interaction with the lattice, and an excited energy level structure rather similar for shallow and deep impurities approaching the valence or the conduction band for acceptors or donors, respectively (Sect. 5.4).

The first indication that laser emission can occur by optically excited impurity states was proposed after analyzing the THz emission from a Ge laser doped with a Ga acceptor concentration of $\sim 10^{14} \text{ cm}^{-3}$. Stimulated emission on transitions between excited states and the ground state of the acceptor was found. Spectral measurements revealed a characteristic discrete structure, which was attributed to impurity transitions related to the G (1.6 THz), E (1.9 THz) and C (2.2 THz) lines of Ga. It was assumed that stimulated emission leads to the depletion of the ground

state resulting in a population inversion and amplification on transitions from the excited states to the ground state [292].

Follow-up calculations suggested that population inversion and laser emission might be obtained from Si doped by group-V impurities [293], and this was experimentally discovered by Pavlov et al. in 2000 [294]. For all group-V donors in Si, there exist particular excited states which are long-lived with respect to the $1s(E)$ and $1s(T_2)$ states. At low lattice temperatures ($T < 30$ K), the interaction with phonons is small for the donor centers and the population of the excited states is controlled by acoustic-phonon-assisted relaxation (Fig. 4.14a). Population inversion of the intracenter states can be achieved by either optical excitation into the conduction band with radiation from a CO_2 laser or by resonant optical excitation into one of the excited states with radiation from a free electron laser (FEL). In the first case, optically excited electrons lose their energy through their interaction with acoustic or optical phonons in the conduction band (Fig. 4.14a). In the second case, the electrons are pumped directly into one of the excited impurity states. Various Si lasers with different dopants (P, Sb, As, Bi) were demonstrated (for a review, see [295]). Typically, these lasers are small square-shaped crystals with a length

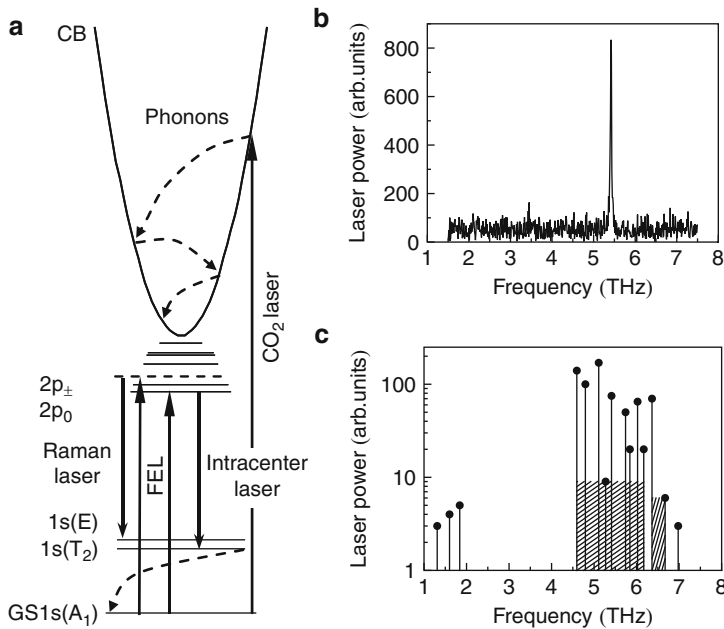


Fig. 4.14 (a) Lasing scheme of an Si:P laser. The straight upward arrows indicate optical excitation with a CO_2 laser or an FEL. The straight downward arrows indicate laser transitions. Dashed arrows illustrate non radiative decay of electrons due to interaction with phonons. The non radiative decay from upper to lower excited states is not shown for clarity. (b) Typical spectrum of optically pumped Si:P laser. (c) Emission lines of all intracenter Si lasers and frequency ranges of Si Raman lasers (shaded areas)

of a few millimeters, which need to be cooled to below 30 K with liquid helium, or in a mechanical cryocooler. The crystals are doped in the range from 10^{15} to 10^{16} cm^{-3} . The laser transitions occur between particular impurity states and 15 laser transitions have been observed so far (Fig. 4.14b,c). Most of these are in the range from 3.5 to 7 THz, but some are between 1 and 2 THz. Frequency tunability can be achieved by applying a magnetic field or a compressive force, and multi-frequency operation has been obtained from Si crystals doped with two different donors [296]. A much broader frequency coverage is provided by Raman lasing from n-doped Si crystals under optical excitation with an FEL [297].

Optically excited Si lasers have existed for some years. Electroluminescence from a variety of electrically pumped Si structures has also been obtained (e.g., [298]). THz lasers based on Si are particularly desirable because of their compatibility with Si technology [299] but a compact, electrically pumped THz laser remains elusive.

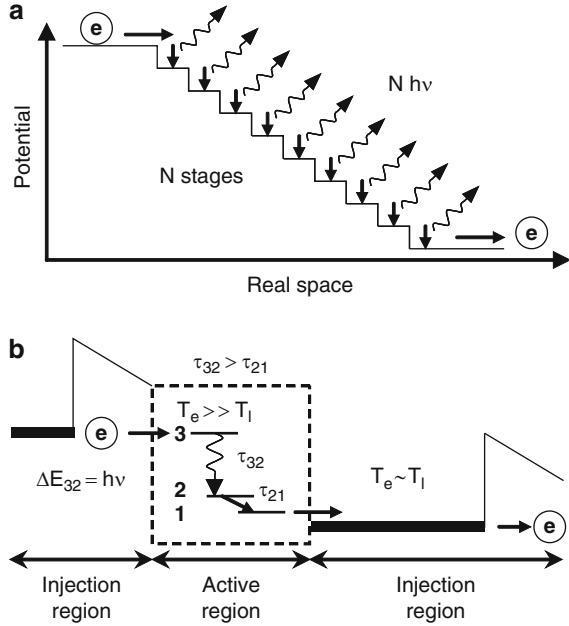
4.6 Quantum Cascade Laser

The quantum cascade laser (QCL) is a very promising source for many applications in the THz spectral region. QCLs have three marked differences compared to diode lasers. Firstly, only electrons take part in the laser process. Therefore, QCLs are unipolar devices. Secondly, the whole laser process occurs between conduction band states. These are called subbands and accordingly QCLs are intersubband devices. Thirdly, a cascading scheme is employed where one electron can generate many photons. The idea of a unipolar laser based on intersubband transitions can be traced back to the work of R. F. Kasarinov and R. A. Suris in 1971, who formulated a theory of light amplification in a superlattice under an applied electric field [300]. In 1989, M. Helm et al. succeeded in observing intersubband luminescence from a superlattice excited by resonant tunnelling [301]. An IR QCL operating at $4.2 \mu\text{m}$ was realized in 1994 at Bell Laboratories by J. Faist et al. [302]. It took another eight years before A. Tredicucci and his co-workers made the THz QCL [108].

4.6.1 Principle of Operation

The operating principle of a QCL is illustrated in Fig. 4.15a. An electron is injected into the laser structure and cascades down a potential staircase, which is created by a series of quantum wells exposed to an electric field (for a more detailed description, see [303–305]). At each step, a photon is emitted. A simplified scheme of one step or module is shown in Fig. 4.15b. The electron is injected from the left into level 3 of the active region. The laser transition occurs from level 3 to level 2, from where the electrons are quickly extracted into level 1 and the injection region of the next step. In this very simple scheme, the condition for population inversion is $\tau_{32} > \tau_2$, where $1/\tau_{32}$ is the non radiative scattering rate from level 3 to level 2 and $1/\tau_2$ is the

Fig. 4.15 (a) The operating principle of a cascade laser with N stages. At each stage, the electron drops down (straight arrow) and emits a photon (curved arrow). (b) Principle of a single stage of a cascade structure: it consists of an injection region, an active region, and a relaxation region which serves as the injection region for the next module (adapted from [303])



total rate out of level 2, including the transition to level 1 (characterized by $1/\tau_{21}$). The lifetimes in levels 3 and 2 are of the order of sub-ps to ps. The peak gain of an intersubband transition is proportional to the population inversion between levels 3 and 2 and the oscillator strength of that transition. It is inversely proportional to the linewidth of the transition. If the overlap between the wavefunctions in level 3 and level 2 is large, the transition is said to be vertical, and when the overlap is small the transition is said to be diagonal. Vertical transitions have a larger oscillator strength than diagonal ones.

A variety of quantum well structures exist which empower lasing and the active region is usually made of several quantum wells, not just one. They are mostly based on GaAs/AlGaAs superlattices. The most widely used designs can be divided into three groups: chirped superlattice (CSL), bound-to-continuum (BTC), and resonant-phonon (RP) design (Fig. 4.16). The CSL is based on the coupling of several quantum wells, which form a so-called miniband when a voltage is applied [108]. Two minibands (approximate width 15–20 meV), one for the upper laser state and one for the lower laser state, separated by a minigap, are formed in the conduction band of the superlattice. The laser transition takes place from the lowest state of the upper miniband to the uppermost state of the lower miniband. This is similar to a band-to-band transition in a conventional bandgap laser, with the distinction that both minibands and the minigap are part of the conduction band. Population inversion can be achieved because scattering between electrons in one miniband is favored over scattering of electrons from one miniband to the other.

The BTC design is based on a chirped superlattice but the upper laser state is a single (bound) state in the minigap instead of a miniband [306]. In the CSL,

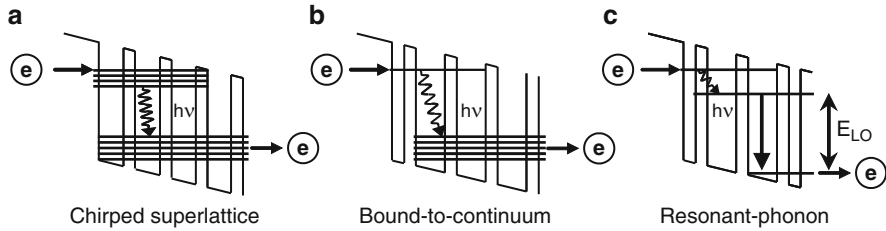


Fig. 4.16 Simplified conduction-band diagrams of one module of several different QCL designs: (a) chirped superlattice, (b) bound-to-continuum, and (c) resonant-phonon. Note that each module typically consists of up to approximately ten quantum wells and that a QCL consists of approximately 100 cascaded modules (adapted from [303])

the lower laser state is at the top of the lower miniband. The laser transition in this design is more diagonal in real space and the oscillator strength is therefore somewhat smaller than in the CSL, but non radiative scattering is reduced. The injection process is more selective for the upper state because the injector states couple more efficiently with this state than with the lower miniband. In summary, BTC lasers have superior performance compared to CSL lasers with respect to operating temperature, laser threshold, and output power.

The RP active medium makes use of the LO phonon in GaAs ($E_{LO} = 36$ meV), which couples the lower laser state resonantly with the upper laser state of the adjacent quantum well [307]. By this means electrons in the lower laser state will scatter very quickly into the injector state by emitting an LO phonon. The challenge of this approach is to depopulate, selectively, the lower state and not the upper state, which is close-by. This is achieved by designing the quantum wells in a manner where there is a broad tunneling resonance between the lower laser state and the adjacent injector state. The laser transition in the RP design has a smaller oscillator strength than the BTC laser. But a module of an RP laser is about half as thick as that of a BTC laser, which results in more gain-producing laser transitions for a given thickness of the active medium.

The active region of THz QCLs is based on GaAs/AlGaAs superlattices grown by molecular beam epitaxy (MBE). The superlattice is 10–15 μm thick. One module consists of about four (RP design) or eight (CSL, BTC design) quantum wells, yielding a module thickness of ~ 55 nm (RP design), or ~ 110 nm (CSL, BTC design). This amounts to about 1000 quantum wells per QCL. These all need to be grown with extreme accuracy on the scale of an atomic layer. Most QCLs are linear ridges (a few mm long and 150–250 μm wide), which form Fabry–Pérot resonators. One of the end facets may have a mirror coating. In addition lasers with sub-wavelength microdisk or microring resonators have been fabricated [308–311].

The thickness of the active medium is up to 10 times smaller than the laser wavelength in the active medium. In addition, free carrier absorption increases with the square of the wavelength and is correspondingly large at THz frequencies. This requires special waveguide designs which should maximize the overlap between the laser mode and the active medium, on the one hand, and on the other hand minimize

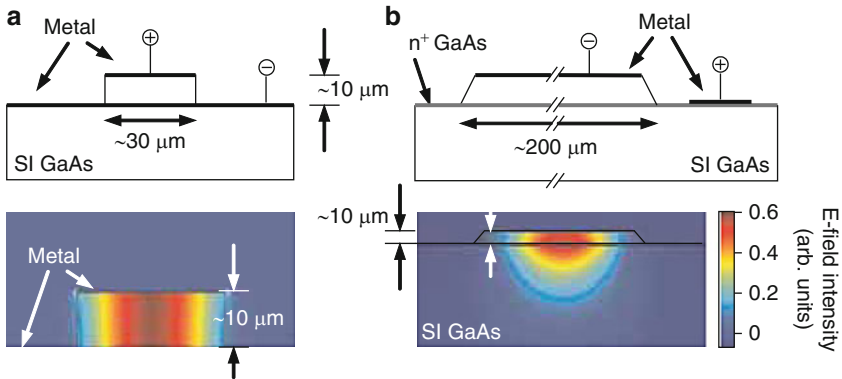


Fig. 4.17 Cross section of a QCL with metal–metal waveguide (a) and with surface-plasmon waveguide (b). At the bottom, a typical two-dimensional mode intensity pattern for both waveguides is shown (adapted from [305])

the absorption loss (characterized by the loss coefficient α_w) in the waveguide. Also, the loss at the mirrors α_M needs to be taken into account. The mode overlap is characterized by the confinement factor Γ , which is a measure for the fraction of the laser mode, which is confined in the active medium ($\Gamma = 0, \dots, 1$). If Γ is equal to 1, the mode is completely confined in the active medium. The gain g_{th} required to reach the lasing threshold must equal the total losses, i.e. $g_{th}\Gamma = \alpha_w + \alpha_M$. Essentially, two types of waveguide are in use: the surface-plasmon or single-plasmon (SP) waveguide and the metal–metal (MM) waveguide (Fig. 4.17). The SP waveguide consists of a thin, highly doped layer between the active medium and the substrate and a metal layer on top of the laser ridge (Fig. 4.17b). The ridge is about 200 μm wide. It has relatively low loss but suffers from an incomplete overlap of the laser mode with the active medium, because the laser mode extends significantly into the substrate. The confinement factor Γ of this type of waveguide is around 0.5 so that half of the power is located in the waveguide, and the remaining power leaks into the substrate. The MM waveguide has a metal layer instead of the doped layer between the active medium and the substrate (Fig. 4.17a). This results in better thermal coupling to the substrate and a confinement factor of one. However, the MM waveguide has a higher loss than an SP waveguide. A variety of modifications to the two main waveguide structures discussed so far are in existence. For example, the top metal layer can be a grating in order to provide a distributed feedback structure. This results in a single emission line [312]. Also, the top layer can be a photonic bandgap structure in order to improve the beam profile [310].

4.6.2 Performance

QCLs with emission frequencies between 1.2 and 4.9 THz have been attained (Fig. 4.18a). The lower frequency limit can be extended to 0.8 THz by applying a magnetic field. However, none of these lasers covers a very broad frequency range.

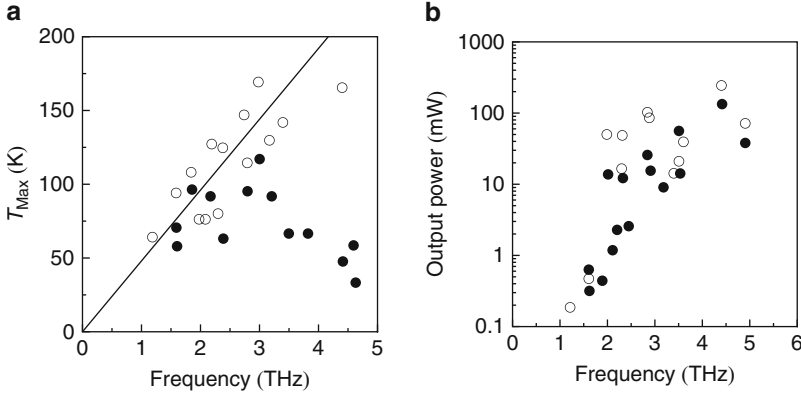


Fig. 4.18 Typical best performance values of QCLs. (a) Maximum operating temperature and (b) maximum output power as a function of frequency. Hollow circles indicate pulsed laser operation while filled circles indicate cw operation. The straight line in (a) corresponds to $h\nu = k_B T$ (adapted from [305] and various sources)

Typically, only about 10 GHz is covered by a single mode, due to temperature or current induced frequency tuning. The operating temperature of a cw QCL can be as high as 130 K. The output power in cw mode can reach up to ~ 140 mW. In pulsed operation, a maximum temperature of 186 K was obtained (Fig. 4.18b). However, since for most of the lasers emission ceases above 70 K, they are operated in a liquid He flow cryostat or in a mechanical cooler. Two major processes terminate the laser action in a QCL. The first is thermal backfilling of the lower laser state with electrons from the injector which gain energy by thermal excitation, or by re-absorption of non equilibrium phonons. The second is LO phonon emission of electrons in the upper laser state, which leads to non radiative decay from this state. A typical performance of a BTC laser is shown in Fig. 4.19. For many applications, a Gaussian beam profile and an M^2 -value close to 1 are essential (2.11). The beam profile of a laser with an SP waveguide consists of two lobes. One is emitted from the laser ridge and the other from the substrate (Fig. 4.20) [189, 313]. In addition, a fringe pattern appears around the lobes. This structure is more pronounced for lasers with short wavelength output, while at wavelengths above about 100 μm the lobes start to merge due to diffraction effects. By integrating the signals of the two lobes and the fringe pattern separately the confinement factor for a 4.3 THz laser was determined to be 0.49, which is in good agreement with the computed value of 0.47 [313]. In the direction vertical to the layers of the superlattice, the width of the beam pattern (ridge lobe as well as substrate lobe) is determined by a single aperture. The fringe pattern corresponds to a diffracting aperture, which is about the size of the substrate and ridge. In the direction orthogonal to this, the FWHM is determined by λ/w with λ the emission wavelength and w the width of the ridge. With such a laser, an M^2 -value of less than 1.2 has been demonstrated [314].

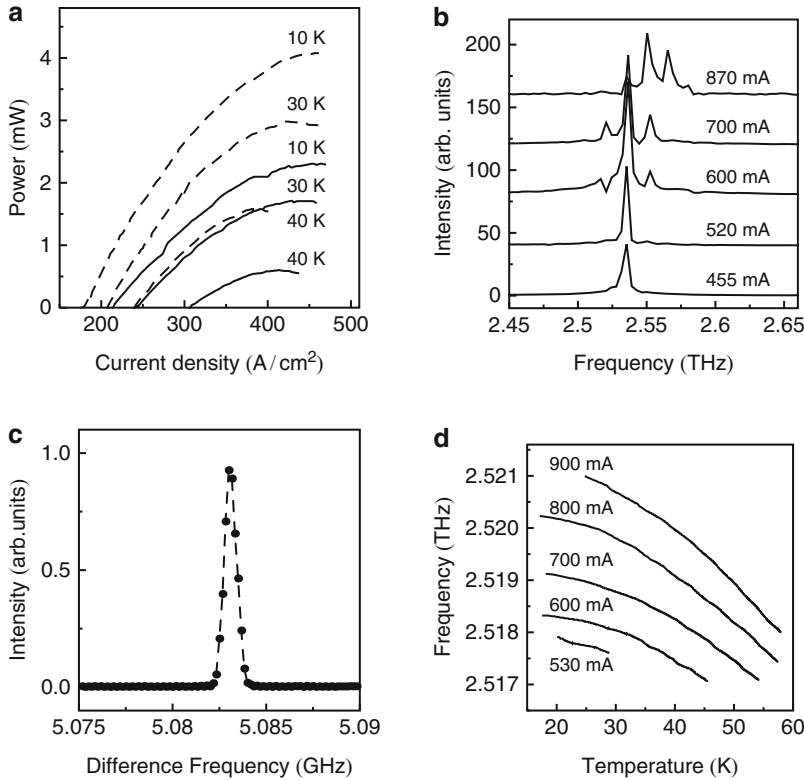


Fig. 4.19 Typical performance of QCLs. **(a)** Light-current characteristics of two chirped superlattice QCLs, recorded in cw operation at different heat sink temperatures. The *solid lines* are measured with a 1.96 mm long and 150 μm wide laser with both facets uncoated. The *dashed lines* represent data collected from a 2.23 mm long device with a coated back facet. Due to the slightly different device sizes, the scale of the current refers only to the *dashed lines* (adapted from [315]). **(b)** Emission spectrum of a 2.5 mm long bound-to-continuum QCL, with a Fabry-Pérot resonator, as a function of current. With increasing current the gain profile shifts toward higher frequency and more laser modes appear. The spectra are offset for clarity. **(c)** Linewidth of the emission of a bound-to-continuum QCL with a DFB grating. The profile was measured by mixing the emission from an optically pumped gas laser with that from the QCL at 2.5 THz. The frequency of the QCL was not stabilized. The resulting difference frequency is approximately 5.083 GHz, and the FWHM is 0.7 MHz measured with an integration time of 0.3 s (adapted from [316]). **(d)** Frequency tunability of the same QCL as in c) as a function of temperature and current (adapted from [316])

The beam pattern of a QCL with an MM waveguide is much more divergent than that of an SP QCL. The lobe structure is not observable. Instead, a pronounced ring-like pattern exists which results from the far-field interference of the coherent radiation emitted by all facets of the laser ridge. These lasers have subwavelength cross-sections and are much longer than the wavelength. They are called “wire lasers” and their beam profile can be described by an antenna model [317, 318]. Therefore, several methods have been developed in order to improve the beam

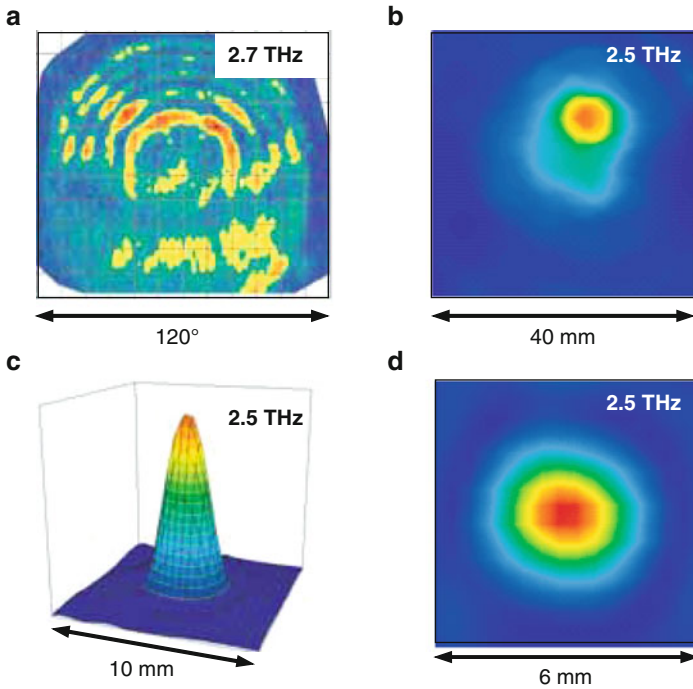


Fig. 4.20 Typical beam profiles of QCLs with metal–metal waveguide (a) and surface-plasmon waveguide (b–d). In (a), a pyroelectric detector was scanned on a spherical surface around the QCL at 80 mm distance. The diagram is an equiangular projection (adapted from [318]). The pronounced fringe pattern is due to interference effects. The diagram in (b) was measured by scanning a Golay cell detector in a plane orthogonal to the emission direction at 60 mm distance from the QCL. The emission has a pronounced peak and a plateau. These correspond to emission from the laser ridge and the substrate into which part of the laser mode extends. In (c) and (d), the beam profile of the same laser as in b) is measured at the position of the minimum beam waist after transformation of the output profile by a lens. The corresponding M^2 -value is approximately 1.2 (adapted from [315, 316])

pattern of QCLs with MM waveguides. These are a silicon lens at the output facet [319], a horn antenna at the output facet of the ridge [320], a wide area surface emitting structure [321], or a photonic bandgap structure [310]. Yet none of these approaches has resulted in an improved beam profile when compared to the SP waveguide QCL.

Beating of two laser modes of a QCL with a multi-mode Fabry–Pérot resonator yields an upper limit of about 20 kHz (FWHM) for the linewidth measured within about 4 ms [313, 322]. This linewidth increases to several MHz within one minute. Phase-locking of two longitudinal modes of a 2.7 THz QCL has been demonstrated and the beat linewidth was less than 10 Hz. Under frequency stabilization, the line profile was found to be Lorentzian with a minimum linewidth of ~ 6.3 kHz [323]. The frequency of a QCL has been locked to that of a THz gas laser [324] as well to that of a multiplied microwave source [325]. The FWHM was less than

100 Hz and the lock condition could be maintained indefinitely. Another method is based on frequency modulation of the laser emission across a molecular absorption line. A Gaussian line shape with 300 kHz width has been demonstrated [326]. This stabilization scheme is robust and versatile, because it requires only an additional detector and a small gas absorption cell, while being applicable even at the highest THz frequencies, due to the rich absorption spectra of molecules such as CH_3OH or H_2O . Phase-locking of a longitudinal mode of a 2.7-THz QCL to a mode-locked erbium-doped fiber laser yielded an RF beat-note signal with 80 dB of (S/N)-ratio and a bandwidth of 1 Hz [327]. These results demonstrate that THz QCLs can be frequency controlled to very high accuracy.

The frequency of a QCL can be tuned by changing current or temperature. The tuning is determined by the change of the optical length of the resonator due to changes of its refractive index. Another effect which changes the laser frequency is the quantum confined Stark effect. With increasing electric field the separation of the laser states changes, the peak of the gain profile shifts toward higher frequencies, and frequency pulling leads to a shift of the emission frequency. Excluding mode hopping, the QCL laser frequencies were reported to be linearly tunable by the current as well as the laser temperature. Frequency-tuning coefficients $(1/f) \times (\Delta f/\Delta I) \approx -(10^{-2} \text{ to } 10^{-1}) \text{ A}^{-1}$ have been found for RP QCLs [322, 328]. By mixing the radiation from CSL and BTC QCLs with that from a THz gas laser, tuning coefficients of about 10^{-3} A^{-1} have been found [322]. For these types of laser positive tuning coefficients have also been observed ($3 \times 10^{-3} \text{ A}^{-1}$). Temperature-related shifts have been reported ranging from $(1/f) \times (\Delta f/\Delta T) \approx -(10^{-5} \text{ to } 10^{-6}) \text{ K}^{-1}$ [328] for temperatures below 50 K. The changes in the emission frequency were attributed entirely to temperature changes of the refractive index of the active medium, regardless of whether they originated from variation of the current or from variation of the temperature [322]. A peculiarity has been observed with an RP QCL with a sub-wavelength microdisk resonator. It changed the sign of the tuning rate from a negative (-0.57 A^{-1}) to a positive ($+0.14 \text{ A}^{-1}$) value with increasing electric field. This was attributed to a shift of the maximum of the gain curve by the quantum-confined Stark effect [308]. In general, RP QCLs seem to have a broader tunability with current than lasers with CSL or BTC design.

Tuning by current or temperature covers a range of several GHz. Integration of a QCL into an external cavity enables the frequency tuning range to be extended, possibly across the whole gain bandwidth. With a BTC QCL in an external cavity coarse tuning up to ~ 90 GHz was obtained near the center frequency of 4.8 THz, and continuous mode-hop-free tuning was observed over ~ 10 GHz [172]. An elegant way to achieve broadband frequency tunability has been demonstrated for a “wire laser.” Placing a dielectric or metallic moveable object close to the wire laser manipulates a large fraction of the mode which propagates outside the cavity, thereby tuning its resonant frequency. Continuous frequency tuning of ~ 137 GHz corresponding to 3.6% from a single laser device at ~ 3.8 THz has been demonstrated [329].

Since the demonstration of the THz QCL in 2002, significant progress has been made. It has been shown that the fundamental requirements for many applications

can be met and proof-of-concept experiments have been performed in imaging [330], spectroscopy [331], and for use as an LO in a heterodyne receiver [313, 332]. Cw operation at liquid nitrogen temperature or in a small Stirling cooler, with sufficient output power and a good beam profile, is especially needed and was recently demonstrated [333]. A further progress is the integration of a Schottky diode mixer and a QCL in a monolithically integrated THz transceiver, which has the basic functions of a discrete-component transceiver. However, at this stage of development its performance is not sufficient for application in a practical system [334]. A major challenge remains to achieve all the specifications required for a particular application in a single device. Once these issues are resolved, THz QCLs will be a very valuable source which may find wide-spread applications.

4.7 Photomixing for the Generation of Terahertz Radiation

Electromagnetic waves with different frequencies generate sum and difference frequencies in an element with a nonlinear I-V curve or any other nonlinear element. This frequency mixing is well known in radio technology, where mixing typically occurs in a diode. Heterodyne detection is described in Sect. 5.5. In this section, the focus is on the generation of THz waves as a difference frequency in a nonlinear element. The term “photomixing” describes the generation of THz photons by using two IR or two visible laser photons, respectively, for which the laser frequency difference corresponds to a THz frequency. This is also called difference frequency generation (DFG). Using fixed frequency THz (gas) lasers and microwave frequency mixing allows the generation of tunable THz radiation, due to the tunability of the microwave generators.

Initially, the THz nonlinear elements investigated were GaAs-Schottky and metal-insulator-metal (MIM) diodes. High-frequency lasers such as CO₂ lasers were used with MIM devices and THz gas lasers and microwaves with Schottky diodes. Photomixing has gained more interest due to the availability of high power, room temperature visible and near-IR lasers. Nonlinear materials based on semiconductors, most prominently low-temperature-grown (LTG or LT) GaAs [335, 336], and on crystal materials such as ZnTe, GaP, and GaSe [337] have been used. Photomixers based on semiconductors are normally integrated in an antenna structure to couple out the THz waves efficiently (Sect. 3.15). The low conversion efficiency from the high-frequency light into the THz frequency range requires lasers with an output power up to a few 100 mW, and thermal management to remove heat from the photomixer. This is more important for semiconductors than for crystal materials [337] because local heating of the semiconductor can lead to migration of dopants, undesired annealing, and subsequent change of the doping level and recombination centers [336]. Often a bias voltage is applied to the THz antenna structure to further increase the THz output power [338].

The first step in simplifying the technology is the use of two-color lasers, so that a single laser medium generates two visible or near-IR laser photons simultaneously.

Now a single laser beam is focused on the nonlinear element for photomixing [339, 340]. A little later it was demonstrated, with an off-the-shelf laser diode at 835 nm [341], that the laser medium of a two-color laser could generate THz waves due to the highly nonlinear dynamics of the charged carriers. The optical sidebands at the laser frequencies were tunable by the external cavity from 0.3 to 2.5 THz in cw systems. This is extended to 4.2 THz in pulsed operation. A cw source based on a dual color vertical cavity surface emitting laser has been shown to deliver 2 mW at 1.9 THz. Mixing of two laser modes was accomplished in an intracavity nonlinear crystal, and frequency tunability was achieved by changing the spacing of the two laser modes [342].

The inverse process was shown by injection of a near-IR laser into a THz QCL cavity. Near-IR sidebands were generated, spaced by the THz frequency. Furthermore, a two-color IR QCL generated THz radiation by internal photomixing [343] at low temperature. This has been improved up to room temperature operation for a two-color QCL designed at 8.9 and 10.5 μm . At frequencies from 4.5 to 5 THz, a power of 0.15 μW at a duty cycle of 1.5% was detected [344].

Direct THz emission from single laser devices requires simultaneous optimization of the two-color high-frequency laser structure, the THz wave mode confinement and the outcoupling. The latter is of utmost importance because the two-color device structure is typically much smaller than the THz wavelength to be emitted. An important issue is also the doping concentration. A high doping concentration, required for visible and IR lasers, can prevent THz radiation because of the resultant plasma frequency.

4.7.1 *Low-temperature-grown Gallium Arsenide*

For a fast response of the emitter material in the THz regime, a relaxation time of the photo carriers of less than 1 ps is required. On a GaAs substrate a layer of about 1 μm thickness is grown by molecular-beam epitaxy (MBE) with an arsenic over pressure with a nonstoichiometric content of 1% arsenic at low temperatures of 200–250°C [336]. The Ga-source in the MBE chamber is closed and a thermal annealing step (10 min at 600°C) in an arsenic atmosphere completes the LTG-GaAs. The excess arsenic atoms primarily move to Ga lattice positions, forming antisite defects as well as metallic As clusters and precipitates. The low temperature results in arsenic clusters, which are considered as recombination centers for charge carriers created by the near-IR lasers. The abundant arsenic clusters lead to very short recombination times of 0.2 ps. The material can therefore follow THz frequencies, which allows the fabrication of THz detectors, mixers, and even THz emitters.

Frequently antenna structures, or optical elements such as lenses (Sect. 3.15), are integrated to improve the coupling of the THz waves to free space propagation [338, 345]. These components enable the free space propagation of the THz electrical field. In addition to a short photo-excited carrier lifetime of less than 0.2 ps, which is

less important for emitters in contrast to detectors, the substrate has to fulfill further requirements. A high mobility of LTG-GaAs in the range of $1,000 \text{ cm}^2/(\text{Vs})$ and a high breakdown field of several 100 kV/cm under dark conditions are beneficial. It should be noted that LTG-GaAs photomixers work at wavelengths around 800 nm . At telecom wavelengths of $1.5 \text{ }\mu\text{m}$ GaAs becomes transparent. Other materials such as InP can be used together with cheap telecom components [346].

4.7.2 Schottky Diode

Mixing the radiation of a fixed frequency THz gas laser and some type of frequency tunable microwave source in a Schottky diode provides the possibility of generating tunable THz radiation. This technique dates back to the 1970s [347–349]. In the following years, it has been improved significantly, and was successfully used to study high-resolution spectra of many molecules, free radicals, and ions [350, 351] (Sect. 6.7.3). The principle of the technique relies on the ability of Schottky diodes to generate sidebands when irradiated with radiation from two sources. One source is a THz gas laser operating at ν_{THz} and emitting more than 10 mW of power. The other is a microwave source operating at ν_{MW} . The Schottky diode generates the sideband frequencies $\nu_{\text{SB}} = \nu_{\text{THz}} \pm m\nu_{\text{MW}}$ with $m = 1, 2, 3, \dots$. The $+$ sign refers to the upper sideband and the $-$ sign to the lower sideband. Physically, the generation process is the same as in the case of a Schottky diode when used as a heterodyne mixer, but with two differences. First, the microwave radiation is quite powerful, typically tens of mW , as compared to the signal radiation and secondly, the sideband frequency is close to the laser frequency, while for heterodyne detection (Sect. 5.5) the frequencies of the signal and LO radiation are close to each other. In some sense, sideband generation is the inverse process to heterodyne mixing. The sideband radiation is emitted from the Schottky diode in the opposite direction to the incoming laser radiation. Schottky diodes with a corner cube antenna similar to the one described in Sect. 5.5.2 have been used. One problem is that laser radiation is emitted as well as reflected from the Schottky diode, and most of the power in the emitted beam is at the fundamental laser frequency. This necessitates sophisticated schemes to separate the sideband radiation from the laser radiation. It should be noted that not only the first-order sideband ($m = 1$) but also higher order sidebands can be used. However, the power emitted in the sideband decreases strongly for higher order sidebands [350]. Nevertheless, almost the entire frequency range from 0.5 to 3 THz is accessible by sideband generation [350]. It should be noted that this coverage cannot be achieved with a single Schottky diode, because the corner cube antenna does not provide sufficient bandwidth. Several diodes with optimized antennas have to be used. The output power emitted in the first order is about three to four orders of magnitude lower than the power from the gas laser, and it tends to decrease with increasing frequency due to efficiency loss in the Schottky diode. As an example, $10.5 \text{ }\mu\text{W}$ have been obtained at 1.6 THz [352] (Fig. 4.21). With a Schottky varactor diode, in a reduced waveguide mount, as much as $55 \text{ }\mu\text{W}$ have

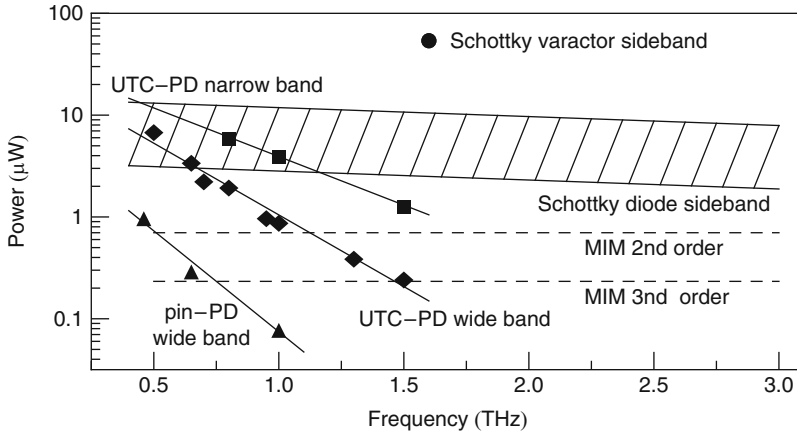


Fig. 4.21 Output power of different photomixers. Several mixers are required to cover the whole frequency band if narrow band antennas are used. The abbreviations are explained in the text (from various sources)

been produced at 1.6 THz [353]. However, despite these impressive results it seems that this type of photomixing is no longer widely pursued.

4.7.3 Metal–Insulator–Metal Diodes

A somewhat similar technique to sideband generation is difference frequency generation by mixing the radiation of two lasers operating in the IR or visible spectral region in a nonlinear element. While the concept is quite simple and dates back to the 1970s [354], it was 10 years later that K.M. Evenson and his co-workers implemented this method successfully for spectroscopic applications [355] (Sect. 6.7.4). They used two CO₂ lasers and a metal–insulator–metal (MIM) diode (sometimes also called MOM, metal–oxide–metal diode) for generation of the difference frequency. The MIM diode consists of an Ni metal base, which is contacted by an electrochemically sharpened tungsten whisker. Both metals are separated by a thin (~ 1 nm) naturally grown oxide layer. Applying a small bias voltage leads to tunneling of electrons from one metal to the other. Although the nonlinearity of the I–V curve of an MIM diode is much smaller than that of a Schottky diode, it is a more efficient mixer at $10\text{ }\mu\text{m}$ wavelength, because the tunneling process is much faster than thermionic emission in a Schottky diode [356]. The conical tip of the whisker acts as an antenna for in-coupling of about 200 mW of CO₂ radiation from each laser [357], while outcoupling of THz radiation is achieved with a long wire or corner cube antenna (Sect. 5.5.2). For generation of tunable THz radiation, a CO₂ waveguide laser, which is frequency tunable by $\Delta\nu_2 = \pm 120$ MHz, is mixed with another laser at a frequency $\nu_1 = \nu_{\text{CO}_2}$. The waveguide is tuned to

a frequency ν_2 . The THz radiation can be swept due to second-order generation: $\nu_{\text{THz}} = |\nu_1 - \nu_2| \pm \Delta\nu_2$.

In a configuration for third-order generation, with cobalt as a metal base, both CO_2 lasers are at fixed frequencies, and frequency tunable microwave radiation $\Delta\nu_3 = \Delta\nu_{\text{MW}}$ from a synthesizer is coupled into the MIM diode to allow for frequency tuning of the THz radiation. In this case, the MIM diode generates sidebands of the difference frequency from the CO_2 lasers, i.e. $\nu_{\text{THz}} = |\nu_1 - \nu_2| \pm \Delta\nu_3$. Since third-order generation provides much better frequency coverage, along with a higher absolute accuracy, it has found applications in high-resolution molecular spectroscopy (Sect. 6.7.4). With this approach, an almost complete (>95%) coverage of the THz region up to 6.5 THz is possible and, when one of the CO_2 lasers is replaced by a $^{15}\text{NH}_3$ laser, the coverage can be extended to 9.1 THz [358]. But, as with the Schottky photomixer, the corner cube antenna needs to be adapted. The output power is up to 1 μW in the second order and one third of that in the third order (Fig. 4.21). It is worth noting that MIM diodes can be fabricated as planar structures at the expense of reduced high-frequency performance [359].

4.7.4 Photodiodes

A more recent approach is photomixing of the radiation from two semiconductor lasers which operate at telecommunication wavelengths ($\lambda = 1.55 \mu\text{m}$) in a photodiode (PD). THz generation up to 1 THz has been reported using a pin-PD. However, these diodes have a limiting trade-off between bandwidth and output power and around 1 THz a few nW have been obtained [360]. The efficiency has been significantly improved by pin-PDs with waveguide coupling to near-IR radiation. These are based on InGaAs/InGaAsP heterostructures which are grown on semi-insulating InP substrates. Several nW up to several μW output power are reported for pin-PDs [346]. It is worth noting that a photonic mm-wave source with a pin-PD provides the phase reference around 100 GHz for the LO of each of the 64 antennas of the ALMA observatory. To increase the frequency of operation and the output power, a uni-traveling-carrier (UTC) PD is a better candidate [361]. Its bandwidth is at least three times higher and its output power at least two orders of magnitude greater than that obtained with a pin-PD. This improved performance is due to the fact that only electrons are the active carriers, which travel through the junction. In a UTC-PD electron-hole pairs are generated upon light absorption in a thin InGaAs absorption layer. The photo-generated electrons diffuse or drift into an InP collection layer, while the photo-generated holes respond very rapidly within the dielectric relaxation time by their collective motion. This differentiates it from a pin-PD, where the response time is limited by the low velocity hole transport, which is an order of magnitude smaller than the electron velocity. Another advantage of the UTC-PD is a higher saturation current because the space charge effect is less pronounced. The third advantage is that high speed, and high saturation current, can be obtained with low or even zero bias voltage, which reduces power consumption

and simplifies heat sinking. The UTC-PD is embedded in a planar antenna, and the chip with diode and antenna, is placed on the rear side of an Si lens, which provides the lens-antenna system for outcoupling of the THz radiation (Sect. 3.15). The radiation from the two 1.55- μm lasers is combined into a single fiber and coupled to the UTC-PD by a lens. Output powers of up to $10\text{ }\mu\text{W}$ below 1 THz have been obtained with an UTC-PD incorporated in a logarithmic-periodic antenna or a twin-dipole antenna, respectively (Fig. 4.21). It should be pointed out that when using a logarithmic-periodic, or any other frequency independent antenna, a single UTC-PD covers the whole frequency range. With a resonant antenna, the covered bandwidth is about 20% of the center frequency. In combination with a traveling wave coupling concept, up to 1 mW at 1 THz is predicted [362]. This indicates that there is considerable room for improving UTC-PDs with great potential for future applications.

4.8 Harmonic Generators

Sources based on harmonic generation extend microwave technology into the THz frequency range. This type of source relies on a cascaded multiplier chain, which is driven by a high-power (about 100 mW) source (Fig. 4.22). Impact ionization avalanche transit time (IMPATT) diodes and Gunn diodes deliver approximately 50 to 100 mW, but their tuning ranges are limited. Monolithic microwave integrated circuit (MMIC) power amplifiers are an alternative. They are available up to $\sim 100\text{ GHz}$ and deliver up to 150 mW at $\sim 100\text{ GHz}$ and 1 W at $\sim 35\text{ GHz}$ with a 10% bandwidth around the center frequency. Cooling may increase the output power by up to a factor of 10 [363]. The power amplifier is fed by a signal from a standard microwave signal generator, for example an yttrium iron garnet (YIG) oscillator, and its output is multiplied by a cascade of doublers or triplers. It should be kept in mind that for an ideal multiplier the output power drops with both frequency and the number of multiplications n while linewidth, drift, and tuning step sizes increase by n and the phase noise by $20\log(n)$ [364]. The output power can be smoothly varied, for example by varying the gain of the last amplifier or by a variable attenuator placed between the last amplifier and the following multiplier stage.

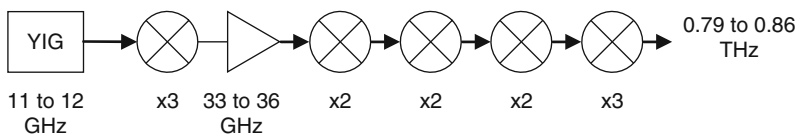


Fig. 4.22 Sources based on harmonic generation extend microwave technology into the THz frequency range. This type of source relies on a cascaded multiplier chain, which is driven by a high-power mmW source

The nonlinearity of Schottky barrier diodes can be used to generate harmonics of the pump source frequency. This is possible by using either their nonlinear capacitance or nonlinear resistance. The first type is called a varactor (variable capacitor) diode, while the second type is referred to as a varistor (variable resistor) diode. The structure of a Schottky barrier varactor is qualitatively the same as for a Schottky diode mixer. The two most important parameters for application in a harmonic generator are the conversion efficiency, i.e. the ratio between the power generated at the harmonic frequency and the pump power, and the output power. Varactors are operated under reverse bias compared to mixer diodes. This gives the best capacitance variation and limits the current. In order to achieve high conversion efficiency and high output power the design of a varactor diode differs from that of a mixer diode in several aspects: the doping density in the epilayer is relatively low (10^{16} to 10^{17} cm⁻³), the epilayer is thicker than in mixer diodes and the junction area is relatively large. By these means, the breakdown voltage of the varactor is increased compared to mixer diodes and the diode is able to handle large pump power. Varactors can be driven to the point where the junction begins to conduct. Beyond this point, the conversion efficiency drops and the output power saturates. In an ideal diode, with a uniformly doped epilayer, the voltage has a square law dependence on the charge and the current. If the diode is sinusoidally excited at the fundamental frequency, and open-circuited at all harmonic frequencies, the current can have no other harmonic components beyond the second. Higher harmonics have to be generated by higher order mixing. For example, the third harmonic output can be generated by mixing the fundamental frequency with the second harmonic. In order to have a large second harmonic current, a short circuit across the varactor at the second harmonic is required. This is called an idler circuit. To generate higher harmonics, idlers must be provided at all intermediate harmonics. It should be noted that this consideration is strictly valid only for varactors with an ideal square law charge–voltage curve. Real varactors deviate from the ideal square-law and higher harmonics can be generated even without an idler. Nevertheless, the use of idlers improves the performance of varactors as well as other reactive frequency multipliers.

In a similar fashion to Schottky mixer diodes, the advent of planar multipliers in conjunction with advanced design tools and fabrication technologies has led to multipliers that operate over a complete waveguide band without mechanical tuning (Fig. 4.23). Depending on the particular design, they reach conversion efficiencies of about 30% at 0.3 THz, which drop to a few percent at 1.9 THz [365, 366]. The output power decreases from 10 mW at 0.3 THz to a few μ W at 1.9 THz (Fig. 4.24). A somewhat different varactor diode is the heterostructure barrier varactor (HBV). In this device, an undoped high band gap semiconductor (e.g. AlGaAs or InAlAs) is sandwiched between two low band gap semiconductors (e.g. GaAs or InGaAs) [367]. Several such units can be stacked on top of each other. The potential barrier between the materials prevents an electron flow through the structure without applied bias voltage. If a bias voltage is applied, electrons are accumulated at one side of the barrier and depleted at the other side, resulting in a

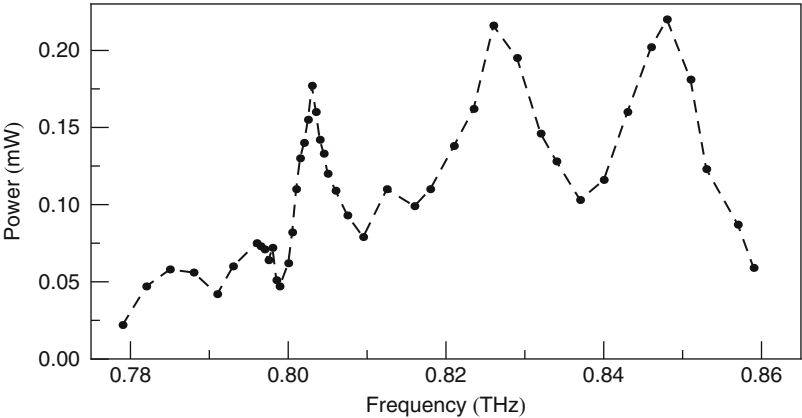


Fig. 4.23 Output power of the multiplier source depicted in Fig. 4.22. Note that the peak output power can be increased at the expense of frequency coverage (*dashed line as guide to the eye*)

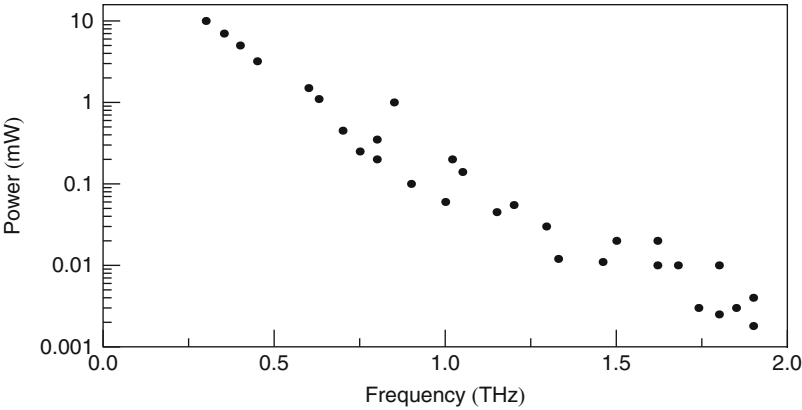


Fig. 4.24 Output power of multiplier-based harmonic generators. The data are compiled from the literature and include different fundamental sources, as well as different multiplication schemes. All data are for 296 K operation. Reducing the operating temperature to 120 K can improve the output power by a factor up to five

decreased capacitance. An applied THz field modulates the electron distribution and harmonics are generated. Due to its symmetrical design, the HBV has an antisymmetric I-V curve and a symmetric capacitance–voltage curve, allowing only odd harmonics to be generated. In contrast to a Schottky varactor, there is no idler that needs to be terminated. Quintuplers up to 0.5 THz output frequency have been fabricated [368].

Yet, another approach is the superlattice (SL) multiplier. It is a periodic structure, which consists of alternating thin layers of an n-doped semiconductor (GaAs) and an insulator (AlAs). This results in the formation of minibands separated

by a mini-band energy gap. Applying a voltage to the SL leads, at a particular value, to a reflection of electrons at the miniband edges. This can be seen in the I-V curve as a negative differential conductance, which is suitable for frequency multiplication. Two mechanisms contribute to the multiplication process. Below about 1 THz the intra-miniband relaxation time ($\sim 10^{-13}$ s) is short enough and the THz field modulates the group velocity of miniband electrons [369]. At higher frequencies, harmonics are generated by Bloch oscillations of the electrons [370]. Since the I-V curve is antisymmetric, only odd number harmonics are generated if no bias voltage is applied. If a bias voltage is applied, the symmetry of the I-V curve is broken and odd and even harmonics are generated. SL multipliers have been successfully employed in high-resolution spectroscopy [371]. Harmonics up to the 11th order at 2.7 THz were generated with a BWO as input source. The output power was $0.23 \mu\text{W}$ at 0.8 THz with about 50 mW input power. This is less than with a Schottky multiplier-based source. However, a unique advantage of the SL is that it is the only multiplying element generating all harmonics at the same time.

The multipliers have in common that they are mounted in a structure with two waveguides. One waveguide is for coupling the input frequency to the multiplier and another one for outcoupling of the harmonic signal. Care has to be taken to prevent leakage of input radiation into the output waveguide. Multiplier-based sources are now commercially available, with a performance which meets the requirements of many applications in spectroscopy and imaging.

4.9 Backward Wave Oscillator

An alternative to extending the frequency of microwave sources to the THz range by harmonic generation, as discussed in the previous section, is to design fundamental oscillators. The most successful source of this type is the backward wave oscillator (BWO). The first BWOs were invented independently in France [55] and in the USA [56] in the 1950s. The French inventors called their device a *carcinotron*, a name still quite widely used. A plausible explanation of the name *carcinotron* is that for security reasons the nature of the operating principle was not to be apparent from the nomenclature. “Carcinos” is French for the crayfish whose usual movement is a slow walk but, when startled, uses rapid flips of its tail to swim backwards.

To understand the operation of a BWO, it is convenient first to consider how the energy of an electron beam can be used to provide gain for a high-frequency RF signal. This can be perceived by first studying the mechanism of the “traveling wave amplifier,” shown schematically in Fig. 4.25a.

Consider an RF signal injected at the source end. The wave associated with this signal has its phase velocity reduced by the “slow-wave” structure to approximately the drift velocity of the electron beam. With a uniform periodic structure, the phase velocity will be constant and at any point in time the axial electric field distribution due to this will vary sinusoidally with distance, as shown in Fig. 4.25b. Electrons entering the slow-wave structure are exposed to an electric field. When the field is

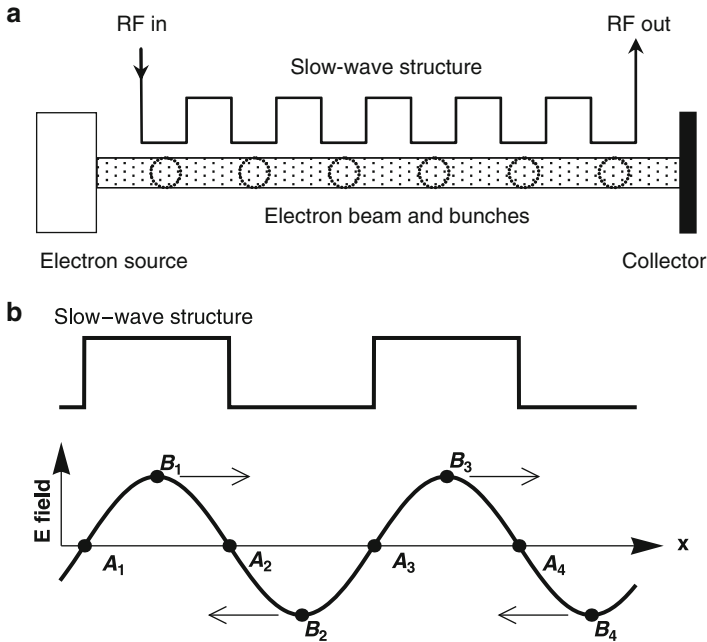


Fig. 4.25 (a) Schematic design of a traveling wave amplifier. (b) Electron bunching in a nearly synchronous traveling wave field. The process is discussed in the text

zero at points A_1 , A_2 , A_3 , A_4 , etc., the electrons will drift with the wave and their phase, relative to the wave, will be constant. Electrons at points like B_1 and B_3 will be accelerated and drift toward the constant phase electrons at A_2 and A_4 . Electrons which are at B_2 and B_4 will be decelerated toward the constant phase electrons at A_2 and A_4 , etc. This causes the electrons to move into bunches around the phase stable electrons at A_2 , A_4 , etc. When there is exact synchronism between the electrons and waves, it is clear that no energy will be extracted from the beam. However, if the electron drift velocity is slightly faster than the wave, the centers of gravity of the bunches will be in the decelerating regions between A_2 and B_2 , A_4 and B_4 , etc. In this manner, energy is transferred to the wave, which grows exponentially in amplitude with the distance traveled by the electron beam. If, for any reason, there is feedback in the system, the traveling wave amplifier will oscillate, and it may be necessary to include some loss mechanism into the slow-wave structure to avoid this.

Essentially, the mechanism of a BWO closely resembles the traveling wave amplifier, except that the slow-wave structure is deliberately designed to provide feedback. The design of a high-frequency BWO is shown schematically in Fig. 4.26 [372]. Fundamentally, a BWO resembles a very sophisticated high vacuum diode. The cathode, which is warmed by a low voltage heater, emits electrons that are accelerated by a high voltage field and travel toward the anode. The electrons

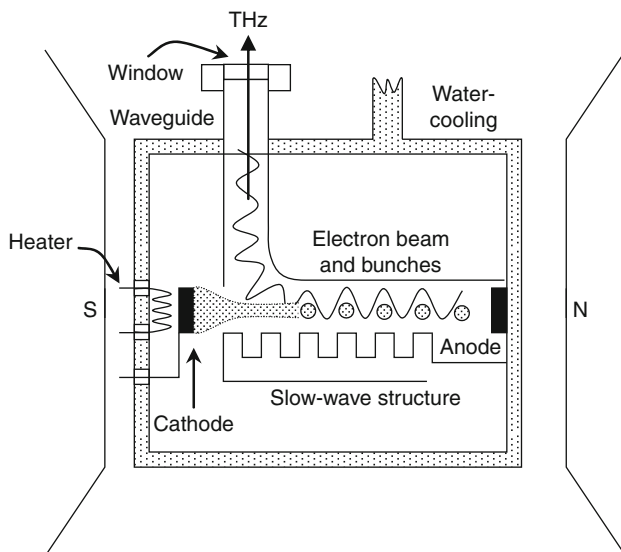


Fig. 4.26 Design of a high-frequency BWO (adapted from [372])

are collimated by a very uniform external magnetic field and pass over a comb-like slow-wave structure to produce the required bunching for the transfer of the kinetic energy of the electrons to an electromagnetic wave that builds up from noise fluctuations. The slow-wave structure is designed in such a way that, when the gain is sufficient for oscillation, the generated wave is 180° out of phase with the electron bunches. Hence, this type of source was given the name “Backward Wave Oscillator” by its American inventors.

Although the basic design of the BWO seems straightforward, they are highly sophisticated devices with stringent tolerances, particularly for high-frequency operation. This can be understood by considering the requirements to accomplish a net transfer of energy from the electrons to the electromagnetic wave. In order to achieve the necessary phasing, the spatial periodicity of the slow-wave structure must be on the scale of the distance traveled by the electron bunch in one oscillation of the electromagnetic wave. For an accelerating voltage of a few thousand volts and a frequency of 0.3 THz, this distance is $\sim 100\ \mu\text{m}$. With modern fabrication techniques, these dimensions are not difficult to achieve but, as the proportions are reduced for higher frequencies, severe complications arise. Apart from the fact that the “comb” spacing is now close to $10\ \mu\text{m}$, and all the other dimensions are also reduced, there are problems with heat dissipation from the cathode, which has to be close to 1200°C to achieve the high current densities required. Furthermore, circuit losses increase rapidly with increasing frequency. Apart from the difficulties of manufacture, an inevitable result of the reduced size of the high-frequency BWOs is lower power output and this can be clearly seen in Fig. 4.27.

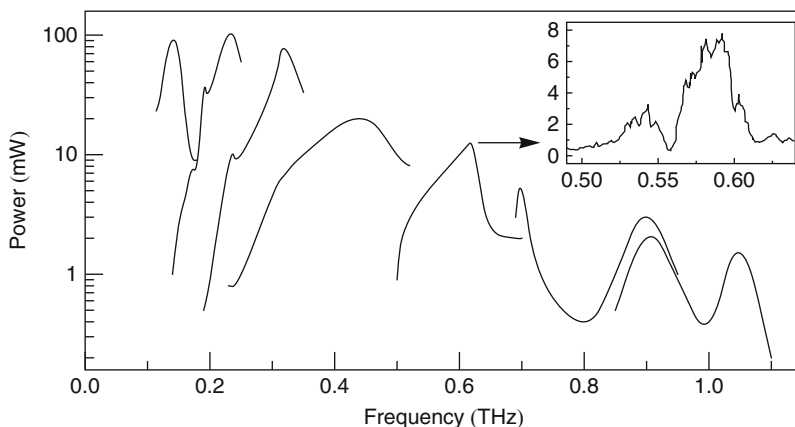


Fig. 4.27 Power versus frequency for a range of BWOs. Each curve in the main figure represents the output of an individual BWO. The inset shows the detailed output of one BWO. The apparent noise fluctuations in these outputs is in fact a reproducible signal level that is unique for each BWO (adapted from [372])

A major advantage of BWOs is their tunability. The output frequency is approximately proportional to the square root of the voltage across the tube, and the tuning rate is ~ 10 MHz/V for low-frequency devices, rising to ~ 100 MHz/V for those operating at 1 THz and above. As can be seen from Fig. 4.27, the output power of each BWO varies quite rapidly with frequency and in practice the useful tuning range is approximately $\pm 10\%$ from the center frequency. The maximum frequency achieved is ~ 1.4 THz. Because of the many constraints involved, it seems unlikely that this value will be significantly increased. Although the output power appears “noisy” this apparently random structure, which is unique for each BWO, is highly reproducible.

The linewidth of a BWO is critically dependent on the applied voltage and, with the most stable power supplies available, is typically 10–20 MHz. However, the intrinsic linewidth is less than 1 kHz and frequency stabilization can be achieved by using phase-locking. Figure 4.28 shows a block diagram of a system for achieving a bandwidth of 10 kHz at 0.633 THz, giving a potential resolving power of more than 6×10^7 [373]. The arrangement uses two phase-lock loops (PLL). A small fraction of the BWO beam is coupled out and fed into a mixer. The signal is mixed with the 7th harmonic of an 80–90 GHz Gunn oscillator, which in turn is stabilized by a second PLL using the 13th harmonic of a tunable 5–7 GHz frequency synthesizer as a reference.

The output of this frequency stabilized BWO provides a powerful alternative to solid-state sources for producing tunable radiation at very high frequencies by harmonic generation. This particular system has been designed to drive a frequency tripling Schottky barrier diode. The input power to the diode is 7 mW and the output at 1.9 THz is more than $1.5 \mu\text{W}$.

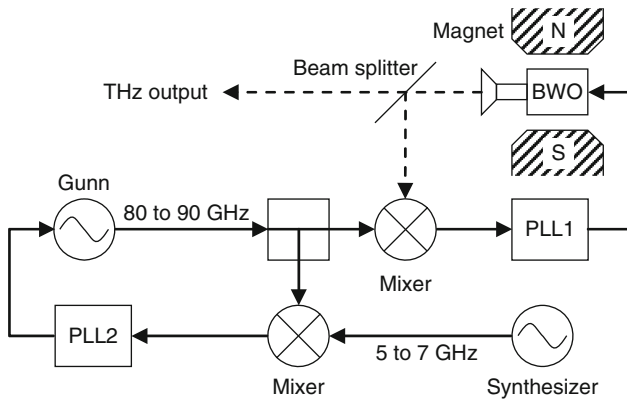


Fig. 4.28 Stabilized BWO operating at 0.633 THz with two phase-locked loops PLL1 and PLL2 (adapted from [373])

The necessity for a very precise magnetic field has the result that BWOs fall into two types. Low-frequency devices for operation below ~ 180 GHz are packaged with a small samarium-cobalt magnet. The construction can then be very compact, weighing about 1 kg, and requiring only air cooling. For higher frequencies, the magnetic field required increases from 0.6 to 1.2 T and either massive permanent magnets, or sometimes electromagnets, are needed, with very precise mechanical adjustment systems. This can only be achieved by separating the magnet from the oscillator, which now requires water-cooling.

4.10 Smith–Purcell Sources

E.M. Purcell had conjectured that if an electron passed close to the surface of a metal diffraction grating, moving at right angles to the rulings, the periodic motion of the charge induced on the surface of the grating should give rise to radiation. In 1953, in conjunction with S.J. Smith, he confirmed this effect in the visible region, using 300–350 kV electrons from a Van der Graaff generator and an electron accelerator tube [374]. They showed that the wavelength λ emitted was given by

$$\lambda = \frac{g}{n} \left(\frac{1}{\beta} - \sin(\theta) \right), \quad (4.11)$$

where g is the grating period, $\beta = v/c$ is the electron velocity relative to the speed of light, θ is the angle of emission measured from the normal to the surface of the grating, and n is the spectral order. (4.11) can be derived by the simple Huygens construction illustrated in Fig. 4.29.

Consider two points, separated by one period of the grating, that emit light simultaneously in the direction θ . The light interferes constructively if the difference

Fig. 4.29 Huygens construction for Smith–Purcell radiation with an electron traveling close to the grating with a grating constant g . THz emission is emitted at an angle θ measured from the normal relative to the grating surface

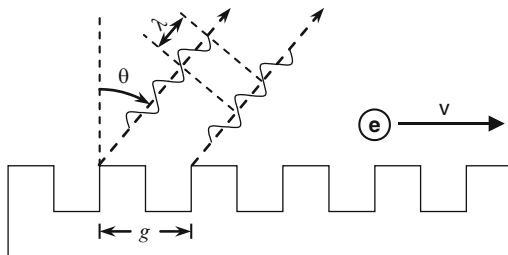
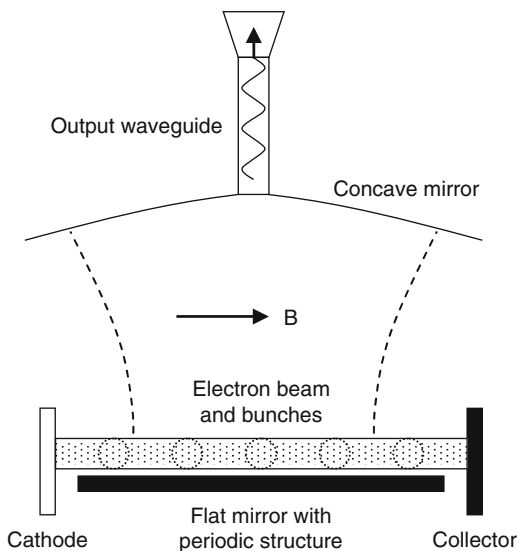


Fig. 4.30 Design of a mmW orotron in an external magnetic field with induction B



in path length to a distant observation point is an integral multiple of the wavelength: $g \sin(\theta) = n\lambda$. However, the light is not emitted simultaneously but at times separated by the transit time of the electron g/v . Effectively, this introduces an additional path length of g/β , resulting in (4.11).

The use of Smith–Purcell (S-P) radiation to provide mmW and THz radiation was demonstrated in the USSR in 1969 [375] and independently in Japan in the same year [376]. In the USSR experiment, where the device was called an orotron, a ribbon-like beam, produced by a diode gun, was passed across a ruled grating. Feedback of radiation, produced perpendicular to the surface, was achieved by placing a concave mirror above the grating, which provided a Fabry–Pérot cavity. Power was coupled out through a waveguide set in the center of the mirror, as shown in Fig. 4.30. To focus the electron beam, flat pole pieces of an electromagnet were placed alongside the beam. The electron beam was pulsed with a pulse length of $1\ \mu\text{s}$ and, using different period gratings, power varied between 200 W at 25 GHz and several hundred mW at 350 GHz. As can be seen, the orotron has a close similarity to a BWO, with the essential difference that the resonant cavity is not restricted to dimensions corresponding to the source wavelength. Successful development of

cw orotrons at longer wavelengths was achieved at lower frequencies [377], but in Russia there has been a significant drive into the THz region [378]. This has led to the commercial availability of long pulse length orotrons at frequencies up to 0.37 THz. For example, devices with output powers of 100 mW with pulse lengths approaching 10 ms at 0.3 THz can now be obtained.

Research at Dartmouth College, USA, headed by J.E. Walsh, has led to a somewhat different type of S-P source. Initially, this concentrated on high-power pulsed systems employing relativistic electrons with outputs in the THz region. However, the availability of a scanning electron microscope (SEM) suggested the possibility of a cw source. Although the total current of the SEM beam is low, the beam is of high quality and can be focused and steered with great accuracy over the surface of an S-P grating. The design of Walsh's experiment is shown in Fig. 4.31.

Initially, this design produced spontaneous radiation, in agreement with (4.11), over the wavelength range $350\text{--}900\text{ }\mu\text{m}$ (0.33–0.86 THz) by using a number of grating periods and varying the beam energy from 20 to 40 keV. Power emitted was approximately 1 nW. Subsequently, various alterations were made to both the power

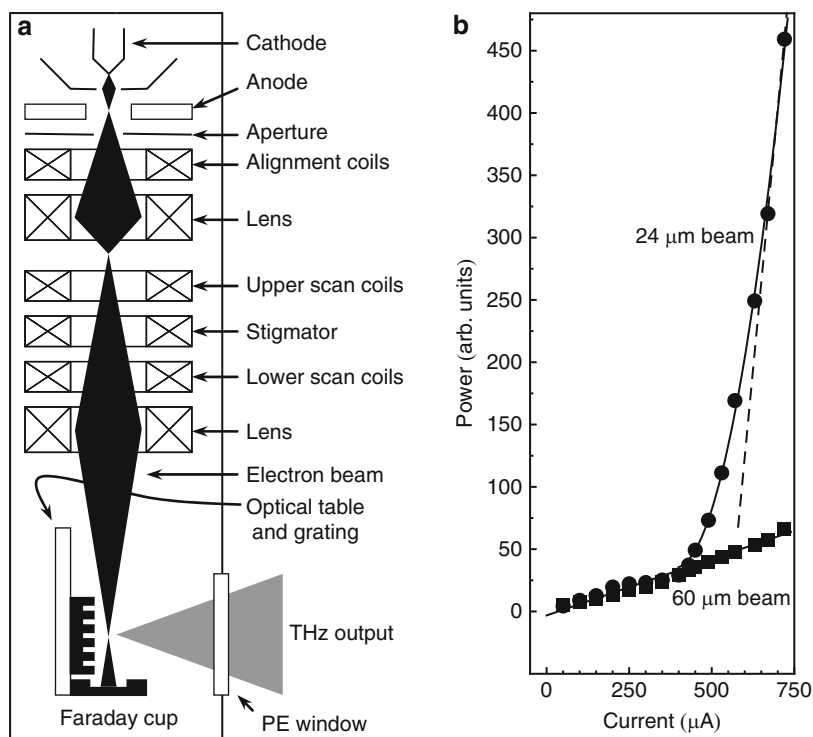


Fig. 4.31 (a) Smith–Purcell THz source based on a scanning electron microscope. (b) Detected power versus beam current for two beam diameters at 35 keV electron energy. The sharp transition to superradiant emission at the higher current density is clearly shown (adapted from [379])

supply and the alignment and focusing coils of the SEM to provide a very narrow high-current density beam as it passed over the grating. With the modified design, the current density was sufficient for a dramatic change to occur in the output.

As the current was increased, the power emitted suddenly changed from a linear to a superlinear dependence as illustrated in Fig. 4.31b. There have been a number of suggestions as to how the required feedback is produced as there are no mirrors to provide external feedback. The most convincing explanation of what occurs has been given by Andrews et al. [380], who pointed out that, in addition to the S-P radiation, there are evanescent modes that travel along the grating and decay exponentially above it. When the phase velocity of the evanescent wave is synchronous with the electron velocity, the electron beam amplifies the evanescent wave. The group velocity of the evanescent wave can either be positive, as in a traveling wave tube, or negative, as in a BWO, depending on the dispersive properties of the grating and the electron velocity. External feedback is not required. The frequency of the evanescent wave is below that of the S-P radiation, so it does not radiate except at the ends of the grating. However, it is the bunching of the electrons by their interaction with the evanescent wave that enhances the S-P radiation and provides the much-increased output power. In confirmation of this explanation, the evanescent wave at the predicted lower frequency was later reported [381].

Experiments with a purpose-built system that greatly improved the attributes of the SEM has resulted in increased output power that has reached the microwatt level at frequencies close to 1 THz [382]. Experiments are continuing on a variety of structures in several laboratories in the hope of higher power levels. A possible way forward is the suggestion from a Russian group of replacing the pencil-like beam used in the Dartmouth College and subsequent experiments, by a sheet beam [383]. This could permit a significant increase in the operating current and in the output power. The same group has also demonstrated S-P radiation enhancement by the use of a two-mirror open cavity similar to that of an orotron, and the initial results look very promising [384].

A very useful application of S-P radiation is as a diagnostic tool for the study of high-energy electron beam bunches, typically 1–10 ps long with a bunch spacing of a few ns. If a blazed diffraction grating (Sect. 6.2) is placed adjacent to the electron beam, SP signals are emitted over a wide angular spread. By studying this angular spread and the power emitted, both the length and shape of the electron bunch can be obtained. This method for the study of electron beams has proved to be very versatile and has been employed over beam energies from 2 MeV up to 28.5 GeV [114].

4.11 Gyrotrons

Gyrotrons are electronic tubes that are normally employed to provide very high power in the mmW region of the spectrum. However, a few THz gyrotrons have been constructed and the outline of one designed at the Massachusetts Institute of Technology (MIT) is illustrated in Fig. 4.32. In essence, it is a simple construction requiring only an electron source and a resonant cavity contained within a strong

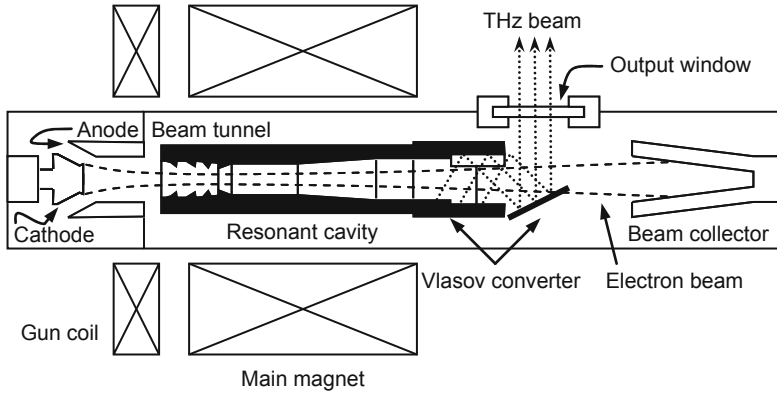


Fig. 4.32 Design of the THz gyrotron at MIT discussed in the text (adapted from [385])

magnetic field. Electrons emitted from the cathode acquire transverse and axial velocity due to the electric and magnetic fields in the electron source region.

Electrons gyrating in an external longitudinal magnetic field do so at an angular frequency given by

$$\omega = \frac{eB}{m}, \quad (4.12)$$

where B is the magnetic induction and e/m is the ratio of charge to mass of an electron. At low electron velocities, where m_0 is the rest mass of an electron, (4.12) becomes in terms of frequency

$$v = \frac{eB}{2\pi m_0}. \quad (4.13)$$

However, when the electron velocity v becomes significant, its mass increases by the relativistic factor (4.15). Although the accelerating voltages in gyrotrons are comparatively low, typically 10–100 kV, this increase in electron mass is sufficient to produce instabilities. As the electrons travel toward the resonant cavity into an increasing magnetic field, the electron orbital momentum increases. The relativistic electrons are initially randomly distributed in their cyclotron orbits but phase bunching occurs due to the mass change of the electrons. This can be understood by realizing that electrons which lose energy to the electromagnetic field become lighter and build up a phase lead while those that gain energy have a phase lag. Both groups of electrons tend toward an equilibrium phase and with sufficient bunching of the electrons they radiate coherently. When the wave frequency is slightly greater than the cyclotron frequency, a large percentage of the electron beam energy is converted to electromagnetic radiation.

There is an analogy between the gain process in the gyrotron and the traveling wave amplifier discussed in Sect. 4.9, but here it is the change of mass of the

electrons that leads to instabilities. These produce the self-oscillation that then transforms the kinetic energy of the electrons into radiation.

The great advantage of gyrotrons compared with other high-frequency electronic tube sources is that they emit radiation at a frequency determined by the strength of the applied magnetic field and not the dimensions of a resonant structure. At microwave frequencies, there is a further advantage that their efficiency can be very high, approaching 40%. Their main disadvantage as THz sources is explained by reference to (4.13). This yields a cyclotron frequency of 27.992 GHz per Tesla. For a device operating at 0.3 THz, a field of more than 10 T is needed. Fortunately, it is possible to obtain useful powers and efficiency at harmonics of the fundamental frequency and as early as 1974 a cw power of 0.5 kW was obtained at 0.33 THz from a gyrotron operating on the 2nd harmonic [386]. Superconducting solenoids were used to provide the required high magnetic field. Much of the early research on gyrotrons was performed in the USSR in a group led by A.V. Gaponov who, in the late 1950s, was one of several authors to discuss independently the possibility of emission from monoenergetic electrons in a magnetic field [62]. Interestingly, the first person to recognize the potential of radiation from electrons rotating in a uniform magnetic field was an astrophysicist, R.Q. Twiss [61]. He used a quantum mechanical approach, as did J. Schneider, who ended his paper with “It does not appear unlikely that this effect could be used for a new type of maser . . .” [387]. R.H. Pantell [63] and A.V. Gaponov used a classical approach. It took many years of research and development to reach the very powerful and efficient sources of mmW radiation that are now used in controlled fusion research and other investigations.

Progress toward higher frequencies has been comparatively slow. In recent years, a main research center for the development of THz gyrotrons has been at Fukui University in Japan, following on from a joint research program with the University of Sydney, Australia. The Fukui team, led by T. Idehara, has developed a number of high-frequency gyrotrons reaching to above 1 THz with pulsed magnetic fields [388]. This laboratory is planning a 1 THz device using a 20 T superconducting magnet with a cw output power of 250 W on the 2nd harmonic [389]. The Institute of Applied Physics of the Russian Academy of Sciences in Nizhny Novgorod has a long and distinguished history of gyrotron development dating back to the early 1970s [386], including research on THz devices [390]. They are collaborating with the Fukui University group on the development of relatively small gyrotrons to produce cw output powers of 1,000 to 1 W at frequencies from 0.3 to 1.5 THz [383].

As mentioned earlier, another group that has had significant success with high-frequency gyrotrons is at MIT led by R.J. Temkin [385]. One requirement was for a high-power THz source for dynamic nuclear polarization enhanced nuclear magnetic resonance for studying biological samples. Their gyrotron (shown in Fig. 4.32) provides a cw output of more than 8 W at 0.46 THz, when operating on its 2nd harmonic, in a nearly Gaussian radiation pattern. This high quality beam is achieved by the use of a mode converter of the type first described by S.N. Vlasov and his colleagues [391]. These so-called Vlasov launchers, or converters, are commonly used, either internally or externally, to convert the high-order circular waveguide modes into a low-order waveguide mode, or a Gaussian mode in free

space. In the MIT design (Fig. 4.32), where the axis of the gyrotron tube and the electron beam are within the bore of a 9.2 T superconducting magnet, the mmW power is extracted via the mode converter, which consists of a Vlasov launcher and two mirrors. This is in a room-temperature section which intersects at a right angle to the magnet bore and provides efficient separation of the microwave beam from the electron beam. An in-depth analysis of this mode converter has been made [392]. BWOs and gyrotrons can be considered as complementary sources. The former cover the THz region to above 1 THz but with low power at the high-frequency limit. The latter give very high power at low frequencies and even at their present highest frequency of ~ 0.5 THz for cw devices the output power is 1,000 times more than that of a BWO.

There are two distinct designs for gyrotrons. Most of the structures use a high-Q cavity and operate on a single frequency with linewidths that can be less than 100 kHz. Despite this high-Q a limited tuning range of 2 GHz at 0.46 THz has been obtained by small variations of the magnetic field, producing switching between overlapping higher axial modes [393]. Alternatively, using low-Q configurations gyrotrons can be step-tuned over a wide frequency range to produce relatively low power. However, this arrangement means that the output is switching between a wide range of higher order modes and very sophisticated mode converters are required to simulate an approximation of a Gaussian beam. An example of the wide tuning range that can be obtained with tunable systems was GYROTRON V at the University of Sydney [394]. This could be step-tuned from 150 to 300 GHz at its fundamental and its first harmonic (0.3 to 0.6 THz), with power outputs of up to 20 W and hundreds of mW, respectively.

4.12 Terahertz Sources Based on Relativistic Electrons

Synchrotron radiation is generated by the acceleration of charged particles, which move close to the speed of light. The emission spectrum contains frequencies from mmW up to X-rays. Its name was derived from a General Electric synchrotron where it was discovered in 1946/1947 [395]. The generation mechanism of synchrotron radiation is illustrated in Fig. 4.33.

For simplicity, only one electron is considered. The electron moves in a bend with a speed close to that of light and emits electromagnetic radiation due to the acceleration. An observer outside the inertial system of the electron sees an electromagnetic pulse with duration in the order of femtoseconds to picoseconds. The spectral content of the pulse is obtained by a Fourier-transform of the modulus squared of the amplitude of the pulse. This results in a broadband spectrum. Towards short wavelengths the power drops as $\exp(\lambda)/\lambda$ and at the long wavelength side of the spectrum the power is proportional to λ^{-3} . The total power radiated by a single electron, which is accelerated vertically to its direction of motion, is given by the relativistic version of Larmor's formula:

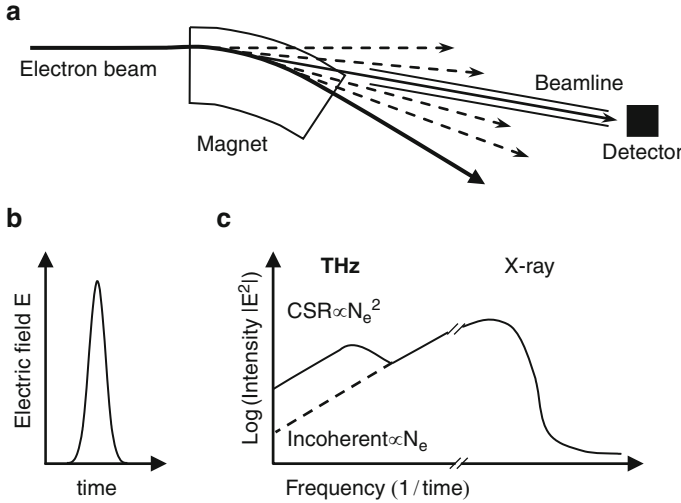


Fig. 4.33 (a) Electron beam traveling through a magnet emitting synchrotron radiation collected at the beam line detector. (b) An observer of an electron traveling in an arc “sees” an electric field pulse. (c) Power spectrum given by the Fourier-transform of the electric field pulse (adapted from [396])

$$P = \frac{e^2 a^2}{6\pi \epsilon_0 c^3} \gamma^4, \quad (4.14)$$

where a is the acceleration and γ is the ratio of the mass m of the electron to its rest mass m_0 ($m_0 c^2 = 0.51 \text{ MeV}$):

$$\gamma = \frac{m}{m_0} = \frac{1}{\sqrt{1 - (v/c)^2}}. \quad (4.15)$$

As can be seen, the radiated power increases dramatically as the electron velocity v approaches the speed of light. It is this effect which makes synchrotron radiation much more powerful than non relativistic sources based on accelerated electrons. Due to the relativistic effects, synchrotron radiation is not emitted into $4\pi \text{ sr}$, but rather into a narrow cone with a frequency-dependent opening angle of a few tens of milliradians. Because of its high intensity, high collimation, and broad emission spectrum, synchrotron radiation is a powerful tool in fundamental as well as applied research. In the following discussion, THz sources that are based on a particular type of synchrotron radiation will be considered, namely coherent synchrotron radiation (CSR) (for a review, see [396]). In contrast to FELs, which are also based on synchrotron emission, these sources emit across a broad, i.e. larger than 1 THz, frequency range.

In real machines, electrons move in bunches rather than alone. CSR occurs when the electrons in a bunch emit synchrotron radiation in phase. This happens at

wavelengths equal to, or longer than, the longitudinal bunch length (in the direction of the electron motion). In this particular case, the electrons emit coherently and the intensity is now proportional to the square of the number of electrons in the bunch, rather than to the number of electrons, as in incoherent synchrotron radiation. Considering the large number of electrons in a bunch (typically 10^{10}), the intensity gain for CSR is huge. For a given wavelength λ , the power P emitted by a bunch of electrons can be derived from the power P_e emitted by a single electron according to [397]:

$$\frac{dP}{d\lambda} = (N_e + N_e(N_e - 1)f_\lambda) \frac{dP_e}{d\lambda}, \quad (4.16)$$

where N_e is the number of electrons in the electron bunch. f_λ is the so-called form factor, which describes the degree of coherence. This is one for a point-like source with full coherence, and zero in the incoherent case of a long bunch. The first term in the bracket (4.16), which is proportional to N_e , describes the incoherent synchrotron radiation while the second term, which is proportional to $N_e^2 f_\lambda$ for $N_e \gg 1$, represents CSR. For $N_e f_\lambda \gg 1$ CSR dominates in the total emission. The form factor is given by the modulus squared of the Fourier-transform of the longitudinal charge density $n(z)$ of the bunch:

$$f_\lambda = \left| \int_{-\infty}^{\infty} n(z) \exp\left(2\pi i \frac{z}{\lambda}\right) dz \right|^2. \quad (4.17)$$

In the case of a Gaussian electron density distribution in a bunch with an rms length σ_{rms} , the form factor is given by $f_\lambda = \exp[-(2\pi\sigma_{\text{rms}}/\lambda)^2]$. This relation states that CSR is most easily obtained from short bunches and at long wavelengths. The dependence of the radiated power on the wavelength takes into account other effects, such as shielding by the vacuum chamber in which the electron is moving. CSR was experimentally discovered using electron bunches in a linear particle accelerator (linac) [398]. In 2002, a team at the Berliner Elektronen Speicherring–Gesellschaft für Synchrotronstrahlung (BESSY) succeeded in generating powerful and stable CSR with an electron storage ring [103], and in the same year high-power THz radiation from a linac was demonstrated [399].

4.12.1 Coherent Synchrotron Radiation from Electron Storage Rings

In an electron storage ring, electrons circulate close to the speed of light. The ring consists of a vacuum chamber and bending magnets, which keep the electrons moving in a circle. Undulators can also be incorporated (Sect. 4.12.3). The electrons are grouped in bunches, with a length in the range from ~ 0.1 mm to several millimeters. Many bunches separated by a few centimeters can circulate in a ring. For example, the BESSY II electron storage ring has a 240-m circumference. Up to 400 electron bunches, each with up to $\sim 10^{10}$ electrons, can be stored. The revolution

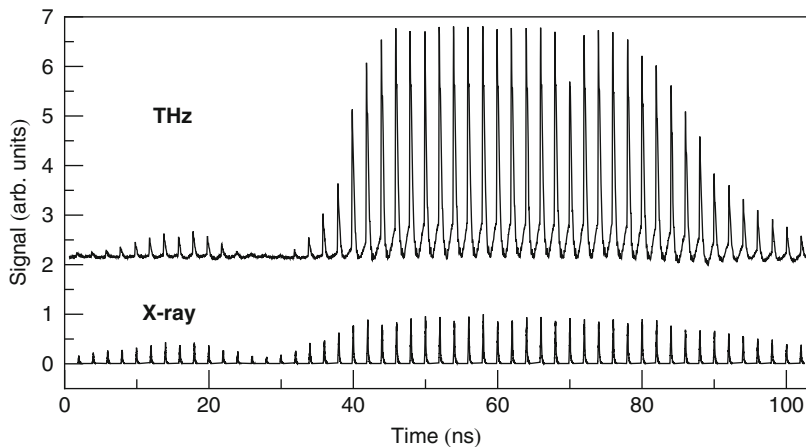


Fig. 4.34 Fill pattern (rms bunch length of 5 ps) measured with a superconducting hot electron bolometer at THz frequencies (raised by two units) and with an APD at X-ray energies for a total current in the ring of 5.8 mA and a maximum current in a bunch of $\sim 200 \mu\text{A}$ (adapted from [400])

frequency in the ring is 1.25 MHz and the separation between two adjacent bunches is 2 ns (60 cm) (Fig. 4.34).

Although CSR was predicted to occur in high energy storage rings in 1954 [401], stable and steady-state CSR was measured in 2002 as mentioned above [103]. Stable, steady-state CSR, as opposed to bursting CSR (see below), means that the CSR power emitted from a particular bunch does not fluctuate between different revolutions in the storage ring. Stable CSR can be achieved by an appropriate setting of the magnetic fields interacting with the electrons. This allows lowering the so-called momentum compaction factor α , which describes the relative change of orbit length from the nominal length for electrons whose momentum deviates from the average momentum. The bunch length is proportional to $\sqrt{\alpha}$. For example, at BESSY II α can be reduced to $\sim 10^{-6}$, corresponding to an rms bunch length of 700 fs ($\sigma_{\text{rms}} \approx 0.2 \text{ mm}$) at a bunch current of $\sim 0.3 \mu\text{A}$. A power increase of more than 10^5 times compared to incoherent synchrotron radiation has been obtained [402]. Increasing the bunch current beyond a certain threshold leads to self-induced fluctuations of the electron density in the bunch. The resulting microbunches have a characteristic length shorter than the bunch length and a much larger form factor than a Gaussian-shaped bunch. As a consequence, microbunches emit not only significantly more powerful CSR but also higher frequencies. However, the microbunches are not stable and the CSR emission has a bursting character with power fluctuations up to 100%, if a single bunch is considered. The fluctuations might be reduced by averaging, but for many applications stable CSR is preferred. The transition from stable CSR to bursting CSR is illustrated in Fig. 4.35. It displays the signal from a fill pattern of the storage ring consisting of electron bunches with an rms bunch length of 5 ps and different electron population. An avalanche photodiode (APD) detected $\sim 8 \text{ keV}$ X-rays, and the THz signal, from the same

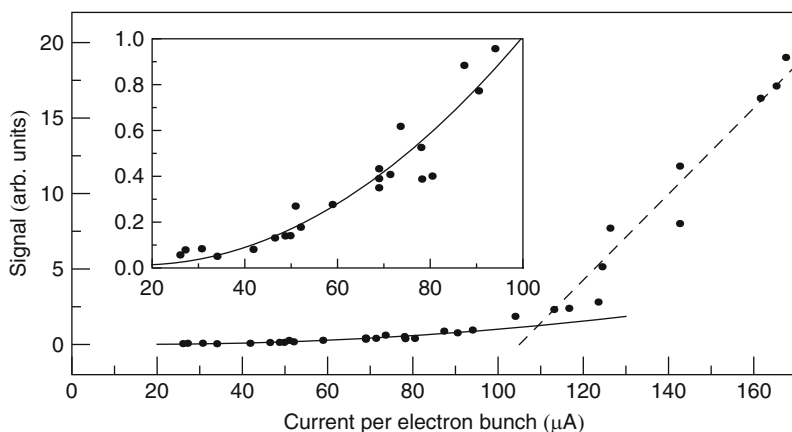


Fig. 4.35 THz signal as a function of the bunch current. The transition from stable to bursting CSR occurs at about $90 \mu\text{A}/\text{bunch}$. The inset is an expanded view of the quadratic power increase expected for stable CSR (adapted from [400])

fill pattern, has been measured with a superconducting hot electron bolometer (Sect. 5.5.4). Below $\sim 100 \mu\text{A}$, CSR was generated by stable electron bunches and the power increased as the square of the current. Above this threshold, bunches became unstable and microbunching occurred as shown in Fig. 4.35 [400].

In Fig. 4.36, typical CSR spectra above and below the bursting threshold are shown. The bunch length was 3 ps. In this case, the bursting threshold was at about $10 \mu\text{A}$ per bunch. The power of stable CSR peaks at about 0.4 THz, while that of bursting CSR is almost three orders of magnitude more powerful, with a maximum around 0.6 THz. The average total power was several mW for stable CSR. It should be noted that the power at the source point is several orders of magnitude higher. But since the beamlines are usually not optimized for THz radiation, their attenuation is large. The first storage ring optimized for generating CSR is the Metrology Light Source of the Physikalisch-Technische Bundesanstalt in Berlin. Stable or bursting CSR from 0.1 to 1.5 THz up to an average power of about 60 mW can be obtained [403].

The spectral coverage can be extended by so-called “femto-slicing” [404]. This method requires a femtosecond laser pulse which is, for a short distance, co-aligned with an electron bunch. The electric fields of both interact in such a way that the electrons are laterally pushed out of the bunch center. The longitudinal electron density modulation induced by the fs laser can be considered as a hole in the electron bunch with a length of ~ 130 fs. Correspondingly, the CSR, emitted by the density modulation, peaks around 1.5 THz and extends to about 3 THz. The low-frequency cut off is determined by the metal vacuum chamber in which the electron beam is traveling. This acts as a waveguide with a cut off frequency below 100 GHz. Ideally, synchrotron radiation is linearly polarized, with the electric field vector in the plane of the electron beam. The polarization can be modified by insertion devices. For example, if the source point is located in the homogeneous region of the

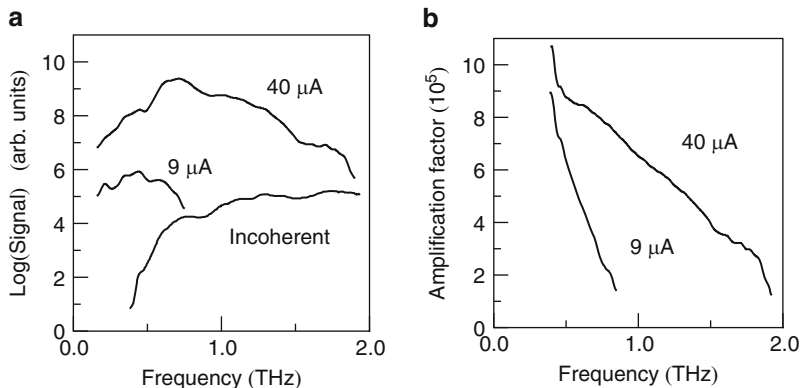


Fig. 4.36 The current dependence of incoherent and coherent synchrotron radiation. **(a)** Coherent output power for bunch currents of 9 and 40 μA , and for incoherent THz synchrotron radiation with a total ring current of 10 mA. **(b)** Amplification factors calculated by dividing the coherent synchrotron radiation spectrum by the incoherent one normalized to the same current per bunch (adapted from [402])

magnetic field of the bending magnet it is linearly polarized, while radiation from the inhomogeneous portion of the magnetic field yields radially polarized light.

The broadband emission spectrum, in conjunction with very stable average output power, high peak power, large brilliance, and a high level of polarization, makes CSR an excellent source for Fourier-transform spectroscopy. First experiments in THz near-field imaging [405], as well as superconductor and semiconductor spectroscopy [406], have been performed, and currently several synchrotron facilities offer THz CSR for user experiments. Another important application for CSR is electron beam diagnostics. This is based on the fact that the CSR spectrum depends on the shape of the electron bunch from which the THz radiation is emitted. Basically, the bunch shape in the time domain can be retrieved from the emitted CSR spectrum by means of Fourier transformation. By applying this method, it has been shown that CSR bunches at BESSY II have a sharp leading front edge, while the tail is less steep [407].

4.12.2 Coherent Synchrotron Radiation from Linear Accelerators

Another way to produce CSR is to use a linac [398, 399], where short electron bunches are created and then accelerated to close to the speed of light. A magnet is installed in a special section of the linac, which bends the otherwise straight path of the electron bunch. The light emitting point is located on the bend. Linacs can create very short bunches (~ 100 fs) resulting in CSR up to a few THz. But the pulse repetition rate is usually much less than 1 kHz. Thus, the average power is quite low compared with that obtained from a storage ring. A way to overcome this is to create

photoelectrons from a GaAs cathode with a high repetition rate (75 MHz) laser. Although the bunch length of the photoelectrons is initially ~ 40 ps, they become tightly bunched to ~ 500 fs in the accelerator. This repetition rate is comparable to that in storage rings. However, to support a similar current and beam energy as that in a storage ring would require huge input power to the linac. This limitation has been overcome by recovering the energy of the electrons prior to discarding them (the so-called energy recovery linac). The CSR spectrum observed with such a system has a maximum at 0.6 THz and extends up to 3 THz. The measured peak power was 0.5 W/cm^{-1} but it was deliberately reduced by a factor of approximately 550 [399]. These developments demonstrate the potential of linacs to generate very powerful broadband coherent THz radiation in the kW range.

4.12.3 *Free-Electron Lasers*

Free-electron lasers (FELs) are high-power sources for a wide range of frequencies and have numerous applications in the THz region. Because they require high energy and high current electron beams for their operation, a number of FEL user facilities have been set up where visiting researchers can pursue their experiments [408].

The concept of FELs dates back to well before the invention of “conventional” lasers, with an article in 1951 by H. Motz entitled “Applications of radiation from fast electron beams” [409]. Experiments performed by Motz and his colleagues produced both coherent emission in the mmW region and incoherent visible light. Later, in the search for higher power than was currently available at mm wavelengths, R.M. Phillips invented the “ubitron” in which the letters UBI stood for “Undulating Beam Interaction” [410]. The ubitron is now generally recognized as the precursor of the FEL. It produced powers of 150 kW at 60 GHz, but the search for high-power electron beam sources for the THz region diminished with the invention of gas lasers, and the rapid progress in new solid-state sources. Interest was revived with the work of John Madey and his colleagues, who coined the term “free-electron laser” and produced stimulated emission at 30 THz [411]. Development of FELs has progressed rapidly and systems now exist over the entire frequency range from soft X-rays to microwaves.

The advantage that FELs have over conventional lasers is that they are widely tunable sources. This is unlike the situation in nearly all solid, liquid, or gas lasers, which are dependent on transitions between specific energy states. The reason why FELs are tunable can be readily understood by the principle of their operation.

An FEL consists of a relativistic electron beam that travels through a periodic magnetic structure called an undulator (or wiggler – the terms are interchangeable), which has either helical or planar geometry. The basic layout for a planar undulator is illustrated in Fig. 4.37. The electrons can be created and accelerated in various ways. The two usual methods of providing electrons are either from a thermionic cathode or by photoemission, produced by a conventional laser, with a photon energy greater than the work function, fired at a metallic surface. The accelerators

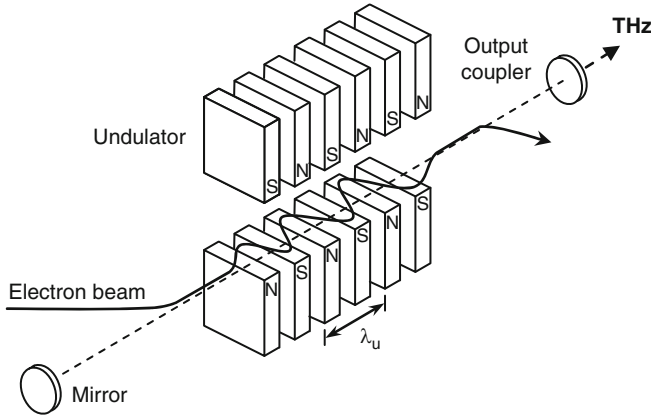


Fig. 4.37 Schematic outline of a free-electron laser

used can be electrostatic but are more often induction radio frequency linacs or microtrons, sometimes in combination with storage rings. This combination of electron source and accelerator produces a high-energy beam, which needs to have a specific energy and a low angular spread, that is injected into one end of the undulator.

The undulator induces transverse oscillations of the electrons, as shown in Fig. 4.37, and this results in incoherent spontaneous radiation as in a synchrotron. The power emitted is proportional to the electron current and has a radiation wavelength given by

$$\lambda_L = \frac{\lambda_u}{2\gamma^2} \left(1 + \frac{K^2}{2} \right), \quad (4.18)$$

where λ_u is the period of the undulator, the distance over which the magnetic field undergoes a complete cycle. γ is the relativistic factor according to (4.15). K is an undulator parameter, which depends on the average magnetic induction B_u , and is given by

$$K = \frac{\omega_u}{ck_u} = \frac{eB_u \lambda_u}{m_0 c 2\pi}. \quad (4.19)$$

FELs are able to operate over a very wide range of wavelengths, and this can be understood by realizing the strong dependence of λ_L on the electron energy. For example, with typical undulator parameters $\lambda_u = 10$ cm and $K = 1.4$, $\lambda_L = 1$ mm, $10 \mu\text{m}$, and 100 nm are obtained for electron energies of 5 MeV, 50 MeV, and 500 MeV, respectively. FELs for the THz region typically require electron energies in the 5 to 30 MeV range.

The gain in an FEL develops from coherent electron bunching. As the relativistic electrons travel through the undulator, they are accelerated from side to side and radiate spontaneously in the forward direction. This is the same process that occurs, for example, in synchrotron radiation sources. The mirrors at either end of the

undulator form a low-loss cavity and ensure that most of this spontaneous radiation is trapped inside the cavity. This stored radiation builds up over many transits as electrons continue to be fed into the undulator and, eventually, the process becomes very similar to the mechanism of BWOs, gyrotrons, etc. Within the undulator, the electromagnetic radiation is traveling with the electron beam, and initially there are many radiation wavelengths inside the cavity. However, when the wavelength of the laser satisfies the resonance condition of (4.18), the electrons interact with the radiation field and form bunches spaced at the laser wavelength. This bunching mechanism can be understood by realizing that both the radiation's electric field vector, and the electrons' undulator velocity vector, rotate through 2π within each radiation wavelength. This determines that half the electrons lose energy to the radiation field and the other half gain energy. Those which gain energy move ahead and those which lose energy move back to produce bunching at the laser wavelength.

Apart from being tunable over a wide range, by varying the electron beam energy and the magnetic induction, FELs can be scaled up to very high powers. In conventional lasers and electron tube sources, most of the input energy required to drive the devices is retained within the laser cavity, causing severe heat dissipation as the power is increased. In FELs only a small fraction of the electrons' energy is converted into electromagnetic radiation: the rest is carried outside with the electrons exiting from the undulator. In some FELs, the emitted electrons are increased in energy and returned to the entrance of the undulator (electron recovery FEL).

The gain in THz FELs is normally quite low. To overcome the unavoidable losses, lasers are long, typically 5–10 m, with high reflectivity mirrors and only 0.5–2.5% of the power is coupled through the output mirror. Inevitably, because of the requirement of an electron accelerator, with its associated shielding, the complete FEL facility is large and expensive. This has led to the construction of a number of FEL user facilities throughout the world. The more important of these with outputs in the THz region are listed in Table 4.2. Included in the table are explanatory notes relevant to each facility. Features relevant to the various types of facility are discussed in the following paragraphs.

The time structure of the pulse depends on that of the electron beam source which, for the THz region, can be a linac, microtron or electrostatic accelerator. Radio frequency (RF) linacs are the most widely used source for THz FELs. These employ a series of cavities that contain rapidly varying electromagnetic fields to accelerate streams of electrons. The beams that they produce are composed of a sequence of macropulses, typically 5–10 μs long, each of which consists of a train of micropulses that are spaced by the RF period of the linac. This is usually in the ns range, with the width of each micropulse being between 5 and 20 ps. The required RF power is tens of MW.

The output of this type of FEL is illustrated in Fig. 4.38. Because of the low gain of most THz FELs, several passes through the cavity are needed before oscillation begins. After 2–3 μs , the signal builds up and this is followed by a constant output region and finally, at the end of the current pulse, the output decays as the cavity empties itself of photons.

Table 4.2 THz FEL user facilities with listed wavelength range λ , micropulse energy E_μ and duration t_μ , macropulse duration t , and average power $\bar{P}$. (1) Accueil CLIO, LCP-Bât. 201P2, Centre Universitaire, Paris-Sud, France – 91405 Orsay Cedex; (2) Forschungszentrum Dresden-Rossendorf, POB 51 01 19, 01314 Dresden, Germany; (3) Osaka University Graduate School of Engineering, Osaka, Zip 573-0128 Yamate Tsuda 2-chome, No. 9, No. 5, Japan; (4) FOM-Institute for Plasma Physics, Rijnhuizen, Postbus 1207, NL-3430 BE Nieuwegein, The Netherlands; (5) Budker Institute of Nuclear Physics, Lavrentyev Prospect 11, 630090, Novosibirsk, Russian Federation; (6) UCSB Center for Terahertz Science and Technology, University of California, Santa Barbara CA 93106, USA

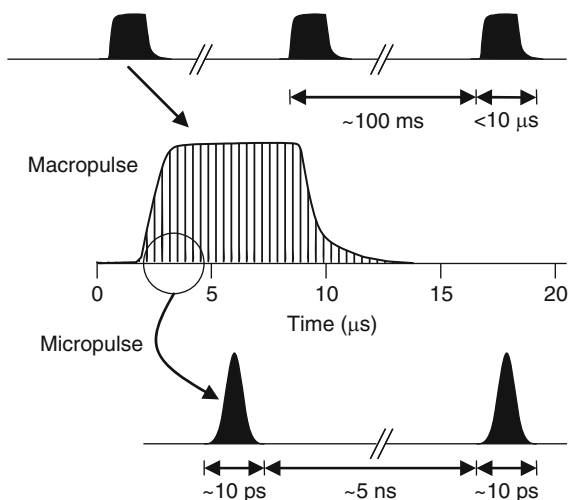
	Country	Name	λ (μm)	Type	E_μ (μJ)	t_μ (ps)	t (μs)	$\bar{P}$ (W)
1	France	CLIO	3–150	linac	10–100	0.5–6	9	1
2	Germany	FELBE	18–280	SC linac	0.01–3	1–25	>100	0.1–40
3	Japan	iFEL	20–100	linac	50	5	20	3
4	The Netherlands	FELIX	4.5–250	linac	1–50	6–100	<10	0.5
5	Russia		120–230	ER linac	50	50	n.a.	400
6	USA	FIR/MM	63–2500	electrostatic	n.a.	n.a.	1–20	0.25

(1) CLIO has a two-part undulator which allows two wavelengths to be produced simultaneously at 10% to 20% of its normal power. (2) FELBE has a superconducting linear accelerator operating in cw mode with a pulse repetition rate of 12 MHz. The electron beam is also available for a second undulator covering the wavelength range 4 to 22 μm . (3) The FEL system at iFEL consists of two RF linacs and five undulators to cover the wavelength range 0.23–100 μm . iFEL 4 is used for 20–60 μm and iFEL 5 for 50–100 μm . (4) FELIX has two conventional accelerator sections. At the exit of the first section, the beam can be passed through the undulator of FEL-1, covering 25–250 μm . Alternatively, the beam passes through both accelerator sections and then through the undulator of FEL-2 to provide 3–40 μm radiation. A third beamline, which can be used with either of the other two, is available for low-absorption experiments within its cavity, where the power is 1–2 orders of magnitude higher than that available externally. This section was called FELICE, free-electron laser for intracavity experiments to distinguish it from FELIX free-electron laser for infrared experiments. The wavelength range of FELICE is from 5 to 40 μm , but this may be extended further into the THz region. It is planned to move FELIX-FELICE from their present home in Rijnhuizen to the High Field Magnet Laboratory at Nijmegen (also in The Netherlands) by the end of 2012. There they will join the FEL FLARE. This is a powerful pulsed source operating between 200 GHz and 3 THz. (5) The Novosibirsk FEL uses a cw energy recovery linac that passes the electron bunch through the undulator at repetition rates up to 22.5 MHz, producing THz output in single pulses 50–100 μs long. (6) The FIR-FEL provides a very uniform 1–20 ms pulse with a power of 1–6 kW, dependent on the wavelength. A second laser (MM-FEL) operates over the range 338–2500 μm , producing 1–6 μs long pulses of up to 15 kW

Microtrons are a useful alternative to linacs, producing a similar temporal structure. They are composed of a single accelerating cavity coupled to a magnet that causes the electron beam to recirculate through the cavity several times. Microtrons are used at ENEA in Frascati, Italy, for very successful THz FELs, which have been used for a number of applications, including biology [412]. Electrostatic accelerators have not been widely used for FELs, but the success of the facility at UCSB shows how well-matched this type of electron beam is to the THz region.

The facilities in Russia and Germany are of particular interest, as they can both be operated in cw mode, making them somewhat analogous to a synchrotron from the user's point of view. FELBE is particularly versatile, as it can also be used to provide a long macropulse, at repetition rates up to 25 Hz or single-pulse operation.

Fig. 4.38 Typical time structure of an FEL consisting of a train of macropulses, which contain several thousand micropulses each with a length of ~ 10 ps



At Novosibirsk a second FEL is being commissioned. The present FEL uses a 12 MeV electron beam but in the new one a four-orbit accelerating structure will raise this to 40 MeV. The new laser is designed to cover the wavelength range from 5 to 240 μm .

Electrostatic generators provide a continuous pulse of typically 5–20 μs . The only facility (Table 4.2) providing this type of beam is at the University of California, Santa Barbara, where there are two FELs covering the range 63–2500 μm at power levels of 1–15 kW with pulse lengths between 20–60 μs .

THz FEL facilities have become important sources for a very wide range of activities in physics, chemistry, biology, and medicine. Because of the pulse structure of RF linac and microtron-driven FELs, these have proved invaluable for fast pump probe-type experiments. An example is the lifetime measurements of excited states in semiconductors performed at FELIX [413].

4.13 Parametric Oscillators

Optical parametric oscillators (OPOs) are room temperature, frequency tunable, and coherent wave emitters in the visible and near-IR region providing several watts of continuous output power, or even higher power for pulsed operation. These OPOs use a three-wave mixing process in a nonlinear material (Sect. 2.5), typically LiNbO_3 , due to its transparency at wavelengths between 0.55 and 4 μm and due to the availability of high-power pump lasers in the near-IR. THz OPOs (TPO) were demonstrated in the 1970s [414–416]. Since the end of the 1990s Kawase and co-workers have demonstrated several tunable TPOs of different design [417, 418]. Figure 4.39 shows a tunable TPO for which the THz signal wave is emitted under

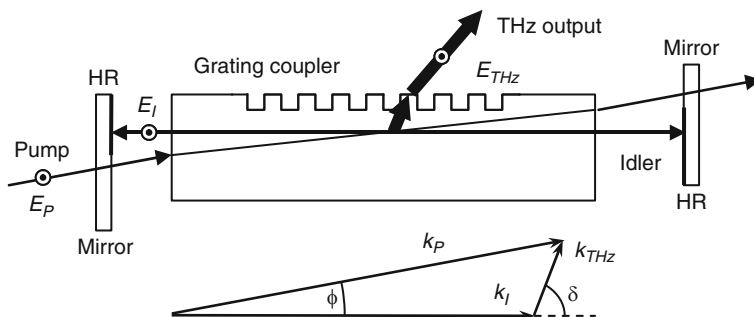


Fig. 4.39 A cavity arrangement for a THz OPO is shown for which the pump wave (E_P) passes only once through the nonlinear crystal. The THz signal wave (E_{THz}) is emitted under an angle δ and coupled out by a monolithic grating while the resonant idler wave (E_I) oscillates between two mirrors with high reflectivity coatings (HR). The phase matching condition for broadly tunable, coherent THz emission is fulfilled by rotating the crystal and resonator as a unit with respect to the pump beam by an angle ϕ , for example, by a motorized rotation stage (adapted from [418])

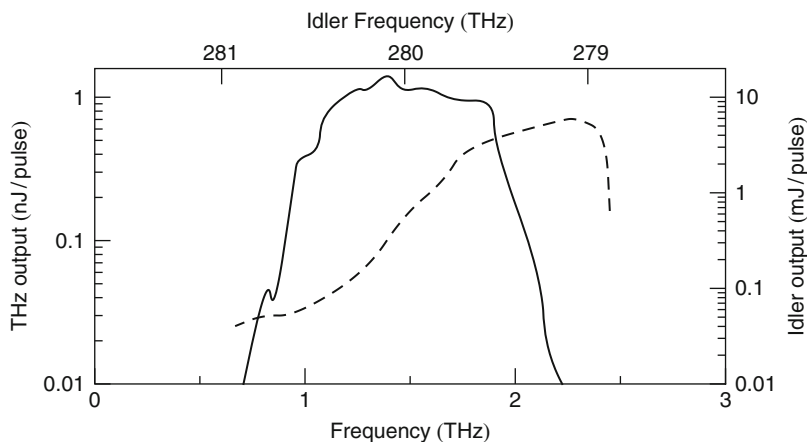


Fig. 4.40 THz output power (solid line) and idler power versus idler frequency (dashed line) (adapted from [418])

an angle and coupled out by a monolithic grating. Frequency tuning is achieved by, mechanical rotation of the laser cavity with respect to the pump beam leading to frequencies from 0.7 to 2.2 THz (Fig. 4.40). A nonlinear LiNbO_3 crystal was used, pumped with a ns Q-switched Nd:YAG laser operated at a wavelength of $\sim 1 \mu\text{m}$. A Si-prism array or a monolithic grating coupler was used to provide efficient outcoupling and to avoid internal reflection of the generated THz wave, which is emitted under an angle with respect to the pump beam directed along the longest crystal axis. Injection seeding with another near-IR laser provided a Fourier-transform limited linewidth of about 100 MHz. Compact THz OPOs have been demonstrated by using a quasi-cw diode laser as the pump laser for a Nd:YAG

crystal. Mounting the crystal in intersecting cavities for pump and parametric wave lowered the threshold pump pulse energies below 1 mJ, leading to a down-conversion efficiency of 50% [419]. A prerequisite for efficient transfer of power from the high-power near-IR laser to the THz wave is “phase matching”, which is described in the following paragraphs.

The pump laser photon is split into two photons with lower energy or expressed in waves: the pump wave (P) separates in the signal wave (S), here the THz wave, and the idler wave (I). The process has to obey two conditions, conservation of energy, $h\nu_P = h\nu_S + h\nu_I$, and momentum or wavevector, respectively: $\mathbf{k}_P = \mathbf{k}_S + \mathbf{k}_I$. For optimal transfer of energy from the pump wave to the THz signal wave, “phase matching” is required, which is a function of the refractive index n and the wave frequency $\omega = 2\pi\nu$. Using $|\mathbf{k}| = \omega n(\omega)/c$ (2.78) the phase mismatch is expressed as:

$$\Delta k = k_P - k_S - k_I = (\omega_P n(\theta, T, \omega_P) - \omega_S n(\theta, T, \omega_S) - \omega_I n(\theta, T, \omega_I))/c \quad (4.20)$$

for a temperature T and as a function of the angle θ , which describes the orientation with respect to the optical axis, and to the slow and fast axis of the nonlinear crystal for extraordinary and ordinary rays, respectively (Fig. 4.41). In typical dispersive media, the phase-matching condition is not easily fulfilled because the refractive index n varies with frequency. Especially the frequency of the pump wave, typically in the visible and near-IR region, is several orders of magnitude higher than THz frequencies. However, in birefringent media the refractive index is also a function of the angle between the wavevector of the propagated light, the crystal axis, and the polarization of the pump light. The tuning range is therefore determined by the possibility of phase-matching by birefringence.

Phase-matching, i. e. $\Delta k = 0$, is the main task to convert energy efficiently from the pump beam into the lower energy waves. If the condition is not fulfilled, then the pump wave and the polarization of the medium have different phases. If the waves propagate further than the coherence length $L_c = \pi/\Delta k$ partial and destructive interference is observed (Fig. 4.41a). The conversion efficiency is given by the factor $\text{Sinc}^2(\Delta k L/2)$ with the crystal length L and the phase mismatch Δk .

For frequency doubling and sum frequency generation phase-matching can be obtained for certain directions in a nonlinear crystal corresponding to the points of intersection of two refractive index ellipsoids of ordinary and extra ordinary rays (Fig. 4.41b). The condition can be met by rotating the crystal with respect to the laser beam, leading to non critical phase-matching if this angle is 0° or 90° , or critical phase-matching for all other angles.

Tunable optical parametric oscillation requires three index of refraction ellipsoids. The points of intersection become curves. In this case, the first and the last points of the curve determine the minimum and maximum angle for which the crystal can be tilted to achieve continuous frequency tuning. The latter is also achieved by heating the crystal, and subsequently changing the refractive index which is a function of temperature. There are some limitations to temperature tuning. The high-power near-IR laser for pumping is focused in the center of the

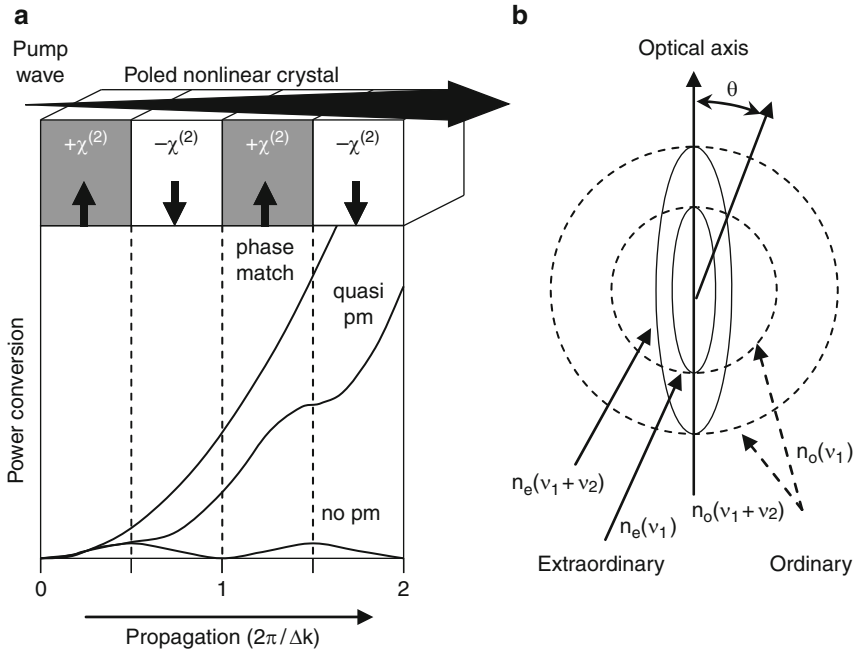


Fig. 4.41 (a) Diagram illustrating the amplification of a wave along the propagation direction in different types of nonlinear crystals if the phase is matched (pm), not matched, or quasi-phase matching is introduced. The crystal shown is a poled crystal which allows quasi-phase matching, if sections adapted in size to the coherence length of the two waves are intentionally polarized in opposite directions to introduce phase jumps at the coherence length. (b) Illustration of the index of refraction ellipsoids for sum-frequency generation in a nonlinear crystal if poling is not used. The nonlinear crystal needs to be rotated by an angle θ with respect to the crystallographic axes so that the waves experience the same refractive index for phase matching at the intersection points. THz optical parametric oscillation requires three index of refraction ellipsoids so that these points of intersection become curves. In this case, the first and the last points of the curve determine the minimum and maximum angle for which the crystal can be tilted for continuous THz frequency tuning

nonlinear crystal propagating in a Gaussian beam. This leads to “thermal lensing” because the laser heats the crystal locally but inhomogeneously, changing the refractive index, and effectively forming a lens within the crystal similar to the Kerr-effect technique used for generating mode-locked fs pulses (Sect. 6.6). The focal length f of such a thermal lens formed by a Gaussian beam with waist w inside the nonlinear crystal of length L is obtained as

$$\frac{1}{f} = \frac{\alpha P}{\pi \kappa_c} \frac{dn}{dT} \int_L \frac{dz}{w^2(z)} \quad (4.21)$$

with the power P of the pump laser, the absorption coefficient α of the nonlinear crystal and its thermal conductivity κ_c , and with the change of the refractive

index dn/dT , which is a function of temperature T . To minimize the effect of thermal lensing, nonlinear crystals are often mounted in a temperature regulated and stabilized small oven, providing heat well above 100°C and 150°C during operation. LiNbO_3 crystals doped with MgO are less susceptible to thermal lensing, which allows lower temperatures close to 60°C simplifying OPO operation.

A better method than mechanical crystal rotation is quasi-phase-matching (QPM) by engineering a periodically “poled” nonlinear crystal before using it in an OPO (Fig. 4.41a). THz emitters using poled crystals have been demonstrated [420]. QPM corrects the relative phases of the participating waves by switching the sign of the nonlinear coefficient $\chi^{(2)}$ (Sect. 2.5) periodically with the coherence length L_c . This is achieved by applying an electric field, with alternating opposite directions, over a section matching the poling period to the length $\lambda = 2\pi/\Delta k$ with respect to the incoming perpendicular pump wave. The poling induces periodic domains with a period Λ corresponding to the coherence length. LiNbO_3 can be poled in sections as short as $4\text{ }\mu\text{m}$ for 1-mm thick crystals. In practice, several poling periods are placed on the same crystal perpendicular to the incoming beam direction. Translation of the crystal with respect to the pump beam allows a global change in the usable frequency range. Higher order QPM is often used since the practical poling periods often exceed the coherence length. Then the phase matching condition is fulfilled at a higher order m , i. e. every 3 or 5 coherence lengths. The QPM condition becomes $\Delta k = \pm 2\pi m/\Lambda$ and the nonlinear coefficient d_{ij} of the nonlinear tensor (Sect. 2.5) is modified into an effective nonlinear coefficient $d_{\text{eff}} = 2d_{ij}/(\pi m)$. The QPM condition is independent of polarization so that the largest tensor element can be chosen. For LiNbO_3 , this is d_{33} instead of d_{31} , being 6.5 times larger. The OPO phase matching condition is modified for QPM: $k_p = k_s + k_i \pm \pi/\Lambda$. Coarse and fine frequency tuning is achieved by the poling period L and changing the crystal temperature, respectively.

THz OPOs are room temperature devices, commercially available, and attractive sources, because the pump laser properties are immediately converted to the OPO output. Therefore, OPOs benefit from the improvements of pump lasers, which are developed for the telecommunications industry. Higher power, shorter pulses at high repetition rates, and stable operation in a small package can be expected for THz OPOs.

Chapter 5

Detectors

5.1 Introduction

Detectors are transducers which convert an incoming signal into some convenient form which can be observed, recorded, and analyzed. The signal – an electromagnetic wave, has an amplitude and a phase. Both contain information. According to the nature of the electromagnetic wave detectors can be separated into two classes. If only the amplitude is detected, it is an incoherent or direct detector. When both the amplitude and phase are detected, the detector is a coherent one. Coherent detection is not a direct process. In a heterodyne receiver, which is the most common coherent detector, detection is in two stages, with the incoming signal being “mixed” with another signal, and it is the combination of both which is detected.

The development of faster and more sensitive detectors during the past sixty years has been a major reason for the opening up of the THz spectral region. As might be expected, it has been a two-way approach, with the extension of microwave techniques, particularly heterodyne systems, to higher frequencies, and of IR detectors to longer wavelengths.

There is a wide range of physical effects, which can be used in detection systems. The majority of these can be conveniently categorized into four main groups:

- *Thermal detectors*: In these devices, the radiation is absorbed to produce heat. The change of temperature produces some physical change in the detector material, which can then be measured. The most widely used thermal detectors are bolometers, in which the rise in temperature causes the electrical resistance to change. Alternatives are pyroelectric detectors that have an in-built electrical charge, which alters when the temperature increases, and the Golay cell detector, which is a very sensitive gas thermometer, where the expansion of a gas-filled cell is measured with an intricate optical system.

The general feature of thermal detectors is that they are normally comparatively slow because a bulk of material needs to be heated up. However, there

are important exceptions. Both superconducting and semiconducting electron bolometers, for example, can be very fast. A second feature is that they usually cover a wide spectral range. Ideally, the absorption is independent of frequency but, particularly at lower THz frequencies, this is often difficult to achieve and a major problem of the THz region is measurement of absolute power.

- *Photodetectors*: As the name implies, these respond to individual photons. At THz frequencies, the photon energy is small, and this typically corresponds to the energy gap between shallow impurity states and the conduction or valence band in semiconductors. The general features of these so-called extrinsic photoconductors are that they are fast, as only electrons are involved, and they have a sharp cutoff of response at a frequency where the photon energy becomes less than the ionization energy of the impurity state. To avoid thermal ionization, these detectors require cooling, usually close to liquid helium temperature. THz photodetectors are nearly always photoconductors, with the change in the number of electrons producing a measurable change in resistance. At frequencies above about 10 THz intrinsic photodetectors can be employed. In these, the excitation by the radiation causes transitions between the valence and conduction band of the semiconductor but, as the energy gap is relatively small, cooling is still required.
- *Rectifiers*: Rectifiers for the THz region are high-frequency versions of those used in the microwave or mmW spectral regions. In these, currents are induced at the radiation frequency and, by a nonlinear voltage current relationship in the device, a DC component, related to the amplitude or power of the AC input, is obtained. There are, however, very real difficulties in extending rectifying detectors to higher frequencies because a combination of resistance, capacitance, and inductance causes increased losses. Point contact Schottky diodes (essentially the same detectors as used in 1920s' crystal radios) have been operated to well beyond 3 THz but these devices are delicate and lack the necessary reliability for many applications. Progress in extending planar diode structures to higher frequencies has been steady, and relatively low-loss detectors are now available to above 1 THz. Other rectifiers are based on particular properties of superconductors such as photon-assisted tunneling in superconductor–insulator–superconductor devices. However, these detectors have to be cooled to about 4 K.
- *Mixers*: There are great advantages in employing rectifying detectors in heterodyne systems. In a heterodyne system, the detector is used as a mixer. It receives signals from both the source under investigation and a local oscillator operating at a frequency near to that of the signal and generates the difference frequency between both. This process is called mixing or down-conversion and the reasons for this are explained in detail in Sect. 5.5.

The detectors used in time-domain spectroscopy are a special case and these are discussed in Sect. 6.6.

5.1.1 Detector Cooling

As mentioned in the above discussion of the various types of detectors for the THz region, many of them require cooling to approximately the temperature of liquid helium (4.2 K at 1,000 hPa) or less. For many devices, the most convenient arrangement is to use a commercially available cryostat such as that shown in Fig. 5.1. The design is comparatively simple. The outer casing is a cylindrical vacuum vessel in which the liquid helium is held in a metal container suspended by a low thermal conductivity stainless steel tube. This tube passes through the center of a second container filled with liquid nitrogen. Thermally attached to this second container is a thin walled metal cylinder cooled by thermal conduction to close to 80 K, which acts as a radiation shield between the outer casing and the helium container.

The base of the helium container is a thick plate made of oxygen-free high conductivity (OFHC) copper that has a large thermal conductance at 4 K. Both the base of the outer casing and of the radiation shield are removable, allowing easy access for the mounting of detectors. In Fig. 5.1, a typical arrangement for a

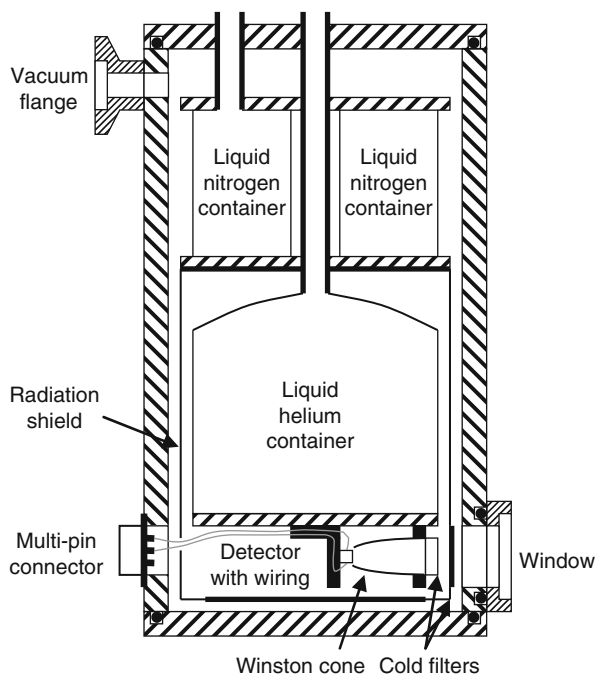


Fig. 5.1 Liquid helium cryostat consisting of a liquid nitrogen container connected to a metal screen, which serves as a radiation shield. A detector and a condensing Winston cone including cold filters, limiting the frequency band incident on the detector, are mounted on the bottom plate of the liquid helium container

detector such as a bolometer or a Ge:Ga photoconductor is shown. The detector is mounted on an L-shaped piece of OFHC copper and in front of it is a Winston cone (Sect. 3.10) for condensing the incoming radiation onto the detector. Filters can be mounted on both the liquid nitrogen-cooled radiation shield and, at the entrance of the Winston cone, to cut off most of the thermal background radiation. Some cryostats are designed with a system for interchanging several cold filters during experiments.

In the cryostat shown in Fig. 5.1, the volume of the helium container is approximately 1 liter. The hold-time, which depends on the specific design, can be well over 48 hours before all liquid has evaporated. By reducing the pressure above the liquid, by connecting a vacuum pump to the top of the filling tube, the helium temperature can be lowered. When it reaches ~ 2.2 K, liquid helium becomes superfluid, and operating below this temperature can be advantageous for two reasons. First, many bolometer-type detectors and extrinsic photoconductors have increased detectivity as the temperature is reduced and, second, superfluid helium is completely still. There is no bubbling of the liquid as all the evaporation is at the surface. Bubbling often produces excess noise due to both small temperature changes and mechanical vibration.

Larger laboratory cryostats have helium lifetimes of a week or more. For particular applications, much more capacious helium containers have been used. An extreme example was for the European Space Agency's "Infrared Space Observatory" (ISO). This was launched in 1995 with detectors covering the range $2.5\text{--}240\text{ }\mu\text{m}$. The cooling system used 2300 liters of superfluid helium which lasted for 30 months. Within the cryostat of the Herschel Space Observatory are 2300 liters of liquid helium for primary cooling. Individual instrument detectors are equipped with additional, specialized, cooling systems to achieve the very lowest temperatures, down to 0.3 K.

An alternative to the use of liquid helium is a mechanical cooling engine. The Joule–Thompson effect describes how a compressed gas expanding into a vacuum lowers its temperature due to adiabatic cooling. Typically, helium gas is pressurized to 1.6 MPa (16 bar) and expanded into a vacuum, thus cooling the attached copper plate to cryogenic temperatures. With the more efficient Gifford–McMahon process temperatures down to ~ 2 K can be reached. At 4.2 K, a thermal load of 1.5 W can be extracted. Pulse-tube coolers spatially separate the mechanical compression system and the cold section more effectively than conventional cryogenic coolers. Such pulse-tube coolers are desirable for vibration sensitive systems.

5.2 Detector Theory

The theory relevant to each type of detector, and to the specific devices employed at THz frequencies, will be described in later sections. However, it is convenient at this stage to include a discussion of the features that affect the performance of any detector and which allows direct comparison between them.

5.2.1 *Detector Parameters*

The performance of a THz detector depends on a number of parameters, some of them inter-related. The more important of these parameters are

- *Band of response*: The spectral range over which the detector responds.
- *Responsivity*: As the great majority of detectors produce an electrical output, this is specified as the voltage or current output per watt of incident power. For photodetectors, it would be more logical to quote the responsivity in terms of the photon flux but data sheets and most publications are written in terms of power.
- *Noise characteristics*: Noise from a wide variety of processes also produces an electrical output from the detector. Often noise properties of a detector are characterized by the noise equivalent power NEP, which is the signal power required to yield a signal-to-noise (S/N)-ratio of unity at the output of the detector in a 1-Hz bandwidth. Note that the NEP is a measure of the (S/N)-ratio and not just noise. The other important parameter is the frequency spectrum of the noise.
- *Dynamic range*: The range from where the signal produces an output equal to that caused by all sources of noise, to the point where saturation causes nonlinearity.
- *Speed of response*: This expresses the speed at which a device responds to a change in incident power and may be limited by any associated circuitry. It is a parameter that can cause confusion, as there is more than one definition. The time constant τ is the traditional method for defining the speed of a detector. It is the time required for the output signal of a detector to rise to $(1 - 1/e)$ of its final value, when it is abruptly exposed to a continuous field of electromagnetic radiation, or fall to $1/e$ of its peak value when the field is cut off. It is important to note that, depending on the detector mechanism, the rise and fall times may or may not be equal. An alternative parameter to the time constant is the rise time, T , which is the time interval required for the detector's output signal to rise from 10% to 90% of its final value. This is related to the time constant by $T = 2.2\tau$. In spectroscopic experiments, when the source is continuous, it is normal to modulate the radiation to provide an AC signal. With slower detectors, it is important to use an appropriate modulation frequency. The relationship between modulation frequency and time constant is $f_C = 1/(2\pi\tau)$ where f_C is the frequency where the signal has fallen by 3 dB from its peak value. This assumes that the detector has a single time constant, i.e. that the rise and fall times are equal. It is important to note that this may not be the optimum modulation frequency, as this will also depend on the variation of detector noise with frequency.
- *Sensitive area*: This parameter is normally the physical area of the detector.
- *Acceptance solid angle*: This parameter is normally governed by physical limitations. With cooled detectors, these limitations are often deliberately created in order to increase the detector's sensitivity.

Other parameters which may need to be considered include sensitivity to environmental factors such as vibration, radiated electrical and magnetic fields, X-rays, etc. If the detector is to be used for long periods where it cannot be replaced, for example in a satellite, any deterioration due to ageing becomes important.

5.2.2 Relationship Between Detector Parameters

It is instructive to consider the (S/N) -ratio of a direct detector with only photons from the signal source arriving at it. The signal power is P_S . For Poisson statistics, the root mean square (rms) fluctuation in the number of photons arriving in one second at the detector is

$$\langle (\Delta n)^2 \rangle = \frac{P_S}{h\nu}. \quad (5.1)$$

The rms noise power at the output of the detector can be obtained by multiplication with $(h\nu)^2$ and $2B$ in order to convert the 1 s averaging time into a bandwidth B

$$\langle (\Delta P_N)^2 \rangle = 2h\nu P_S B. \quad (5.2)$$

In most cases, incoherent detectors are square law detectors, i.e. the output power P_{out} is proportional to the square of the input power

$$P_{\text{out}} = \eta P_S^2, \quad (5.3)$$

where η is the fraction of the incident power which produces a signal at the output of the detector. The (S/N) -ratio in the signal-noise-limited case is given by

$$(S/N)_{\text{SL}} = \frac{P_{\text{out}}}{\langle (\Delta P_N)^2 \rangle} = \frac{\eta P_S}{2h\nu B}, \quad (5.4)$$

and the NEP can be obtained by setting $(S/N)_{\text{SL}} = 1$:

$$\text{NEP}_{\text{SL}} = \frac{2h\nu B}{\eta}. \quad (5.5)$$

There are a number of noise sources which can contribute to the NEP. The more important of these are:

- The detector receives background radiation from its surroundings (P_B). This has a statistical fluctuation in intensity and produces noise power.
- Electrical fluctuations in the detector produce noise (P_D). At a minimum, this is Johnson noise but, as most detectors are biased, this is normally increased by the current flowing through the detector.

- Detectors normally require subsidiary electronics to amplify and display the signal. In certain cases, it is not possible to reduce the input noise of the electronics (P_A) below that of the detector.
- Any fluctuation in the observed signal, for example atmospheric turbulence in THz astronomy, introduces noise (P_S).
- Unwanted environmental noise, for example, thermal fluctuations (P_E), is often a major factor when employing very sensitive detectors.

The total noise P_t is given by

$$P_t^2 = P_B^2 + P_D^2 + P_A^2 + P_S^2 + P_E^2. \quad (5.6)$$

Similarly, the total NEP is given by a quadratic combination of the NEPs from various noise sources. In an “ideal” detector

$$P_B^2 > P_D^2 + P_A^2 + P_S^2 + P_E^2, \quad (5.7)$$

and the NEP is then determined by the fluctuations in the background power falling on the detector. In this ideal situation, only photons from the signal source and photons from the background are present. Using the same arguments as above yields for the rms noise power

$$\langle (\Delta P_N)^2 \rangle = 2h\nu(P_S + P_B)B, \quad (5.8)$$

and the (S/N) -ratio and NEP in the background limited case are

$$(S/N)_{BL} = \frac{\eta P_S^2}{2h\nu(P_S + P_B)B}, \quad (5.9)$$

$$NEP_{BL} = \sqrt{\frac{2h\nu(P_S + P_B)B}{\eta}} \approx \sqrt{\frac{2h\nu P_B B}{\eta}} = \sqrt{\frac{2h\nu I_B}{\eta}} \sqrt{AB}. \quad (5.10)$$

This approximation is valid for $P_S \ll P_B$, which is the case in many applications. I_B is the background irradiance, i.e. the power of the background radiation per area A . It is worth noting that the NEP is proportional to the square root of the bandwidth. This results from the square-law detection process. If, for example, the bandwidth is doubled, the output noise power is also doubled (5.2), whereas the signal power has only to be raised by a factor of $\sqrt{2}$ (5.3) in order to obtain the same (S/N) -ratio at the output of the detector.

For many purposes, the NEP of a detector is a quite convenient parameter. However, it has the psychological disadvantage that the *better* the detector the *smaller* the NEP, and therefore the term “Detectivity” (D) was introduced [421]. This is the reciprocal of the NEP and hence is given by inverting (5.5) or (5.10). In a background limited detector, it can be seen from (5.10) that the NEP increases as

the square root of the area, and it is an added convenience to take this into account. The term D^* (pronounced as Dee-Star) is defined as

$$D^* = \frac{\sqrt{AB}}{\text{NEP}_{\text{BL}}} = \sqrt{\frac{\eta}{2h\nu I_{\text{B}}}}. \quad (5.11)$$

With A normally in cm^2 and B in Hz the units of D^* are $\text{cm}\sqrt{\text{Hz}}/\text{W}$. The advantage of using D^* comes from the fact that it is independent of the size of the detector as long as the rms noise of the detector is proportional to its area. This is true for any background limited detector but also in many other cases, for example when the NEP is limited by amplifier noise. Definitions change over the years. Originally, D was the “Detectivity” and D^* the “Specific Detectivity” but, as D is largely redundant, “Detectivity” now normally means D^* .

If an ideal background limited detector (one that has total absorption at all wavelengths) is placed in a blackbody enclosure at a temperature T , the fluctuations in power, and therefore the NEP, are given by [422]:

$$\text{NEP} = 4\sqrt{\sigma k_{\text{B}} T^5 AB}, \quad (5.12)$$

where σ is Stefan’s constant. If $A = 1 \text{ cm}^2$ and $T = 300 \text{ K}$ this yields an NEP of $5.55 \times 10^{-11} \text{ W}/\sqrt{\text{Hz}}$ and a detectivity of $1.80 \times 10^{10} \sqrt{\text{Hz}}/\text{W}$.

D^* is the most frequently used figure-of-merit for comparing detectors. Because the responsivity of many detectors varies with frequency, all detectors have a limiting speed of response and, as detector noise is also frequency dependent, D^* normally has extra information added to it. As an example, for a photodetector that is frequency dependent in its response, the D^* might be written in several ways: $D^*(\lambda_{\text{p}}, f, B)$. Here, λ_{p} is the wavelength of peak response, f is the modulation frequency of the radiation falling on the detector, and B is the bandwidth of the amplifying system. Unless otherwise stated, the bandwidth is assumed to be 1 Hz and this parameter is frequently omitted. An alternative figure-of-merit for photodetectors is $D^*(T, f, B)$. Here, D^* is the average detectivity over the wavelength range over which the detector responds to radiation from a blackbody at temperature T . Historically, the modulation frequency used for assessing photodetectors was 800 Hz, as this was relatively easy to obtain and fast enough to minimize the effect of current noise in biased detectors, which typically varies as $1/f$.

As early as 1947, Golay proposed a figure-of-merit for detectors, which took into account the reduction of background radiation by reducing the acceptance solid angle of cooled detectors [32]. In 1959, Jones [421] defined a figure-of-merit D^{**} appropriate for BLIP (Background Limited Infrared Photodetector) detectors, which removed any need to specify the solid angle when characterizing the D^* of a detector.

From geometric considerations, assuming negligible radiation is reaching a BLIP detector except that coming from the required field of view, the dependence of D^* on the viewing angle is given by

$$D^*(\Theta) = D^* \frac{2\pi}{\sin(\Theta)}, \quad (5.13)$$

where $D^*(\Theta)$ is the D^* value obtained when the aperture of the receiving system has a half-angle of Θ and $2\pi D^*$ is the D^* value when the detector is exposed to radiation from the entire forward hemisphere. The relationship between D^{**} and D^* is

$$D^{**} = D^* \sin(\Theta), \quad (5.14)$$

and the units of D^{**} are $\text{sr cm} \sqrt{\text{Hz/W}}$.

A noticeable feature of photodetectors is the apparent steady increase in detectivity toward lower frequencies, followed by a sharp fall at the cutoff frequency. Traditionally detectivity is written in terms of watts and, for an ideal photoconductor with uniform absorption across the band of response, the increase is because the number of “photons per watt” decreases linearly with frequency. For thermal detectors, the responsivity is reasonably constant over a wide frequency range but may, for example, be optimized for a specific frequency or frequency range with an antireflection coating.

When determining the ultimate performance of detectors, (5.12) must be modified for the parameters of the detector and the conditions under which it is used. The equation is only true for devices receiving background radiation at all frequencies. For photodetectors this is not the case, as photons beyond the cutoff frequency will not affect the performance. More importantly, THz detectors are often cooled, sometimes to temperatures below 1 K, and, at the same time, the field of view of the background is reduced to a small solid angle. Moreover, in many astronomical studies and experiments on the molecular emission in the upper atmosphere, the background temperature is much lower than room temperature.

The reduction in background radiation, and therefore background noise, is dramatic, as illustrated in Fig. 5.2. The values shown are for an “ideal” photodetector and assume that the only noise is due to the background and that the detector has a quantum efficiency of one. Quantum efficiency is the probability that a photoelectron is produced when a photon is absorbed by the detector. Not surprisingly, the curves in Fig. 5.2 have a close resemblance to inverted blackbody curves plotted in terms of photons, with the minimum D^* corresponding to the peak of the blackbody emission. In practice, unless the detector is observing a source close to its own temperature, noise from the signal will also be present and, although this may only be received within a small solid angle, it can be the dominant factor. Detectors where the performance is limited only by background radiation are said to be BLIP devices, with the photon source then being the dominant source of noise.

Thermal detectors can also reach very high detectivities by cooling to very low temperatures and using cooled filters to allow only the wanted frequencies through to the detector. But it must be emphasized that with both thermal detectors and photodetectors, as the noise from background radiation is reduced, other noise sources often dominate the performance. The great advantage of heterodyne systems, particularly when observing narrow-band sources as, for example, in studying

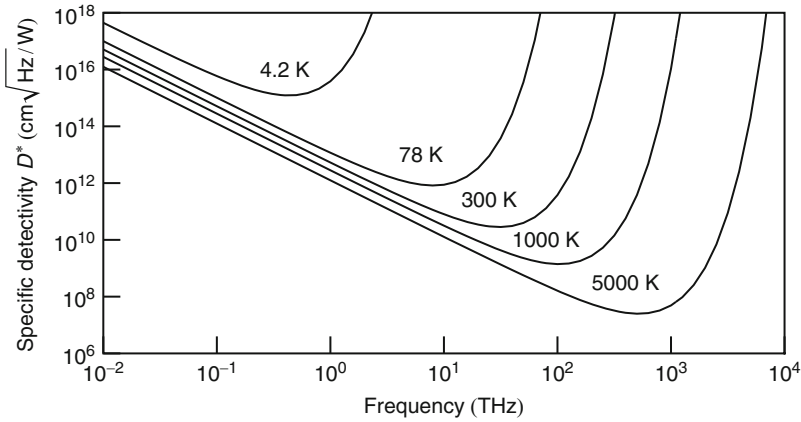


Fig. 5.2 Photon noise limited D^* for an ideal photovoltaic detector at its peak frequency of response (assumed to be the cutoff frequency) for a range of background temperatures, a 2π field of view, and with the generation-recombination noise neglected

molecular emission in the upper atmosphere, is that all the noise components are restricted to the observing bandwidth. Detectivity can then improve by many orders of magnitude.

5.2.3 Sources of Excess Noise

An ideal detector is one that is only limited by noise from the background but, particularly when this noise source is reduced by cooling, other noise contributors may become dominant. Some of these have been mentioned in Sect. 5.2.2 and any special noise contributors for specific detectors will be discussed in later sections, but it is convenient, at this stage, to give a brief explanation of the major types of noise that occur.

- *Johnson noise*: Any resistive device produces a fluctuating voltage across its ends due to the random movement (Brownian motion) of its charge carriers. Johnson noise voltage is given for a device of resistance R by

$$V_J = \sqrt{4k_B T R B}, \quad (5.15)$$

where T is the absolute temperature, and B is the bandwidth of the detection system. The Johnson noise for a 50Ω device at 300 K is $0.9 \text{ nV}/\sqrt{\text{Hz}}$.

For cooled detectors, care is often required in reducing the input noise of any amplifying system to below the Johnson noise from the detector. Cooling from room temperature at 300 K down to 3 K reduces the Johnson noise by a factor of 10.

- *Current noise*: When a current is passed through resistive devices, there is usually increased noise. At low frequencies, typically 10^3 – 10^4 Hz, this has a $1/f$ characteristic while at higher frequencies shot noise dominates. At low frequencies, typically less than 100 Hz, current noise often rises very rapidly. It is still not a fully understood noise source and in early semiconductor detector research it was a major problem, associated with surface conditions of photoconductors, imperfect ohmic contacts, and dislocations within the crystals. These effects seem now to have been largely overcome with modern technology. With photoconductors, current noise in the biasing resistors needs to be considered. Metallic wire resistors seem not to produce excess noise but solid carbon resistors are particularly bad. However, it is now possible to find even very high value resistors which show little excess noise.
- *Generation–recombination noise*: This noise is present in semiconductors and arises from lattice vibrations (phonons), which excite electrons and holes. These electrons and holes “wander” around the lattice and eventually recombine so that at any one time there is a fluctuation of charge carriers about a mean value. This causes a change in the resistance of the semiconductor, which shows up as a voltage variation when current flows, and adds to any other noise process. Detailed study of generation–recombination noise [423] has shown that, in a photoconductive detector at the same temperature as its background, its noise fluctuation is equal to that of the background and therefore the total noise contribution in an ideal photoconductor is $\sqrt{2}$ higher than that shown in Fig. 5.2.
- *Thermal fluctuation noise*: If a detector is in thermal contact with a heat sink, as is the case with bolometers, there will be thermal noise due to fluctuations in the temperature of the heat sink. This will be a contributor to the total noise voltage when a current is passed through the detector. The ultimate limit of thermal fluctuation noise is given by the passage of quantized carriers, for example, phonons, which affect the thermal conductance of the bolometer. These generate a white noise voltage of

$$V_{TF} = \sqrt{4k_B T^2 G R B}, \quad (5.16)$$

where G is the heat conductivity between the detector and the heat sink (Sect. 5.3). Thermal fluctuation noise is normally only significant in very sensitive cooled bolometers, which are designed in such a way that all other noise sources are smaller than the thermal fluctuation noise.

- *Amplifier noise*: With very low noise detectors, the input noise of the first stage of any electronic equipment employed to amplify the detector signal can be significant. Typically, the minimum input noise voltage of a room-temperature amplifier is about $1 \text{ nV}/\sqrt{\text{Hz}}$, and this may be higher than the noise voltage from a cooled detector. Fortunately, it is possible to cool the early stages of an amplifier system and what used to be a significant problem in the past can now be overcome by careful design.

5.3 Thermal Detectors

Room-temperature thermal detectors are used throughout the THz range. The earliest thermal detectors for IR frequencies were thermocouples, thermopiles, and bolometers. Thermopiles with useful detectivity and relatively uniform response throughout the THz region are employed for power measurement and are described in Sect. 5.3.3. The problem with thermopiles in the THz region is the difficulty of making them with a uniform absorbance close to unity across the whole detection band, without increasing their thermal mass very significantly. The most useful thermal detectors for room-temperature operation are the Golay pneumatic detector and pyroelectric devices, which are discussed in detail in Sects. 5.3.1 and 5.3.2, respectively. As mentioned earlier, bolometers operating at very low temperatures are the most important thermal detectors for the THz region. The more important of these are described in detail later in this chapter, but it is convenient to consider the parameters that affect their performance at this stage.

Any thermal detector can be represented by Fig. 5.3. When signal radiation falls on the detector its rise in temperature $\Theta(t)$ is determined by the heat transfer equation

$$P(t) = P_C + \eta P_\omega(t) = C \frac{d\Theta(t)}{dt} + G\Theta(t), \quad (5.17)$$

where $P(t)$ is the total power absorbed by the detector. P_C is a constant part which is absorbed in thermal equilibrium. Usually, this is due to background radiation. $P_\omega(t)$ is the time-varying signal power incident on the detector and η is the fraction of power absorbed, which clearly should be as close to unity as possible. In practice, the signal radiation will be modulated at some appropriate angular frequency ω in order to produce an AC signal from the detector, so

$$P_\omega(t) = P_0 \exp(i\omega t). \quad (5.18)$$

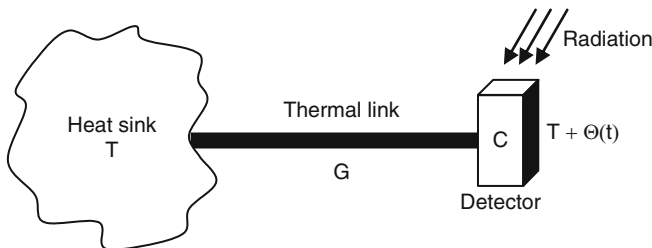


Fig. 5.3 Thermal detector circuit. The detector has a heat capacitance C and is connected by a thermal conductance G to a heat sink of sufficient mass to remain at a constant temperature T . The detector will also be at temperature T , but background radiation from its surroundings will cause fluctuations in this temperature. Incident radiation increases the detector temperature by $\Theta(t)$

Solving (5.17) for the change in temperature Θ due to P_ω yields

$$\Theta = \frac{\eta P_\omega}{\sqrt{G^2 + \omega^2 C^2}}. \quad (5.19)$$

Equation (5.19) illustrates several desirable features of thermal detectors, which are qualitatively self-evident. The ratio of temperature rise to input power should be as large as possible. To achieve this, the heat capacitance needs to be small but consistent with the requirement that the fraction of the incoming radiation absorbed is near unity. The thermal coupling to the surroundings should also be low so that $\omega C \ll G$. If this last requirement is not met and ω is increased until the term ωC exceeds G , then Θ will fall inversely as ω and the thermal response time of the detector can be defined as

$$\tau = C/G. \quad (5.20)$$

In room-temperature thermal detectors, response times will typically range from a few seconds to a few milliseconds. To achieve faster response and maintain a detector's voltage responsivity, the thermal mass must be reduced, but this leads to problems if the detector's area is to be retained. First, the device becomes more fragile and second, it is increasingly difficult to preserve high absorption of the incoming light.

As mentioned in the introduction to this chapter, bolometers are the most widely used thermal detectors because, by cooling to low temperatures both of these problems can be largely overcome. Cooling introduces further advantages, as can be seen from the following analysis. Bolometric detectors are made from materials which show a change of electrical resistance when electromagnetic radiation is absorbed. Since a bolometer needs to be biased it is subject to time-varying electrical heating, because its resistance R is a function of temperature. It is useful to introduce the temperature coefficient of the resistance $\alpha = R^{-1}(dR/dT)$, which is a measure of the steepness of change of resistance with temperature. Assuming that such a bolometer is biased with constant current I , the heat transfer equation needs to be modified to

$$P(t) = C \frac{d\Theta}{dt} + G\Theta - \Theta \frac{dP_E}{dT}. \quad (5.21)$$

Here, $P_E = I^2 R(T)$ is the electrical power dissipated in the bolometer. The change of electrical power due to the temperature-dependent resistance is

$$\frac{dP_E}{dT} = I^2 \frac{dR}{dT} = \alpha I^2 R = \alpha P_E. \quad (5.22)$$

Substituting this into (5.21) and rearranging leads to

$$P(t) = C \frac{d\Theta}{dt} + (G - \alpha P_E)\Theta. \quad (5.23)$$

In analogy with the solution of (5.19), the temperature change is found to be

$$\Theta = \frac{\eta P_{\omega}}{\sqrt{G_{\text{eff}}^2 + \omega^2 C^2}} \quad (5.24)$$

and

$$G_{\text{eff}} = G - \alpha P_E. \quad (5.25)$$

G_{eff} is the effective thermal time constant. Semiconducting bolometers have a negative α and therefore $G_{\text{eff}} > G$, while superconducting bolometers have a positive α which means that $G_{\text{eff}} < G$. This has particular consequences for the design of such bolometers, which will be discussed in forthcoming sections. In addition, the response time is modified due to the electrical heating. In analogy with (5.20), the effective time constant τ_{eff} is obtained

$$\tau_{\text{eff}} = \frac{C}{G_{\text{eff}}} = \frac{\tau}{1 - \alpha P_E / G}. \quad (5.26)$$

Input radiation, which produces a change of temperature Θ , will produce an output voltage

$$V(\Theta) = I \alpha R \Theta, \quad (5.27)$$

and with (5.24) a time-varying voltage is obtained

$$V(t) = \frac{\eta P_{\omega} I \alpha R}{\sqrt{G_{\text{eff}}^2 + \omega^2 C^2}}. \quad (5.28)$$

Note that the thermal conductance has been replaced by the effective thermal conductance. The voltage responsivity is

$$\mathcal{R} = \frac{V}{P_{\omega}} = \frac{\eta I \alpha R}{G_{\text{eff}}} \frac{1}{\sqrt{(1 + \omega^2 \tau_{\text{eff}}^2)}}. \quad (5.29)$$

By choosing an appropriate material and by cooling to 4 K or less, the heat capacitance can be much reduced and the temperature coefficient of the resistance greatly increased, whilst retaining high absorption. A further advantage is that it becomes much easier to control the thermal impedance, thus allowing the speed of response to match the requirement of a particular experiment. F. Low, who invented one of the early liquid helium-cooled bolometers made from p-Ge [66], could control the response time from a few seconds to less than 10 μs . A significant problem with the Ge bolometer was in optimizing both the absorption and resistance coefficients simultaneously, which led to the introduction of composite bolometers where the semiconductor, either Ge or Si, is used to measure the temperature change of a suitable absorbing material which is attached to it (Sect. 5.3.5). This concept was expanded to superconducting bolometers, which are described in Sect. 5.3.6. In other semiconductors, notably lightly doped n-InSb cooled to 4 K, the coupling

of electrons to the lattice is on a timescale of 10^{-6} to 10^{-7} s. At frequencies below about 1 THz, free electrons in n-InSb become absorbing. These electrons are then the “material” of the bolometer with the crystal lattice as the heat sink. This type of detector is described as an electron bolometer (Sect. 5.4.2).

5.3.1 The Golay Detector

As mentioned in Sect. 1.2, the idea of this detector, named after its inventor, dates back to before the Second World War, when both IR detection and radar were being considered for the location of aircraft, but publication was delayed until 1947 [32]. Commercial versions of this detector became available quite rapidly. For many years these were the mainstay of THz research and are still widely used more than sixty years later.

The mechanism of a modern version of this detector is shown in Fig. 5.4, which indicates the main features of the detection mechanism. The gas cell contains xenon, a gas of low thermal conductivity, and it is sealed at one end by a window transparent to the frequency range of interest, and, at the other end, by a very light flexible mirror. Within the cell there is a thin absorbing metallic film with an impedance approximately matching that of free space. Golay found empirically that, allowing for internal reflections within the cell, a resistance of $270\ \Omega$ per square produced optimum performance. One of the great advantages of this detector is that it provides nearly constant absorption over the entire IR and THz region. When radiation passes through the window and is absorbed by the metallic film the gas is heated and there is a small movement of the flexible mirror. The optical system shown then converts this movement into an electrical signal. Light from the source is condensed by the lens system through a grid and is then reflected by the flexible mirror back through the grid onto the detector. Any movement of the mirror distorts the reflected image of the grid and changes the amount of light reaching the light detector.

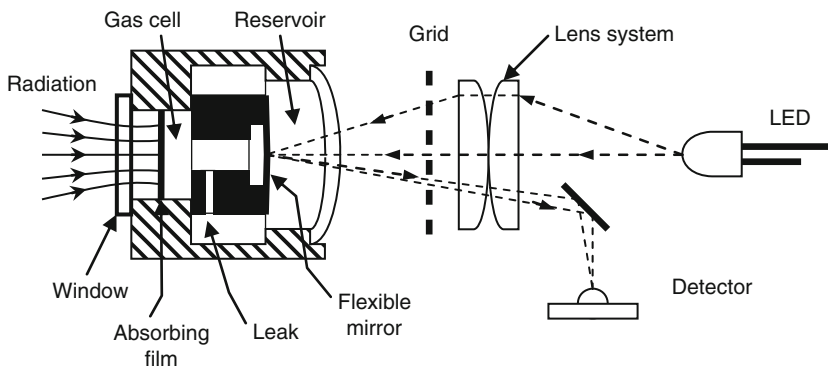


Fig. 5.4 Design of a Golay detector (adapted from [32])

This optical arrangement is very sensitive to change of shape of the flexible mirror and movements of less than 10 nm are detectable. To compensate for changes in the ambient temperature, and to make the device into an AC detector, there is a small leak from the gas cell into a reservoir. This leak leads to a time constant of a few Hz and because of this, and the fact that the noise from the detector rises rapidly below 10 Hz, the optimum modulation speed for incoming radiation is in the range 10–20 Hz.

For THz frequencies, the most useful window materials are high-density polyethylene, high-resistivity Si, crystalline quartz, and diamond. The high-frequency cut-on is determined by the window material, while the low-frequency cutoff is determined by diffraction at the entrance aperture of the cell. Due to the very delicate construction of the gas cell and optical system, the detectors must be isolated from any mechanical vibration. Rotary pumps are a particular problem. Golay detectors are rather easily damaged by excess input energy which can over-expand the cell before the leak to the reservoir can act.

Golay detectors are remarkably sensitive. Selected devices have an NEP of $2 \times 10^{-10} \text{ W}/\sqrt{\text{Hz}}$, which is less than a factor of four greater than an ideal room temperature thermal detector. One of their major advantages is that their responsivity can be calibrated very precisely with blackbody sources at frequencies corresponding to the near-IR and, because of the absorption process, this calibration remains reasonably accurate to frequencies down to at least 0.3 THz.

An interesting alternative design for a Golay-type sensor uses Micro-Electro-Mechanical Systems (MEMS) technology for its construction. The device is made from micro-machined Si components, and its prototype design is shown schematically in Fig. 5.5.

Apart from being much smaller than a conventional Golay, the important feature is the read-out arrangement that uses an electronic rather than an optoelectronic design. The deflection is detected by a tunneling displacement transducer, which measures variations in a quantum mechanical tunneling current between the moving

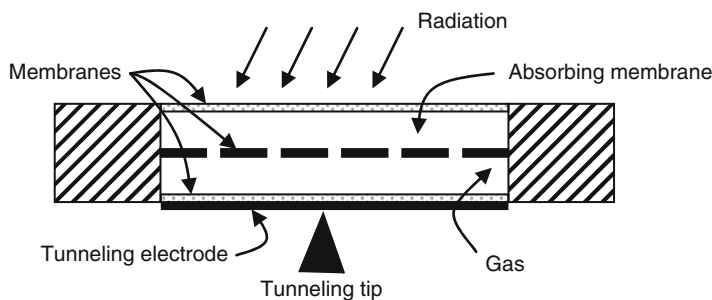


Fig. 5.5 Golay-type sensor made from micro-machined Si components. The incident radiation heats the gas in the sensor, deforming the membranes and, subsequently the tunneling electrode, which is deposited on one of the outer membranes. The measured tunneling current between the moving membrane and the fixed tip is a function of the incident radiation (adapted from [424])

membrane and a fixed tip adjacent to the membrane [424]. The tunneling transducer offers the required sensitivity with very simple circuitry, and it is easy to envisage the development of more compact cells using the same type of structure. In a 1996 paper, it was suggested that the availability of suitably thin Si wafers would restrict manufacture to dimensions greater than ~ 0.3 mm [425]. However, as the authors pointed out, pixels with dimensions close to 1 mm would be useful for detector arrays employed in the THz region. Other authors have suggested that a detectivity of 3.6×10^9 cm $\sqrt{\text{Hz/W}}$ could be expected for a 1 mm² area, micro-machined version of the Golay detector [426]. In practice, research appears to have been concentrated on the development of detectors of a few mm² area, and a high-yield wafer process for the bulk manufacture of miniaturized Golay cells has been described [427]. An alternative approach, which could lead to much denser arrays, uses robust and flexible nano-composite membranes to seal a uniform array of micro-cavities [428]. In this experiment, the deformation of the membrane with temperature was measured by optical interferometry. In a 64×64 array of $80 \mu\text{m}^2$ cavities spaced at $150 \mu\text{m}$, no cross-talking was observed between neighboring cells.

5.3.2 *Pyroelectric Detectors*

A suggestion for using the pyroelectric effect to detect electromagnetic radiation was made in 1938 [64] and the earliest detectors were constructed in the 1950s [429]. Detailed analysis of the characteristics of the pyroelectric detector was made in the early 1960s [430] and by the end of that decade devices employing the effect were in widespread use. Pyroelectric detectors are commercially available either as single devices or as arrays for the entire IR and THz spectral regions. They have many advantages, including being relatively cheap and rugged, with room-temperature operation. Their most useful property is that, with appropriate design of an associated amplifier, they can have response times varying between tens of milliseconds to less than a nanosecond.

Certain materials possess a unique axis along which a permanent electric dipole moment exists. This moment implies that surfaces cut normal to this axis should be electrically charged, but this effect is normally not observed because it becomes neutralized by stray charges. However, if the temperature of the material is changed, for example by pulsed radiation, this causes the lattice spacing to alter, with a consequent change of the dipole moment and electrical charge. This charge change can then be detected by producing a capacitor from the material. This is achieved by using a small slab of the pyroelectric material and applying a pair of electrodes to opposite faces. Nearly all pyroelectric materials are very good electrical insulators so the charge remains relatively stable and quite slow changes in temperature are measurable. The fundamental electrical time constant τ_E of a pyroelectric detector is governed by its own capacitance and resistance. This is normally longer than the thermal time constant, which is dependent on the same parameters as a conventional

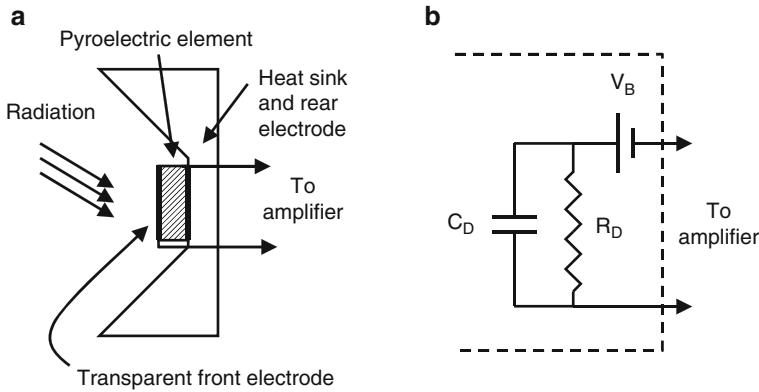


Fig. 5.6 (a) Pyroelectric detector; (b) equivalent electric circuit diagram

bolometer. However, almost invariably τ_E is reduced, either by putting a resistor in parallel with the detector or by the input impedance of the following amplifier. Figure 5.6 shows a typical design for a pyroelectric detector with its thermal and electrical circuits. To keep the thermal mass small, the detector is normally thin. If the actual absorption coefficient is not sufficiently high for the frequency range of interest, the face of the detector is covered with a suitable absorbing material, which can be an absorbing paint. The front electrode is made transparent by using a very thin metallic layer and, if a longer thermal time constant is required, an insulating material can be placed between the rear electrode and the heat sink. Assuming that the detector has an area A and that the modulated input produces a temperature change Θ , then the voltage developed across the amplifier input (Fig. 5.6b) is given by

$$V = \frac{\omega p A \Theta R_S}{\sqrt{1 + \omega^2 \tau_E^2}}, \quad (5.30)$$

where p is the pyroelectric coefficient measured in units of $C/(\text{cm}^2\text{K})$ and $\tau_E = R_S C_E$ is the electrical time constant of the detector and amplifier in parallel.

Substituting for Θ from (5.19), the voltage responsivity of a pyroelectric detector is given by

$$\mathcal{R} = \frac{V}{P_\omega} = \frac{\eta \omega p A R_S}{G} \frac{1}{\sqrt{(1 + \omega^2 \tau_E^2)(1 + \omega^2 \tau^2)}}, \quad (5.31)$$

where τ is the thermal time constant (5.20). At very low frequencies, (5.31) shows that $\mathcal{R}$ will be small and increase with frequency. As the frequency is increased, there will be a range where $\mathcal{R}$ is essentially constant, assuming the two time constants are not equal, and at high frequencies $\mathcal{R}$ will fall inversely with the frequency.

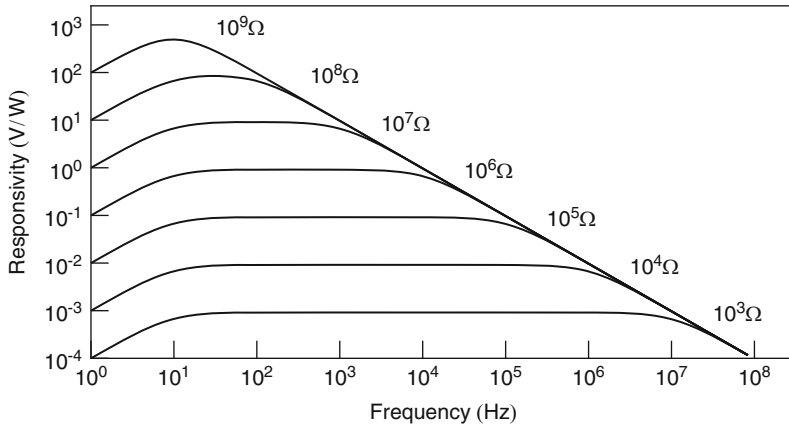


Fig. 5.7 Voltage responsivity of a pyroelectric detector as a function of frequency and load resistance

The high-frequency approximation of (5.31) is

$$\mathcal{R} = \frac{\eta p A}{\omega C C_E}. \quad (5.32)$$

As can be seen from Fig. 5.7, which is a plot of (5.31) for a typical pyroelectric detector, a desired bandwidth can be achieved by placing an appropriate value resistor in parallel with the detector. It is this ability to use a relatively slow device to provide a variable speed detector that distinguishes pyroelectric devices from other thermal detectors, and makes them particularly valuable throughout the IR and THz regions. It is still, of course, important for the input radiation to produce the maximum response and, from (5.31) and (5.32), it is clear that the term p/C must be optimized. The thermal mass and capacitance are proportional to the volume specific heat C of the pyroelectric material and its dielectric constant ϵ , and the quantity $p/(\epsilon C)$ is therefore a convenient figure-of-merit for selecting detector material.

The most sensitive pyroelectric detectors are made from the triglycine sulfate (TGS) group of materials, and these have detectivities quite close to that of Golay detectors. However, in practical applications, because Golays are more susceptible to environmental noise, pyroelectrics are often comparable and even superior. TGS was suggested as a suitable material for pyroelectric detection in 1960 [431] and a detailed study of its detectivity was published in 1965 [432]. Although the detectivity of TGS proved to be superior to alternative materials, there were problems. Pyroelectric materials have a Curie point, the temperature above which their electrical charge disappears. When TGS detectors are heated and then cooled to below the Curie point, the charge may or may not reappear. More alarmingly, even when the material is kept below the Curie temperature the charge can disappear spontaneously, or reverse polarity. Several attempts were made during the 1960s

to overcome these difficulties by the addition of inorganic dopants to the solution from which TGS crystals were grown but these had limited success. However, in 1971 the problem was solved by P. J. Lock, with the introduction of the organic compound aniline into the TGS solution [433]. This aniline doping had a second advantage, as it reduced the dielectric constant of the crystals, thus improving the responsivity. Devices made from this modified TGS were described as ATGS detectors. Further improvements followed as alternative dopants were investigated. The best results were obtained with L- α -Alanine TGS, which was abbreviated to LATGS, but this still had the disadvantage of a low Curie temperature of 49°C. However, the use of deuterated material raises this temperature to 57°C to 60°C, dependent on the deuteration level. DLATGS is now the preferred material from the TGS group and DLATGS detectors are widely available from commercial sources.

Apart from its still quite low Curie temperature, DLATGS has the disadvantage of being a delicate crystal that can be easily damaged by mechanical shock or overheating. More rugged pyroelectric devices are made from lithium tantalate (LiTaO₃), strontium barium niobate, or ceramic members of the lead zirconate (PZT) group. Of these, LiTaO₃ has the highest detectivity and a relatively uniform response throughout the IR and THz regions. Its Curie temperature is very high, at ~650°C, and, in the authors' experience, LiTaO₃ detectors will withstand high energy laser pulses without damage.

Figure 5.8a is a comparison of the performance of a Golay detector with that of DLATGS and LiTaO₃ pyroelectric detectors in the THz region. A great advantage of pyroelectric devices is their small physical size. For example, a $3 \times 3 \text{ mm}^2$ detector with an integrated amplifier fits into a standard transistor package (Fig. 5.8b). A minor disadvantage of TGS-based detectors, with their low Curie temperature, is some variation in responsivity with ambient temperature changes, or when the average signal power falling on them is at a high enough level to cause significant heating. In situations where this would present a problem temperature-stabilized devices are available, with a thermoelectric cooler and thermistor included in the detector package. Pyroelectric detectors are versatile devices as, providing the detector material is absorbing or coated with an absorbing layer, they can detect from the visible to the THz spectral region. Furthermore, there is virtually no limit to a detector's area. Very large area, and very cheap, pyroelectric detectors can also be made from films of the polymers polyvinyl fluoride [434] and polyvinylidene fluoride [435].

With the development of charge-coupled device (CCD) technology, it became possible to simplify the read-out electronics and to design compact 2D arrays of pyroelectric detectors. While these were initially for thermal imaging in the 8 to 14 μm range, they have proved to be very versatile. For example, pyroelectric cameras, which have typically 124×124 LiTaO₃ elements within an active area of $12.4 \times 12.4 \text{ mm}^2$, were primarily designed for real-time imaging of laser beams in the ultraviolet, visible, and IR spectral regions. However, they have been employed for sources over a range from X-rays to below 3 THz, in addition to thermal imaging and various other applications.

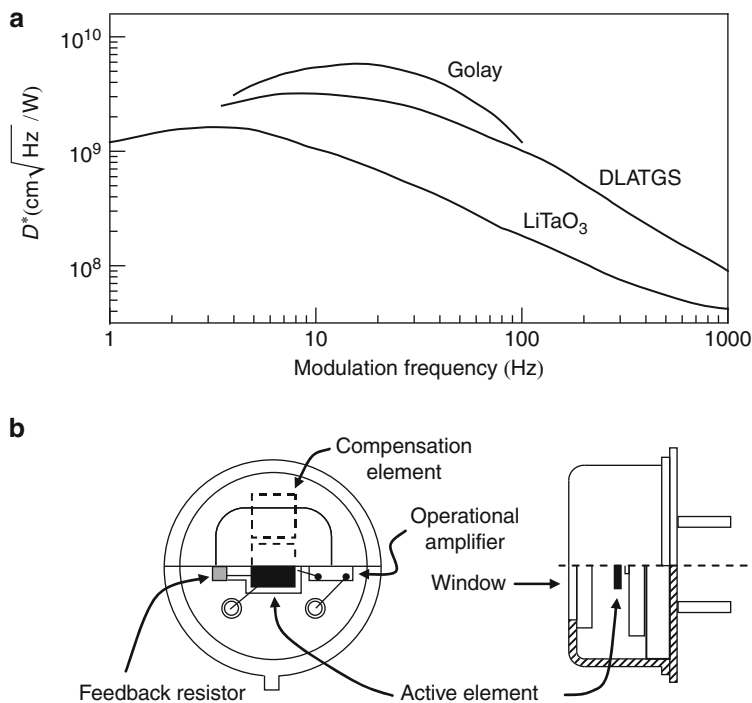


Fig. 5.8 (a) Variation of D^* with frequency for three thermal detectors operating at 295 K: 5 mm diameter Golay, $3 \times 3 \text{ mm}^2$ DLATGS pyroelectric, $3 \times 3 \text{ mm}^2$ LiTaO₃ pyroelectric (adapted from various sources). (b) Detector in standard transistor package

The noise from pyroelectric detectors is predominantly Johnson noise from the resistor in parallel with the detector, although in some applications it may be difficult to avoid extra noise from the input stage of the following amplifier. An advantage compared to bolometers is that excess electrical noise at low frequencies is absent because there is no current flowing through this type of detector.

Although pyroelectric detectors can respond over the entire THz region, the response relies on the absorbance of the detector material or absorbing layer. This is not easy to achieve, and Fig. 5.9 [436] shows the spectral response of two pyroelectric detectors normalized to the spectrum measured with a Golay detector. There is clear evidence of interference effects due to internal reflections. However, this seems to be an extreme example. Other users have reported much less variation [114].

It is important when using pyroelectric devices to match the bandwidth of the detector exactly to the required speed of response if optimum detectivity is to be achieved. Making the device faster and then limiting the bandwidth after the detector will always produce an inferior performance. This is clear from inspection of Fig. 5.7, which shows that the voltage responsivity varies inversely with frequency and, assuming only Johnson noise, the detectivity will decrease as the square root of

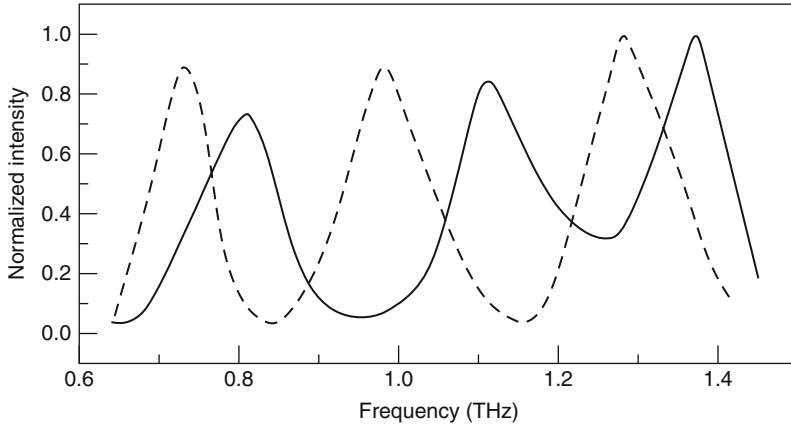


Fig. 5.9 Spectral response of two pyroelectric detectors normalized to the spectrum measured with a Golay detector. The peak structure is caused by interference effects in the pyroelectric crystal due to internal reflections (adapted from [436])

the bandwidth. The requirement of any amplifying and display system employed is that it should have sufficient bandwidth to utilize fully the required speed of response of the detector and have lower input noise than that produced by the detector resistor.

5.3.3 Thermopiles

The thermoelectric effect can be used for detection of THz radiation. When two different metals are in contact, the mismatch of their Fermi levels results in an electromotive force across the contact. The magnitude of this force depends on the temperature. If a second junction from the same metals is held at a constant reference temperature and if both junctions are connected by an electrical circuit, a voltage ΔV will develop between them:

$$\Delta V = \delta\alpha\Delta T. \quad (5.33)$$

Here, ΔT is the temperature difference between the two junctions and $\delta\alpha$ is the difference of the Seebeck coefficients of the two metals. The exact value of $\delta\alpha$ depends on the metal pair. Typically, it is in the order of $50\mu\text{V/K}$. By connecting several of these thermoelectric junctions or thermocouples in series, the response is increased and the device is then called a thermopile.

In order to use thermopiles for the detection of radiation, they have to be combined with an absorber. Ideally, it has complete absorption across the whole frequency band of interest. In addition, a reference temperature needs to be

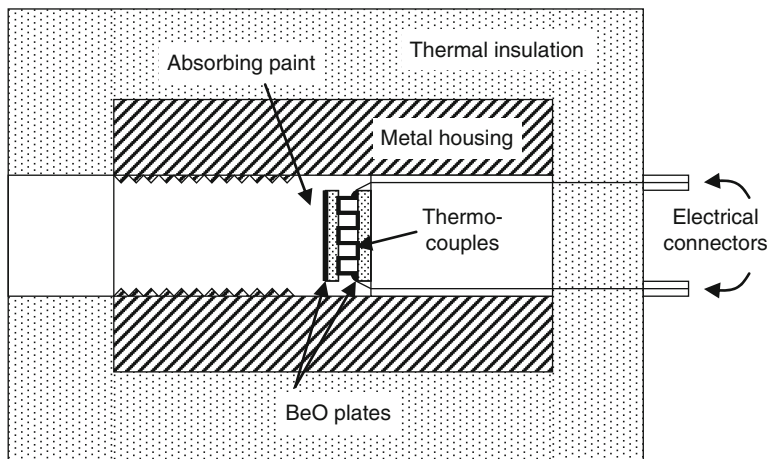


Fig. 5.10 Cross section of a thermopile. Radiation incident on the BeO plate with absorbing paint is converted to heat which sets up a temperature gradient between the front and rear BeO plates. A series of thermocouples mounted between the two plates creates a voltage which is proportional to the temperature gradient and induced by the incident radiation between the two plates (adapted from [437])

provided, which should be highly stable. A thermopile design which is used at THz frequencies is depicted in Fig. 5.10 [437]. The sensing element is a Peltier cell made from two BeO ceramic plates. The front plate, which is quite thin ($600\text{ }\mu\text{m}$) in order to reduce the thermal mass and to achieve a reasonable speed of response ($\sim 4\text{ s}$), is covered by a highly absorbing paint. Several thermocouples are connected in series between the two plates. They are mounted in a metal housing which in turn might be, but is not necessarily, incorporated in a housing with thermal insulation. The minimum detectable power within an integration time of 4 s is $\sim 10\text{ }\mu\text{W}$ per cm^2 of detector area. During a measurement, it is important to keep the temperature of the device constant. Even a slight temperature drift can account for a baseline drift of several tens of μW . The absolute accuracy is between $\pm 5\%$ and $\pm 20\%$, depending on the frequency [437]. Limiting factors are uncertainties in the absorption of the paint, reflection from the BeO plate due to mismatch of the refractive index ($n_{\text{BeO}} \approx 2.5$), and standing waves.

5.3.4 Power Meters

Absolute power measurements are challenging at THz frequencies. Diffraction leads to uncertainties of the beam propagation, and reflectance from the detector leads to standing wave effects. Both affect the coupling of THz radiation into the power meter. At THz frequencies, the detectors discussed in the previous sections are only

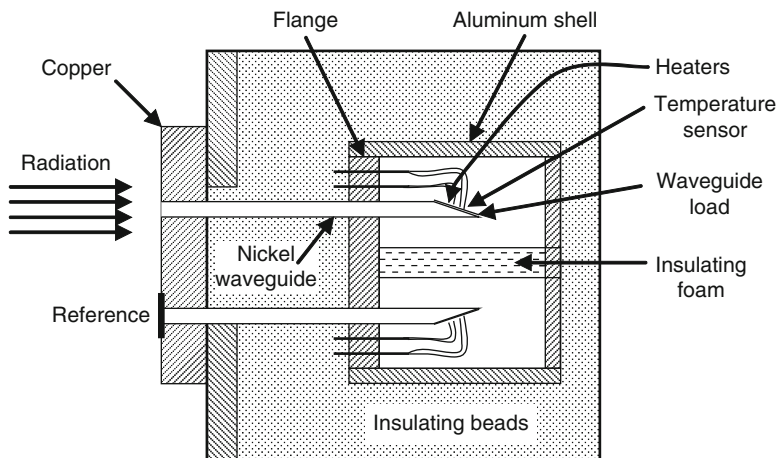


Fig. 5.11 Diagram of the power sensor head, showing the waveguide loads, heaters and temperature sensors (adapted from [440])

of limited use for absolute power measurements if accuracy better than a few percent is required. This is mainly because the responsivity of a particular device is usually not known with the required accuracy.

Good results can be achieved with calorimetric power meters. The basic design of a THz calorimeter [438] is based on the twin-load type of calorimeter developed for millimeter wavelengths [439]. Figure 5.11 shows a schematic design. Two nominally identical waveguides are mounted in a chamber that is thermally well insulated. Both waveguides have thin walls in order to minimize the heat flow. They are terminated with absorbers which can be electrically heated. Often the absorbers are specially shaped, for example, in the form of a pyramid, in order to increase the absorption. For the measurement the THz signal is coupled into one waveguide, while the absorber of the other waveguide is electrically heated until the temperature difference between the two loads vanishes. With some corrections, the electrically dissipated power is then equal to the THz power. Both the electrical power and the temperature difference can be easily measured. Differences in the distribution of electrical and THz power in the absorbers can lead to contradictory temperature rises at the measurement points, although the same amounts of power are absorbed. This so-called equivalence error needs to be minimized by precise thermal configuration of the absorber, including the electrical heating and the temperature sensors. With careful design this error can be as low as 0.2% [441]. Another potential source of error is the waveguide attenuation, which has to be determined for each frequency. Overmoding is less of a problem because the absorbers can be made very broadband and the matching is relatively independent of the mode. The asymmetry between both waveguides and absorbers can be taken into account by changing the function of both waveguides. State-of-the-art calorimeters achieve an uncertainty of $\sim 3\%$. They can accurately measure power from $1\ \mu\text{W}$ up to $200\ \text{mW}$ at frequencies up to

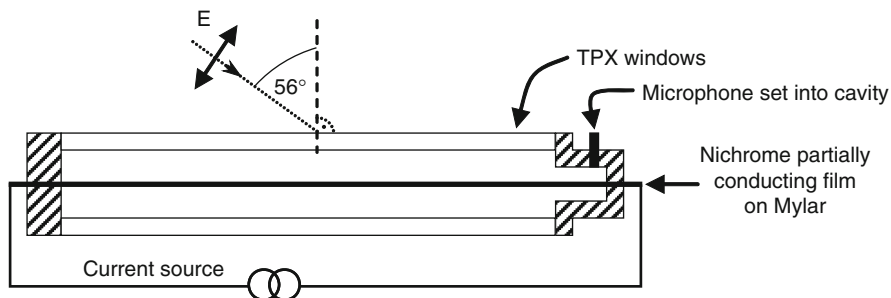


Fig. 5.12 Thomas Keating Ltd. power meter (adapted from [442]). Accurate measurements require polarized radiation at the Brewster angle 55.5°

2 THz and possibly above. The rms noise is $0.15 \mu\text{W}$, with a time constant of 1 s [440, 441].

Another type of power meter is based on the photoacoustic effect (Fig. 5.12) [442]. Two closely spaced parallel windows made from TPX form a sealed gas cell. A thin metal film which absorbs a known fraction of the incident THz radiation is placed between them. Chopping the incident radiation leads to a modulation of the film's temperature, which in turn induces a modulation of the pressure in the cell. This modulation is detected by a pressure transducer (microphone) and measured with a lock-in amplifier. The modulated pressure change is closely related to the total absorbed power. Assuming that ohmic heating of the metal film and radiation heating are equivalent, the power meter can be calibrated by passing a modulated current through it to heat the film. The TPX windows are characterized with regard to their absorption and reflection losses as a function of frequency. These losses are taken into account via a look-up table in the software of the power meter. For practical measurements, the power meter needs to be aligned so that the beam is incident at the Brewster angle, with the plane of polarization in the plane of incidence. If the incoming radiation is not linearly polarized, the two orthogonal linearly polarized components need to be measured separately, for example by using a wire grid. Power which is not absorbed in the metal film is partly transmitted or reflected. Parameters which affect the accuracy of the measurement are air–pressure fluctuations at the reference frequency, IR and visible light which fall onto the power meter, and deviation from square wave modulation. A power meter of this type is commercially available and calibrated for frequencies between 30 GHz and 3 THz (Thomas Keating Ltd.) but is also usable at higher frequencies. Such a power meter has a typical NEP of $5 \mu\text{W}/\sqrt{\text{Hz}}$, and this allows measurements down to the $10 \mu\text{W}$ level. The maximum power is 200 mW.

A further method of measuring power in the THz region is to use detection based on the photon drag effect, which was first observed independently in the USSR [443] and the UK [444], using pulsed and Q-switched CO_2 lasers at $10 \mu\text{m}$. Photon drag is a simple manifestation of photon pressure, because it generates an electric

field in a semiconductor by transferring momentum from an incident light beam to charge carriers. Unless saturation occurs the magnitude of the electron field is a linear function of the light intensity. It is of interest to note that the observation of photon drag has ignited fresh interest in a controversy as to what is the momentum of light in a dielectric, which began in the early years of the twentieth century and still continues [445].

A typical photon drag device consists of a cylindrical or rectanguloid rod of semiconductor material, with sufficient doping to absorb radiation over the length of the rod at the wavelength of interest. When light is directed onto the end face of the rod, an electric field is induced, which can be observed as a voltage if electrodes are placed at either end of the device. Normally, the detector has a length to diameter ratio sufficient to ensure a uniform field, regardless of where the light beam strikes the end of the rod.

At short IR wavelengths, where $h\nu > k_B T$ at room temperature, there are competing absorption processes within semiconductors that can lead to significant variations in the voltage produced, even over a quite short wavelength range. For example, the photon drag signal from p-type Ge varies by more than a factor of three between 9 and 11 μm . But over much of the THz region where $h\nu < k_B T$, and the only source of absorption is free carriers, the voltage produced can be predicted from the detector parameters. This led to the suggestion that photon drag detection could provide calibration devices at THz wavelengths [446]. The main disadvantage of photon drag detectors is their low responsivity, typically $<1 \mu\text{V/W}$ at 10 μm , but when using laser sources there is normally quite sufficient power for a good (S/N)-ratio. Their advantages are room temperature operation and the ability to give a linear response up to high intensities. At 9.2 μm linearity up to 30 MW/cm^2 has been reported [447]. A further asset is their speed of response, which is less than 1 ns.

Photon drag detection has been discussed in considerable detail by Gibson and Kimmitt, with a special section on "Calibration Standards for 100–2,000 μm " [448]. Voltage responsivities are much higher than at 10 μm . For example, an n-type Ge detector with a receiving area of 16 mm^2 has a responsivity of $\sim 40 \mu\text{V/W}$ at 0.3 THz, but this falls rapidly above 0.6 THz. For a more uniform response in the THz region p-type Si is a better alternative, with a relatively uniform response of $6 \mu\text{V/W}$ up to 2 THz. However, it is important to note that, by a suitable choice of semiconductor, good response can be obtained over the entire IR and THz region. Although primarily of importance for pulsed or Q-switched sources, detectors can also be used for power measurement with cw sources. For example, the power output of an HCN laser at 337 μm has been measured at signal levels below 0.2 mW [449].

Attractive features of these detectors are their ease of manufacture and their versatility. A useful feature is that for higher power sources it is simple to make larger detectors, thus avoiding surface damage or any possibility of nonlinearity. Online monitors are commercially available for CO_2 lasers that absorb 10–20% of the source energy and give a continuous reading of the laser power. Similar devices have been used in the THz region. However, these 10- μm monitors are AR coated to avoid unwanted reflections, and this is difficult to achieve over a wide

wavelength range. With polarized sources, devices with Brewster angled end faces can be envisaged but, with the high refractive index of most semiconductors, this could mean inconveniently long structures.

5.3.5 Semiconducting Bolometers

Semiconducting bolometers are among the most important of THz detectors. The first bolometer of this type was made from Ge. It was invented by F. Low in 1961 [66]. In 1971, M.A. Kinch introduced the Si bolometer [450]. These rely on free-carrier absorption by electrons closely coupled to the lattice. A design for such a bolometer is shown in Fig. 5.13. It consists of a small ($\leq 1 \text{ mm}^3$) chip of doped semiconductor, usually Si or Ge. The detector element is suspended in vacuum by two thin lead wires between the electrical contacts, which provide the electrical connection as well as the thermal link to the heat sink.

The optimum level of doping is determined by two requirements. On the one hand, the temperature coefficient of the resistance should be large while, on the other hand, the bolometer should have a resistance which allows for an efficient coupling to a low noise amplifier. Since the resistance of intrinsic or lightly doped extrinsic material is large at low temperatures, the semiconductor must be heavily doped, that is, close to the metal–insulator transition, with a majority donor impurity and a compensating minority acceptor impurity. Electrons from the donors fill some acceptor states and, in the presence of an electric field, the dominant conduction mechanism is hopping of carriers from one dopant atom to another. It is worth noting that the temperature coefficient and the resistance depend strongly on the degree of compensation, because the hopping conduction mechanism is determined by the distance between dopant atoms. Typical doping concentrations are 10^{16} cm^{-3} for Ge and 10^{18} cm^{-3} for Si. The hopping conduction mechanism gives rise to a temperature-dependent resistance of

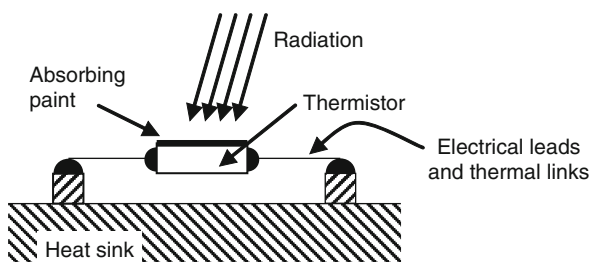


Fig. 5.13 Simple design of a semiconducting bolometer. The thermistor is suspended between electrical contacts by very thin electrical leads which at the same time serve as thermal links to the heat sink

$$R(T) = R_0 \exp \left(\sqrt{\frac{T_0}{T}} \right). \quad (5.34)$$

The constants T_0 and R_0 are of the order of 2–10 K and 0.1–0.5 Ω . For Ge thermistors, excellent agreement between (5.34) and experiment was found [451, 452]. The corresponding temperature coefficient of the resistance is

$$\alpha(T) = \frac{1}{R} \frac{dR}{dT} = -\frac{1}{2} \sqrt{\frac{T_0}{T^3}}. \quad (5.35)$$

Note that α is negative and that it has strong temperature dependence. This is in contrast to superconducting bolometers, which have positive α allowing for electrothermal feedback (Sect. 5.3.6).

As the doping is a crucial parameter of the bolometer, much development has been devoted to optimizing it. Semiconductors doped in the melt have variations in the acceptor and donor concentration, which are often unacceptably large, meaning that there are large variations of the temperature coefficient and resistance in a single crystal. Neutron transmutation doping (NTD) of ultra-pure Ge yields Ga, As, and Se. The weak absorption of neutrons by Ge leads to a uniform doping concentration. For Ge with natural isotopic abundances, NTD-Ge has a compensation of 0.32. Therefore, R_0 and T_0 can be changed independently only by changing the dimensions of the crystal or by isotopically enriching Ge. Special care has to be taken with respect to the contacts. Soldered In contacts, which were often used for Ge bolometers in the early days, introduce excess noise ($1/f$ -noise). Metallized ion-implanted and annealed contacts give the best results.

In some bolometers, the semiconductor chip combines the functions of radiation absorption and thermometry (Fig. 5.13). However, there is a significant surface reflectivity for Ge ($\sim 36\%$) as well as for Si ($\sim 30\%$), which decreases the quantum efficiency. Antireflection coatings are possible but rarely used because they are not sufficiently broadband. The absorption can be improved by an appropriate paint on the surface of the semiconductor chip. By this means the detector can be made uniformly absorbing throughout the THz frequency range and even beyond. The disadvantage is that the paint contributes to an excessive heat capacitance. This problem can be overcome by using a composite bolometer in which the functions of absorption and thermometry are separated (Fig. 5.14) [70, 453]. In this case the semiconductor chip is bonded or glued to a thin plate of sapphire or diamond with an absorbing paint. Since the heat capacitance of sapphire is about 8% of Si (2% of Ge) and the heat capacitance of diamond is ~ 5 times less, the absorber can be made quite large (typically 1–10 mm²) without affecting the thermal time constant of the bolometer. Absorbing paint has a relatively large heat capacitance and if applied to a large absorber it would degrade a bolometer's performance. Instead, a thin metallic layer of, for example, a nickel–chromium alloy or bismuth, is deposited onto the back surface of the absorbing plate. A frequency independent absorbance can be achieved if the thickness of the layer is adjusted to yield a

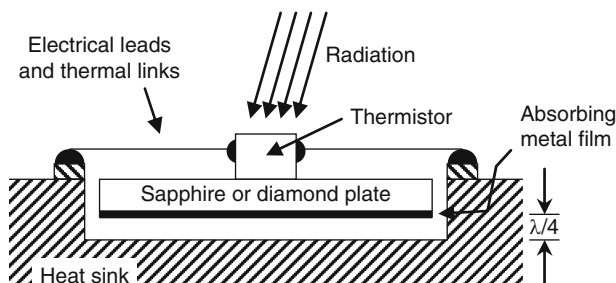


Fig. 5.14 Scheme of a composite bolometer. The thermistor, which is glued to a plate of sapphire or diamond, senses any temperature change of the plate. A thin layer of metal which matches the impedance of free space absorbs the incoming radiation. The thermistor has typically a size of 0.25 mm^2 , while the crystal plate can have dimensions of a few millimeters

resistance of $377/(n - 1) \Omega$ per square (n : refractive index of the absorbing plate, vacuum impedance: 377Ω). In this case, the impedance of the absorbing plate is matched to the combination of the film and free space [453] and no reflection occurs at the surface. This is necessary in order to avoid interference fringes and to achieve a frequency independent absorbance. As an example, for a diamond absorber ($n = 2.37$) covered by a metal film with a sheet resistance of 275Ω , the absorbance for normal incidence is $4(n - 1)/(n + 1)^2 = 0.48$. Larger values can be obtained over a limited frequency range by using a metal film with lower sheet resistance, which produces interference fringes. Placing the bolometer in an integrating cavity also increases the absorption. An additional advantage of the composite bolometer is its larger effective area, to which the focal spot of the THz radiation can be more easily adapted.

Coupling of radiation to the bolometer is usually done with a light collecting cone, either one of a Winston type or a straight one (Sect. 3.10). Since the heat capacitance of a bolometer scales with its area, small bolometers are preferable. In order to obtain a large throughput, bolometers are usually illuminated with a wide solid angle. Because the absorbance of bolometers can be significantly less than one, they are often placed in an integrating cavity in order to enhance the absorption. The amplifier should not contribute significantly to a bolometer's NEP. For heat sink temperatures down to 0.1 K this can be achieved by amplifiers with junction field effect transistors at 300 K temperature. However, wires which connect the bolometer to the amplifier are sources of microphonic noise. Movements of the wires lead to changes in the capacitance, which in turn give rise to an additional current and a noise voltage across the bolometer that is proportional to the current and resistance of the bolometer. This can be significant because bolometers are high resistance devices. Furthermore RF pick-up by the wires leads to additional noise. The first amplifier stage is usually mounted close to the bolometer inside the cryostat to overcome these problems.

Bolometers are commercially available either as single (composite) devices or in small arrays. The geometry, size, operating temperature, time constant, and

thermal conductance can be adjusted by the manufacturer to meet the specific requirements of the experiment. The typical performance of a composite Si bolometer operated at 4.2 K is: $\text{NEP} = 10^{-13} - 10^{-12} \text{ W}/\sqrt{\text{Hz}}$, frequency coverage 0.15–15 THz, depending on filters, and a response time down to 10 μs . A notable application of semiconducting bolometers is for astronomy with SPIRE, the Spectral and Photometric Imaging Receiver, aboard the Herschel Space Observatory. This is based on bolometric detector arrays operating around 0.6, 0.8, and 1.2 THz. Each individual bolometer consists of an Si-nitride membrane with metallic micromesh serving as the absorber. This provides a significant reduction (~ 20 times) in the heat capacitance as well as a reduced cross-section for cosmic rays [454]. It is worth noting that by choosing an appropriate mesh this type of bolometer can be made polarization sensitive. The NTD-Ge thermistor is placed on one side of the absorber. With a prototype array, an NEP of $5 \times 10^{-17} \text{ W}/\sqrt{\text{Hz}}$ and a time constant of 5–6 ms has been demonstrated [455].

5.3.6 Superconducting Bolometers

The transition from the normal to the superconducting state can take place in a very narrow temperature range (Fig. 5.15). This transition can be used for a very sensitive thermometer. The earliest bolometers based on superconductors date back to the middle of the past century, when Andrews and his co-workers used tantalum and niobium nitride [34, 35]. Since that time, progress in superconducting material

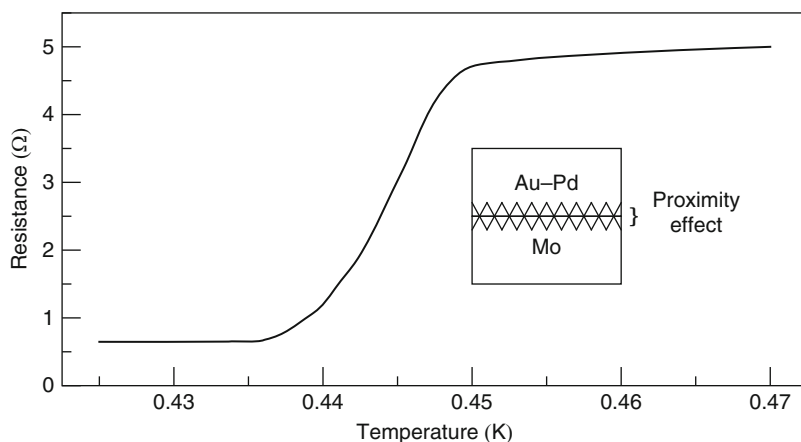


Fig. 5.15 Transition from the normal to the superconducting state for an Au-Pd/Mo bilayer. An Au-Pd layer of ~ 8 nm is deposited on top of a ~ 60 nm Mo layer. The transition takes place in a very narrow temperature range of about 0.01 K. Due to the proximity effect at the interface between both layers, it is possible to adjust the critical temperature of the Mo layer by sputtering of a Au-Pd alloy (adapted from [456])

development, as well as micro- and nanofabrication technologies, has made a wide variety of superconducting bolometers possible. Nowadays superconducting metal thin films are used for bolometers. They can be fabricated by deposition techniques such as magnetron sputtering or laser ablation. Structuring of these films can be done by optical or e-beam lithography. In addition, by choosing the appropriate material or material combination, as well as by controlling the metallurgy of the metal, a wide range of temperature coefficients can be obtained.

Although superconducting bolometers have been around for a long time, they were not widely used, due to the lack of comparable performance to semiconductor bolometers, and their complexity of operation. One major disadvantage was caused by the constant current biasing scheme, which was employed for a long time. In superconducting bolometers, the temperature coefficient of the resistance α is positive. This implies that, for a current-biased bolometer, the power which is dissipated by a constant current bias source increases with increasing temperature. If the current exceeds the limit $G_{\text{eff}} = G - I^2 R \alpha = 0$, thermal runaway occurs. This limits the useful bias range and the linearity of the detector.

Thermal runaway can be avoided by an appropriate feedback technique, which can be obtained by using a constant bias voltage. If the power absorbed by the bolometer increases, its temperature and resistance also increase. For constant voltage bias, the current decreases. For a steep superconducting transition, the decrease in bias power compensates for the increase in radiation power. This process is called electrothermal feedback, and a device based on this scheme is called a voltage-biased superconducting bolometer (VSB). The stable self-biasing yields an increased linearity and dynamic range compared to a current-biased bolometer. Another advantage is a greatly reduced response time. The thermal conductance of a superconductor can be described by

$$G = nKT^{n-1}, \quad (5.36)$$

where K is a material and geometry-dependent parameter and n depends on the thermal impedance between the film and the substrate. n is 4 if the Kapitza resistance dominates, and n is 5 or 6 if electron–phonon decoupling in the film dominates. When the temperature of the substrate is well below that of the film, the equilibrium temperature is $P_E = GT/n$. Taking these equations, one obtains from (5.26)

$$\tau_{\text{eff}} = \frac{\tau_0}{1 + \alpha T/n}, \quad (5.37)$$

where τ_0 is the intrinsic time constant of the film, the response time in the absence of bias heating. While τ_0 is usually determined by the design of the thermal coupling to the heat sink, its ultimate limit depends on the processes in the superconductor, for example electron–electron interaction, electron–phonon coupling, and phonon escape from the superconductor into the substrate. Typical response times of a VSB are in the order of a millisecond to hundreds of milliseconds. In special cases,

response times well below 1 ns have been achieved. Since α is essentially the inverse of the width of the superconducting transition, superconducting materials with a steep transition are preferred, which is why bolometers of this type are often referred to as transition edge sensors (TES). The αT product can be as high as 1,000. Therefore, the response time of a bolometer with electrothermal feedback can be two orders of magnitude faster than that of a bolometer without feedback. At first glance, it might be surprising that the response time can be less than the intrinsic time constant of the film. However, with voltage bias and electrothermal feedback, the energy is removed by a reduction of bias heating, instead of by an increased heatflow to the substrate, which is a slower process.

For the operation of a VSB, the bath temperature T_{Bath} is chosen to be below the critical temperature of the superconductor. The optimum is a factor of 2 below the bath temperature [457]. The voltage bias heats the bolometer to the critical temperature T_C according to

$$P_E = \frac{V_B^2}{R} = G(T_C - T_{\text{Bath}}). \quad (5.38)$$

If radiation heats the bolometer, its resistance increases. Due to the constant bias voltage, the electrical power P_E dissipated in the bolometer is lower and its temperature drops. This feedback keeps the operating temperature of the bolometer nearly constant.

Superconducting bolometers have a low impedance. Superconducting quantum interference device (SQUID)-based amplifiers are well suited for readout of their signals. A SQUID is a superconducting loop which contains two tunnel junctions, each with a resistor and a capacitance in parallel. When an external magnetic flux passes through the loop a change of its current–voltage curve is detected. Figure 5.16 shows a simplified SQUID readout scheme. The current to be measured passes

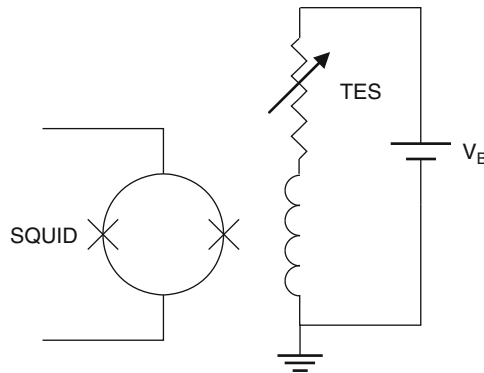


Fig. 5.16 Electric circuit of a voltage biased transition edge sensor (TES) with SQUID readout

through a coil which is in series with the VSB. A SQUID in close proximity detects the magnetic flux generated by the coil. Changes of the magnetic flux result in current or voltage changes at the output of the SQUID. Since SQUIDs are low noise devices ($<1 \text{ pA}/\sqrt{\text{Hz}}$), and can be easily lithographed during the processing of the VSB, they are commonly used amplifiers for VSBs.

Many designs and materials have been developed for superconducting bolometers. In general, these are composite bolometers. The superconductor serves as a thermometer which is attached to an absorber. The first prototype VSB was made from a 40-nm thin tungsten film with a T_C of about 95 mK. The bolometer was etched to a rectangle ($1.8 \times 0.9 \text{ mm}^2$) on a $300 \mu\text{m}$ thick Si substrate. The readout was performed with a coil in series with the bolometer and SQUID. An electrical NEP of $3.3 \times 10^{-17} \text{ W}/\sqrt{\text{Hz}}$ and a time constant of $10 \mu\text{s}$ were measured. The latter value is about 100 times faster than the intrinsic time constant [458]. In another approach, a fully lithographed VSB was fabricated from thin metal layers (for example Au/Ti/Au or Au-Pd/Mo). By this means, the superconducting transition temperature could be adjusted using the proximity effect to somewhat above the sink temperature of a ^3He sorption refrigerator. The VSB is lithographed onto a Si_3N_4 membrane and located in the center of a spider web [459] or similar structure [456]. The fully lithographed fabrication is a major advantage of superconducting VSBs compared with semiconducting bolometers. Monolithic arrays of VSBs with thousands of pixels, a filling factor close to one (i.e. the focal plane can be Nyquist sampled), and an NEP of $\sim 10^{-17} \text{ W}/\sqrt{\text{Hz}}$ have been demonstrated [460, 461] or are under development for astronomical applications [456, 462]. A major challenge for large arrays is the readout at cryogenic temperatures. SQUID arrays with either time-domain multiplexing [463] or frequency-domain multiplexing [464] offer a solution. Whilst astronomical background limited observations require this low NEP, the situation is different for laboratory applications, for example in a Fourier-transform spectrometer, where the background radiation is much higher. In this case, desirable features are operation in a standard liquid helium cryostat, large dynamic range to detect the central peak as well as the outer parts of an interferogram, and a speed sufficient for rapid-scan spectroscopy. This has been achieved with an Nb film thermometer ($T_C \approx 8 \text{ K}$) on an Si_3N_4 membrane, which is covered on the rear side by a metal absorber [465]. A particular application of a superconducting TES is time resolved detection of short THz pulses from a synchrotron. In this application speed, with a response time well below 1 ns, is the main requirement. A thin film (4–6 nm) NbN TES embedded in a planar antenna serves well for this purpose. Due to the material properties of the NbN film, and its small dimensions, it has a fast response time of about 40 ps [466]. This allows detection of coherent THz synchrotron radiation and the resolving of single pulses (Sect. 4.12.1) [400]. It is worth noting that in this case the response time is limited by the material properties, and that voltage biasing does not improve the speed of the device because its time constant is larger than the response time of the bolometer. Due to their fast response, these bolometers are also excellent mixers in a heterodyne system (Sect. 5.5.4).

5.3.7 Room Temperature Microbolometers

The limitations due to the thermal mass mentioned in Sect. 5.3 resulted in there being relatively little interest in room temperature thermal detectors for some years. However, it was pointed out, in a theoretical paper toward the end of the 1970s, that if detectors were made with very small volume this would overcome the limitations of larger devices [467]. But for use in the THz spectral range, where the detector dimensions would be much too small compared with the radiation's wavelength, an antenna would be required to couple-in the light. Following this suggestion, a novel room-temperature detector for the wavelength range 10–1,000 μm was constructed. This consisted of a $5 \times 4 \mu\text{m}^2$ area bismuth film with a thickness of 55 nm. Bismuth has a relatively large temperature coefficient of resistance. At low modulation frequencies, a value of $\sim 1.6 \times 10^{-10} \text{ W}/\sqrt{\text{Hz}}$ was obtained at 119 μm and, due to its tiny size, the device remained within a factor of 10 of this value up to modulation frequencies of 25 MHz [468].

In a later paper, an array of 400 bismuth-film microbolometers with a total area of 1 cm^2 has been described [469]. Each detector had its own single mode antenna, and the motivation for designing such an array was to produce a detector that would give improved detectivity for distributed sources such as hot plasmas, or for terrestrial radiometric mapping, and for cosmic-background measurements. With a point source, there is no great advantage in using an array. But with an extended source, each detector antenna sees a different part of the source and the signal increases in proportion to the number of detectors employed. Because each detector element employed was very small, with an area of $3 \mu\text{m} \times 3 \mu\text{m}$, a response time of $\sim 200 \text{ ns}$ was obtained, with a D^* of $4 \times 10^8 \text{ cm}\sqrt{\text{Hz}}/\text{W}$ in the frequency range 0.1–0.3 THz.

Although the bismuth-film device consisted of many individual detectors, these were coupled together to make a single receiver. But the requirement for sensitive thermal imagers for both military and civil applications, coupled with the advent of CCD technology, led to renewed interest in very small detector elements. The initial interest was in arrays of photodetectors. But for the important 8 to 14 μm region, which corresponds both to the peak emission region for objects close to room temperature and to an atmospheric window, photodetectors need to be cooled to $\sim 120 \text{ K}$ or less.

The most obvious detectors for room-temperature arrays were pyroelectrics (Sect. 5.3.2) in a hybrid arrangement [470], but it was recognized that microbolometers could also be competitive. In the late 1980s Honeywell began research on vanadium oxide microbolometer array technology. Although this type of array is primarily designed for use in the 8–14 μm atmospheric window, some arrays operate reasonably well at THz frequencies [471].

A typical design for a single microbolometer is shown schematically in Fig. 5.17. The bolometer is in the form of a very thin (50–200 nm) membrane. This only absorbs a percentage of the light, but the underside of the membrane and a mirror deposited on the upper surface of the substrate wafer form a resonant cavity for any

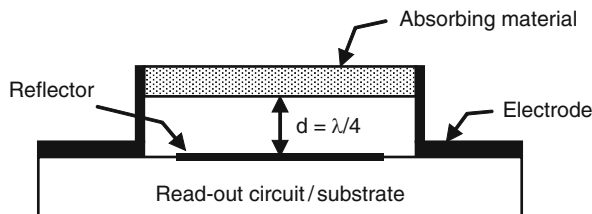


Fig. 5.17 Concept of a microbolometer

radiation transmitted through the membrane. For the 8–14 μm region, the cavity spacing is typically 2–2.5 μm . In some detectors, any loss of light is further reduced by adding an antireflection layer to the upper surface of the membrane.

A number of different materials have been employed for the detecting membrane, but those most favored for the 8–14 μm atmospheric window are amorphous Si and oxides of vanadium, of which vanadium pentoxide (V_2O_5) appears to be the most widely used. Typical pixel dimensions vary in area between $50 \times 50 \mu\text{m}^2$, with 320×240 pixels in the array, to $25 \times 25 \mu\text{m}^2$ with 640×480 pixels. However, larger pixel numbers of 1024×760 are envisaged [472]. Microbolometer arrays optimized for operation in the THz range are currently under development [473] and commercially available.

Clearly, detectors designed for 8–14 μm are not ideal for the THz region but the use of a 160×120 pixel commercially available camera with vanadium oxide microbolometers, spaced at 46.25 μm , was used for real-time imaging experiments at 2.5 and 4.9 THz, using an optically excited gas laser or a quantum cascade laser (Sect. 7.4.4). With this, good transmission images were obtained. A somewhat similar experiment, but with a 2.8 THz quantum cascade laser, has been reported [474]. Apart from substituting a Picarin lens for its antireflection-coated Ge lens, the thermal-imaging camera was not modified. However, the modified camera could only detect temperature changes of approximately 3 K over the 1–5 THz region, which appears very inferior to the unmodified camera's 0.1 K over the 8–14 μm wavelength range. Nevertheless, as the integrated power available over 8–14 μm is some 15 times greater than that at 1–5 THz, this represents an acceptable performance in the THz region.

5.4 Photoconductive Detectors

The essential difference between thermal detectors and photodetectors is that in the former the incoming energy produces a measurable temperature change, while in the latter the energy of the photons interacts with electrons within the material. In photoconductive detectors, as the name implies, a measurable resistance change is produced and, if a current is passed through the detector element, this can be measured.

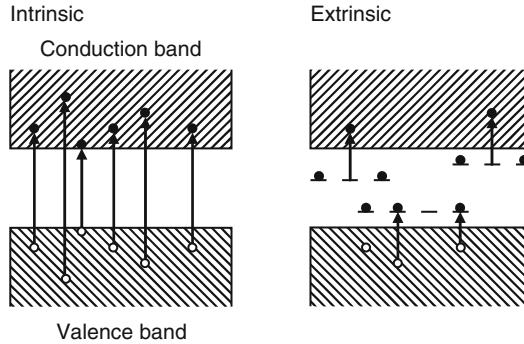


Fig. 5.18 Concept of intrinsic and extrinsic detection

The two main photoconductive processes are illustrated in Fig. 5.18. In a semiconductor there is a gap between the valence and conduction band. If a photon with energy greater than the gap energy impinges on the semiconductor, an electron can be excited from the valence to the conduction band, producing an extra electron. At the same time, a hole is created in the valence band and both the electron and hole will cause increased conductivity. This is termed intrinsic excitation. If the photon has not sufficient energy to overcome the gap energy, it is still possible to cause a photoconductive process by adding impurities to the semiconductor. As illustrated, these can produce either donor or acceptor states, which can be quite close to the conduction or valence band, and a low energy photon can excite an electron out of a donor state or into an acceptor state. This is extrinsic excitation and is the process employed for THz photodetection, except at the very highest frequencies, close to 10 THz, where intrinsic detectors are available.

The parameters that affect the performance of photoconductive detectors can be easily seen from quite elementary considerations. Because almost all THz photoconductors are extrinsic, it is convenient to assume that only one type of carrier, either electrons or holes, is involved. Other assumptions on which the following arguments are based are that each absorbed photon produces exactly one free carrier in the conduction or valence band, the impurity atoms are distributed uniformly throughout the host crystal, there are no trapping effects of charge carriers, and the crystal has ohmic contacts.

The first studies of extrinsic photoconductivity were made in the early 1950s [475]. The energy levels of many impurities in Si and Ge were already well documented from earlier research, leading to the development of the transistor. Although it was known that in Ge a number of impurities were at levels close to the conduction or valence bands, the main goal at that time was to find sensitive detectors for the thermal imaging atmospheric window at 8–13 μm . It was in the late 1950s and 1960s that applications for lower frequency detectors arose and research on very shallow level extrinsic photoconductors began [54]. Initially, this research concentrated on Ge and Si but was later extended to InSb and GaAs.

The shallowest impurity centers in the Group IV semiconductors Ge or Si are those associated with impurities from Group V, for example P, As, Sb or Group III, e.g. B, Ga, or In. In these atoms, the number of electrons in the outermost shell differs by one from the four in Ge or Si, and the extra electron (or hole) is balanced by an extra unit of charge on the atomic core. This system of free carrier plus fixed core charge can be treated as a hydrogen atom embedded in the dielectric medium of the Ge or Si lattice, giving rise to localized electronic states within the band gap. The effective mass approximation can then be used for substitutional impurities. An atom that lacks one valence electron with respect to the host lattice can easily attract an electron from the valence band, leaving a mobile hole, and therefore acquiring a negative charge. This mobile hole moves in the coulombic potential analogous to the hydrogen atom, except that in the immediate vicinity of the origin the potential departs from a simple coulombic form. The energy levels of the excited states indexed by n , neglecting core effects, are for a spherical band

$$E_{\text{IMP}}(n) = -\frac{R_{\text{H}}^*}{n^2} \quad R_{\text{H}}^* = \frac{m^* e^4}{8(\epsilon_r \epsilon_0)^2 \hbar^2} = \frac{1}{\epsilon_r^2} \frac{m^*}{m_e} R_{\text{H}}. \quad (5.39)$$

The effective Rydberg energy R_{H}^* only deviates from the Rydberg energy $R_{\text{H}} = 13.6 \text{ eV}$ of the hydrogen atom by a factor due to the density of states, effective mass of the valence band m^* , and the relative dielectric constant ϵ_r of the host material. The effective Rydberg energy for heavy holes in Ge is determined as 15 meV, using $\epsilon_r = 16$ and $m^*/m_e = 0.28$, for an approximation. The effective Bohr radius r_{B}^* follows as

$$r_{\text{B}}^*(n) = \frac{4\pi \epsilon_r \epsilon_0 \hbar^2}{m^* e^2} = \epsilon_r \frac{m_e}{m^*} r_{\text{B}}. \quad (5.40)$$

According to (5.40), the hydrogen Bohr radius $r_{\text{B}} = 0.0529 \text{ nm}$ is modified by the effective mass and the relative dielectric constant of the host material, and an effective Bohr radius of 3 nm is obtained. A more accurate treatment gives an ionization energy of 11.2 meV, using screened potentials and a proper weighting to obtain the hole effective mass [476, 477]. This value for the ionization energy is in good agreement with measured values for B, Al, Ga in Ge, but there is significant variation for heavier impurities or nonhydrogenic acceptors and donors. This is due to the influence of the atomic core on the wave function of the ground state. In Si, where the effective mass of the free carriers is heavier, the frequency equivalent to the ionization energy is above 10 THz, but it is worth noting that extrinsic Si detectors are important devices for the 10–60 THz region. Inspection of (5.39) and (5.40) shows that to detect lower frequency radiation m^* must be reduced. However, this leads to a larger Bohr radius and the consequences of this are important in InSb (Sect. 5.4.2) and GaAs (Sect. 5.4.3).

The minimum energy for detection predicted by (5.39), assuming an n-type impurity, is that from the donor center to the conduction band. It is only observable

at low temperatures when a considerable fraction of the impurity centers is not ionized. The system is comparable to that of a hydrogen atom with a series of energy levels between the ground state and the conduction band. Excitation from the ground state to the excited states of an impurity center leads to a line absorption spectrum at lower frequencies than those for the continuous absorption process, due to the donor states to conduction band transitions. This line absorption process is normally much weaker, but in an extrinsic device it provides a detection process for specific frequencies below the normal absorption edge. With p-type impurities a similar line spectrum occurs, due to levels between the acceptor state and the valence band. A specific example is given in Sect. 5.4.1.

The requirement for extrinsic conduction to occur is

$$h\nu \geq E_{\text{IMP}}. \quad (5.41)$$

The detector response rises steeply above the low-frequency limit ν_0 for excitation, reaching a maximum at about $1.2\nu_0$, and falls off at higher frequencies with approximately $(h\nu - E_{\text{IMP}})^{-2/3}$. Essentially, this reflects the absorption cross-section from the ground state into the conduction or valence band. At THz frequencies, E_{IMP} is very small. At 3 THz, for example, $E_{\text{IMP}} \approx 11$ meV, and detectors need to be cooled to 4 K or less to reduce thermal excitation.

Impurities will only behave as isolated donors or acceptors if their concentration is sufficiently small. In Ge, the value of the Bohr radius indicates that interaction effects should become appreciable between 10^{16} and 10^{17} cm^{-3} , but experimental results showed that it was necessary to keep the impurity concentration at much less than this. The reason is that hopping conduction, the tunneling of electrons from neutral states to neighboring ionized states, produces a current in the absence of a photon flux. This causes an unacceptable increase in conductivity for shallow hydrogenic impurities in Ge if the carrier concentration is above $2 \times 10^{14} \text{ cm}^{-3}$ [451]. As the absorption coefficient depends on the number of impurity centers, the optimum doping can depend on the particular application for a detector.

The design of good extrinsic photoconductors for THz frequencies is complex, requiring the optimization of a number of parameters. Only the more important features are included here. Assume that an extrinsic photoconductive detector is biased from a constant current source, i.e. $R_L \gg R_D$ as shown in Fig. 5.19.

When the detector is illuminated by photons with energy greater than E_{IMP} , the current will increase. This increase will depend on $\dot{N}_\nu$, the number of photons of frequency ν absorbed per second, the quantum efficiency η , the number of extra carriers produced per photon, and the photoconductive gain G_{PC} , which is the number of electrons that flow through the detector for each absorbed photon. This current increase is given by

$$I = \eta e \dot{N}_\nu G_{\text{PC}}. \quad (5.42)$$

Clearly, the quantum efficiency is less than one but, in theory, the gain can be above unity because, as long as an excited carrier does not recombine, electrons will continue to flow through the circuit. The photoconductive gain G_{PC} is

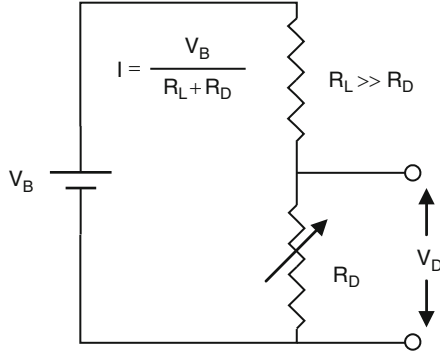


Fig. 5.19 Photoconductive detector biased from a constant current source. With $R_L \gg R_D$, the signal voltage is directly proportional to any change in the resistance of the detector when radiation falls upon it

given by

$$G_{PC} = \frac{\tau}{t_t}, \quad (5.43)$$

where τ is the lifetime of the excited carriers and t_t is the transit time between the electrodes. In intrinsic photoconductors, lifetimes can be very long, leading to extremely high gains, but in extrinsic devices lifetimes are short and the gain is normally less than one [478].

The transit time is governed by the carrier mobility μ , the electric field E across the detector, and the distance d between the electrodes, and is given by

$$d = \mu E t_t. \quad (5.44)$$

Combining this with (5.42) and (5.43) yields:

$$I = \frac{\eta e \dot{N}_v \tau \mu E}{d}. \quad (5.45)$$

The current responsivity is important for comparing the quality of detector material. It is given by

$$\mathcal{R} = \frac{I}{P_\omega} = \frac{\eta e \tau \mu E}{h \nu d} = \frac{\eta e}{h \nu} G_{PC}. \quad (5.46)$$

If speed of response is not an important parameter what is required is high mobility, long carrier lifetime with high breakdown electric field, and good quantum efficiency. Optimizing these parameters, particularly in the most important extrinsic detector for THz radiation, Ga-doped Ge, has produced very high detectivity devices. However, inspection of (5.46) shows that the responsivity is proportional to

the lifetime of the excited carriers. When detectors such as Ge:Ga are employed to observe short pulses, τ can be reduced by the addition of compensating impurities when growing the crystal material, but only at the expense of reduced responsivity. The detector is said to have a “gain-bandwidth” product because increasing one parameter reduces the other one. A similar interaction occurs in most detectors. For example, in bolometers τ can be increased by closer coupling to the heat sink, but only with a proportional reduction in responsivity, as shown by (5.20), (5.26), and (5.29).

The NEP of a background-limited photoconductor can be found to be [479]

$$\text{NEP}_{\text{BL}} = \sqrt{\frac{2(2\dot{N}_v)(h\nu)^2 B}{\eta}}. \quad (5.47)$$

The additional factor of $\sqrt{2}$ compared to (5.10) arises from the distribution in the mean free paths of the photo-generated carriers because of the fluctuating character of the recombination process. This results in a fluctuation of the signal pulse widths at the output of the detector. In the derivation of (5.10), it is tacitly assumed that all signal pulses have the same width. A figure-of-merit is the detective quantum efficiency η_D . It is defined as

$$\eta_D = \left(\frac{\text{NEP}_{\text{BL}}}{\text{NEP}} \right)^2. \quad (5.48)$$

The detective quantum efficiency is a measure of how far the performance of a photoconductor deviates from the ideal performance, for which $\eta_D = 1$. Note that η_D transforms all noise sources into a hypothetical loss in quantum efficiency, although the noise might originate from quite different sources.

In many applications, the voltage response of a detector is an important parameter. The voltage V generated by the detector, biased at constant current I is

$$V = I R_D. \quad (5.49)$$

The detector resistance R_D is given by

$$R_D = \frac{d}{e\mu A N_C}, \quad (5.50)$$

where N_C is the carrier concentration before illumination, d is the distance between the electrical contacts, and A the contact area which is given by the other detector dimensions.

Assuming that the resistance change is small when photons fall on the detector, the signal voltage obtained by combining (5.45), (5.49), and (5.50) is

$$V = \frac{\eta\tau E \dot{N}_v}{A N_C}. \quad (5.51)$$

Study of (5.50) and (5.51) indicates how the performance of a detector can be optimized. A high electric field is beneficial. The limitation is set by the breakdown electric field at which impact ionization of the impurities occurs. The carrier mobility and the carrier lifetime should be high. Since a high impurity concentration is necessary in order to obtain high quantum efficiency, and because scattering by neutral impurities limits the mobility, there is little room to increase the mobility. The lifetime can be maximized by reducing the number of compensating impurities of opposite charge. As mentioned earlier, it is essential that N_C , the number of excited carriers before illumination, be kept low and this is achieved by cooling the detector and reducing the background radiation with cold filters. However, it is important to realize that the detector resistance is inversely proportional to the number of carriers. As the important noise contributions increase as the square root of the resistance, the detectivity increases as $\sqrt{N_C}$. Reducing the dimensions of a detector, especially the inter-electrode distance, will increase the responsivity but the area obviously needs to be sufficient to receive the incoming radiation, and the thickness required depends on the absorption coefficient. Surrounding a small detector with an integrating metal cavity (Sect. 3.10) is advantageous, as this increases the absorption and allows the use of very small devices.

The quantum efficiency of extrinsic photoconductors is lower than that for intrinsic ones and the reasons for this are not fully understood. The definition of this parameter can cause some confusion. Strictly, it is the ratio of absorbed photons producing ionization to absorbed photons, but many authors use the term for the ratio of absorbed photons to the number of photons incident on the detector crystal, thus including the absorption coefficient and reflection losses. When these losses are included, the accurate term is “responsive quantum efficiency” [480], and this definition is used in the following sections.

5.4.1 *Extrinsic Germanium Detectors*

Although a number of shallow impurity levels in Ge have been investigated for THz detection, including Ge:Sb, Ge:As, Ge:P, Ge:In, Ge:B, Ge:Al, Ge:Ga, and Ge:Be, only the last two are widely employed. However, it is of interest that one of the blocked impurity band devices that is being studied (Sect. 5.4.3) for lower frequency detection is Ge:Sb, the very first of the THz photoconductors [54]. Ge:Ga has been the most widely studied and widely used THz photoconductive detector since its introduction in 1965 [68]. Single Ge:Ga detectors and detector arrays have been employed for THz astronomy since the 1980s. The maximum pixel numbers are 16×25 and 32×32 for stressed and unstressed arrays, respectively. In its conventional form, it has good response from about 2.4 to 7 THz, but it was later realized that its ionization energy could be reduced by applying a uniaxial compressive force, and in this mode the low-frequency limit is extended to close to 1.5 THz, but with lower response at high frequencies [69]. The responses are illustrated in Fig. 5.20. A very small response at lower frequencies (not shown in

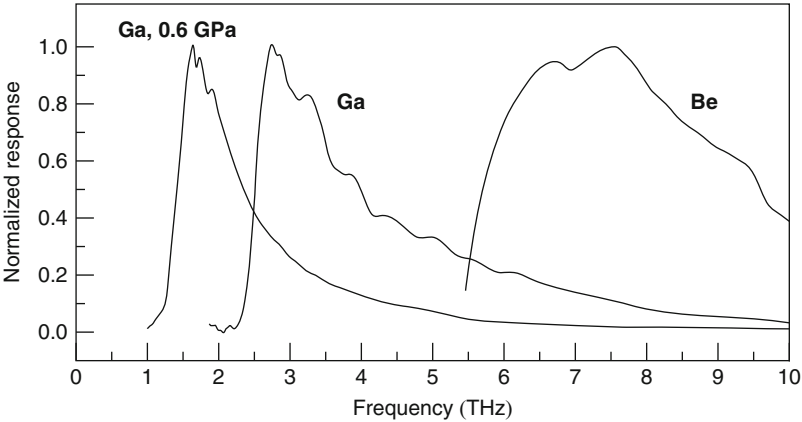


Fig. 5.20 Response of Ge extrinsic photodetectors doped with Ga and Be and under uniaxial stress at 0.6 GPa (adapted from various sources)

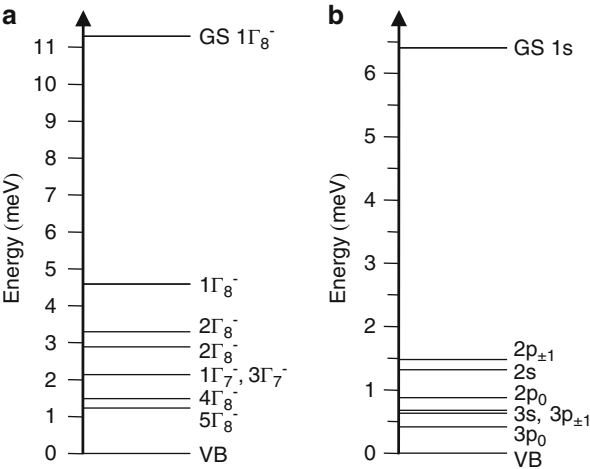


Fig. 5.21 Energy level diagram for unstressed (a) and stressed (b) Ge:Ga (VB: valence band, GS: ground state, states are labeled according to the literature)

Fig. 5.20) is observable, due to the other hydrogenic levels described in the previous section. The energy level diagram for unstressed Ge:Ga is shown in Fig. 5.21. It is worth noting that transitions from the ground state to excited states provide useful detection at lower frequencies than those corresponding to absorption from the ground state into the valence band. The requirements for a good extrinsic photoconductor are that it should have a high D^* value and that its response time can be varied to match the requirements of a particular experiment. It is the ability to control these parameters that has led to Ge:Ga becoming the preferred hydrogenic-type impurity for the THz region.

As shown by (5.46), the current responsivity of a photoconductor is proportional to the response time. This parameter is dependent on the degree of compensation of the detector material, the ratio of acceptors to donors. Donor states provide recombination centers, thus reducing the lifetime of the holes created by excitation of electrons from the valence band to the acceptor states. For the highest responsivity extremely pure starting material is required and, thanks to the work of E.E. Haller and his colleagues over many years at the University of California, Berkeley, Ge with less than 10^{10} cm^{-3} residual donors became available [451]. Material of similar quality has also been produced in Japan [481]. The particular advantage of doping with Ga, compared with other possible p-acceptors such as Al or B, is that it is easier to control both the amount of Ga and its uniformity throughout the crystal. The presence of other shallow level acceptors are not harmful to the performance of Ge:Ga photoconductors, as they have approximately the same ionization energy and absorption cross-sections as Ga. For many THz applications, Ge:Ga provides a fast detector with response times of 50–500 ns, dependent on its compensation, and such material is available commercially. The requirement for detectors with very low compensation and very low NEP values comes mainly from the astronomy community, where speed of response is a secondary consideration. At the other extreme, material with a response time of 3 ns has been used to study FEL pulses [482].

Equation (5.46) shows that the responsivity of a photoconductor increases with both mobility and electric field. While the mobility increases in low compensation Ge:Ga due to reduced scattering, the maximum electric field that can be applied is reduced because in pure material the carriers can pick up energy at a greater rate from the electric field, and this leads to impact ionization and voltage breakdown at lower applied fields. This is more than compensated for by an increase in allowed field from ~ 1 to 3–4 V/cm. The optimization of this type of detector is complex but a very remarkable performance has been achieved under low background conditions, with NEP values close to $10^{-17} \text{ W}/\sqrt{\text{Hz}}$ [483]. Commercial systems, typically used in higher background laboratories, reach values of 10^{-12} – $10^{-13} \text{ W}/\sqrt{\text{Hz}}$. The detective quantum efficiency of Ge:Ga photoconductors is in the range 0.1–0.2. This is much less than for intrinsic photoconductors. As mentioned earlier, the physical mechanism that is responsible for this is not clearly understood. Possibly trapping and re-emission of photoionized carriers in one of the bound excited states contributes additional noise and degrades the performance.

Application of uniaxial compressive force along the [100] axis of Ge:Ga crystals produces a reduction in the binding energy of the Ga acceptor states and, as also mentioned previously, extends the low-frequency response to about 1.5 THz. The required force leads to a pressure of $\sim 0.6 \text{ GPa}$ (6 kbar). The responsivity of the stressed Ge:Ga detector is approximately a factor of 10 larger than that of the unstressed one. A factor of 2–3 can be explained by the decrease in photon energy at smaller frequencies, and the rest is caused by an increased free carrier lifetime and mobility [484].

Electrical contacts are applied to two opposite faces of the detector. These should be mechanically stable at cryogenic temperatures and not add excess noise.

The latter requirement is not straightforward because impact ionization can occur relatively easily, due to inhomogeneities at the interface between the contact and the bulk Ge. Modern contacts are ion implanted and annealed. The implanted region is doped well above the metal-to-insulator transition concentration and metallized (for example with Pd/Au layers). By this means, a uniform metal to semiconductor transition is obtained. It is often convenient to use rather small detectors to achieve maximum responsivity and, as mentioned earlier, to optimize the photon absorption these can be placed in an integrating cavity. An alternative approach is to use the blocked impurity band devices that are described in Sect. 5.4.4.

The detectivity of Ge:Ga peaks at about 3 THz but falls quite rapidly at higher frequencies. For optimum response above 6 THz, Be-doped Ge photoconductors have been developed. The first studies on Ge:Be were produced in 1967 and in 1983 a detailed examination was made, when it was recognized as the most likely candidate for sensitive detection in the 6–10 THz region [485]. Beryllium is a double acceptor in Ge, with energy levels at 24.5 and 58 meV above the valence band. As with Ge:Ga, reducing compensation to very low percentages is required for optimum performance. Typical Be concentrations are between 0.5 and $1 \times 10^{15} \text{ cm}^{-3}$, with residual acceptor and donor concentrations as low as 10^{10} – 10^{11} cm^{-3} . Very high current responsivities of 40 A/W have been achieved at ~ 6.5 THz, with a quantum efficiency approaching 50%. Under reduced backgrounds, NEPs are close to $10^{-16} \text{ W}/\sqrt{\text{Hz}}$, giving a significant advantage over Ge:Ga in the 6–10 THz region. An alternative to Ge:Ga is antimony doped Ge, Ge:Sb. Sb is a donor impurity with a smaller energy gap below the conduction band than Ga above the valence band, leading to a lower cut-on frequency. Using material with antimony concentrations of 1.8 and $3.5 \times 10^{14} \text{ cm}^{-3}$, and very low compensation, detectors with almost identical performance to the best Ge:Ga in terms of responsivity, detector quantum efficiency, and dark current have been fabricated [486]. Although the extended range compared with Ge:Ga is not large, it covers significant frequencies, for example, 2.5 THz, where Ge:Sb has about a five times increase in detectivity compared with Ge:Ga.

5.4.2 Indium Antimonide Detectors

Stressed Ge:Ga reaching to about 1.5 THz represents the lower frequency limit of conventional extrinsic detectors based on Si and Ge. For smaller ionization levels, a semiconductor with lower effective mass carriers is required. In the early 1960s, the most likely candidate appeared to be n-InSb, but it was realized that even in the best material available the donors would not act as isolated centers. This is clear when the parameters for n-InSb are inserted into (5.39) and (5.40). When this is done the very low effective mass of $0.013m_e$ shows that the ionization energy of the donors would be about 1 meV, but the Bohr radius is ~ 60 nm. To achieve any significant isolation of the donors, the required purity would need to be less than 10^{13} cm^{-3} . At that time, the purest material had more than 10^{14} cm^{-3} donors and even today, with donor concentrations of less than 10^{13} cm^{-3} , there is little freeze-out of donors in n-InSb at

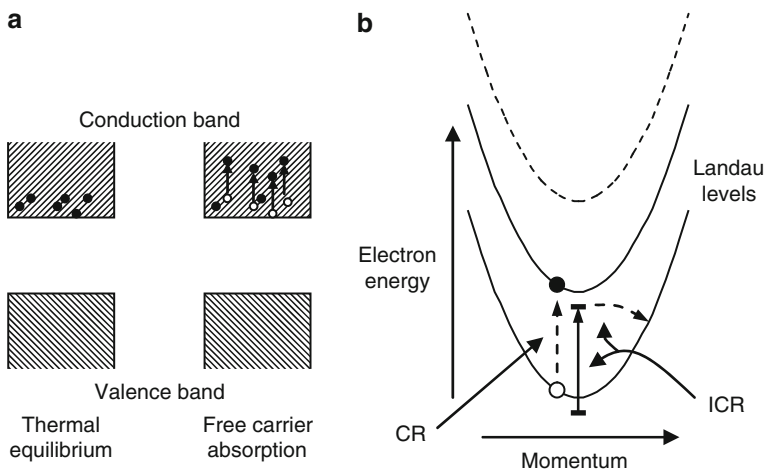


Fig. 5.22 (a) Detection process due to an electron bolometer effect, (b) Cyclotron resonance (CR) and impurity cyclotron resonance (ICR)

low temperature. Without freeze-out the impurity levels merge with the conduction band, producing free electrons, and it was found that under appropriate conditions these electrons would absorb THz radiation and provide a detection process, due to an electron bolometer effect. This process is illustrated schematically in Fig. 5.22a.

The absorption of free carriers is well documented [148]. The important features are that at high frequencies the absorption varies as ω^{-2} , where ω is the angular frequency, but at low frequencies the absorption is independent of frequency. The requirement for this frequency independent regime to be reached is that $\omega\tau_e < 1$, where τ_e is the carrier scattering time. For n-InSb $\omega\tau_e = 1$ at about 0.2 THz, but the electron absorption is reasonably high up to 1 THz. When absorption occurs the electrons are heated. This changes their mobility and, because the conductivity of a semiconductor is proportional to the mobility, the resistance also changes. The semiconductor has become a bolometer, with the thermal mass being that of the electrons, and hence the term electron bolometer. For such a detection process to be useful, it is necessary to consider the thermal conductance to the heat sink, which in this case is the lattice of the semiconductor. Under most circumstances, the coupling between the electrons and the lattice is very strong, so any detection process would be very rapid, and consequently a detector would have a low responsivity. However, in pure, high mobility semiconductors, such as InSb at low temperatures, the coupling is weak and the response time is typically about 0.5 μs [487].

A practical difficulty that occurs with the n-InSb electron bolometer is the low resistivity of the detector material. The mobility of the electrons in high purity n-InSb at 4 K is typically 10 to 50 $\text{m}^2/(\text{Vs})$, leading to a resistivity of only a few $\Omega \text{ cm}$. A conventionally shaped detector made from such material, with a useful area and sufficient thickness to be absorbing, has a resistance of about 100 Ω . The

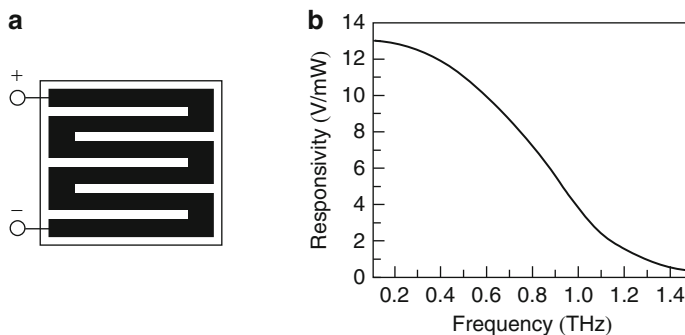


Fig. 5.23 (a) InSb detector element shaped in a meander structure to increase the resistance; (b) InSb hot electron bolometer response (adapted from [488])

noise from electron bolometers is close to that of Johnson noise and, remembering that the detector is operated at 4 K or less, this is much lower than the input noise of the best room temperature amplifiers. To overcome this problem, commercial InSb detector elements are shaped in a meander structure, as shown in Fig. 5.23a, thus increasing the resistance to 5–10 k Ω .

The NEP of InSb detectors at 0.3 THz is quite similar (10^{-12} to 10^{-13} W/ $\sqrt{\text{Hz}}$) to that of a Ge:Ga photoconductor at 3 THz, with the same speed of response, but the performance falls very rapidly at higher frequencies. As there is no energy gap in electron bolometers, there is no low-frequency limit until plasma resonance causes the device to become reflecting. InSb electron bolometers have been employed to well below 30 GHz. The responsivity details for a detector of the design illustrated in Fig. 5.23a is shown in Fig. 5.23b.

An earlier method of overcoming the low resistance of the electron bolometer was to use a cooled step-up transformer [489]. Although very effective in increasing the noise and the signal until the noise was above that of the amplifier, this reduced the bandwidth to a very limited range. While this is appropriate for many spectroscopic applications, it precludes the use of the device in fast-pulse situations. E. H. Putley, who was the first to detect THz radiation with n-InSb, used an approximately cubic-sized crystal with a volume of about 100 mm³ (Fig. 5.24). To increase the detector resistance, he employed a moderate magnetic field [65]. The field caused a separation between the impurity states and the conduction band, thus reducing the number of free carriers and increasing the detector resistance. There was also increased responsivity at higher frequencies, as illustrated in Fig. 5.24. Initially, this improved frequency coverage was thought to be due to photoionization produced by a limited freeze-out effect, but it was later realized that it was caused by a combination of free-carrier absorption and a very broad cyclotron resonance process. This was confirmed when much higher magnetic fields were employed [67].

At high magnetic fields and low temperature, the conduction band of InSb separates into Landau levels, as illustrated in Fig. 5.22b. Depending on the detector

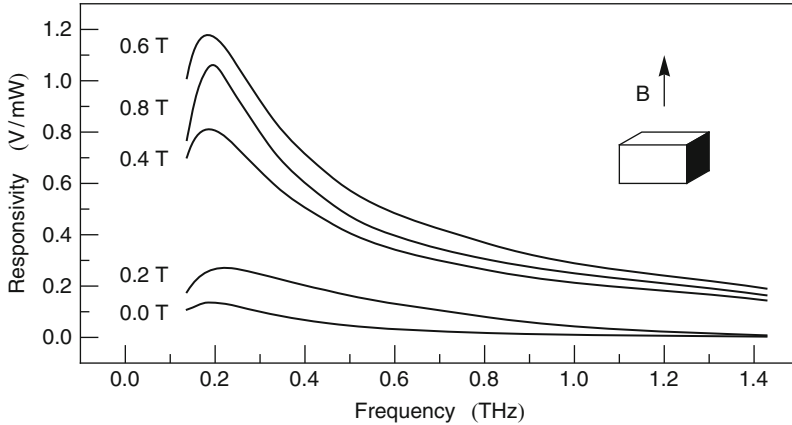


Fig. 5.24 Putley detector (crystal parallelepiped): response of an InSb detector in a magnetic field as a function of frequency and magnetic induction B at optimum current density. The high responsivity is due to the increased detector resistance, caused by the reduction of the number of electrons in the conduction band as the magnetic field is increased. However, above about 0.6 T the detector becomes optically thin and the responsivity then falls (adapted from [490])

purity, temperature, and the magnetic field intensity, two transition processes occur. The first is cyclotron resonance (CR) between the Landau levels and the second is due to transitions between the impurity levels associated with the Landau levels. This latter process, impurity cyclotron resonance (ICR), does not lead directly to a detection effect, but emission of a phonon permits a transition into the lowest Landau level, producing a change of conductivity.

At lower magnetic fields, both processes are observable but at higher fields, where all the electrons are in the impurity levels, only the phonon-assisted detection occurs. This is shown clearly in Fig. 5.25. The gap in response between 4.8 and 6.5 THz is due to an intense reststrahlen band that dominates all other absorption processes.

The two types of cyclotron resonance effect have been studied in detail over the frequency range 0.6–2.7 THz, with magnetic fields between 0.3 and 14 T [491]. The contribution of CR and ICR to the detection process was shown to be quite complex but, interestingly, over this frequency range the peak responsivities of the two processes are approximately equal. Other experiments have shown that the absorption coefficients at the peak of the response are very high, thus allowing the employment of very thin detectors. With these NEP values of less than $4 \times 10^{-13} \text{ W}/\sqrt{\text{Hz}}$ have been obtained [492]. In high magnetic fields, the dominant $1s-2p^+$ line splits into three lines, due to the Zeeman effect, and the broad continuum disappears. The $1s-2p^+$ transition becomes dominant and this provides a photoconductive line tunable over approximately 1–5 THz, using fields up to 8 T. At fields above 1.5 T the line becomes narrow. For example, at 3 T the line-width is $\sim 10 \text{ GHz}$, giving comparable resolution to that obtained with spectroscopic systems using a broadband source.

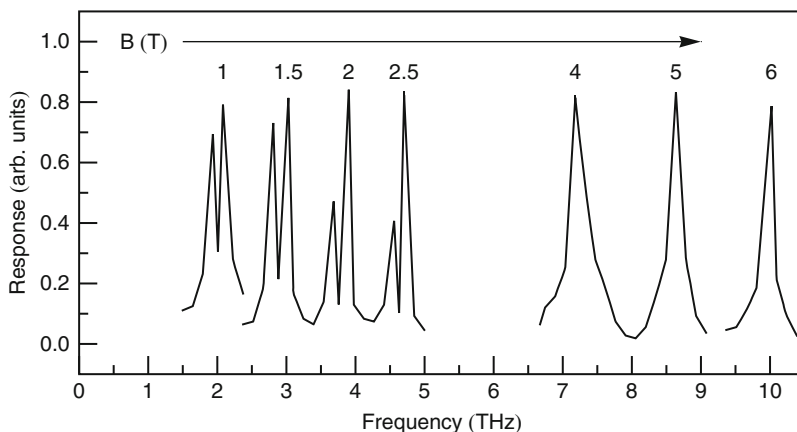


Fig. 5.25 Responsivity of an InSb detector as a function of magnetic induction B . The change from the region where both CR and ICR detection occurs, as shown in Fig. 5.22b, to only ICR at high magnetic fields, is clearly observable (adapted from [488])

Although the tunable version of the n-InSb detector has not been widely employed, it could be useful for studying the output of the various tunable sources now being developed for the THz region. When the detector was first invented, most THz spectroscopy was being performed with grating monochromators and with these the removal of unwanted higher order radiation was a considerable problem. Tunable InSb devices were a complete solution over a useful spectral range [493]. However, with the increased use of Fourier-transform instruments, this application is largely redundant. A version of this detector, which uses a nonuniform permanent magnetic field, is commercially available [494] to give enhanced response above 2 THz, thus providing an overlap with stressed Ge:Ga photoconductors.

InSb electron bolometers have been used for heterodyne detection with considerable success, but their speed of response normally limits the intermediate frequency bandwidth to about 1 MHz, and other THz devices provide much greater bandwidths (Sect. 5.5). When limited bandwidth is acceptable, InSb mixers have yielded excellent double-sideband receiver noise temperatures of 250 K at 0.5 THz, and 510 K at 0.8 THz [495]. By operating InSb at somewhat higher temperatures, the response time of n-InSb is decreased but the detectivity is much reduced. In a heterodyne system, the loss of detectivity can be overcome, to some extent, by increasing the local oscillator power. By operating a mixer at 18 K, a bandwidth of 10 MHz has been achieved [496].

There appears to have been only limited use of array technology for the InSb detector but in one interesting experiment a linear array of 20 elements was employed. This was to study the electron cyclotron emission from a tokamak plasma and an unusual feature was the use of a refrigerator, rather than liquid helium, to hold the detectors at 4 K [497].

5.4.3 Gallium Arsenide Detectors

The effective mass ratio for electrons in n-GaAs is ~ 0.07 , and it was recognized in the 1960s that it should be a suitable candidate for filling the “gap” in the THz spectrum between extrinsic Ge photoconductors and InSb electron bolometers. Photoconductivity in the material was first observed using DCN and HCN lasers at 1.54 and 0.89 THz, respectively [498]. The first spectra of absorption and photoconductivity with $n = 10^{15} \text{ cm}^{-3}$ showed both rising to a peak at a frequency corresponding to an energy gap of 4.6 meV (1.11 THz), with virtually no structure due to excited state transitions. However, later experiments, with higher purity material, showed significant structure with a pronounced peak corresponding to the $1s\text{-}2p^+$ transition at 4.4 meV (1.06 THz), and other peaks that could be identified with specific transitions in the hydrogenic model. These peaks are superimposed onto a broad continuum produced by transitions from the $1s$ and $2p$ states into the conduction band (Fig. 5.26) [498]. Unlike Ge:Ga, the photoconductive response increases at lower frequencies than those corresponding to the ionization energy between the ground state of the donor and the conduction band, which is 5.7 meV corresponding to a frequency of 1.38 THz.

At the frequency of peak response, 1.06 THz, NEPs of $4 \times 10^{-14} \text{ W}/\sqrt{\text{Hz}}$ have been obtained with n-GaAs with donor concentrations of $2 \times 10^{14} \text{ cm}^{-3}$ and low compensation by acceptors [500]. Response times of 250 ns have been measured in detectors with low compensation and 23 ns in those with much higher acceptor concentration [498]. In practice, it is difficult to achieve good detectivity and high speed simultaneously, due to the resistance of GaAs detectors, which is typically in the G Ω range.

An alternative to InSb and GaAs for THz detection is the alloy semiconductor $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, which is primarily of importance as an intrinsic photoconductor in

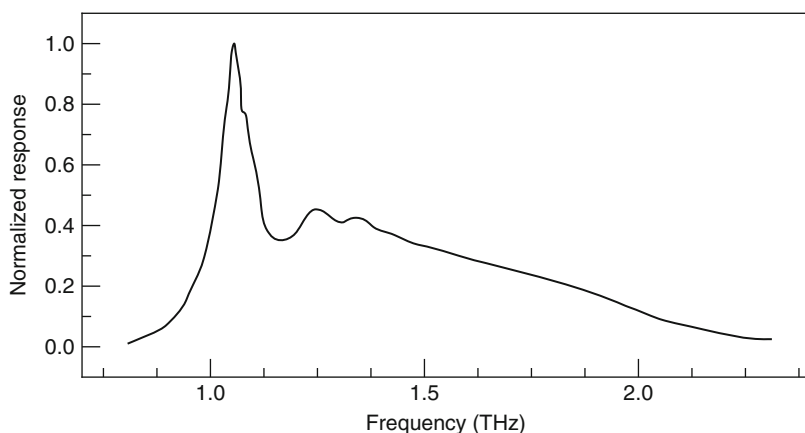


Fig. 5.26 Photoconductivity of GaAs detector material (adapted from [499])

the 20 to 100 THz spectral region. By choosing the appropriate alloy composition, the effective mass ratio of the electrons can be varied over a wide range. In the most detailed study of this alloy as a THz detector, material with effective mass ratios between 0.028 and 0.045 at 4.2 K was used, and both impurity photoconduction and hot-electron bolometer effects were observed [501]. Unfortunately, lattice absorption severely limits detection between 0.85 and 1.4 THz, but at lower frequencies the detectivity was similar to n-InSb and the speed of response was much faster. In a heterodyne experiment at 0.7 THz, a bandwidth of more than 50 MHz was observed. A practical problem with $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ is obtaining material with suitable alloy composition. InSb and GaAs, with the appropriate parameters for THz detection, are much easier to produce.

5.4.4 *Blocked Impurity Band Detectors*

Although Ge extrinsic photoconductors have been employed with great success in many applications, the fact that their absorption cross-section is quite low is a disadvantage, as this means detectors need to have a thickness of several millimeters to obtain optimum performance. While the shape of a detector is of secondary importance when it is a single device in the laboratory, this is not the case when it is used in space applications. A major problem of large-volume extrinsic detectors is that they are also good detectors of high-energy radiation such as X-rays and gamma-rays. Apart from swamping the IR signal, this radiation can cause long-term changes to a detector's performance, or even permanent damage. A further problem arises when long small-area detectors are used in arrays. Each detector has its own bias and, because this produces a fringing electric field, this can cause charge carriers to drift into neighboring detectors. This causes cross-talk and consequently distortions in the image detected by the array.

As explained in Sect. 5.4, the doping level in extrinsic photoconductors must be kept low, due to shot noise induced by hopping and impurity band conduction. When the detector is biased, these conduction processes cause a current to flow, even when there is no light falling on the detector. If there was not this unwanted effect, doping levels could be increased by approximately two orders of magnitude, before overlapping of the Bohr radii of the donors or acceptors produced high conductivity. Such an increase in doping would mean that extrinsic Ge detectors would need to be only $\sim 100\ \mu\text{m}$ in depth, and Si ones only $\sim 10\ \mu\text{m}$, because the absorption of such a thin highly doped layer is sufficiently high. In 1980, an alternative design was proposed [96] that allows doping to be increased. This employed what was termed a “blocked impurity band” (BIB) structure, and the design for an n-type detector is illustrated in Fig. 5.27.

It is convenient to consider a particular detector, Ge:Sb, because it has the advantage of reaching the lowest frequency of all the extrinsic Ge detectors. The essential

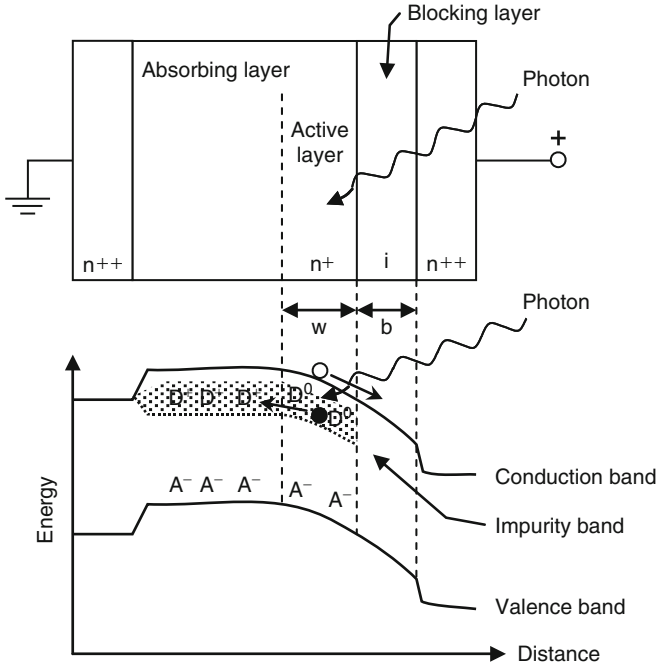


Fig. 5.27 Schematic of an n-type BIB detector with the band diagram below, shown for a device with an applied electric field. Heavily doped contacts are labeled n^{++} , the depletion width w , the blocking layer width b , ionized acceptors A^- , ionized and neutral donors D^+ and D^0 , respectively (adapted from [502])

difference between the BIB configuration, compared with a conventional Ge:Sb photoconductor, is the insertion of a very pure intrinsic layer of Ge between the doped detector material and the positive electrode, as shown in Fig. 5.27. The Sb doping level in the absorbing layer is typically $>10^{16} \text{ cm}^{-3}$, which would normally produce a very large dark current, due to conduction within the impurity band when the detector is biased. However, the intrinsic layer of undoped material prevents current flowing through it because it has no impurities to allow conduction to occur. Therefore, the intrinsic layer is also called the blocking layer. At low temperatures, and with no electric field applied, the doped region of the detector has a density N_A , of ionized residual (unintended) acceptors, an equal number of ionized donors, and a much larger density of neutral donors. When a reverse bias voltage is applied to the structure, electrons move in the impurity band toward the positive contact, but are stopped by the blocking layer. Ionized donor states are filled, leaving behind a depleted region of negative space charge created by ionized acceptor states. The depletion width is the active region of the detector. Absorption of a photon by a neutral donor in the depletion region generates an electron in the conduction band and a hole in the impurity band. These are moved in opposite directions by the

electric field, which leads to a current flow through the detector. In effect, a BIB detector resembles a diode when it is reverse biased, because current can only flow in one direction.

A large depletion width is required for a good BIB detector. In an n-BIB, this depleted region has a width, w given by

$$w = \sqrt{\frac{2\epsilon\epsilon_0(V_B - V_b)}{eN_A}} + b^2 - b, \quad (5.52)$$

where V_B is the applied bias voltage, V_b is a built-in bias of the detector, and b is the blocking layer width. V_b is normally very small compared with V_B and can be neglected. The most important parameter of (5.52) in producing a significant depletion layer width is N_A , and reducing this concentration to the required level has proved to be very difficult in both Ge and GaAs. This is in contrast to Si where, for example, in Si:Sb and Si:As, with doping levels of more than 10^{17} cm^{-3} , the compensating acceptor concentration is as low as $5 \times 10^{12} \text{ cm}^{-3}$. Si BIB detectors have to a large extent replaced conventional devices for most applications but the lowest frequency that they reach is $\sim 7 \text{ THz}$. The success of Si BIBs has been due to a major research and development effort that produced pure Si by chemical vapor deposition.

For Ge:Sb BIBs, the requirement on doping levels is more severe, due to the shallower levels involved [502]. Bulk Ge, with extremely low minority carrier concentration, has been available for many years but, despite considerable effort, especially with liquid phase epitaxy, detectors of comparable quality to conventional extrinsic Ge photoconductors for THz frequencies have yet to appear. This is due to the difficulty of obtaining sufficiently pure material for the blocking layer. The rewards for reaching similar detectivities to bulk Ge:Sb are immense. Apart from being much less susceptible to cosmic rays, the low-frequency response is extended from 2 THz in conventional Ge:Sb detectors to $< 1.5 \text{ THz}$, because the much increased doping level reduces the energy gap. Furthermore, the greatly increased absorption makes them ideal for use in high-density detector arrays with small pixel size. Ge:Ga has also been extensively studied for BIB detectors [503], but again with limited success compared with the outstanding performance of conventional extrinsic Ge:Ga photoconductors. However, Ge:B BIB detectors with NEPs of $5.23 \times 10^{-15} \text{ W}/\sqrt{\text{Hz}}$ have been reported, and these also have a response to below 1.5 THz [504].

There has also been considerable research devoted to GaAs BIBs, which should provide detection to well below 1 THz . Because of the even shallower impurity states than in Ge BIBs, the material requirements are even more demanding and, although GaAs BIBs have been made, there are severe problems to be overcome before the promise of these detectors is fulfilled. However, an alternative approach to BIB design has been suggested that may allow a thicker blocking layer and a consequent reduction in unwanted dark current [505].

5.5 Heterodyne Detection

Today, the heterodyne principle is the basis of nearly all radio and television receivers as well as wireless telecommunication. It can be dated back more than a hundred years when in 1901 R. A. Fessenden, a Canadian born radio engineer, filed the first patent regarding the heterodyne principle, which was granted in 1902 [506]. Remarkably, this invention appeared at the turn of the century, when technology such as vacuum tubes or continuous wave oscillators was still not available. During World War I, work in several countries led to the first practical heterodyne receivers. After the war, this quickly led to radio receivers for the wider public. In a 1925 publication [507], W. Schottky called the heterodyne technique “ein unschätzbares Werkzeug der drahtlosen Technik” (translation from German: an invaluable tool of wireless technique). For a review on the history of heterodyne technology, see [508]. The word heterodyne was coined by Fessenden. It is derived from two Greek words: *heteros*, meaning other or different, and *dynamis*, power. If currents of two different but closely spaced frequencies “beat” in the same electric circuit they will produce new frequencies, sum and difference frequencies, as well as multiples thereof. This process is called heterodyning. A heterodyne receiver is optimized for the detection of the difference frequency, which occurs at a fraction of the beating frequencies, typically a few GHz for a THz heterodyne receiver. It consists of two basic subsystems: the front-end and the back-end (Fig. 5.28). The front-end deals with the THz radiation. Its major components are:

- A local oscillator (LO) that delivers the reference frequency to the mixer.
- A mixer onto which the signal radiation and the radiation from the LO impinge. The mixer delivers an output at the difference of the signal and the LO frequencies, the so-called intermediate frequency (IF).
- Optical elements that couple the signal radiation and the LO radiation onto the mixer.

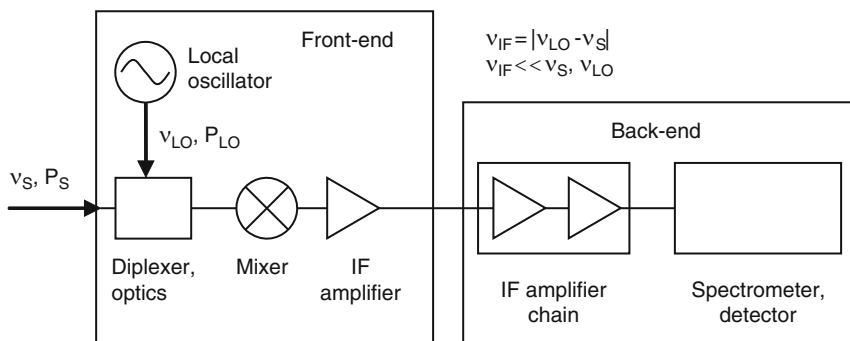


Fig. 5.28 Scheme of a heterodyne receiver

The back-end deals with the downconverted IF signal. Its components are:

- An IF processor that amplifies the difference signal generated by the mixer (note that the first amplifier that follows the mixer is usually assigned to the front-end).
- A spectrometer or detector for the detection of the IF signal.

In this section, only the THz part of a heterodyne receiver, namely the front-end with the mixer as its core element, is considered. Back-ends will be discussed in Sect. 6.8.2. Three characteristics of the heterodyne technique are responsible for its ubiquitous range of applications. First, since the signal which carries the information is downconverted, low-frequency amplifiers can be employed. This allows the use of heterodyne receivers even at extremely high frequencies, where direct amplification is not possible due to the lack of high-speed amplifiers. The second advantage is its high-frequency selectivity. For telecommunication, this means that it is possible to make very efficient use of a given frequency band, i.e., many transmitting channels can be fitted into it. For spectroscopy, it means that high spectral resolution is possible. The third advantage is also based on the narrow bandwidth detection process. The noise can be reduced by choosing a detection bandwidth similar to the signal bandwidth and, because the signal is narrow band, the noise will be correspondingly low.

5.5.1 Heterodyne Detection Theory

Mixing takes place in a device with a nonlinear characteristic such as a nonlinear current–voltage (I-V) curve. In principle, any type of nonlinearity can be used for mixing. Here, a mixer with a nonlinear I-V curve, which can be expanded into a power series, is considered

$$I(V) = k_0 + k_1 V + k_2 V^2 + \dots = \sum_{i=0}^{\infty} k_i V^i. \quad (5.53)$$

The voltage change is induced by the electric field of the LO and the signal:

$$V = V_{\text{LO}} \sin(\omega_{\text{LO}} t) + V_{\text{S}} \sin(\omega_{\text{S}} t) \quad (5.54)$$

with $\omega_{\text{LO}} = 2\pi \nu_{\text{LO}}$ and $\omega_{\text{S}} = 2\pi \nu_{\text{S}}$. Combining (5.53), (5.54) and, after some algebraic and trigonometric manipulation, this leads to

$$\begin{aligned} I = & k_0 + k_1 (V_{\text{LO}} \sin(\omega_{\text{LO}} t) + V_{\text{S}} \sin(\omega_{\text{S}} t)) + \frac{k_2}{2} (V_{\text{LO}}^2 + V_{\text{S}}^2) \\ & - \frac{k_2}{2} (V_{\text{LO}}^2 - V_{\text{S}}^2) \cos(2\omega_{\text{LO}} t) + k_2 V_{\text{LO}} V_{\text{S}} \cos((\omega_{\text{LO}} - \omega_{\text{S}}) t) \\ & - k_2 V_{\text{LO}} V_{\text{S}} \cos((\omega_{\text{LO}} + \omega_{\text{S}}) t) + \dots \end{aligned} \quad (5.55)$$

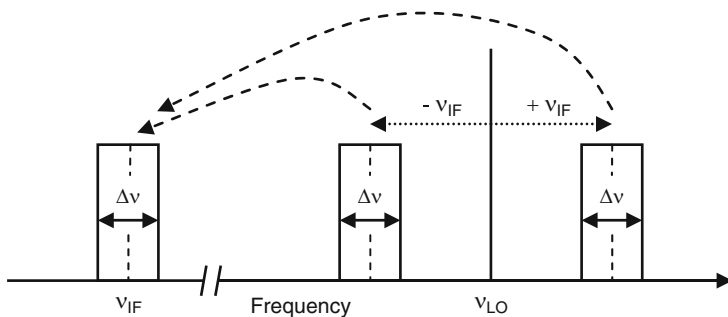


Fig. 5.29 Illustration of the frequency conversion process in a mixer

The mixer produces a frequency spectrum of the form

$$\nu_k = |\ell \nu_{LO} \pm m \nu_S|, \quad \ell, m = 0, 1, 2, \dots \quad (5.56)$$

Since the power of the signal is usually much smaller than the power from the LO, and because the power in the higher harmonics is approximately proportional to $1/\ell^2$, only three frequency components are important

$$\begin{aligned} \nu_{LO} + \nu_S & \quad \text{sum frequency,} \\ |\nu_{LO} - \nu_S| = \nu_{IF} & \quad \text{intermediate frequency,} \\ 2\nu_{LO} - \nu_S & \quad \text{image frequency.} \end{aligned} \quad (5.57)$$

Essentially, two frequencies are transferred to the IF, $\nu_{LO} - \nu_{IF}$ and $\nu_{LO} + \nu_{IF}$ (Fig. 5.29), called the lower sideband (LSB) and the upper sideband (USB), respectively. If a mixer is operated in a way that both sidebands are present at the IF, it is called a double-sideband (DSB) mixer. If, on the other hand, one sideband is suppressed, for example by a filter in the signal path, only one sideband is down-converted and it then becomes a single-sideband (SSB) mixer.

A different type of operation occurs when the mixer reacts to the LO and signal field individually. The mixer can be considered as a switch, which interrupts the voltage waveform induced by the LO field. The voltage at the IF is the product of switching and signal waveform. If the switching waveform does not have a 50% duty cycle, the product includes all harmonics, and the desired intermediate product needs to be separated by appropriate filtering.

The difference between the two methods of operation is fundamental. The first type of mixer is a square-law or power mixer. It responds to the total power of the incident fields $P \propto |E_{LO} + E_S|^2$ but not to the individual fields of LO and signal, because the physical mechanism on which the mixer is based is too slow to follow the fields. Hot-electron bolometers are an example of this type of mixer. A switching mixer responds to the LO and signal fields individually. Schottky diodes and SIS junctions are examples of this type.

It is worth noting that any mixer can also be operated as a rectifier. In this case (5.55) is simplified, because no LO radiation is present and the output signal of the rectifier is proportional to the square of the amplitude of the signal radiation. Although rectifiers are not usually as sensitive as other direct detectors, they may have advantages for particular applications. In particular, Schottky diodes and superconducting hot electron bolometers are used for time-resolved detection of picosecond short pulses such as coherent THz synchrotron radiation, because they are intentionally designed to have a short response time [400, 509] (Sect. 5.5.4).

Two parameters are used to characterize a mixer: the conversion loss L_M and the mixer noise temperature T_M . The conversion loss is defined as the power at the IF, which is delivered to a load, divided by the signal power available at the input terminal. In most mixers the conversion loss is less than one. Theoretically, it is possible to achieve gain (conversion loss greater than one), for example by parametric amplification in a nonlinear capacitance. However, in practice, only superconductor–insulator–superconductor (SIS) mixers have shown gain due to quantum mechanical phenomena. Several loss mechanisms contribute to the total conversion loss. These are losses due to the physical peculiarities of the mixing process (L_P), losses due to power loss into other frequency bands than the IF band (L_B), losses due to absorption in the electrical circuit (L_A), and losses due to reflection at the input port and the output port (L_R). The total loss is a product of all individual loss factors:

$$L_M = L_P L_B L_A L_R. \quad (5.58)$$

A mixer can be considered as a two-port device (Fig. 5.30) with a particular noise temperature T_0 , which appears at its output, and a particular conversion loss L_M . In the Rayleigh–Jeans approximation, the noise power P_L within a bandwidth B delivered to the load R_L is given by the noise from the mixer and the noise generated by the signal divided by the conversion loss of the mixer (Fig. 5.30a)

$$P_L = k_B \left(T_0 + \frac{T_S}{L_M} \right) B. \quad (5.59)$$

Next a mixer is considered where all noise is attributed to the input of the mixer, i.e. the mixer noise is virtually assigned to the signal noise (Fig. 5.30b). Then the power delivered to the load can be written as

$$P_L = k_B \frac{T_S^*}{L_M} B \quad (5.60)$$

with $T_S^* = T_S + L_M T_0$. The mixer noise temperature, T_M is defined as the increase of the signal noise temperature T_S , which is required to generate the same noise power at the input of the load as would be generated by a noisy mixer, i.e. $T_M = L_M T_0$. Note that T_0 is usually not the physical temperature of the mixer. It depends on the

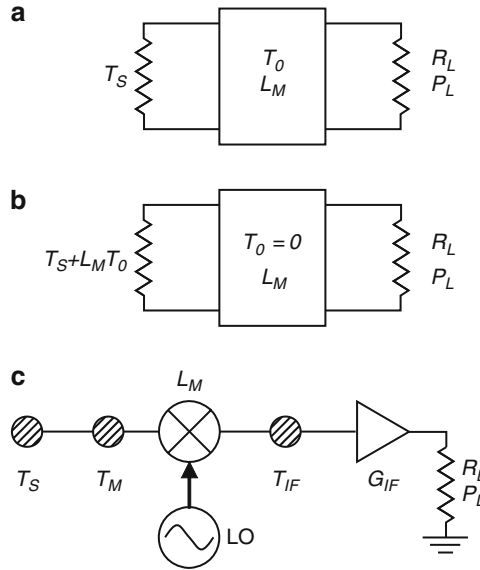


Fig. 5.30 Two-port model of a mixer with the noise assigned to the output (a) and input (b) of the mixer. In (c), a noise model of a mixer with IF amplifier is depicted. The noise sources are assigned to the input of each component and are symbolized by hatched circles

noise processes involved. In a heterodyne receiver, the mixer is followed by an IF amplifier. It is characterized by a noise temperature T_{IF} , which is referred to the input of the amplifier, and a gain G_{IF} . The noise temperature at the output of the IF amplifier is $G_{IF}T_{IF}$. The noise power delivered to the load by a combination of mixer and IF amplifier is (Fig. 5.30c)

$$P_L = k_B \left(G_{IF} \frac{T_S + T_M}{L_M} + G_{IF} T_{IF} \right) B = k_B G_{IF} \frac{T_S^*}{L_M} B \quad (5.61)$$

with $T_S^* = T_S + T_M + L_M T_{IF}$. In analogy with the mixer noise temperature, the receiver noise temperature T_{REC} is defined as the increase of signal noise temperature, which is required to generate the same output noise as would be generated by a noiseless receiver, i.e.

$$T_{REC} = T_M + L_M T_{IF}. \quad (5.62)$$

The fact that there can be two input frequency bands but only one output band complicates the situation. The mixer or receiver noise can be assigned either to one sideband or equally to both sidebands. In the first case, the noise temperature is the SSB noise temperature, while in the second case it is the DSB noise temperature. Assuming that the signal noise and all other noise contributions are the same in both sidebands, then $T_{SSB} = 2T_{DSB}$.

Normally, a heterodyne receiver has more amplifiers than just the one considered so far. However, in analogy with (5.61) and (5.62), the noise contribution from the amplifier, which follows the first one is $L_M T_{IF2}/G_{IF}$. Since the first IF one typically has a gain of about 30 dB, the noise contribution from the second amplifier is a factor of 1,000 less than that from the first one. Therefore, it is usually sufficient to have a low noise first amplifier, while the second and any subsequent amplifiers can have much higher noise. Good cryogenic IF amplifiers, operating at liquid helium temperature, have noise temperatures of less than 5 K across a band of a few GHz, along with a gain of ~ 30 dB. Assuming a conversion loss of 13 dB, the noise contribution from the IF amplifier is less than 100 K.

The receiver noise temperature determines the sensitivity of the receiver, i.e. its minimum detectable power or minimum detectable temperature in a given integration time. The signal can be characterized by a temperature, which adds to the receiver noise temperature at the output of the receiver. The minimum detectable temperature difference ΔT_{MIN} of a system with a noise temperature T_{REC} is equal to the rms noise temperature of the system. It is given by

$$\Delta T_{\text{MIN}} = \frac{k_{\text{REC}} T_{\text{REC}}}{\sqrt{\tau \Delta \nu}}, \quad (5.63)$$

where $\Delta \nu$ is the predetection bandwidth defined by the bandwidth of the IF amplifier chain, τ is the postdetection integration time, and k_{REC} is a receiver-dependent constant, typically between 1 and 2. This equation is called the “radiometer equation.” Although the exact derivation is beyond the scope of this book (see for example [120]), a few qualitative arguments can give some insight. The heterodyne receiver generates $\Delta \nu$ independent noise events per second. This results in $\tau \Delta \nu$ noise events in an integration time τ . The standard deviation Δx of a Gaussian distribution of a quantity x is $\Delta x/x = 1/\sqrt{n}$, where n is the number of data points. In the case where $x = T_{\text{REC}}$ and $n = \tau \Delta \nu$, and with $\Delta x = \Delta T_{\text{MIN}}$ as the minimum detectable temperature change, (5.63) is obtained, but without the constant k_{REC} which depends on the details of the receiver and its mode of operation [120]. The minimum detectable temperature change is the same as the noise equivalent delta temperature (NEDT), which is defined in Sect. 7.3.1, where the relation between NEP and ΔT_{MIN} is also discussed.

The receiver noise temperature can be measured by placing alternately two blackbody emitters with different temperatures, also called the hot and cold loads, in front of the receiver. Often the hot load is a blackbody material at a temperature of $T_H = 296$ K, while the cold load is a blackbody cooled to $T_C = 77$ K by liquid nitrogen. At the output of the receiver, both loads create output signals P_H and P_C which, in the Rayleigh–Jeans region and providing that the receiver is linear, are proportional to the sum of the system noise temperature and the blackbody temperature, and the following ratio is obtained:

$$Y = \frac{P_H}{P_C} = \frac{T_{\text{REC}} + T_H}{T_{\text{REC}} + T_C}. \quad (5.64)$$

The ratio is the Y -factor after which this method is named [510]. Equation (5.64) can be solved for the receiver noise temperature

$$T_{\text{REC}} = \frac{T_{\text{H}} - Y T_{\text{C}}}{Y - 1}. \quad (5.65)$$

This method does not require any absolute calibration of the output signal of the receiver. Note that the receiver noise temperature depends on the way that the physical temperatures of the hot and cold loads are converted into radiated power or temperatures. The derivation is based on the Rayleigh–Jeans approximation and is valid if $h\nu \ll k_{\text{B}}T_{\text{C}}, k_{\text{B}}T_{\text{REC}}$. Otherwise, Planck’s equation must be used (Fig. 4.1). Also, for high accuracy, the general form of the dissipation–fluctuation theorem should be used to derive the noise temperature from the measured Y -factor [511]. The accuracy is often limited by the quality of the blackbody, which in many cases has an emissivity of less than one. A method to determine the noise temperature as well as the conversion loss of a mixer makes use of (5.62). Measuring the receiver noise temperature as a function of the noise temperature of the IF amplifier yields a linear dependence, with the conversion loss as the slope and the mixer noise temperature on the ordinate at $T_{\text{IF}} = 0$ K.

In addition to the mixer, heterodyne detection requires an LO, the combination (diplexing) of LO and signal radiation, and some kind of detector for the IF signal. The LO might be any coherent source such as those described in Chap. 4. The power needs to be sufficient to drive the mixer into the optimum operating region. This depends on the type of mixer. For a Schottky diode mixer, it is typically in the order of 1 mW, while for SIS and superconducting HEB mixers it is in the order of 10 μ W and 1 μ W, respectively. The requirement of the LO power stability depends to some extent on the mixer. For example, Schottky diodes are relatively insensitive against power fluctuations, because the operation point depends only weakly on the power, while the performance of a HEB mixer, which is not driven into saturation, is much more vulnerable to power fluctuations. In addition, any amplitude modulated noise is downconverted by the mixer and appears at the IF output. Another consideration is the frequency, which should correspond to the input frequency band of the amplifier. This is sometimes a problem for fixed frequency sources such as gas lasers. In this case, the amplifier needs to match the IF, as set by the gas laser, and the signal frequency. Frequency tunability of the LO is also required, for example, to compensate for the Doppler shift of astronomical sources. In general, the linewidth of the LO should be much less than the spectral resolution required by the application, typically 1 MHz or less. Phase noise of the LO can be troublesome. It does not appear at the signal or image frequency like amplitude modulated noise, but close to the LO frequency (Sect. 6.8.1). Phase noise arises from low-frequency processes in the LO such as $1/f$ noise in solid-state devices. Filtering this out is rarely practical because it is so close to the LO frequency.

An important issue is the diplexing of the LO and signal radiation. Obviously as little signal power as possible should be lost in the diplexer. Therefore, thin (a few μ m) beam splitters (e.g. from Mylar) are used. The transmission of these

is high (97% for a 3 μm thin beam splitter at 2 THz and parallel polarization) and little signal is lost. On the other hand, this means a high loss of LO power. If the LO has only low output power, as is often the case for multiplied microwave sources above ~ 1 THz, this loss is not acceptable. In such a situation, a Martin–Puplett diplexer (Sect. 6.4) is frequently used or, less often, a ring Fabry–Pérot interferometer (Sect. 6.5).

The type of detector used for measuring the IF signal depends on the application. If only the relative power has to be measured, for example when the radiometric sensitivity is measured with the Y -factor method, a rectifying diode is sufficient. Sometimes, especially if the power has to be known accurately, a microwave power meter or a calibrated diode is used. If spectral analysis of the IF signal is required, the detector is replaced by a spectrometer. The simplest one is a spectrum analyzer. Other more sophisticated spectrometers are described in Sect. 6.8.

So far the discussion has been restricted to a single mixing element, pumped at a frequency several GHz away from the signal frequency. If the mixer has one input port for signal and LO radiation, which is the most common design at THz frequencies, it is called a single-ended mixer. In the mmW range, other concepts have been developed which are also gaining ground at THz frequencies. The most important are balanced mixers and subharmonically pumped mixers. If the LO and signal ports are isolated from each other, the mixer is called a single-balanced mixer. This can be achieved using two mixers. The advantages are the canceling of LO noise at the output and superior power handling. Other balanced configurations are also possible. The subharmonically pumped mixer requires two diodes in antiparallel configuration. In this case, the mixer's conductance varies at twice the LO frequency. No conversion occurs at the fundamental frequency, and the LO noise contribution is less. For more details, the reader is referred to dedicated textbooks (e.g. [510]).

5.5.2 Schottky Diode Mixer

When a metal and a semiconductor are brought into close contact, a potential barrier is formed (see Fig. 5.31 for an n-type contact). Charge neutrality at the interface results – to first order – in a parabolic shape of the potential barrier, and a region in the semiconductor which is depleted of electrons, termed the depletion region [512]. The barrier height is determined by the position of the Fermi level inside the band gap of the semiconductor. In general, the exact position of the Fermi level depends not only on the materials but also on the details of the interface, for example, defects, which may lead to pinning of the Fermi level [513]. In principle, the current flow through a Schottky diode consists of different components. These are thermionic emission over the barrier, tunneling through the barrier, and generation-recombination inside or outside the depletion region (Fig. 5.31). The latter effect can be neglected in THz mixer diodes because there are almost no holes available

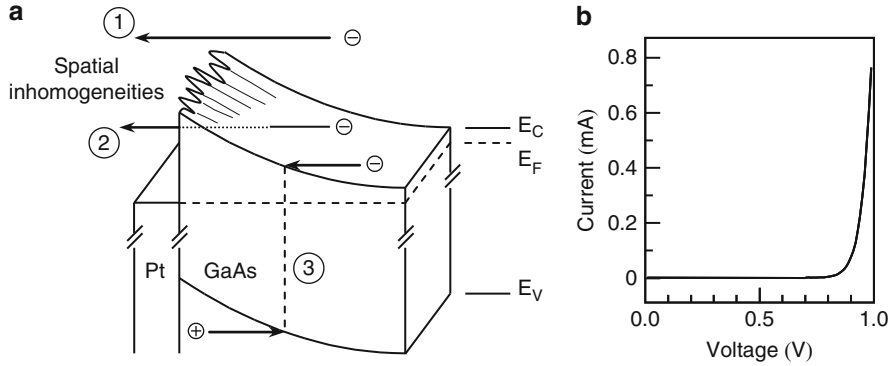


Fig. 5.31 (a) Metal–semiconductor contact with an applied voltage. The current transport consists of three components: emission over the barrier (1), tunneling through the barrier (2), and generation–recombination of electrons and holes (3). For clarity, the spatial inhomogeneities of the potential barrier are exaggerated. (b) Typical I–V-curve of a Schottky diode

for recombination. Whether thermionic emission or tunneling dominates depends on the doping concentration and on the temperature [512]. For THz mixer diodes, the doping density is above 10^{17} cm^{-3} . Therefore tunneling contributes to the total current. In the case of pure thermionic emission, the current flow can be described by the following equation

$$I = I_S \exp\left(\frac{e(V_B - IR_S)}{\eta k_B T}\right) \left(1 - \exp\left(-\frac{eV_B}{k_B T}\right)\right). \quad (5.66)$$

Here, T is the temperature, V_B is the applied forward voltage, R_S is the series resistance, η is the so-called ideality coefficient, and I_S is the saturation current. In the case of thermionic emission, it is given by

$$I_S = A^{**} A T^2 \exp\left(-\frac{e\Phi_B}{k_B T}\right). \quad (5.67)$$

A^{**} is the effective Richardson constant, A is the anode area, and Φ_B is the barrier height when no bias voltage is applied. In the case of tunneling, I_S is a complicated function of the barrier height, temperature, and doping density. However, for THz mixer diodes the tunneling component is small and the current flow is well described by (5.66) and (5.67). The deviation from pure thermionic emission is taken into account by an ideality coefficient larger than one. In the case of pure thermionic emission, the ideality coefficient equals one. However, tunneling through the barrier, image force lowering, generation–recombination, and spatial inhomogeneities of the barrier may lead to $\eta > 1$ [514]. In the presence of spatial inhomogeneities, the potential at the interface is not constant but fluctuates around an average value (Fig. 5.31).

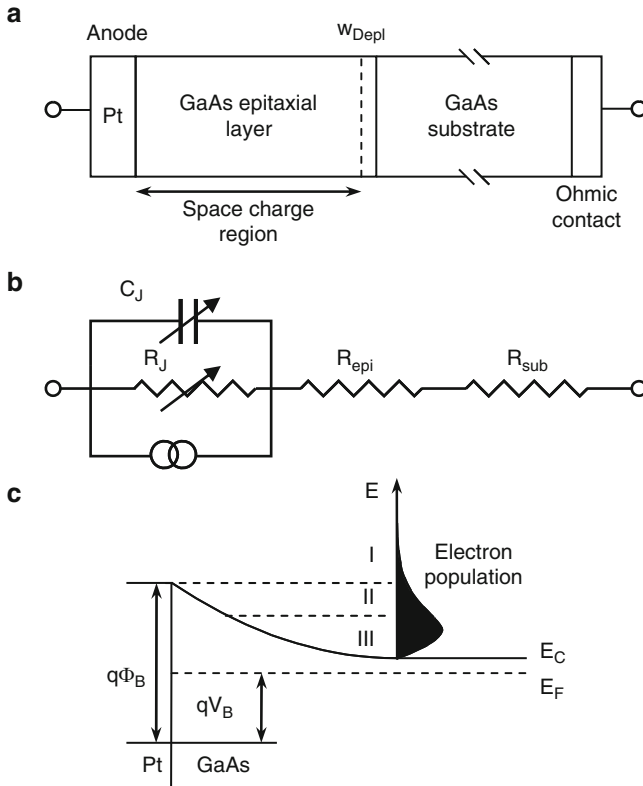


Fig. 5.32 (a) Principal structure of a GaAs Schottky diode for THz mixing applications; (b) its equivalent circuit; (c) energy diagram of a Schottky barrier showing the electron population. The Roman numerals denote three groups of electrons. Their role in the mixing process is discussed in the text

A widely used approach to describe THz Schottky mixer diodes is the equivalent circuit model. It has been extensively studied, and it is possible to predict the mixer performance with models based on a well-defined equivalent circuit. Figure 5.32a,b display the principal structure and an equivalent circuit of the diode, including the noise sources. The effective noise temperature of the diode can be found by combining the noise sources and the diode impedances into a Thevenin equivalent circuit, which yields

$$T_D = \frac{\eta T}{2} \frac{R_J}{R_t} + T_{\text{epi}} \frac{R_{\text{epi}}}{R_t} + T \frac{R_{\text{sub}}}{R_t}. \quad (5.68)$$

Here, the first term represents the shot noise of the junction, which is at temperature T (R_J : resistance of the junction). The second term represents the thermal noise generated by the resistance R_{epi} of the epitaxial layer with an effective temperature T_{epi} . The third term represents the thermal noise from the resistance R_{sub} of the substrate.

The total resistance is $R_t = R_J + R_{\text{epi}} + R_{\text{sub}}$. If the series resistance is low, the noise temperature of the diode is determined by shot noise. Cooling the diode lowers the shot noise significantly. However, for highly doped diodes the ideality coefficient increases significantly with decreasing temperature, and the lowering of the shot noise is a minor effect. Several physical mechanisms can contribute to the noise in a Schottky diode mixer. These include thermal noise, shot noise, hot-electron noise, and noise due to intervalley scattering. Thermal and shot noise can be described by a Maxwell–Boltzmann distribution and an average kinetic energy of the electrons, which is in the order of $k_B T$. The situation changes when an electric field is applied, and a free electron gains a certain amount of energy from the electric field between two collisions with lattice atoms. Due to the relatively low scattering rates in GaAs, the electrons are no longer in thermal equilibrium with the lattice. This effect will increase the diode noise and is called hot-electron noise. It depends on the scattering rates and therefore on the doping density of the GaAs. If the electric field across the diode is increased even further, the electrons may gain sufficient energy to be transferred from the Γ valley into the next higher L valley. This occurs at an electric field of about 3.2 kV/cm. Intervalley scattering causes the diode noise to increase abruptly, due to velocity fluctuations of the electrons, which have different effective masses and velocities in the different valleys.

Until now, only DC transport has been considered. The current transport in a high-frequency field was analyzed in detail by van der Ziel [515]. A brief phenomenological description is given here. As depicted in Fig. 5.32c, the electron population can be roughly divided into three groups. The electrons of group one will move toward the barrier, where they are reflected, or tunnel through the barrier. This yields a net current close to zero, which is not affected by the application of a THz electric field. The electrons of group three are emitted over the barrier and a negligible amount of electrons is reflected. The current of this component is also not influenced by a THz electric field. Only the current generated by the electrons of group two is influenced by the THz field. When the voltage generated by the THz field reaches its maximum, electrons are able to cross the barrier. At a minimum voltage, no electrons from this group cross the barrier. The electrons of group two will suffer from transit time effects because, within half a period of the THz field, they have to transverse the depletion zone in order to cross the barrier. With higher frequency, the bias voltage needs to be increased, in order to keep the depletion width small, and to have sufficient electrons of group two which move across the barrier. The optimum current I_{OPT} , that is the bias current where the Schottky diode yields the lowest noise temperature, is a linear function of the frequency of the incident THz field: $I_{\text{OPT}} = N_e e v$. N_e is interpreted as the number of electrons passing through the Schottky contact at each period of the THz field. It is worth noting that N_e is independent of the frequency and depends only on the doping density and anode diameter. Typical values for N_e range from 1,000 to 5,000 electrons [516, 517].

Other effects which play an increasingly important role at THz frequencies are the skin effect, charge carrier inertia, dielectric relaxation, and plasma resonance. The skin effect constrains the current flow to the periphery of the diode and

the surface of the substrate. This leads to an increase in the series resistance but also yields an inductive component. Above the scattering frequency carrier inertia becomes more important and, above the dielectric relaxation frequency, a major component of the current is due to displacement of the electrons. Carrier inertia yields an inductive component, while dielectric relaxation gives rise to an additional capacitive component. Close to the plasma frequency, the series impedance increases drastically, and the Schottky diode cannot be used as an efficient mixer [518]. A figure of merit, which is useful for comparing different diodes, is the cutoff frequency ν_{CO} . This is the frequency where the series resistance equals the junction capacitance at zero bias voltage C_{J0} , i.e., $\nu_{CO} = 1/(2\pi R_S C_{J0})$ [518]. However, this is not a strict limit. It was shown experimentally that GaAs Schottky diodes can be operated as mixers even above plasma resonance and cutoff frequency [356].

In order to optimize the mixer performance of a Schottky diode for a particular frequency, several parameters can be varied. The main ones are the doping density of the epitaxial layer, the thickness of the epitaxial layer, and the area of the anode. Increasing the epitaxial doping density has the effect that the series resistance of the epitaxial layer and the hot electron noise are reduced, while the ideality coefficient and the shot noise are increased. At THz frequencies, the noise due to the series resistance and hot electrons dominates over the shot noise. Therefore the doping density needs to be high. In addition, transit time effects are less pronounced at a higher doping density and the plasma resonance is shifted toward a higher frequency. The major drawback of an increased epitaxial layer doping is the larger junction capacitance. Perhaps, the most critical parameter is the area of the anode. Any reduction leads to a smaller capacitance but it increases both the series resistance and the hot-electron noise. In general, the anode diameter becomes more critical and needs to be decreased with increasing doping density [519]. Another important aspect is the impedance matching between the diode and the antenna, which is dominated by the anode area.

One of the most important steps toward a practical GaAs-based Schottky diode mixer for THz frequency applications is the so-called “honeycomb” diode chip design [72]. It combines several advantages. First, it is compatible with modern semiconductor processing techniques. Second, it contains up to several thousand diodes on a single chip, and third, the design minimizes parasitic losses such as the series resistance and the shunt capacitance. The honeycomb structure consists of a highly doped ($\sim 5 \times 10^{18} \text{ cm}^{-3}$) GaAs substrate with an ohmic contact on the back surface. A GaAs epitaxial layer is grown on top and is covered by an SiO_2 layer. For applications above 1 THz, the doping concentration is in the range from 10^{17} to 10^{18} cm^{-3} and the dopant is Si. The thickness of the epitaxial layer varies between 0.3 and 1 μm . Holes in the SiO_2 insulating layer define the anode area. Typical diameters are 0.25 to 1 μm . The anode metal is Pt, which shows almost no diffusion into GaAs and forms an abrupt contact (width $\sim 8 \text{ nm}$) [520]. The electrical parameters are $R_S \approx 25 \Omega$ and $C_{J0} = 0.25 \text{ fF}$ to 1 fF. In order to couple the signal and the LO radiation to the mixer, a long wire antenna in a 90° corner-cube reflector is used (Fig. 5.33). DSB receiver noise temperatures measured with these diodes increase fairly linearly, from 3,500 K at 0.8 THz up to 8,600 K

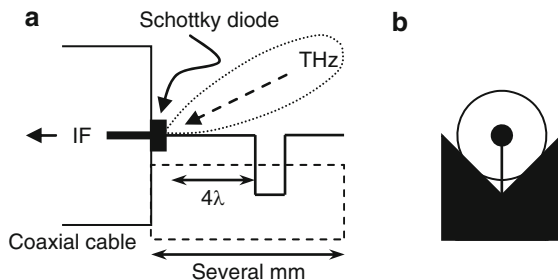


Fig. 5.33 Schematic of the 4λ corner-cube mixer mounting. (a) The *dashed line* indicates the corner-cube reflector and the *solid line* indicates the wire antenna with a typical diameter of $25\ \mu\text{m}$, which contacts the Schottky diode. The bend in the wire is grounded by a contact with the corner-cube reflector. It defines the 4λ length of the antenna. The *dotted line* indicates the main lobe of the antenna pattern. (b) View from the right of (a)

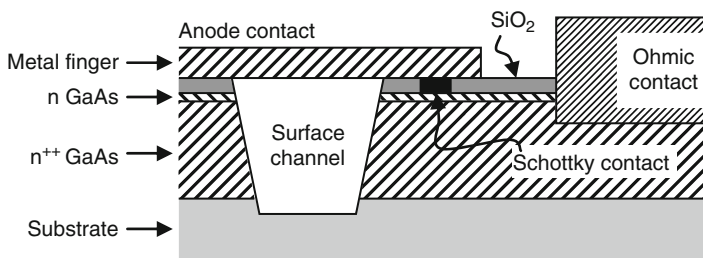


Fig. 5.34 Schematic of a planar Schottky diode. Note the surface channel, which lowers the parasitic capacitance

at 2.5 THz. Above 2.5 THz this increase is much steeper, mainly due to increasing losses of the antenna and reduced performance of the diode itself. The mixer noise temperature is below 5,000 K for all frequencies below 2.5 THz [521]. It is worth noting the excellent agreement between the predicted mixer noise temperatures and the measured ones [517]. Between 0.8 and 2.5 THz the DSB conversion loss of this diode mixer increases from 9 to 12.5 dB and the required LO power from 1 to 10 mW. Cooling of these diodes to 20 K improved the noise temperature by less than 30% [521]. This confirms that cooling has little effect on the shot noise, because of the high doping.

Schottky diode mixers in a waveguide configuration have an improved coupling efficiency compared to the open structure mount. Several approaches have been devised in order to integrate a Schottky mixer into a waveguide mounting. A major problem is the large shunt capacitance caused by the coplanar contact pads. One typical design is shown in Fig. 5.34. The planar diodes are soldered onto a microstrip circuit, for example on a thin quartz substrate, which is mounted into a mixer block. Typical results for mixers operating around 0.6 THz are DSB system noise temperatures of about 1,000 K and a conversion loss of ~ 8 dB [522]. The required LO power is about 1 mW. Besides the shunt capacitance, the supporting high dielectric substrate limits the high-frequency performance because surface

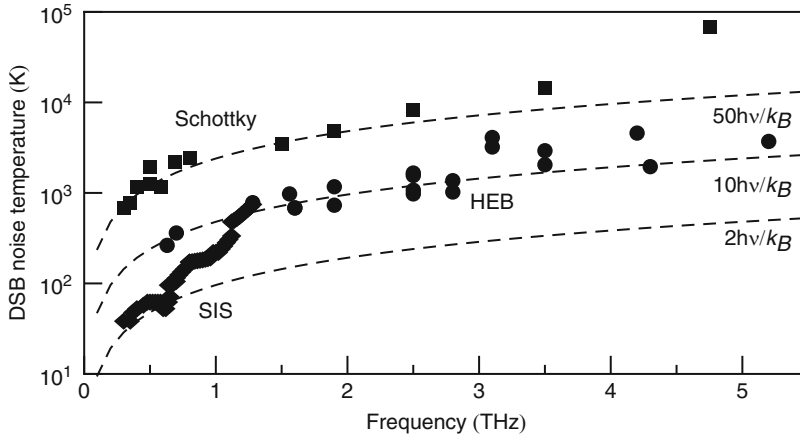


Fig. 5.35 DSB noise temperatures achieved with Schottky diode mixers (squares), SIS mixers (diamonds), and hot electron bolometers (HEB, circles), (adapted from various sources)

modes can lead to substantial losses. To overcome this problem, a process that allows the fabrication of a Schottky diode onto a thin ($\sim 3 \mu\text{m}$) membrane was developed ($0.6 \times 1.4 \text{ mm}^2$ in a GaAs frame) [523]. The fabrication consists of three major processes. First, the Schottky diode and the embedding RF components are processed, then the anode is formed, and finally the membrane is defined. The Schottky diode itself is very similar to the planar diodes, described previously. The DSB noise temperature of a receiver with such a monolithic membrane diode (MOMED) is 8250 K, and the mixer noise temperature is 3500 K [523].

DSB noise temperatures achieved with Schottky diode mixers are presented in Fig. 5.35. Above 3 THz a steep increase appears, mainly due to increasing losses of the antenna, and reduced performance of the diode itself. Schottky diode mixers are mainly employed in heterodyne receivers, where cooling with liquid cryogenics is not an option, and at frequencies below 1 THz. For example, the MOMED mixer is used in the 2.5 THz heterodyne receiver of the microwave limb sounder on the Aura satellite to measure OH in the stratosphere [524]. Receivers operating below 1 THz are often employed in imaging systems for security applications. The Schottky diodes are planar, waveguide-mounted devices of similar design to that shown in Fig. 5.34.

5.5.3 Superconductor–Insulator–Superconductor Mixer

The superconductor–insulator–superconductor (SIS) mixer is a sandwich of two superconductors separated by a thin ($\sim 20 \text{ \AA}$) insulating layer (Fig. 5.36). It is based on photon-assisted tunneling of quasiparticles through the insulating layer. Although the physics of this process had already been investigated and theoretically

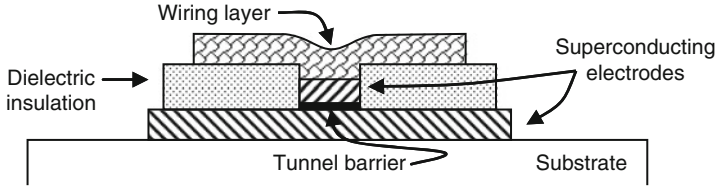


Fig. 5.36 Schematic structure of an SIS mixer. The tunnel barrier is sandwiched between two superconducting electrodes, with the upper electrode connected to a wiring layer

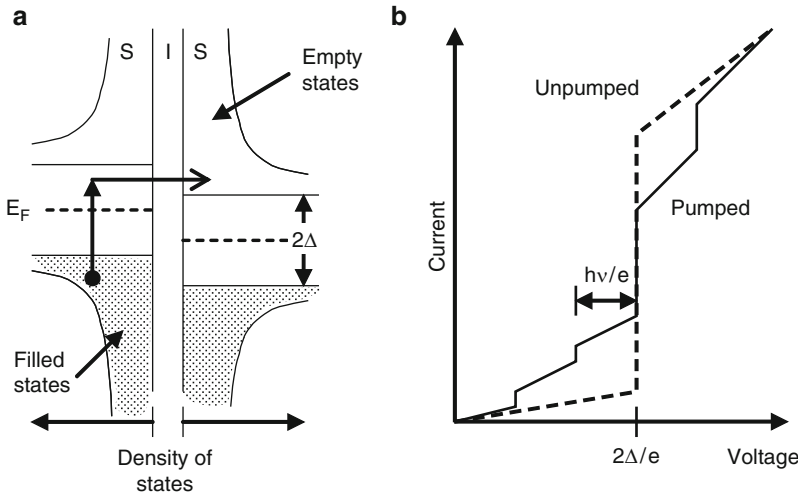


Fig. 5.37 (a) Energy diagram of an SIS mixer with applied bias voltage and illustration of the photon-assisted tunneling process. (b) I-V curve of an SIS junction with LO radiation (pumped, *solid line*) and without LO radiation (unpumped, *dashed line*)

explained in the 1960s [525, 526], it took more than 15 years to make use of the effect. The advent of suitable microfabrication technology enabled the first SIS mixers to be produced in 1979 [89, 90]. Nowadays, SIS mixers are used in virtually all astronomical heterodyne receivers operating below about 1.3 THz.

Below the superconducting transition temperature, two electrons form so-called Cooper pairs. The SIS mixer is based on a tunneling phenomenon, which can be described by using the energy band representation known from semiconductors (Fig. 5.37a). In this model the density of states has an energy gap 2Δ . On either side of the energy gap, the density of states diverges, because of the formation of Cooper pairs. The gap voltage V_G and the gap frequency ν_G are the corresponding voltage and frequency

$$V_G = \frac{2\Delta}{e} \text{ and } \nu_G = \frac{2\Delta}{h}. \quad (5.69)$$

In the situation where the superconductors on both sides of the barrier are different, 2Δ needs to be replaced by $\Delta_1 + \Delta_2$, where Δ_1 and Δ_2 are half the energy gaps of each of the superconductors 1 and 2. The I-V curve of an SIS device is shown in Fig. 5.37b. When a bias voltage is applied to the device, a steep increase of current occurs when it reaches the gap voltage. At this particular voltage, the divergent densities of states of both superconductor layers cross, and Cooper pairs on one side of the insulating layer break up into two electrons (so-called quasiparticles). If the insulating layer is sufficiently thin, these quasiparticles tunnel from one side of the insulator to the other, where they recombine. In an ideal junction at zero temperature, this would result in an infinitely sharp increase of the current, due to the infinite density of states. In a real device, the nonlinearity is less pronounced, primarily due to the operating temperature of about 4 K, and to imperfections at the superconductor–insulator interfaces. However, the nonlinearity is still sharp and occurs on a voltage scale of a few tenths of a millivolt. This can be compared to the nonlinearity of a Schottky diode, which occurs on a scale in the order of ten millivolts. For quantum effects to occur, the nonlinearity should be on a scale for which $eV \ll h\nu$, which corresponds to $V/\nu \ll 4 \text{ mV/THz}$. This shows that SIS mixers are quantum devices. Above the gap voltage, the I-V curve becomes linear, with a slope proportional to the inverse of the normal state resistance. The tunneling current is modified when one or more photons are absorbed (Fig. 5.37b). In the situation where a multiple of the photon energy equals the difference ($eV - 2\Delta$), an enhancement ($eV < 2\Delta$) or a reduction ($eV > 2\Delta$) of the tunneling current occurs. The quasiparticle tunnels with the absorption or emission of one or more photons. This process is called photon-assisted tunneling. It becomes apparent in the I-V curve as a step-like structure. The steps are clearly visible only if the onset of quasiparticle tunneling is sharp on the voltage scale $h\nu/e$. At low frequencies, or for less sharp I-V curves, the I-V curve approaches the classical limit of a DC I-V curve, which is averaged over the LO voltage swing.

The nonlinearity of the I-V curve at the gap voltage is substantially less than the energy of THz photons. Therefore classical mixer theory is no longer applicable. Instead, quantum effects need to be taken into account. A rigorous treatment of the theory is beyond the scope of this book but can be found in [527]. However, some properties of the mixer, such as its noise and gain, can be qualitatively understood from a simple picture. The basic source of noise is shot noise. The rms current fluctuation due to shot noise can be expressed as

$$\langle \delta I^2 \rangle = 2eIB, \quad (5.70)$$

where B is the bandwidth of the receiver and I the current, which flows through the mixer. If the resistance of the mixer is equal to $R_D = dV/dI$, the equivalent output noise power is

$$P_N = \frac{\langle \delta I^2 \rangle}{4dI/dV} = \frac{e}{2} I B R_D. \quad (5.71)$$

In the Rayleigh–Jeans limit P_N can be expressed as $k_B T_N B$ and the output noise of the mixer is

$$T_N = \frac{e}{2k_B} I R_D. \quad (5.72)$$

It is instructive to compare an SIS mixer and a Schottky diode mixer using (5.72). For an SIS mixer, the output noise temperature is less than 10 K, while for a Schottky diode at 300 K operation temperature it is 200 K (100 K at 4 K temperature). Note that (5.72) describes a device biased with DC. Another contribution to the shot noise originates from broadband shot noise in the signal band, which is down-converted into the IF band. These contributions to the mixer output noise are correlated, and therefore destructive interference can occur. For a properly tuned SIS mixer, the SSB noise temperature due to shot noise can be as low as $h\nu/(2k_B)$.

The gain of an SIS mixer can be considered in a similar way. Mixing between the signal and the LO can be viewed as a small amplitude modulation of the LO at a frequency ν_{IF} . This generates a current I_{IF} corresponding to an available IF power $R_D I_{IF}^2/4$. Since R_D has a maximum at each of the steps of the I–V curve, the gain is largest at these bias voltages.

There is a limitation to the useful bias range. This comes from instabilities which are caused by the Josephson effect and occur at low bias voltage. Therefore, the SIS mixer must be biased above a threshold given by

$$V_T = V_{LO} + K \sqrt{\frac{2h\nu\Delta}{4\pi e\nu R_N C}}, \quad (5.73)$$

where K is a constant close to one [527], R_N is the normal state resistance, and C the capacitance of the junction. Below this threshold, the bias point is unstable and the noise is high. Because an SIS mixer is usually biased at about $h\nu/(2e)$ below the gap, the threshold becomes a problem at high frequencies, and a magnetic field is applied in order to suppress Josephson currents.

In Fig. 5.36, a simplified sketch of an SIS junction is shown. A ground plane is laid down on a substrate (for example, quartz or Si). The SIS trilayer is then deposited on top of the ground plane. While for the first SIS mixers, a Pb alloy was used as the superconductor, now Nb or NbTiN are almost exclusively employed. The geometry of the junction is defined by either ultraviolet or e-beam lithography, and unnecessary parts of the trilayer are etched away. Typical junction areas are $\sim 1 \mu\text{m}^2$. An SiO_2 layer insulates the junction and serves as the dielectric for the wiring and RF tuning circuit on top of the junction.

For a low-noise SIS mixer, it is not sufficient to have an SIS junction with a sharp I–V curve. In addition, the impedance at the input and the output of the mixer needs to be matched, because the noise temperature of the mixer is approximately inversely proportional to the coupling efficiency, and the mixer gain is approximately proportional to it. Although the area of an SIS junction is similar to that of a Schottky diode, its parasitic capacitance is much higher

(typically 50–100 fF compared to ~ 1 fF), because the two superconductors form a parallel plate capacitor with a thin insulating layer. The corresponding impedance $1/(2\pi\nu C)$ is only a few Ω , as distinct from the typical $50\ \Omega$ of RF circuits. Consequently, on-chip tuning circuits are required to compensate for the capacitance, and the proper design of these is a key issue for any SIS mixer. The equivalent circuit of the junction and tuning circuit is shown in Fig. 5.38. The signal source is represented by a current source with an admittance Y_S . The SIS junction is represented by a capacitor in parallel with the junction's admittance at the signal frequency Y_{SIG} . Y_L is a parasitic inductance, which arises mainly from the leads to the junction. Y represents a tuning structure. In general, two factors limit the maximum operating frequency of an SIS mixer. When the photon energy approaches twice the gap frequency ($h\nu > 4\Delta$), tunneling in the reverse direction occurs, and mixing is terminated. In practice, this becomes a serious limitation above about 1.5 times the gap frequency. The choice of superconductor is therefore important. Nb has a gap frequency of about 0.7 THz. Replacing one or both Nb films by NbTiN (gap frequency ~ 1.2 THz) allows the operating range to be extended. It should be noted that the energy gap of a superconductor depends on its critical temperature. A superconducting film in general has a lower critical temperature than the bulk material. The critical temperature, as well as the gap frequency, depends on both the thickness and quality of the film. Inhomogeneities in the film, for example, may decrease the critical temperature. Another limitation comes from the material of the tuning circuit. If it is made from a superconducting material, the losses in the circuit increase above the gap frequency because the photons have enough energy to break Cooper pairs. This causes a part of the incoming radiation power to be dissipated in the tuning circuit, which is then no longer available for the mixing process. The use of a high conductivity metal, for example Al, which is a normal conductor at 4.2 K, or a combination of a metal and a higher gap superconductor, such as NbTiN, are alternatives [529, 530]. The signal bandwidth of an SIS mixer is ultimately limited

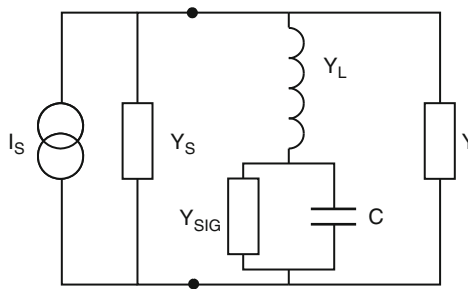


Fig. 5.38 Equivalent circuit of an SIS mixer. The signal source is represented by a current source in parallel with its admittance Y_S . The SIS junction is represented by its admittance Y_{SIG} at the signal frequency in parallel with its geometric capacitance C . The parasitic inductance Y_L arises mainly from the inductance of the leads to the junction, and Y is the total admittance provided by the tuning structure

by its resistance and capacitance product [531]. This product is independent of the junction area, and decreases with thinner barriers because the barrier resistance decreases exponentially while the capacitance increases with the inverse of the barrier thickness. For Nb/ AlO_x /Nb junctions the signal bandwidth is ~ 100 GHz [532]. Using an AlN barrier instead of AlO_x increases this quantity to ~ 300 GHz [533] because the barrier can be made thinner. Below about 1 THz, coupling of LO and signal radiation is usually achieved with a horn antenna and subsequent waveguide, while above this frequency quasi-optical coupling is preferred (Sects. 3.15 and 3.16).

In Fig. 5.35, selected receiver noise temperatures are plotted. Nb-based SIS mixers with Nb wiring yield almost quantum limited, i.e. $T_{\text{REC}} < 3h\nu/k_B$, performance up to the gap frequency of 0.7 THz. Above this frequency devices often have metal tuning circuits, or one Nb film is replaced by an NbTiN film. The conversion loss of a typical SIS-based heterodyne receiver is ~ 13 dB. About 3 dB of the losses are in the optics and ~ 3 dB are in the tuning circuit. The remaining ~ 7 dB are due to the mixing process in the junction itself. The signal bandwidth of an SIS mixer is typically 10–30% of its center frequency, with the larger fractional bandwidth at the low-frequency end. The SIS mixers used in the Heterodyne Instrument for the Far-Infrared (HIFI) on ESA's Herschel Space Observatory are an example. The frequency range from 0.48 to 1.25 THz is covered, without any gap, by just five frequency bands. Each band is equipped with two SIS mixers for two polarizations. Up to 1.12 THz mixers are in a waveguide mount while the 1.12–1.25 THz mixers employ quasi-optical coupling. Although the IF bandwidth of an SIS junction itself is extremely large, due to the inherently fast tunneling process, it is limited in practical applications. A typical IF band is 4–8 GHz.

5.5.4 Hot Electron Bolometer Mixer

Unlike Schottky diodes or SIS mixers, the hot electron bolometer (HEB) mixer is a thermal detector and, as such, belongs to the group of square-law mixers. In order to use a bolometer as a mixer it has to be fast, that is, the overall time constant of the processes involved in the mixing process needs to have a maximum value of the order of a few tens of picoseconds, which corresponds to a bandwidth of up to 10 GHz. This means that the bolometric process is fast enough to follow the IF, but too slow to respond directly to the LO or signal field. In principle, any type of bolometer can be used as a mixer. However, in order to realize sufficient bandwidth, only InSb HEBs (Sect. 5.4.2) and superconducting HEBs have found practical applications. Since the InSb HEB has a bandwidth of typically 2 MHz, its use is very limited. The bandwidth of superconducting HEBs is several GHz and these are frequently employed. Therefore, in this section, the focus will be on superconducting HEBs.

The speed of the bolometer is determined by the ratio $\tau = C/G$ (5.20). Two limiting cases can be distinguished: $2\pi\nu_{\text{IF}}\tau \gg 1$ the bolometer reacts slowly

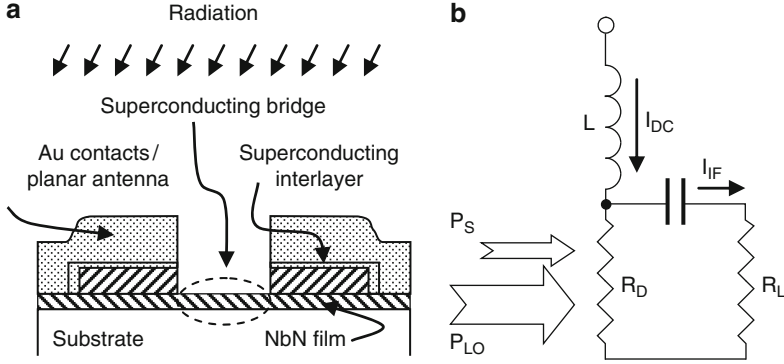


Fig. 5.39 (a) Structure of a phonon-cooled HEB. The superconducting microbridge is indicated by a *dashed ellipse*. The electrical contacts between the microbridge and the Au contacts of the planar antenna are made by a superconducting interlayer. (b) Equivalent circuit of an HEB. R_D is the resistance of the microbridge and R_L is the load resistance

compared to the IF ν_{IF} , and $2\pi\nu_{IF}\tau \ll 1$ the bolometer is fast enough to follow the IF. If the bolometer is to be used as a mixer, the bandwidth must be optimized, which means that for a given IF signal $\tau \ll 1/(2\pi\nu_{IF})$. In other words, high heat conductivity and small heat capacitance are required.

The sensitivity as well as the IF bandwidth required for a mixer, can be achieved with a superconducting microbridge on a dielectric substrate (Fig. 5.39a). The microbridge is made from NbN, NbTiN, or Nb. Several models have been developed to describe and predict the performance of such HEB mixers. A simple model, which nevertheless provides a useful insight into the operating principles, is the so-called uniform heating model. Here, it is assumed that the electrons and phonons in the bridge are heated homogeneously by the incoming radiation (for a complete description of the model, see [534, 535]). Figure 5.39b depicts the equivalent circuit used for illustrating the mixing process. The HEB is biased by a constant voltage. The bath temperature and the absorbed DC and radiation power determine the temperature and resistance of the HEB. The electromagnetic field induced by the power from the LO, P_{LO} at a frequency ν_{LO} , interferes with the field induced by the signal radiation at a frequency ν_S . This gives rise to a modulation of the HEB's temperature, which in turn results in modulation of its current $I_{IF} = \sqrt{P_S P_{LO}} \cos(2\pi\nu_{IF}t)$ through the load resistance R_L . Since this current also passes through the HEB, it gives rise to an additional temperature change. This is the electro-thermal feedback, which also occurs in transition edge sensors (Sect. 5.3.6).

Three parameters are introduced in the uniform heating model: the electron temperature Θ , which equals approximately the critical temperature $T_C \approx \Theta$, the steepness of the resistance $dR/d\Theta$ as a function of the temperature, and the thermal conductance between the electrons and the substrate G_e . In this model, the conversion gain is obtained as

$$G_C \propto \frac{1}{G_e^2} \left(\frac{dR}{d\Theta} \right)^2, \quad (5.74)$$

and the IF bandwidth is proportional to the thermal conductance

$$B_{IF} \propto G_e. \quad (5.75)$$

Several sources determine the noise in a bolometric detector, including thermal fluctuation noise, Johnson noise, and photon noise. Typically, thermal fluctuation noise and Johnson noise are the larger components, and the highest sensitivity is achieved when Johnson noise is minimized and thermal fluctuation noise dominates. For an HEB, thermal fluctuations are determined by fluctuations in the electron temperature. The uniform heating model yields for the noise, referred to the mixer output

$$P_N \propto \frac{\Theta^2}{G_e} \left(\frac{dR}{d\Theta} \right)^2. \quad (5.76)$$

From the above equations, it follows that, in order to have fluctuation noise dominating, $dR/d\Theta$ should be as large as possible, while G_e and Θ should be kept small. However, a small G_e means a small IF bandwidth. This situation is equivalent to a bolometer used for direct detection, where either the sensitivity or the speed can be optimized, but the gain-bandwidth product is constant (Sect. 5.4.1). For a real HEB mixer, the situation is more complicated, because the parameters described above are not constants. G_e is a complicated function of the temperature, and other device parameters such as film quality and geometry. Also, $dR/d\Theta$ depends on the temperature and device geometry.

A more physical model takes into account Cooper pairs, quasiparticles, and phonons in the microbridge, as well as phonons in the substrate. The incoming radiation breaks Cooper pairs and the quasiparticles diffuse to the electrical contacts, or they interact with the phonons in the microbridge, which finally escape into the substrate. This mechanism is depicted schematically in Fig. 5.40. If the length of the microbridge is short compared to the thermal diffusion length $L_{th} = \sqrt{D\tau_e}$ (electron diffusivity D , electron-cooling time τ_e), the cooling process occurs predominantly by out-diffusion of electrons within a time $\tau_{diff} = (L/\pi)^2/D$. HEB mixers based on this principle are called diffusion-cooled. Cooling by phonons dominates for $L \gg L_{th}$. These mixers are called phonon-cooled or lattice-cooled. The time response of a thin NbN film, which is a common material for phonon-cooled HEB mixers, is described by an average electron-cooling time τ_e , which is given by $\tau_e = \tau_{ep} + (1 + C_e/C_{ph})\tau_{esc}$ with the electron–phonon energy relaxation time τ_{ep} and the escape time of phonons from the film into the substrate τ_{esc} [536]. The term C_e/C_{ph} consisting of the electron and phonon specific heat, C_e and C_{ph} , respectively, takes into account the backflow of energy from phonons to the electrons, which is described by the phonon–electron energy relaxation time $\tau_{pe} = \tau_{ep}(C_{ph}/C_e)$. It should be pointed out that this distinction is to some

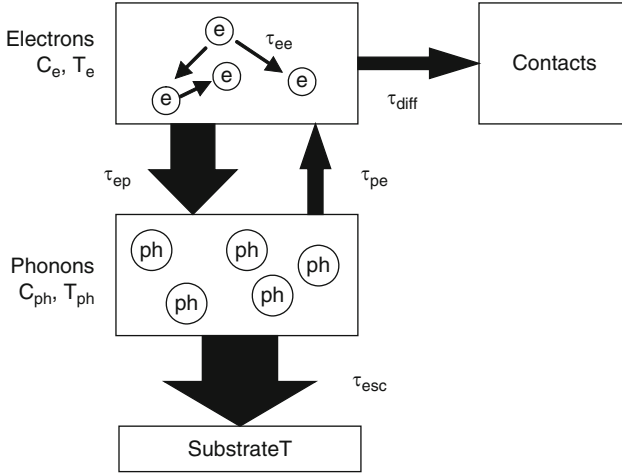


Fig. 5.40 Thermalization scheme of a phonon-cooled HEB mixer. The abbreviations are explained in the text. The subsequent channels of the energy relaxation are shown by bold *arrows*, with the size indicating roughly the relative contribution to the overall relaxation. In a phonon-cooled HEB, the electron–phonon interaction, and the escape of phonons to the substrate, dominate the processes. In a diffusion-cooled HEB, the outdiffusion of electrons to the contacts is dominant (adapted from [536])

extent arbitrary, since phonon-cooling also exists in diffusion-cooled mixers and, vice versa. It is the technical realization, that is, the choice of superconductor and geometry of the bridge, that makes the difference.

For more detailed modeling, the nonhomogeneous temperature distribution along the superconducting bridge needs to be taken into account (so-called hot spot model). This model was originally developed by Skocpol et al. [537], and was later applied to superconducting HEB mixers [538–540]. The power dissipated by the LO radiation and the DC bias current creates an electron temperature profile along the bridge. The region where the actual temperature exceeds the critical temperature and switches into the normal state is called the hot spot (Fig. 5.41). The rest of the bridge remains superconducting. Although the hot spot and the uniform heating model describe the static characteristics of an HEB equally well, the hot spot model is advantageous when it comes to the dynamic properties of the HEB. In the uniform heating model, the reciprocal electron cooling time, modified by Joule heating and electro-thermal feedback, determines the bandwidth of the device, while in the hot spot model this role is taken by the speed with which the boundaries of the hot spot move. When radiation is absorbed, the length of the hot spot increases, and its boundaries begin to move toward the electrical contacts, until the hot spot reaches thermal equilibrium. It is the speed of the boundaries that determines the response time of the HEB. When the power absorbed by the HEB is modulated

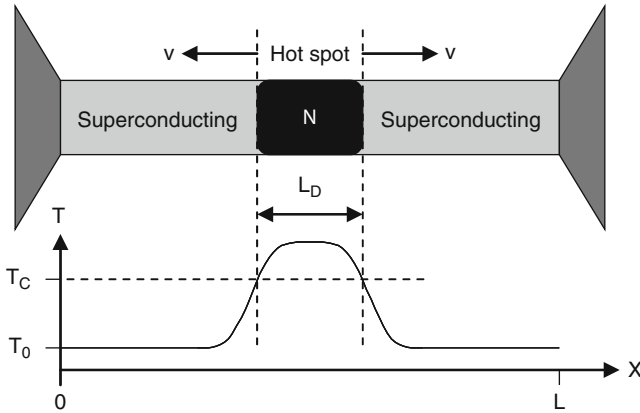


Fig. 5.41 Schematics of the hot spot and the electron temperature profile of a bridge of length L . The hot-spot is formed where the actual temperature in the microbridge exceeds the superconducting transition temperature T_C . The normal conducting hot-spot (N) is represented as the movement of its boundaries with velocity v (adapted from [540])

with a frequency smaller than its response time, the boundaries fluctuate at a rate corresponding to this frequency.

When comparing diffusion-cooled and phonon-cooled bolometers and HEBs, it turns out that the latter has a smaller noise temperature and is less sensitive to LO power fluctuations and variations of the bias voltage. Phonon-cooled HEB mixers are fabricated from thin superconducting NbN films [541] with a nominal thickness of 3–6 nm. The relevant time constants of the films are sufficiently small. The energy of an absorbed photon is distributed among the electron subsystem within a typical time τ_{ee} of about 7 ps. The electrons share the energy with the phonons within the electron–phonon interaction time τ_{ep} . This is a function of the electron temperature Θ (in Kelvin) and has been empirically determined to be $\tau_{ep} \approx 500\text{ps}/\Theta^{1.6}$. For an electron temperature of 10 K, this yields $\tau_{ep} \approx 10$ ps. The heat is transported by the phonons to the substrate. The phonon escape time τ_{esc} depends on the thickness of the film d , the speed of sound in the film u , and the acoustic matching between film and substrate, related to the transparency factor ζ :

$$\tau_{esc} = \frac{4d}{\zeta u}. \quad (5.77)$$

For a 4-nm-thick NbN film $\tau_{esc} \approx 30$ –60 ps. As can be seen, the cooling process is dominated by the escape of phonons into the substrate and, to a lesser extent, by the electron–phonon interaction. In order to obtain a precise description of the cooling process, a two-temperature model needs to be used, where one temperature describes the electron system and the other the phonon system [542].

NbN films for HEBs are deposited on a dielectric (typically Si of $>10\text{ k}\Omega\text{ cm}$) by DC reactive magnetron sputtering [543, 544]. After deposition, the films have a room temperature square resistance of about $500\ \Omega$ at 300 K, which drops abruptly to almost zero (residual resistance less than $1\ \Omega$) at the transition temperature close to 10 K. The processing during the device fabrication leads to a degradation of superconductivity, that is T_C degrades to about 9 K, and the transition width increases to about 0.5 K. The superconducting bridge is obtained by means of electron beam lithography. The length varies between 0.1 and $0.4\ \mu\text{m}$ and the width between 1 and $4\ \mu\text{m}$. Between the contact structure and the NbN film, a superconducting interlayer is used to decrease the contact resistance. HEB mixers can be made either in a waveguide configuration or as quasi-optical mixers. In the case of a quasi-optical mixer, the microbridge is embedded in a planar antenna, which is lithographed from an evaporated gold film. As described in Sect. 3.15, the Si chip with microbridge and antenna is glued to the rear face of an elliptical or extended hemispherical lens. The resistance-temperature curve of an HEB mixer consists of three parts: the superconducting region, the transition region, and the resistive region. These regions can also be found in the I-V curve (Fig. 5.42). The resistance seen in the superconducting state is due to the non-superconducting circuitry.

DSB noise temperatures achieved with HEB mixers range from 400 K at 0.6 THz up to 3800 K at 5.2 THz (Fig. 5.35). Up to 2.5 THz, the noise temperature closely follows the slope $10h\nu/k_B$, with somewhat worse values above 3 THz. This is caused by increasing losses in the optical components and lower efficiency of the hybrid antenna. Nonuniform distribution of the current in the bridge, due to the skin-effect, may also contribute, because it changes the IF matching [545].

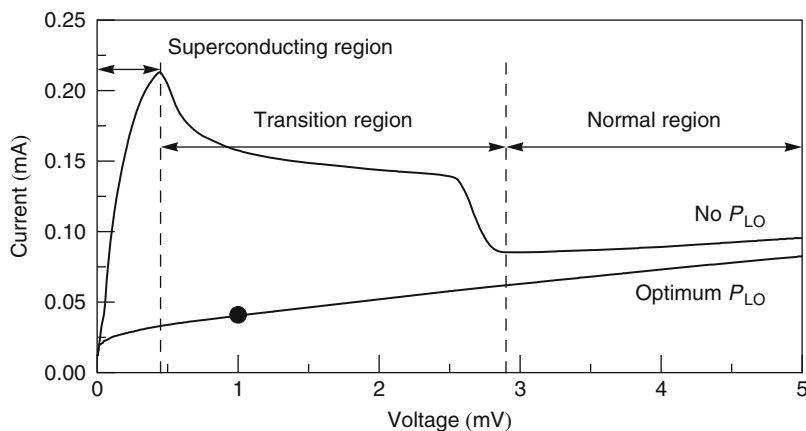


Fig. 5.42 I-V curves of an NbN HEB mixer. The upper line shows the I-V curve when no LO power is applied. The superconducting region, the transition region, and the region where the mixer is in normal conduction are indicated. The resistance seen in the superconducting state is due to the non-superconducting circuitry. The lower line corresponds to LO power for optimum sensitivity. The dot indicates the optimum bias operation point for the HEB mixer

At all frequencies phonon-cooled HEBs have a lower noise temperature than that of diffusion-cooled devices. Compared with Schottky and SIS mixers, HEBs are significantly more sensitive than the former, but somewhat less sensitive than the latter. The gain bandwidth of NbN HEB mixers is 3–4 GHz. Since thermal fluctuation noise and conversion gain have the same roll-off frequency, the noise bandwidth is larger, up to ~ 5 GHz. It is limited by Johnson noise, which increases rapidly at IFs larger than $1/\tau_e$.

A particularly attractive feature of an HEB mixer is the low LO power required for optimum operation. The LO power scales with the volume of the microbridge, and decreases with increasing T_C . The power absorbed in the microbridge is as low as 20 nW for the smallest devices ($0.1 \times 1 \mu\text{m}^2$) and increases to several hundred nanowatts for larger devices. However, these numbers are indirectly measured by comparison with DC power dissipated in the microbridge. In addition, they do not take into account antenna coupling losses, as well as losses in the Si lens and vacuum window. Therefore, a more practical number is $\sim 1 \mu\text{W}$. Since an HEB needs so little LO power, and because it also works very well as a direct detector, care has to be taken to ensure that there is no direct detection contribution to the mixing signal arising from broadband background radiation. Direct detection manifests itself in a shift of the operating point of the HEB, and consequently a change in its sensitivity. In addition, the shift of the operating point appears at the output of the mixer as a contribution to the hot–cold difference signal, wrongly suggesting a better noise temperature. The magnitude depends on the power of the background signal, the signal bandwidth of the antenna, and the volume of the HEB. A low background signal can be achieved by appropriate cold filters in the signal path. Mixers with narrow band antennas, such as the twin slot antenna, are usually less affected by the direct detection effect than mixers with broadband antennas. The smaller the volume of the superconducting bridge the more susceptible it is to direct detection. Nevertheless, it has been shown that, for a mid-size HEB ($0.2 \times 2 \mu\text{m}^2$) with appropriate filtering, the response of the HEB is linear up to about 400 K background temperature [546].

Because of their sensitivity, NbN HEB mixers are nowadays the first choice for heterodyne spectroscopy above ~ 1.3 THz if cryogenic cooling is allowed. A waveguide NbN HEB mixer was used for the first earth-based astronomical heterodyne observation above 1 THz, and NbN HEB mixers are used in the high frequency channel of HIFI on the Herschel Space Observatory, as well as in the heterodyne receiver GREAT on SOFIA. Another interesting application of these detectors, found due to their small response time, is for detection of THz synchrotron radiation. Using a NbN HEB, in direct detection mode, it became possible to measure the fill pattern as well as single THz pulses from a synchrotron source [400].

Chapter 6

Spectroscopic Methods

6.1 Introduction

Historically, there have been eras where a particular spectroscopic method has dominated IR and THz research. At the end of the nineteenth century and into the early twentieth century, prisms were employed, with the alternatives being wire transmission gratings, or interferometers when only simple spectra were involved (Sect. 1.2). This situation changed with the introduction of the very efficient blazed reflection grating in 1910, which had the ability to concentrate a large percentage of the light into a single order [15]. For the next fifty years, grating spectroscopy was the dominant technique at frequencies below the range covered by prisms. The next change occurred with the availability of digital computers from the late 1950s onward. This allowed interpretation of the complex signals produced by interferometers when many frequencies were involved. Interferometers have inherent advantages compared to dispersive systems, and became the primary tool for routine spectroscopy. Time-domain spectroscopic systems, and an increasing number of narrow-band tunable sources, are now supplementing conventional spectroscopic methods in an increasing number of applications.

The change-over from grating instruments to interferometers for THz research was rapid and even led one researcher to suggest that the place for the large and rather beautiful reflection gratings was in a museum showcase [547]. But recent developments have changed the situation again, and there is renewed interest in the use of grating spectroscopy. To understand the reasons for the change from grating spectroscopy to interferometric systems, and the subsequent revival of the former technique, it is convenient to consider the simple dispersion system illustrated in Fig. 6.1.

Assuming that the source is thermal, thus producing a continuous range of frequencies, the dispersive element will select a specific frequency, and this will be focused onto the exit slit and then reach the detector. If the dispersion is produced by a prism or diffraction grating, rotation of this will select, in turn, a succession of frequencies falling on the detector slit. Two features of this optical system are

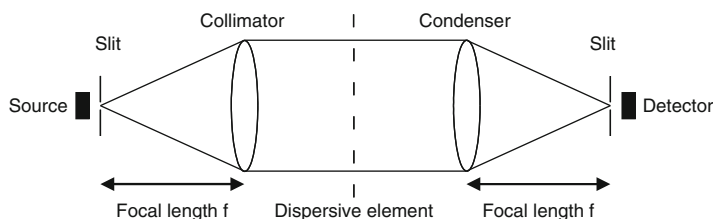


Fig. 6.1 Dispersion system consisting of source, collimating optics, dispersive element, slits, and detector

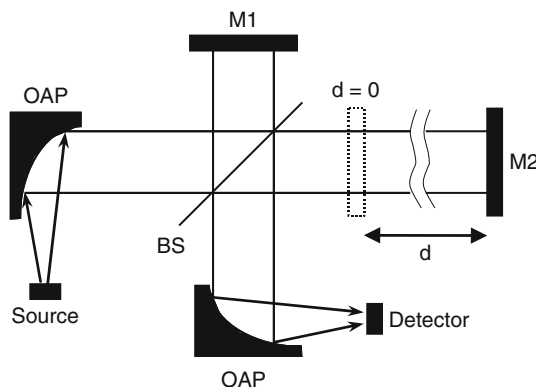


Fig. 6.2 The principle of a Michelson interferometer with beam splitter (BS), mirrors M1 and M2, and off-axis paraboloidal mirrors (OAP). With mirror M2 at position $d = 0$, the zero path length difference position results in the maximum signal

significant. Firstly, the amount of light from the source that reaches the detector will depend on the area of the slits A and the solid angle Ω accepted by the optics. The combination $A\Omega$ of these parameters defines what is known as the throughput or étendue of the system. Secondly, frequencies above and below the central frequency that is passing through the second slit will be focused on either side of this exit slit. In visible and near-IR spectroscopic systems, these frequencies can be detected simultaneously by replacing the exit slit with a CCD line detector. Such an arrangement is said to have a multiplex advantage in that it will detect a number of frequencies simultaneously. As suitable multi-element detector arrays are now available, multiplexing with dispersive systems has been extended to THz frequencies.

Unlike grating instruments, interferometers are inherently multiplexing systems, because information from all frequencies from the source reaches the detector simultaneously, assuming that no frequency selective filters are in place. This can be readily understood from the schematic diagram of a Michelson interferometer, as shown in Fig. 6.2.

If the source produces a single frequency, it is well known that, as mirror M2 is moved, the signal at the detector will fluctuate between a maximum value (at zero path length difference) and zero as the distance d is varied, assuming that the beam

splitter is 50% transmitting and 50% reflecting. If the source is not monochromatic, the detector will receive a summation of all the frequencies involved. The initial drawback of interferometric systems was that, unlike dispersive systems, they do not give the spectrum directly, rather they give its Fourier-transform, and an inverse Fourier transformation is needed to provide the desired spectrum. Both Michelson and Rayleigh were aware that the spectrum from an interferometer could be obtained by Fourier transformation of the interferogram produced by this type of instrument, and Michelson had considerable success in interpreting spectra with an 80-channel mechanical analogue computer [44]. However, it was P. Fellgett, researching on the IR spectrum of stars for his PhD, who actually derived spectra by the Fourier interferometric method [548], and the move from dispersive to interferometric spectroscopy then mirrored the progress in the development of digital computers.

A further reason for using interferometers, particularly in the THz region, where signals are often very weak, can be understood by comparing Figs. 6.1 and 6.2. With a dispersive system, if one wants to select a narrow frequency range, the slit widths must be decreased, thus reducing the throughput of the instrument. The interferometer has cylindrical symmetry, with the light entering and exiting through circular apertures, and to a first approximation its resolving power is dependent on the distance of travel of its moving mirror. Quite large area apertures can be used, giving a very significant throughput advantage. This property of interferometers was first pointed out by Jacquinot [549], and is sometimes referred to as the Jacquinot advantage. For a similar reason, multiplexing instruments are sometimes said to have a Fellgett advantage. From the above comparison of dispersive and interferometric spectroscopy, it can be seen that, even though detector arrays for the former systems overcome their multiplex disadvantage, interferometers still gain due to the increased throughput. It was the combination of both advantages that caused the interferometer to become the dominant spectroscopic technique. However, as early as 1949 it had been pointed out that, by replacing the slits of a dispersive monochromator with a patterned grid structure, it was possible to obtain both a multiplex and throughput advantage, even when using a single detector. M. Golay, who showed that such an arrangement was possible, even built his own simple electronic computer to obtain spectra with improved (S/N)-ratio [45]. From the late 1960s onward, a number of dispersive instruments have been designed using grid structures to increase the throughput (Fig. 6.7). However, these never really challenged the dominance of interferometers for routine spectroscopy. This was because of the complexity of design needed to obtain both the multiplex and throughput advantages simultaneously. More recently, the availability of THz detector arrays has led to much simpler instruments and, as mentioned earlier, there is renewed interest in the use of dispersive systems.

A feature of most dispersive and interferometer systems is that mechanical scanning elements are required. However, with a dispersive instrument and a detector array, a relatively simple static system to cover a useful frequency range can be envisaged. This is a further reason for the revived interest in dispersive spectroscopy. For space-borne applications, systems with no moving parts are especially advantageous (Sects. 6.2 and 6.3).

The principal drawback of a Fourier-transform spectrometer (FTS) and grating spectrometers, as compared to spectrometers based on narrow-band cw radiation sources, are the much lower resolution and the smaller sensitivity because of the limited power per frequency interval. Such spectrometers are operated in a rather similar way to tunable laser spectrometers in other frequency ranges. The frequency of the radiation source is swept across an absorption feature and the transmitted source power is detected by an appropriate THz detector. Often, some modulation technique is applied in order to increase sensitivity and to minimize baseline fluctuations in the recorded spectrum. A first step toward sweeping spectroscopy with narrow-band sources was done in the early 1950s when W. Gordy and co-workers succeeded in extending the emission of microwave sources into the THz region by harmonic generation. In 1954, they reached the THz region by measuring a rotational transition of OCS at 0.389 THz [550]. With further improvements, they were able to extend this to 0.813 THz in 1970 [551]. With modern multiplier-based sources 2.6 THz is now within reach [552].

Pioneering work on spectroscopy with backward wave oscillators (BWOs) was started by A.F. Krupnov et al. [553]. They used BWOs and a Golay cell detector for a sweeping spectrometer that operated to frequencies slightly above 1 THz. Among their first achievements was the measurement of the rotational lines of PH_3 in the region of 0.3 to 1.07 THz. Spectroscopy with BWOs is now routine up to ~ 1.5 THz [554].

In order to explore the frequency range beyond 1.5 THz different mixing approaches were invented. In 1984, K.M. Evenson and co-workers introduced a frequency tunable THz spectrometer based on mixing the radiation from two grating-tunable CO_2 lasers with microwave radiation [355]. Due to the large number of CO_2 laser lines, a significant part of the THz range up to 4.5 THz has been covered [555]. The frequency coverage can be further extended to 9 THz by replacing one of the CO_2 lasers with a $^{15}\text{NH}_3$ laser [358]. An alternative approach is laser sideband generation, which is the mixing of THz radiation from an optically pumped gas laser with a frequency tunable microwave or mmW source in a Schottky diode [350, 351]. Mixing of visible or near-IR radiation from two diode lasers in a nonlinear element such as low-temperature grown GaAs, a technique pioneered by E. Brown and co-workers [335], had led to the development of compact cw spectrometers with continuous frequency coverage beyond 2 THz. These spectrometers are commercially available.

The alternative to conventional spectroscopy is THz time-domain spectroscopy (TDS), which is discussed in detail in Sect. 6.6. This technology became available with the advent of reliable sources of femtosecond, short-wavelength laser pulses. When these pulses are incident on an absorbing semiconductor charge carriers are created. Acceleration of these carriers, in an electrical bias field, produces a transient photocurrent that causes electromagnetic waves to be radiated. The emitted frequency is in effect the inverse of the pulse length and is typically centered between 0.3 and 1.5 THz. Higher frequencies can be obtained by replacing the semiconductor with an appropriate nonlinear crystal, in which optical rectification is produced by the laser pulse. With the photoconductive method, real charge carriers are produced and the upper frequency is limited by their mobility. In optical

rectification, the limitation is the laser pulse length. However, the power levels are often lower than with the semiconductor process.

Remote sensing in astronomy and earth observation are among the most important applications of THz molecular spectroscopy. Essentially, this is done either with grating and FT spectrometers, in situations where low and medium resolution is sufficient, or with heterodyne spectrometers. The latter method is used for very high-resolution spectroscopy (resolving power $>10^6$), where the line shape of the molecule under investigation needs to be resolved. This is particularly important for astronomical observations, where the emission lines are Doppler-broadened, and spectrally resolved measurements (in the order of 1 MHz) enable the acquisition of information concerning the dynamical behavior of the object. In atmospheric research, the spectral resolution has to be even better (less than ~ 100 kHz), because the pressure broadening of the observed line contains the information about the height distribution of the molecule. In astronomy, heterodyne spectroscopy made its breakthrough with the detection of the 21 cm^{-1} HII line at 1420.4 MHz in 1951 [556, 557]. Following this experiment, the study of microwave, and later the submm/THz emission spectra of interstellar molecules, became a rapidly advancing field of astronomy. Heterodyne spectrometers were rapidly developed and improved. Initially, mixer technology relied on semiconducting InSb hot electron bolometers (HEBs) and Schottky diodes, with superconducting SIS and HEB mixers now prevailing for high-end applications. Up to 2 THz, local oscillators are predominantly multiplier-based sources, while above 2 THz optically pumped gas lasers have been used. The latter are likely to be replaced by quantum cascade lasers. Initially, ground-based telescopes operating in the 0.3 and 0.6 THz atmospheric windows were used, but from 1977 to 1993 the Kuiper Airborne Observatory, equipped with THz heterodyne spectrometers, was available. Observations with these instruments opened the view into the THz universe with very high spectral resolution. The development of heterodyne spectrometers eventually culminated in the HIFI heterodyne spectrometer on the Herschel Space Observatory, which has provided a wealth of exciting data [558].

6.2 Grating Spectrometer

Although, as mentioned above, most diffraction grating instruments for THz frequencies employ blazed reflection gratings, it is convenient to use the simple system illustrated in Fig. 6.1 to discuss the essential elements of this type of spectroscopy. Assuming that the dispersive element is a transmission grating, and that it has been rotated to select a central frequency ν from the source, then the second slit will allow a small frequency band $d\nu$ to reach the detector. The ratio $\nu/d\nu$ is the resolving power of the instrument.

It is well known that the theoretical resolving power of a diffraction grating is given by nN , where n is the order number and N is the total number of illuminated lines on the grating. Gratings for THz frequencies are typically up to 200 mm wide and one for use at ~ 3 THz will have approximately 10 lines per mm width. Thus,

the theoretical limit in resolving power is $\sim 2,000$, giving a maximum resolution of 1.5 GHz. There is no advantage in working at higher orders because, for a specific frequency, the number of lines per unit width is proportional to $1/n$. Using larger gratings, and thus more lines, is possible but very inconvenient because the high absorption by water vapor makes evacuated systems essential at most THz frequencies, and instruments then become very bulky.

Unfortunately, even the relatively low theoretical resolution of THz grating spectrometers is difficult to achieve with the thermal sources used for broadband spectroscopy. A reasonable approximation for the spread of frequencies $d\nu$ reaching the detector, as shown in Fig. 6.1 is

$$d\nu = s\nu/f, \quad (6.1)$$

where f is the focal length of the instrument and s is the width of each slit [160]. Equal slit widths and equal focal lengths of the input and output optics are normally required to optimize the product of throughput and resolving power in a grating monochromator. Assuming again that $\nu = 3$ THz, the slit width required for the theoretical limiting resolution of 1.5 GHz is 0.5 mm when the focal length is 1,000 mm. Even with very sensitive helium-cooled detectors, the (S/N) -ratio is low with such narrow slits, particularly when studying materials with significant absorption. But high resolution is not always required and a great advantage of grating spectroscopy is its simplicity. Early THz instruments were inefficient as they used simple transmission diffraction gratings that distributed the incoming radiation into many orders, of which only one was used. Rayleigh pointed out in 1874 that in theory it was possible to design a grating where the majority of the radiation would be diffracted into a single order [16]. Rayleigh appears to have been suggesting a transmission grating made of a large number of very narrow prisms set side by side. As mentioned in the introduction to this chapter, R. W. Wood realized that such a grating could be made, if it was reflecting rather than transmitting, by ruling a metal surface with grooves set at an angle [15].

Wood constructed reflection gratings in which more than 75% of the light at a specific wavelength could be concentrated into a single order. This high efficiency can be understood by reference to Fig. 6.3, which illustrates the diffraction that would be obtained with a transmission grating, except that here the diffracted light is in the reverse direction.

For constructive interference, light is diffracted from a grating with spacing d at an angle β when

$$d \sin(\beta) = n\lambda, \quad (6.2)$$

where λ is the wavelength and n is the order number of the light. In the situation shown in Fig. 6.3b the grating has been rotated through an angle θ , to bring the long faces of the grating perpendicular to the incoming light. All orders are then diffracted back onto the path of the incoming radiation. In effect, the light is “seeing” the grating as a plane mirror when

$$2d \sin(\theta) = n\lambda. \quad (6.3)$$

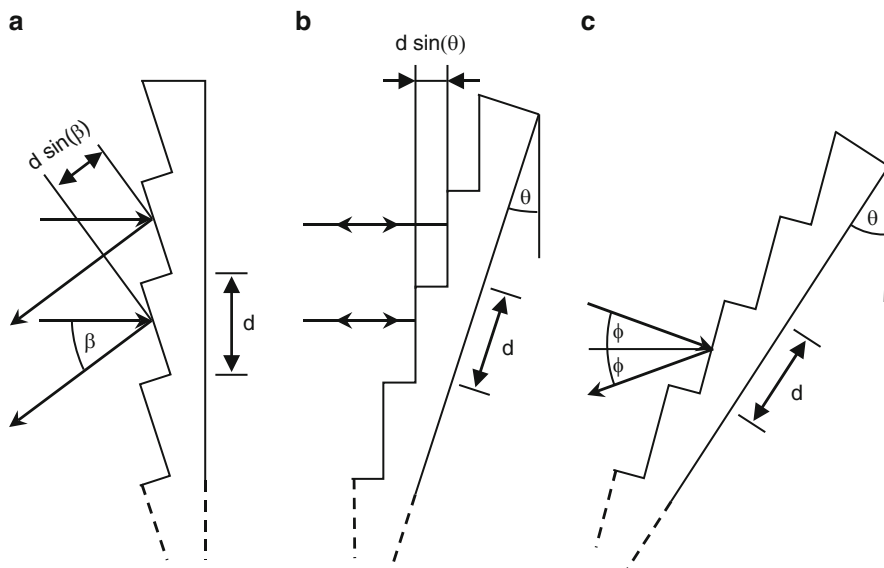


Fig. 6.3 Incident and reflected light from a blazed diffraction grating (a) shows that the diffraction of light by a reflection grating is analogous to that of the more familiar transmission grating, (b) illustrates the situation where the grating is rotated through its blaze angle ($\theta = \theta_B$) when, in theory, all orders will be specularly reflected with close to 100% efficiency, (c) when rotated to an arbitrary angle the grating operates with reduced efficiency but this is still above 50%, from approximately 0.75 to 2 times the blaze wavelength in the 1st order

Wood described his gratings as “echelettes” but they were later called blazed gratings, with θ_B being the blaze angle.

In virtually all diffraction grating spectrometers, the optical components are in fixed positions, and it is the rotation of the grating which disperses a specific wavelength onto the exit slit. The usual arrangement is for the incoming light to fall on the grating at a small angle from the normal, with the outgoing light being dispersed at the same angle on the opposite side of the normal. This is shown in Fig. 6.3c and then (6.3) is modified to

$$2d \sin(\theta) \cos(\Phi) = n\lambda, \quad (6.4)$$

where Φ is the half angle between the incident and diffracted beams. Φ is normally small, typically 5° – 10° , and (6.4) is often referred to as the “grating equation.”

It is relevant to consider the requirements of gratings for the THz region. Assuming, for convenience, that the gratings are to be used at $\theta = 30^\circ$ and that $\cos(\Phi) \approx 1$, then the spacing needed for 0.3–10 THz varies from 1 to 0.03 mm. Gratings with these rulings, of the type shown in Fig. 6.3, can be made using conventional workshop machines. As the reflectivity of both pure metals and alloys is high at THz frequencies, it is straightforward to manufacture quite high quality gratings by choosing a material, which does not tear when cut.

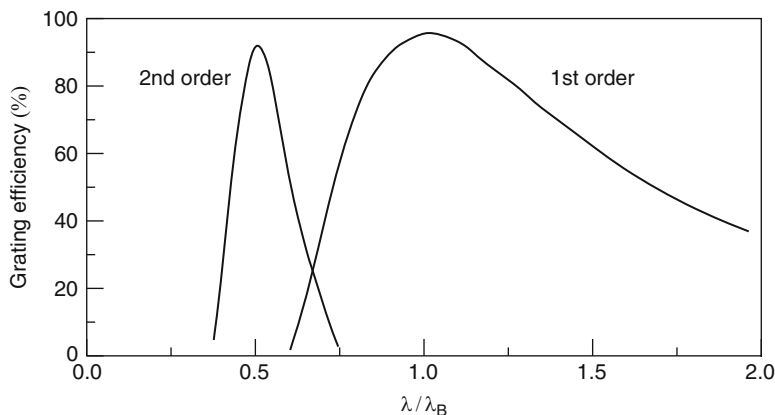


Fig. 6.4 Efficiency of a diffraction grating which is blazed for a wavelength λ_B

While it is clear from (6.3) that all orders will have high efficiency at the blaze angle, most spectroscopic applications require the covering of a significant wavelength range, and it is necessary to consider the grating efficiency as it is rotated away from the blaze angle. The theory of diffraction by blazed gratings is complex but can be simplified using certain assumptions known as the scalar approximation [559]. Figure 6.4 illustrates the theoretical performance of a grating, which is blazed for a wavelength λ_B at which the first-order efficiency peaks. The figure shows that for wavelengths near λ_B most of the incident power falls into the first order and little into the second (or higher) orders. In practice, many gratings are significantly less efficient, and there is normally a marked variation with different polarizations of the incoming light. Experimental results with the grating of Fig. 6.4, using unpolarized light, have shown that more than 50% of the incoming light was diffracted into the first order over a frequency range of more than an octave. While, at the blaze wavelength λ_B reflection gratings have high efficiency at all harmonics of the fundamental frequency, the performance falls off more rapidly on either side of the blaze angle for the higher harmonics. This is illustrated in Fig. 6.4 for the 1st harmonic of the fundamental frequency. At higher orders, the useful wavelength range becomes very small.

The diffraction of parallel and perpendicularly polarized light from a blazed grating has been studied in detail using numerical methods [560]. The authors state that the characteristics of diffraction gratings can be predicted, thus avoiding the necessity of manufacturing the grating and then measuring its performance. Blazed gratings normally have right-angle grooves (Fig. 6.3), but it has been shown that by using rulings with 110° – 120° angles between the facets, polarization effects can be greatly reduced [561].

The variation in efficiency between the two polarizations increases when gratings are used at higher angles. In some circumstances, this can be advantageous. For example, one important application of blazed gratings is in optically excited THz gas

lasers (Sect. 4.4.2). The most frequently employed pump is a CO₂ laser operating between 9 and 11 μm . To select a specific rotational line from the pump laser, one of the laser mirrors is replaced by a reflection grating and this requires a high reflectivity. To achieve a high resolving power, in order to excite only one of the closely spaced rotational levels, a grating with a small period is used at a large angle. An example of such a grating is shown in Fig. 6.5. For the transverse magnetic (TM) polarization, the efficiency is close to 100% over the required wavelength range.

Reflection gratings allow the use of relatively convenient optical designs, and the one most frequently employed is that originally suggested by Czerny and Turner [562] and illustrated in Fig. 6.6a. This arrangement is self-correcting as far as spherical aberrations are concerned, and so allows the use of cheap spherical mirrors rather than the parabolic ones required in most other systems.

An alternative to the Czerny–Turner arrangement is the Ebert–Fastie design shown in Fig. 6.6b. This originally dated back to 1889 [563] but its virtues,

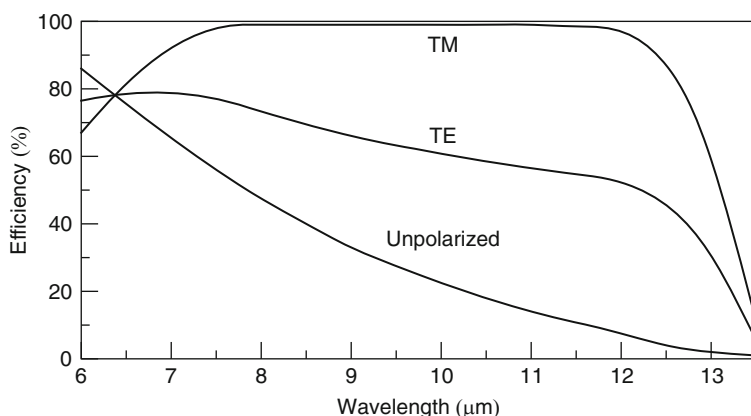


Fig. 6.5 Efficiency of a 150 line/mm 52.7° blazed grating for unpolarized radiation, transverse magnetic (TM) modes without magnetic and transverse electric (TE) modes without electric field in the direction of propagation

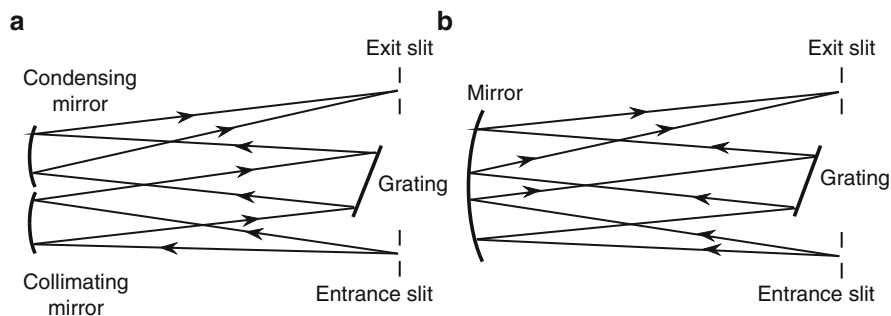


Fig. 6.6 (a) Czerny–Turner monochromator and (b) Ebert–Fastie monochromator

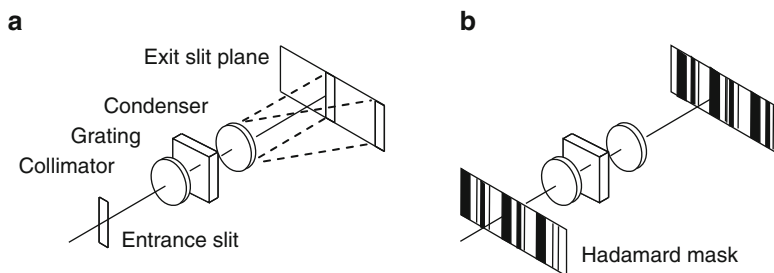


Fig. 6.7 (a) Principle of multiplexing grating spectroscopy, (b) multiplexing by using a Hadamard mask and a linear detector array

particularly for long-wavelength spectroscopy, were not fully realized until 1952 [564]. Because it employs a single, large mirror there is an advantage, compared to Czerny–Turner, in that it can have very long slits that provide increased throughput. The disadvantage is that instruments are usually inconveniently large.

It is possible to incorporate both throughput and multiplexing advantages into dispersive instruments. In a conventional grating spectrometer, with a broadband source and a single detector, only one wavelength is detected at one time. However, in the visible and near-IR, if the detector is replaced by a photographic plate, a CCD or a CMOS camera, and the exit slit is removed, wavelengths on either side of the central wavelength can be detected simultaneously [565]. Many of the THz detectors described in Chap. 5 are now available as linear arrays and, as illustrated in Fig. 6.7, these in effect allow the replication of the photographic plate of short-wavelength systems. Multiplexing, with increased throughput, can be obtained if the single entrance slit is replaced by a patterned mask, and where a linear detector array is employed. This is shown in Fig. 6.7. With the single entrance slit, two different frequencies are shown to illustrate dispersion by the grating. If the single slit is replaced by a patterned mask that is fully illuminated by the source, each transparent slit produces the same spectrum as the single slit, but the image on the detector array is a convolution of the coded-mask pattern and the spectral profile of the light falling onto the mask. In order to retrieve the spectral profile, the spectrum observed on the array is inverted. Encoded mask systems are generally referred to as Hadamard transform spectrometers, because the most useful mask structures have been found to be Hadamard matrices.

If only a single detector is available, it is necessary to encode both slits of a dispersive instrument to obtain the throughput and multiplex advantages simultaneously, as illustrated in Fig. 6.7b. More specifically, the grid structure at the entrance of the spectrometer produces an initial stage of multiplexing, and a second stage follows at the exit plane [566]. While the (S/N) -ratio of dispersive spectrometers of this type is comparable with interferometers, the complexity of the design is significantly greater and relatively few applications have been published. The real interest is in Hadamard instruments with detector arrays. Static spectrometers with a $\pm 8\%$ wavelength range, and a resolving power of more than

1,000, have been designed for the near-IR region [567], and there seems no reason why THz instruments should not follow.

6.3 Fourier-Transform Spectrometer

As mentioned in the introduction the principles of Fourier-transform spectroscopy were discovered by Michelson [568] and Rayleigh [569]. They realized that the interference pattern obtained from a two-beam spectrometer as the path difference between the two beams was altered, was the Fourier-transform of the light illuminating the interferometer. Michelson, Rubens and Hollnagel [52], and Rubens and Wood [14] all derived simple spectra, but with complicated spectra they had no means of performing the necessary Fourier transformation.

It was another forty years before the brilliant work of Fellgett [548] led to the foundation of modern interferometric techniques by the discovery of the multiplex principle. To Fellgett it was clear that the rate of obtaining information from spectrometers was energy limited and that there was much to be gained by utilizing what is, in effect, the spare channel capacity of the detector.

Fourier-transform spectrometers (FTS) are in widespread use. Typically, they employ spectral sources with broad frequency coverage, such as globars, high-pressure lamps, or synchrotron sources. The latter combines a source of high brilliance with broad frequency coverage when used as a standard research tool, typically combined with a microscope for spatial resolution from the optical down to the THz range.

Initially, the basic equations relevant to interference phenomena are discussed. The first interference patterns (colors in thin plates) were observed by R. Boyle and R. Hooke in the seventeenth century. In 1801, T. Young described the phenomenon of interference qualitatively. Finally, in 1818, A.J. Fresnel combined the concept of interference with Huygens principle of elementary waves.

The electric field $E(z, t)$ of a linearly polarized plane wave propagating in the z -direction in its complex representation can be written as

$$E(z, t) = E_0 \exp(i(\Phi(z) - 2\pi\nu t)) \quad (6.5)$$

with the phase $\Phi(z) = 2\pi\nu n z/c + \Phi_0$ at a time t , and refractive index n , electric field amplitude E_0 , and frequency ν . The intensity $I_0 = |E|^2 = E^* E = E_0^2$ of a single plane wave corresponds to an average energy stream perpendicular to a given surface, calculated by the product of the field amplitude E and the complex conjugate E^* . For two different, but parallel, waves 1 and 2, this leads to

$$I = |E_1 + E_2|^2 = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\Phi_1 - \Phi_2). \quad (6.6)$$

The maximum intensity is found for a phase difference of both waves of $\Phi_1 - \Phi_2 = m\pi$ when m is an even number, and the intensity is at a minimum when m is odd.

In this case the intensity is independent of position but, if the second wave is tilted by an angle α , the field E_2 is obtained

$$E_2(x, z, t) = E_{02} \exp \left(i \left(\frac{2\pi\nu}{c} (x \sin(\alpha) + z \cos(\alpha)) + \Phi_{02} - 2\pi\nu t \right) \right). \quad (6.7)$$

The time averaged intensity leads to the interference fringe patterns:

$$I(x, z) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \left(\frac{2\pi\nu}{c} (x \sin(\alpha) + z \cos(\alpha)) + \Phi_{02} - \Phi_{01} \right). \quad (6.8)$$

In the (x, y) -plane, at each constant z , a cosine-modulated pattern, in x -direction, with interference lines along the y -direction, is obtained. This pattern is time independent as long as the two waves are coherent with each other. If the second wave has just a frequency offset of $\Delta\nu$, but is still parallel to the first wave, a periodically changing pattern is obtained according to

$$I = (I_1 + I_2)(1 + K \cos(\Phi_{01} - \Phi_{02} + 2\pi\Delta\nu t)), \quad (6.9)$$

with the contrast

$$K = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} \gamma_{12}, \quad (6.10)$$

for which the degree of coherence γ_{12} is equal to 1 for complete coherence and equal intensities $I_1 = I_2$. The emission of atoms of natural light sources results in wave packets of up to 10^{-8} s, so different light sources can only interfere within such time intervals. Therefore, interference is constructed by the same light source adding spatial coherence. The condition for spatial coherence is

$$a\theta \leq \frac{\lambda}{2} \quad (6.11)$$

with a , the linear extension of the light source, and θ , the opening angle of interfering light bundles. Since θ for lasers is very small, the full aperture of a laser source can correspond to a . For a synchrotron-based THz source, a is very small and θ can be larger. The temporal coherence is given by the length that a wave can travel in time within a spectrometer, to generate interference. The coherence time is expressed by the coherence length $l = L_c/c$. These values depend on the spectral width of the light source defined as $L_c = c/\Delta\nu \approx \lambda^2/\Delta\lambda$. For a single mode laser, this coherence length can, in principle, be up to several km.

Interference phenomena can also be found in thin plates (Fig. 6.8). Using two plates with an adjustable gap leads to a Fabry–Pérot type interferometer (Sect. 6.5). Using beam splitters, ideally splitting the intensity 50:50 in order to achieve maximum contrast, two-beam interferometers of the types built by Mach and Zehnder (Fig. 6.8c), and Michelson are obtained (Fig. 6.2). With a 50:50 beam splitter 50% of the power entering the Michelson interferometer is transmitted

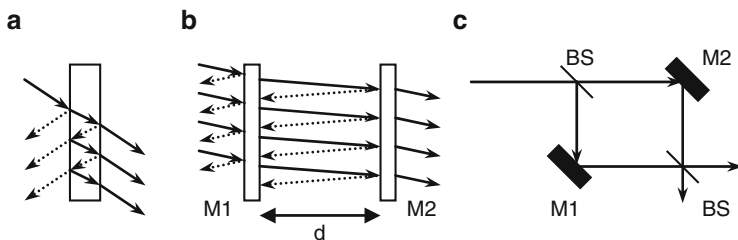


Fig. 6.8 (a) Interference due to a planar plate or etalon, (b) Fabry-Pérot interferometer with two partially transmitting mirrors (M1, M2), (c) two-beam Mach-Zehnder interferometer with two mirrors (M1, M2) and two beam splitters (BS)

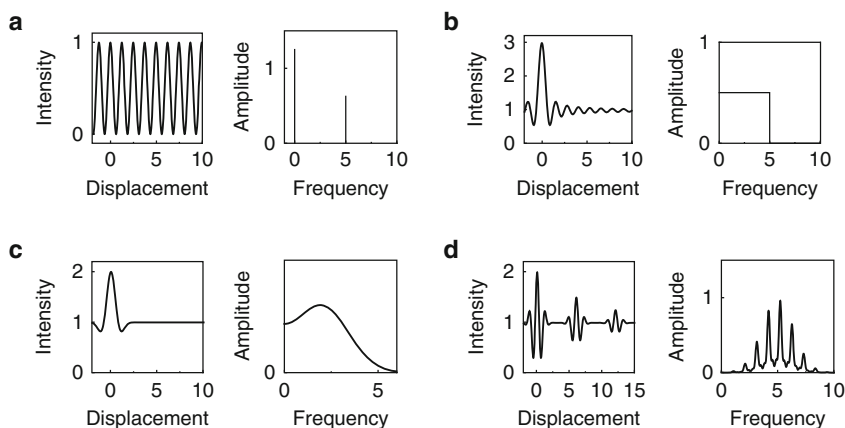


Fig. 6.9 Simple interferograms and their corresponding amplitude spectrum obtained by Fourier transformation for illustration (not to scale). (a) Infinite sine wave and corresponding Fourier amplitude, (b) Sinc function as described in (6.16), (c) interferogram for smooth broadband emission with dominant peak intensity at zero pathlength difference (d) multiple pulses as a result of multiple reflections, for example in an etalon

while the other 50% is directed back into the source. In a Michelson interferometer, the transmission-reflection (TR) product of the beam splitter is 0.25 at maximum. Typically beam splitters are made from thin sheets of Mylar or from Si. Such dielectric beam splitters have efficiencies that vary with frequency. The TR product rarely exceeds 50% of its optimum value of 0.25 over more than $\pm 50\%$ bandwidth around the center frequency. To overcome this limitation, a wire grid can be used as a beam splitter, and this is discussed in Sect. 6.4.

The intensity at the output of a Michelson spectrometer, as a function of the path length difference Δ , is called an interferogram (Fig. 6.9). For a small spectral interval $d\sigma$, with the wavenumber $\sigma = 1/\lambda = \nu/c$, this is given as an even function around the origin by

$$I(\Delta) = \int_0^\infty dI(\Delta) = \int_0^\infty B(\sigma) \cos(2\pi\sigma\Delta) d\sigma. \quad (6.12)$$

This expression is the cosine Fourier-component of

$$B(\sigma) = \int_0^\infty I(\Delta) \cos(2\pi\sigma\Delta) d\Delta. \quad (6.13)$$

A path length difference of Δ is achieved by moving one mirror in the interferometer away from its initial position $d = 0$ (Fig. 6.2), at which the distance between each mirror and the beam splitter is the same. In practice, this path length cannot be infinite and is limited at some distance $\Delta_{\max}$. Typical spectrometers have a path length difference of 5 to 10 cm to obtain a spectral resolution of less than 0.1 cm^{-1} . Systems with path lengths up to several meters are commercially available. They require very stable mirror motion, and excellent mirror and beam splitter alignment.

For a limited path length, (6.12) and (6.13) yield

$$B(\sigma) = \int_0^{\Delta_{\max}} \left(\int_0^\infty B(\sigma') \cos(2\pi\sigma'\Delta) d\sigma' \right) \cos(2\pi\sigma\Delta) d\Delta. \quad (6.14)$$

Evaluating only the cosine terms in the integral over Δ , this becomes

$$R(\sigma, \sigma', \Delta_{\max}) = \frac{1}{4\pi} \left(\frac{\sin(2\pi\Delta_{\max}(\sigma - \sigma'))}{\sigma - \sigma'} + \frac{\sin(2\pi\Delta_{\max}(\sigma + \sigma'))}{\sigma + \sigma'} \right). \quad (6.15)$$

Finally, by neglecting the smaller second term, the instrumental scanning function or slit function is obtained:

$$R(\sigma, \sigma', \Delta_{\max}) = \frac{\Delta_{\max}}{2} \left(\frac{\sin(2\pi\Delta_{\max}(\sigma - \sigma'))}{2\pi\Delta_{\max}(\sigma - \sigma')} \right) = \frac{\Delta_{\max}}{2} \text{Sinc}(2\pi\Delta_{\max}(\sigma - \sigma')), \quad (6.16)$$

which is proportional to the sinus function cardinal or Sinc-function. The following integral remains:

$$B(\sigma) = \int_0^\infty B(\sigma') R(\sigma, \sigma', \Delta_{\max}) d\sigma'. \quad (6.17)$$

For example, if there is only a monochromatic source with intensity $B(\sigma_1)$ at a wavenumber σ_1 the integral simplifies, and the source represents a delta-function. The spectrum becomes $B(\sigma) = B(\sigma_1) R(\sigma, \sigma_1, \Delta_{\max})$, i.e., except for a constant intensity factor $B(\sigma_1)$, the spectral scanning function R is obtained. The function R is the well-known diffraction pattern of a slit centered at $\sigma = \sigma_1$ and is not really a very good representation of a single-line spectrum (Fig. 6.9b). It has a peak width at half its center wavenumber height of $0.7/\Delta_{\max}$. Therefore, the smallest wavenumber interval resolvable is approximately given by $\Delta\sigma = 0.7/\Delta_{\max}$. In practice, the side lobes of the scanning function R cause complications in the interpretation of spectra, and these are suppressed by a suitable weighting function. This process is termed “apodization” and has been investigated by various authors. The simplest function is to multiply the interferogram by the triangular function $1 - \Delta/\Delta_{\max}$. Apodization has the effect of smoothing the instrumental scanning function, at the expense of

increasing its half width by approximately a factor of 1.5. The spectral resolution is then given by $\Delta\sigma \approx 1/\Delta_{\max}$.

The resolving power of an interferometric spectrometer is also limited by the beam spread of off-axis rays. This is due to the imperfect collimation of the finite source aperture. It was shown by Jacquinot [549] that if the source subtends a solid angle Ω at the collimating mirror, then the maximum value of the resolving power R is given by

$$R = \frac{\sigma}{\Delta\sigma} = \frac{2\pi}{\Omega} \approx 6 \left(\frac{f}{d} \right)^2 = 6 (f^\#)^2, \quad (6.18)$$

where d is the source diameter and f the collimator focal length. However, with the small detectors employed, the spectral resolution is mainly governed by the inverse of the mirror translation length. The spread of off-axis rays also leads to a small change in the ray path and causes all the wavenumbers in the computed spectrum to be underestimated by a small factor. The true wavenumber can be obtained from the computed value by multiplying these values by $(1 + \Omega/(4\pi))$. Normally, this small correction can be omitted in the THz range.

If an FTS is used for absolute amplitude measurements, it is necessary to determine the zero path length, with sufficient accuracy to avoid distortion, zero baseline shifts, and negative amplitudes. In practical experiments, the analogue detector signal is digitally sampled and discrete Fourier transformation is employed. Under these circumstances, the Shannon sampling theorem must be obeyed. It states that an analogue signal can be reconstructed, if the highest frequency $f_{\max}$ in the signal and the sample frequency f_{sam} , corresponding to a sample period Δt , fulfils the inequality $f_{\max} < f_{\text{sam}}/2 = 1/(2\Delta t)$. The half-sampling frequency is also known as the Nyquist-frequency. If the detector signal contains frequencies which are higher than the Nyquist-frequency, these frequencies are mapped onto the frequency range below $f_{\max}/2$, leading to the so-called aliasing. By windowing, or anti-aliasing of the detector signal, all higher frequencies than the Nyquist-frequency can be removed, so that the Shannon theorem is fulfilled.

Normal THz FTSs rely on a moving mirror which generates the path difference. It is worth noting that it is feasible to realize an FTS, which does not require any moving component. Although several techniques have yet to be implemented at THz frequencies, there is no reason in principle not to do so, and with technical progress it may well be achieved. Two approaches have been demonstrated. One of these relies on the Talbot effect [570, 571]. At visible frequencies, an FTS without moving parts has been demonstrated by detecting self-images of a periodic structure such as a grating on a CCD [572]. The other approach relies on two phase-locked pulsed sources, which produce frequency combs. This has been demonstrated in the THz and IR range with two microwave sources operating up to 0.3 THz [573] or two mode-locked femtosecond lasers [574–578]. One of the combs is slightly detuned with respect to the other. Due to the difference in repetition rates, the pulses of one beam are continuously shifted against those of the other, akin to the conventional FTS. Pairwise interference and heterodyning generates a comb spectrum of much

lower “beat” frequencies that carries amplitude and phase information. Fourier transformation of this signal yields the spectral information of the original spectrum. This approach has some similarity to the ASOPS technique described in Sect. 6.6. THz spectrum analyzers for the accurate measurement of THz frequencies benefit from frequency combs used for mixing [579, 580].

6.3.1 Dispersive Fourier-Transform Spectrometer

Dispersive Fourier-transform spectroscopy (DFTS) was developed for the THz range by J.E. Chamberlain et al. in the 1960s [74] and has been extensively used by Afsar, Birch, Button, Parker, and co-workers [581–583]. Various setups have been employed to study gases, liquids, and solids under cryogenic conditions, and in magnetic fields. In DFTS, the sample is not placed between the beam splitter and the detector but is placed instead in one of the interferometer arms. For reflecting or opaque samples one of the mirrors is replaced with the sample, which serves as a partial mirror.

If the sample is studied in transmission in one of the interferometer arms, the resulting interferogram is asymmetric and shifted with reference to the zero path length position. The shift is due to the sample’s refractive index. The peak shift reveals an average refractive index over the frequencies that the spectrometer can cover. For a sample thickness d with average refractive n a shift is found of

$$\Delta x = 2(n - 1)d. \quad (6.19)$$

The factor 2 follows from the distance traversed twice in transmission through the sample, and the -1 corresponds to the refractive index of vacuum or air, which the sample replaces. If the sample has internal reflections, the next fringe pattern is at

$$\Delta x_R = 2(n - 1)d + 2nd, \quad (6.20)$$

from which the unknown thickness – without knowing the refractive index n – can easily be found, to a first approximation, as:

$$d = 0.5(\Delta x_R - 2\Delta x). \quad (6.21)$$

Normal FTS can be considered as an autocorrelation and DFTS as a cross-correlation. The advantage of DFTS versus FTS is that it can provide the phase information, which leads to the refractive index of the sample. DFTS allows the complete dielectric constant of a material to be determined in a similar fashion to dielectric spectroscopy using GHz electronics. THz TDS, a technique introduced in the late 1980s, also provides the complete dielectric constant by measuring the electric field instead of intensity, although on a much shorter time scale.

In general, retrieving phase information and the refractive index requires a more stable experimental system for DFTS and the more modern TDS. Phase jitter must be avoided. The following citation emphasizes this point by promoting the careful elimination of systematic errors in a measurement system [583]: “In spite of the wide frequency coverage and the accuracies that can be achieved DFTS has not been widely used. . . . It is true that DFTS is perceived as a difficult measurement technique, and it places more stringent requirements on specimen geometry and interferometer performance than are generally met in conventional FTS. This, though, is not a limitation or difficulty of DFTS. It is, instead, recognition of the fact that to achieve results largely free of systematic error one must have a well-defined and controlled measurement system. Quantitative measurements by conventional FTS could undoubtedly be improved if similar procedures were followed.”

6.4 Polarizing (Martin–Puplett) Interferometer

There is one major problem with FTSs of the type discussed in the previous sections, which is connected with the beam splitter. As mentioned in the previous section, the TR product of dielectric beam splitters rarely exceeds 50% of its optimum value of 0.25 over more than $\pm 50\%$ bandwidth around the center frequency. Therefore, if a large portion of the THz range needs to be covered, this requires an exchange of beam splitters, which in turn affects the quality, and especially the reproducibility, of the spectra. In addition, operation of the spectrometer is more complex and time consuming. A beam splitter which overcomes this problem is a wire grid of the type described in Sect. 3.11. When used as beam splitters wire grids have a very high and uniform efficiency, up to the cutoff frequency as determined by the grid constants. Because two beams are generated by polarization splitting, this type of interferometer is a polarizing interferometer. Often, it is called a Martin–Puplett interferometer (MPI) after D.H. Martin and E. Puplett, who first conceived this widely used instrument [76].

The basic configuration of an MPI is illustrated in Fig. 6.10. It consists of three wire grids. Two of them, P1 and P2, are used as polarizers while the third one, PBS, is the beam splitter. In addition, two rooftop reflectors, M1 and M2, one of them movable, are built-in. Now consider a beam which enters the interferometer at input port 1 and is polarized normally to the paper by P1. The wires of the beam splitter are wound at an angle of 35.4° with respect to the plane of the paper. The incident beam “sees” the wires at an angle of 45° . Therefore, it is separated into two parts having orthogonal polarization with respect to each other, and 45° polarization with respect to the paper. One polarization is reflected by the beam splitter and the other is transmitted. The rooftop mirrors rotate the polarization of each of them and send them back to the beam splitter, where the previously reflected beam will be transmitted, and the previously transmitted beam will be reflected. If the path difference Δ between the two beams is zero, the output beam is horizontally polarized. For $\Delta = \lambda/2$, it is vertically polarized, and for $\Delta = \lambda/4$ and $\Delta = 3\lambda/4$

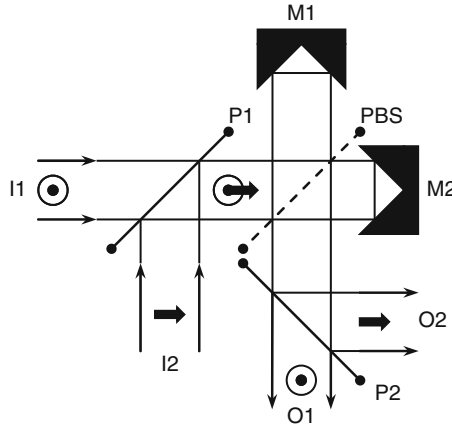


Fig. 6.10 Scheme of a Martin–Puplett spectrometer. I1 and I2 denote two input ports separated by a wire grid polarizer (P1) with the grids parallel to the plane of the paper. Similarly, O1 and O2 denote two output ports of the spectrometer which are separated by another wire grid (P2). The rooftop mirrors (M1, M2) are used as polarization rotators and the wires of the polarizing beam splitter (PBS) are at 45° to the plane of the paper. The complementary nature of the interferograms for the two orthogonal orientations of the polarizer P2 makes it possible to eliminate the high mean-level of a conventional interferogram (Fig. 6.11). This is achieved by alternating the orientation of P1 or P2 instead of chopping or, alternatively, by using two detectors in opposition. One receives the beam transmitted by P2, and the other is reflected by P2 (adapted from [584])

it is circularly polarized. In general, the beam in front of P2 will be elliptically polarized, with the ellipticity and the direction of the major axis determined by the path difference. It is this characteristic which has led to its use as a polarization rotator. The output polarizer P2 can be oriented orthogonally, or in parallel, with respect to P1. It is used to separate the two output ports, and changing its orientation simply interchanges the output ports. For zero path difference, and ideal wire grids, the whole incident power will pass into output ports 1 or 2, depending on whether P2 is polarized parallel or orthogonally with respect to P1. This is another advantage over a Michelson interferometer where, at its maximum, 50% of the power is transmitted. If the input beam is monochromatic with a wavelength λ , the major axis of the recombined beam is rotated as the path difference is varied, and the power at each output port will vary periodically with the period $\Delta = \lambda$. If ideal wire grids are used, and I_0 is the power entering the interferometer, the output power at port 1 for parallel P1 and P2 is given by

$$I_{O1}(\Delta) = \frac{I_0}{2}(1 + \cos(2\pi\Delta/\lambda)), \quad (6.22)$$

while for orthogonal P1 and P2 the power at port 1 is given by

$$I_{O2}(\Delta) = \frac{I_0}{2}(1 - \cos(2\pi\Delta/\lambda)). \quad (6.23)$$

If detectors are placed at both output ports, (6.22) holds for port 1 (P1 parallel to P2) and (6.23) for port 2 (P1 parallel to P2), and vice versa (P1 orthogonal to P2). The sum of the two equations is unity, independent of the path difference. It should be pointed out that interferograms taken at both output ports are complementary (Fig. 6.10). They contain the same spectral information but they are inverted around the mean value $I_0/2$. Equations (6.22) and (6.23), respectively, follow from (6.9) with $I_{O1} = I_{O2} = I_0/2$ and $K = 1$, and the basic mathematical description of a Martin–Puplett interferometer is the same as for a Michelson interferometer.

Besides the increased spectral coverage with a wire grid beam splitter, the MPI has other advantages, because the input and output ports are easier to separate than those of a Michelson interferometer. The complementary nature of the interferograms at the two output ports makes it possible to eliminate the high mean value of the Michelson interferogram, as well as the spurious signals and errors associated with it. By using two detectors, one at each output port, the complementary interferograms are taken simultaneously. For a monochromatic source, the combined interferogram is

$$I_{O1} - I_{O2} = I_0 \cos(2\pi\Delta/\lambda), \quad (6.24)$$

which oscillates around the true zero level (Fig. 6.11 [76]). Instead of using two detectors, one can use only one detector and change the polarization of P1 or P2 by 90° . Another possibility is to use one output port for recording a background interferogram, whilst using the other to record the interferogram of the sample. In this case, the sample needs to be placed behind polarizer P2. Furthermore, the easy access to the input ports has advantages, as one of the two input ports can be used for calibration. For example, by placing two blackbody calibration loads with different temperatures at one of the input ports, the absolute signal power can be

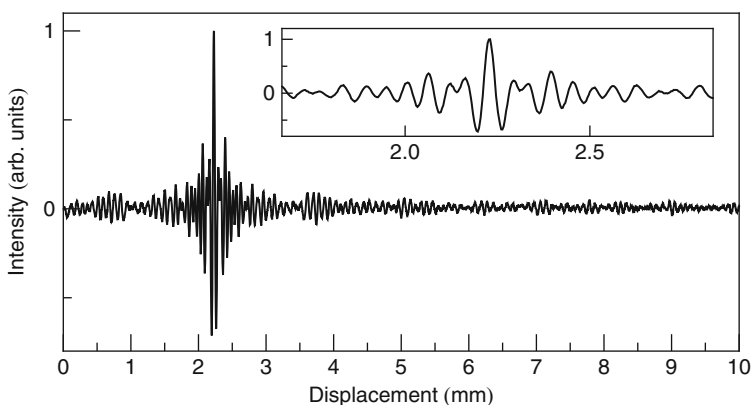


Fig. 6.11 Interferogram of water vapor. For a Martin–Puplett spectrometer as depicted in Fig. 6.10 interferograms oscillate around zero

determined. This is achieved by recording the difference between the interferograms of the signal, and each of the calibration loads. The absolute signal power is given by interpolation between the two. There are alternative designs for MPIs to the one discussed here which all have particular advantages, for example simpler arrangements with less optical components, or more complicated ones, where all the power of an unpolarized beam at the input contributes to the interferogram at the output. A review is given in [585].

The alignment of an MPI is similar to the alignment of other two beam interferometers, with some additional steps regarding the rooftop mirrors and wire grids. A description of how to align it is given in [586], together with a detailed analysis of the effect of misalignments, and imperfections of the optical components, on the instrumental performance. For certain applications, it is important to consider the performance of an MPI for Gaussian beams, keeping in mind that in the case of Gaussian beams the spatial overlap of the two beams, reflected by each of the rooftop mirrors, depends on the path difference Δ . The modulation depth N of a fundamental Gaussian beam is [587]

$$N = \left(1 + \left(\frac{\lambda \Delta}{2\pi w^2} \right)^2 \right)^{-1}. \quad (6.25)$$

This is a measure of the spatial overlap of two Gaussian beams with waist w after traveling a path difference Δ . For $w^2 \gg \lambda \Delta / (2\pi)$ or $N \rightarrow 1$, the losses are negligible. A more detailed analysis of Gaussian beam mode propagation in MPIs can be found in [588, 589].

The MPI is widely used in heterodyne spectrometers, where the radiation from a signal source and a local oscillator (LO) have to be spatially superimposed (diplexed) in order to couple both to the mixer (Sects. 5.5 and 6.8). The frequencies of the signal and LO are typically a few GHz apart. Consider a signal which is coupled to input port 1, and an LO, which is coupled through input port 2 (Fig. 6.10). While at the input ports signal and LO radiation are cross-polarized, they emerge from the Martin–Puplett diplexer co-polarized, provided the path difference is chosen correctly. If the path difference Δ_{Dipl} is

$$\Delta_{\text{Dipl}} \approx (2m - 1)\lambda_{\text{IF}}/2, \quad (6.26)$$

the transmission of signal and LO radiation is maximized. Here, $\lambda_{\text{IF}} = 1/|1/\lambda_{\text{S}} - 1/\lambda_{\text{LO}}|$ is the wavelength associated with the intermediate frequency of signal and LO, and m is the order for which the diplexer is used. A mixer receives signals from the upper and the lower sidebands, which are separated by twice the intermediate frequency. Usually, the signal to be measured is only in one sideband, while the other sideband contributes noise or other unwanted signals. These can be suppressed by using an MPI as a single sideband filter. To discriminate between both sidebands, the path difference of the interferometer needs to be set so that the desired sideband is at a transmission maximum, while the other sideband is at a minimum $2\nu_{\text{IF}}$ away.

Therefore, the periodicity of the interferometer transmission has to be $4\nu_{\text{IF}}$ and the required path difference is

$$\Delta_{\text{SSB}} \approx (2m - 1)\lambda_{\text{IF}}/4, \quad (6.27)$$

which is half the path difference required for signal-LO diplexing. For diplexing, as well as for sideband filtering, the bandwidth of the interferometer is crucial. The 3 dB bandwidth is given by

$$\Delta\nu = \nu_{\text{IF}}/(2m - 1). \quad (6.28)$$

Obviously, it is advantageous to use the interferometer in the lowest order $m = 1$ for which the path difference is smallest. This provides the largest bandwidth, has the least critical tuning, and minimizes losses from incomplete spatial overlap of the beams from both arms of the interferometer, due to the different optical paths over which they travel. The MPI is a widely used device for diplexing and sideband filtering, because of its nearly equal amplitude division and low loss over a wide frequency range, as well as the easily accessible input and output ports. They are used in virtually all airborne, spaceborne, or balloon-borne heterodyne spectrometers for astronomy and atmospheric research. Examples are the satellite observatories SWAS, Odin, Herschel, EOS Chem2, and the airborne observatories KAO and SOFIA.

Up to now only monochromatic signals have been considered. However, when used as a spectrometer the MPI normally has to deal with broadband radiation. Since the MPI is a linear device, each spectral component will propagate through the system independently. Furthermore, the detection system will usually detect the sum of the powers of the spectral components. Therefore, the theory of the FTS, as outlined in previous sections, can be applied to the MPI. A more detailed description can be found in [585]. THz FTSs based on MPIs have found widespread applications in laboratory studies of solids, liquids, gases, and also for plasma diagnostics, synchrotron radiation diagnostics, measurement of cosmic background radiation, and atmospheric emission studies.

6.5 Fabry–Pérot Interferometer

The Fabry–Pérot interferometer (FPI) is a multiple-beam interferometer which is based on two parallel and partially reflective surfaces. It is named after the French physicists C. Fabry and A. Pérot, who invented this device in 1897. The FPI is discussed in many textbooks on optics and only those aspects which are of specific importance for the THz range are considered here. Originally, the FPI was developed for use at short wavelengths where diffraction is not a major concern. This is not the case at THz frequencies where, in addition, lateral offset or walk-off with each successive round trip can lead to significant loss.

Neglecting diffraction and walk-off losses, the transmittance T of an FPI with two identical reflecting surfaces, each characterized by the reflectance R and power absorbance A , follows from the Airy formula

$$T(\nu) = \left(1 - \frac{A}{1 - R}\right)^2 \left(1 + \left(\frac{2\sqrt{R}}{1 - R}\right)^2 \sin^2(\Phi/2)\right)^{-1}. \quad (6.29)$$

Here, $\Phi = 2\pi\nu 2nd/c - 2\Phi_r$ is the total phase shift for one round trip that is between two successive beams. It has two contributions: Φ_r is the phase shift due to reflection on one surface and $2\pi\nu 2nd/c$ is the round trip phase delay, due to the optical spacing nd of the two reflecting surfaces for normal incidence. The transmittance has a maximum when the phase Φ becomes a multiple of 2π radians. The constructive interference maxima are called fringes. If Φ_r is ignored, each fringe corresponds to an integral number of half-wavelengths in the medium between the reflectors. With $\Phi = 2m\pi$ for maximum transmittance, m is referred to as the order of the fringe. The peak transmittance is high if A is small. For normal incidence, the fringes are spaced by

$$\Delta\nu = \frac{c}{2nd}. \quad (6.30)$$

This spacing is called the free spectral range. The finesse of an FPI is defined as the ratio of the spacing between two successive fringes to the FWHM of a fringe:

$$F = \pi \frac{\sqrt{R}}{1 - R}. \quad (6.31)$$

The finesse is the effective number of multiple reflections. It becomes high if R approaches unity. However, this reduces the transmission. Therefore, it is necessary to find a compromise between resolution and transmission, which depends on the type of application or system. The resolving power is the ratio between the frequency at which the FPI operates and the FWHM of a fringe. It can be expressed as the product of the order of the fringe and the finesse:

$$R = \frac{\nu}{\Delta\nu} = mF. \quad (6.32)$$

The resolving power achievable with a practical system is somewhat worse, mainly due to diffraction by finite apertures, and walk-off losses. These effects decrease the degree of interference that can be obtained at the output of the FPI, and are more pronounced in the THz region than at shorter wavelengths. On the other hand, requirements on reflector smoothness, and on the parallelism of the reflectors, are less demanding. An approximate criterion for ensuring negligible degradation of the maximum transmission and the finesse is given by [590]

$$\frac{\delta d}{\lambda} \leq \frac{0.025}{nF}. \quad (6.33)$$

For any practical FPI, it is important to have reflectors with high reflectivity and low absorption loss. Mirrors coated with multilayer dielectrics, which are often used in the visible and IR spectral range, are not practical at THz frequencies because of the required thickness of the layers. Mirrors with thin metal coatings are too lossy to be used, while bulk metal mirrors are opaque. The invention of metal mesh reflectors opened the way to achieving efficient FPIs, and almost all THz FPIs employ metal mesh reflectors. For some special applications mirrors from thin high T_c , YBCO superconductor films on MgO or silicon substrates can be used. A peak transmittance of 0.6 and a finesse of 20 have been obtained [591, 592]. Even with one of the first FPIs in a simple design, it was possible to obtain a resolving power of 2,000 at 1.5 THz [590].

The tunability of an FPI can be provided by a variety of mechanisms, for example, a threaded cylinder, a cantilever, or a translational stage. Besides the standard parallel reflector geometry, other variations have been developed. One, which eliminates walk-off, is the folded FPI or ring FPI [116]. If the distance of the mirrors is chosen for maximum reflectivity, radiation which enters the FPI exits it at the same port but redirected by 90° . In the case of maximum transmission, the radiation emerges at the other port, again redirected by 90° , but in the opposite direction, as in the case of maximum reflection.

FPIs are frequently employed in spectrometers for astronomical observation. For molecular spectroscopy with an FPI, a high resolving power ($\sim 10^5$) is required. With a typical finesse of ~ 50 such an FPI needs to be used in a very high order ($\sim 2,000$), with a correspondingly small free spectral range. To suppress all but one order, additional FPIs, with lower resolution, are mounted in series with the high-resolution FPI [593]. FPIs have been used aboard the IR Space Observatory ISO, covering the range from 1.5 to 26 THz. Here, several FPIs mounted on an interchange wheel, as well as FPIs in series with gratings for order selection, were used [594, 595]. In particular, a ring FPI can be used as a diplexer in a heterodyne receiver where the LO is reflected as described above. The signal radiation, of slightly different frequency, enters the FPI at the other port and is transmitted. Finally, it is worth noting that an FPI can be used as a variable output coupler for THz gas lasers. In this application, it replaces one of the mirrors of the laser. The degree of outcoupling can be varied by adjusting the spacing of the FPI [596].

6.6 Time-Domain Spectrometer

The development of sub-ps mode-locked dye lasers in 1972 [111] and femtosecond (fs)-lasers based on titanium-sapphire (Ti:Sa, a solid state gain media) in 1986 [112], respectively, led to a variety of new techniques in the THz frequency range. Before discussing the details of time-domain spectrometers (TDS) [597, 598], it is appropriate to reflect on short-pulse lasers which were the enabling technology for THz TDS, not only as integral components of such a spectrometer but also as a tool

to investigate and improve the semiconductor and nonlinear materials employed in THz pulse generation and detection.

The technique of Kerr-lens mode-locking developed in the early 1990s [599] led to an easier access to sub-100 fs-laser pulses. The technique describes self-focusing within the gain medium (here the Ti:Sa crystal), which causes bunching of the mode-locked pulse circulating within the laser cavity. In the mid-1990s, the introduction of chirped dielectric mirrors led to even shorter pulses and access to sub-10 fs-laser pulses [600] (Sect. 2.8). Fs-lasers emit pulses in the visible spectral region (typically around 780 nm), although near-IR fs-lasers coupled to fiber technology are increasingly used, due to their widespread availability following their application in telecommunications in the 1,300 and 1,500 nm band. Fourier transformation, from the time domain to the frequency domain, reveals that such short visible pulses correspond to electric field amplitudes, which are spread over a frequency band with a width of a few THz for sub-ps pulses, to several tens of THz for fs-pulses. For a THz TDS, it is necessary to down-convert this broad THz frequency band from the visible to the THz range, and this will be discussed later in this section.

Turnkey fiber-based fs-lasers are compact and commercially available at declining prices. Their ease of operation has prepared the ground for many uses, including THz applications. Mode-locked fs-lasers, in general, produce regularly spaced pulses in time which, viewed in the frequency domain, correspond to electric field amplitudes regularly spaced in frequency, known as “frequency combs” [601]. Moreover, fs-lasers allow the construction of THz broadband spectrometers with only room temperature and solid-state components. In addition, exploiting frequency comb techniques leads to THz TDS without moving parts, thus enhancing data acquisition.

Near-IR fs-lasers with center wavelengths in the 1,500-nm band can use fibers to propagate the pulses efficiently. They can generate pulses of less than 100 fs and typical average power levels of 100 mW, with mode-locked repetition rates ranging from 80 to 100 MHz. Visible lasers using optically pumped Ti:Sa crystals can easily generate pulses of 10 fs and less, with an average power of 500 mW or higher. The repetition rates are in the GHz range. The peak output power of a pulse from an fs-laser can be increased by several orders of magnitude by using a low repetition rate amplifier. For example, a regenerative Ti:Sa amplifier with 1 kHz repetition rate can be used at the expense of increasing the complexity of the experimental setup.

THz TDS consist of polarized fs-lasers and, normally, three additional components (Fig. 6.12). The first component allows the down-conversion of this broad THz frequency band from the visible to the THz range. For down-conversion, an emitter based on a semiconductor with a fast response time (so-called semiconductor switch) can be used. Even with very high bias voltages applied, these semiconductor detectors are designed to only show a current flow if, at the same time, another visible fs-pulse generates sufficient electron-hole pairs in the material. Another method of generating or amplifying a THz wave or pulse coherently is the use of a nonlinear crystal. Initially, crystals such as LiTaO_3 were used, showing fast transients with frequency components beyond 4 THz [602]. In the crystal surface

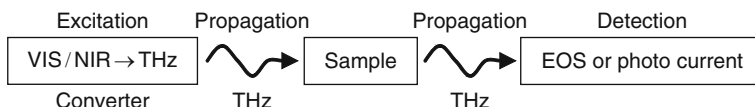


Fig. 6.12 Concept of THz pulse generation and detection: Conversion of the visible or near-IR fs-pulses into THz pulses which propagate, interact with a sample, and further propagate to interact with the detection system. The detection pulse is used to read out the THz response by an electro-optical sampling technique (EOS) in which the detection pulse is modified by the THz pulse. Alternatively, a semiconductor is used in which the detection pulse creates free carriers and leads to a measurable photo current. The propagation can be either in free space or via the surface of a material, for example for THz sensors

layer, on a length scale of the coherence length, THz beams are emitted as electro-optic Cherenkov radiation, with frequency components beyond 1 THz [528].

The second component detects the emitted and freely propagating THz pulse which is, once more, a semiconductor material or nonlinear crystal. The semiconductor material is designed to operate as a very fast switch so that the detector is only active if the fs-pulse or gate-pulse is present. For the ps and sub-ps long THz pulses, and by using gate-pulses with lengths between 5 and 100 fs for commercial lasers, the THz electric field can be viewed as constant in time when the visible gate pulse is present. Then the electric field value, averaged in the time window of the gate pulse, can be measured by the current flow, which is also oriented according to the direction of the electric field vector. This process of optical gating and averaging finds its analogy in the well-known electronic boxcar gating and averaging technique.

Another method, electro-optical sampling (EOS), employs nonlinear crystals, initially LiTaO₃ [603] or ZnTe [604], for detection. In this case, the impinging time-varying electric field changes the crystal's birefringence as a function of time and electric field strength of the THz pulse. A second polarized fs-pulse, for example split off from the original pulse (Fig. 6.13), passes through the crystal at the same time and alters its state of polarization, due to the birefringence experienced. The polarization state can be measured by a polarizer serving as an analyzer. The process of detection of THz pulses in nonlinear crystals relies on the coherent overlap of the optical and THz waves, within the coherence length, which also determines the optimal detector crystal length. This detection process is commonly named coherent detection.

The third component is a method of introducing a variable time delay between the second fs-pulse used for detection and the THz pulse, which is locked in time to the fs-pulse, generating it by down-conversion. The fs-pulse for detection is often much shorter, e.g. 5–100 fs, than the THz transient, so that a variable time delay allows tracing out the electric field of the THz pulse, while sampling the electric field vector, and averaging within each optical gate pulse. The Nyquist theorem then determines which frequencies are still resolved, up to as high as 100 THz for 5 fs pulses.

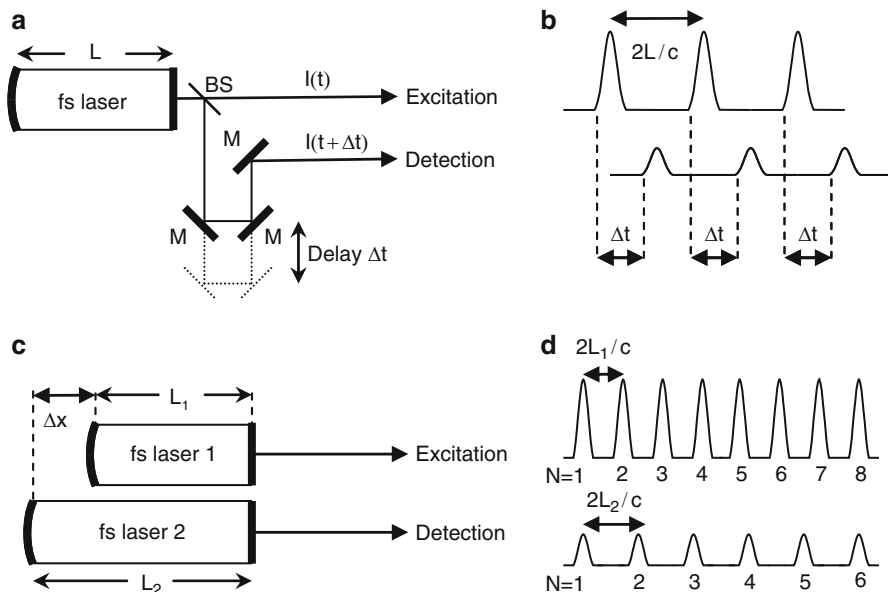


Fig. 6.13 (a) One mode-locked fs laser, emitting pulses regularly spaced in time is split by a beam splitter (BS) into high intensity excitation pulses and low intensity detection pulses. The detection pulse is reflected by a mirror (M) through an element, which introduces a time delay (Δt) with respect to the excitation pulse. (b) Two mode-locked lasers, with a cavity length difference (Δx), generate pulses with different repetition rates. For a certain time interval, here after four and again after eight excitation pulses, coincidence with the third and sixth detection pulse is achieved

The simplest and cheapest way to introduce a variable time delay is to use a mirror mounted on a mechanical translation stage. The mirror should have little dispersion and a high broadband reflectivity to preserve the fs-laser pulse (Sect. 2.8). The translation stages can be simple loudspeaker membranes, fast piezo-driven mirrors, or more sophisticated systems. For short delays it is possible to obtain high speed data acquisition, with modulating frequencies in the kHz range. Step-scan techniques can also be used. In general, delay times up to several ns, corresponding to several tens of cm of inserted path length, increase the spectral resolution in a similar way to that of a standard FTS. However, to achieve high resolution in the order of a MHz, delay lines with a length of ~ 100 m are required. This, and the low spectral brightness of TDS, result in systems, which are typically designed for a resolution of between 1 and 10 GHz.

The prototype model for current THz TDS is based on the so-called Auston-switch, developed in 1984 by D.H. Auston et al. [104]. In analogy with the classic experiment by Hertz, fast photoconducting materials serve as time-varying Hertzian dipoles if excited by fs-lasers. Their apparatus already included generation, transmission, and detection of THz pulses, but was intended to measure properties of semiconductor materials with fast electromagnetic transients, to overcome limitations imposed by transmission line structures.

The device consisted of two identical photoconductors located symmetrically on opposite sides of a 1.15-mm-thin slab of insulating, sapphire material (Fig. 6.14a). The fast photoconductor is a 1- μm thick epitaxial Si film deposited on sapphire (SOS, silicon-on-sapphire) and irradiated by 3×10^{15} Ar^+ ions/ cm^2 with an energy of 2 MeV. The Ar ions introduce a large density of defects, leading to a rapid capture, if carriers are optically excited by a fs-laser pulse. The active region of the photoconductors, into which the visible pulses with intensity $I(t)$ were focused, was a 10- μm wide gap between the Al electrodes biased at $V_B = 45$ V. The 100 fs-pulses were generated by a colliding-pulse, passively mode-locked (CPM) ring dye laser, with an energy of approximately 50 pJ.

The THz transient is transmitted to the detecting film illuminated by another fs-pulse, split from the fs-pulse used for generation, and delayed by a time Δt . The receiving electrode is connected to a low-frequency amplifier giving a voltage V_S related to the average current as the optical delay is scanned. An aperture is included to limit multiple reflections between the dipoles. Fattering and Grischkowsky [605] illustrated the possibility of collimating and focusing the THz transient emission, and producing a beam of freely propagating electrical pulses [606].

The emitter is a Hertzian dipole, with dimensions that are small compared to any of the radiated wavelengths, excited by a 70 fs CPM dye laser. The THz source can

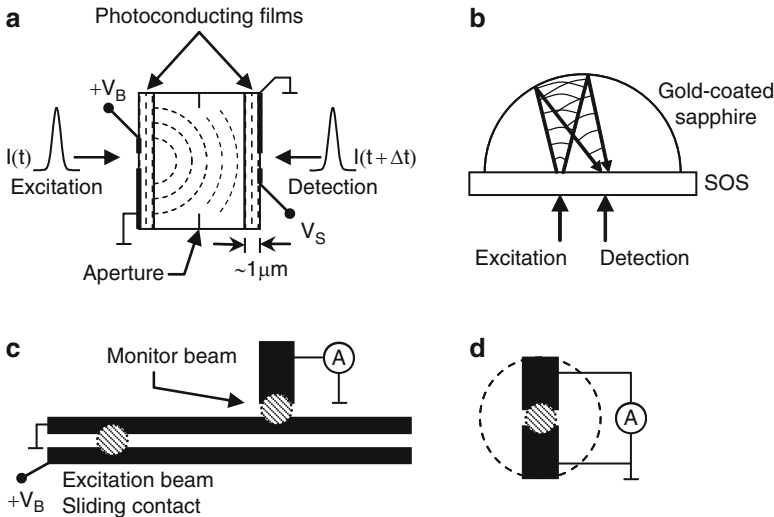


Fig. 6.14 (a) Prototype device of a THz TDS developed by Auston et al. (for a detailed description, see text) (b) Point source THz optics used by Fattering and Grischkowsky consisting of an Au mirror on a solid hemispherical sapphire lens, with a diameter of 9.5 mm, in contact with the rear side (sapphire side) of a 0.43-mm thick SOS chip. (c) Voltage V_B biased coplanar transmission line with the laser excitation beam defining the location of the THz transient dipole. The monitor beam measures the electrical pulse coupled to the line. (d) THz detector placed at a distance of 1.7 mm from the emitter and connected to a current amplifier. The fs-laser detection beam is centered on the 10- μm wide gap within the diffraction limited THz beam indicated by the dashed circle (adapted from [104, 605, 606])

be placed at the focal point of a spherical lens for collimation (Fig. 3.24) but, for the experiment, the focal point was placed mid-way (Fig. 6.14b) between the source (Fig. 6.14c) and THz detector (Fig. 6.14d). The emitter was located on a 20-mm-long transmission line consisting of $5\text{ }\mu\text{m}$ wide, 500 nm thick, and $10\text{ }\mu\text{m}$ spaced parallel electrodes of aluminum. This experiment showed 0.6 ps wide THz pulses [605] containing frequency components beyond 1 THz [606]. Although the authors pointed out that sapphire is not an ideal lens material, due to its absorption and dispersion, they could still show freely propagating diffraction-limited THz beams of pulses, with a maximum emission around 0.5 THz (Fig. 6.15a). After propagating the pulse for 1 m in free space, there was little change in the measured sub-ps pulse shape, when compared to a propagation distance of 0.1 m. However, due to an average beam divergence of 100 mrad, a 20-fold intensity loss was observed [606].

Transferring a THz electric field transient and pulse through a sample with absorption and dispersion and comparing it to the initial pulse, unaltered by the sample, enables broadband THz spectroscopy similar to DFTS, but with a sub-ps to fs time resolution, and often higher (S/N)-ratio in the THz frequency range, due to the optical gating and coherent detection.

In a subsequent experiment [607], two off-axis paraboloidal mirrors, although not optimized for a symmetric beam pattern (cf. Fig. 3.16), were used in combination with MgO lenses, instead of sapphire lenses, to generate, collimate, and refocus THz beams (Fig. 6.15b). The THz TDS was placed in an airtight enclosure, thus allowing spectroscopy of gases. The authors used a 70-fs CPM dye laser with an

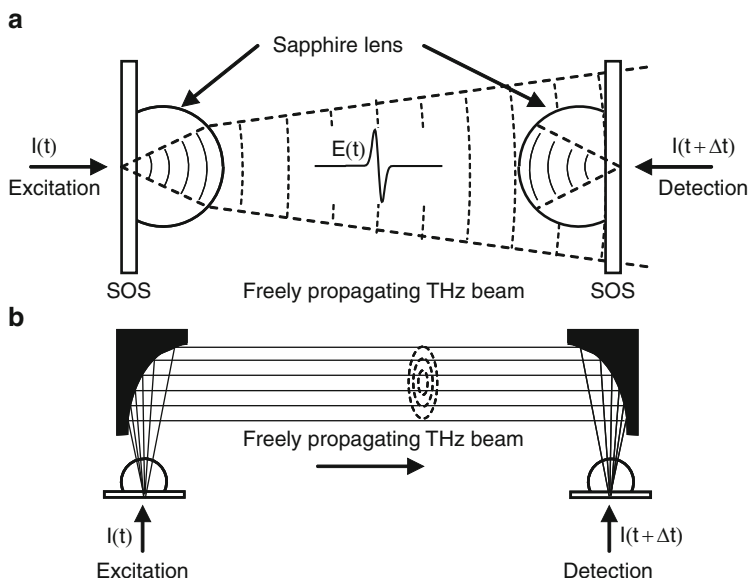


Fig. 6.15 (a) Freely propagating THz beams from a 5-mm diameter coherent source [606] (b) THz TDS setup for measuring the absorption and dispersion in gases simultaneously. The *dashed circles* illustrate the diffraction-limited beams at different THz wavelengths (adapted from [607])

average power of 1.5 mW for excitation with a duty cycle of $\sim 10^{-5}$. The energy per THz pulse is low but, due to the high repetition rate of 100 MHz and the coherent detection, the peak electric field is measured with a (S/N) -ratio of $\sim 3,000$ for an integration time of 125 ms [607], i.e. the peak intensity is measured with a (S/N) -ratio of $\sim 9 \times 10^6$. Since these first experiments, the average THz output power has increased from 50 nW initially, to 50 μ W for a semi-insulating (SI) GaAs emitter, with parallel electrodes spaced by 0.5 mm biased at 200 V [135], and to approximately 100 μ W for cooled emitters [608]. The frequency coverage has been extended to 4–8 THz and even beyond, to 70 THz [609–614] in spectrometers with lasers emitting a few fs long pulses. It should be noted that the power is emitted across a frequency band of several THz and the spectral brightness of a TDS is low when compared with a cw coherent source. The THz pulse energy was increased to about 1 μ J by using high-power excitation lasers with low repetition rates, which resulted in electric field amplitudes exceeding several MV/cm [615, 616]. With cooled emitters and EOS a (S/N) -ratio for the measurement of the peak electric field amplitude in a THz TDS can exceed 10^4 (10^8 for the intensity) using an integration time of 20 μ s, while a sufficiently sampled THz transient is traced out within 20 ms [608].

The concept of THz TDS can also be viewed as a modified pump-probe scheme (Fig. 6.13a,b). Part of the excitation (pump) pulse is split off into a low intensity detection (probe) pulse by a beam splitter. If both pulses are incident on a sample, the excitation pulse will ideally excite a resonance in the material. This is probed by the low intensity detection pulse, which does not affect the material in the first place. By introducing a time delay Δt between detection pulse and excitation pulse, the response of the material at a time Δt after the excitation pulse can be monitored. Continuously varying the time delay allows the response to be recorded in detail. The response can be followed until the next excitation pulse arrives. Pockels cells can pick out pulses, from regularly spaced fs-laser pulses, and effectively reduce the repetition rate and increase peak power, if used as a seed of a regenerative amplifier.

Figure 6.13c,d illustrates a technique called asynchronous optical sampling (ASOPS) [617–620]. Two fs-lasers are used instead of a translation stage for introducing a time delay. One fs-laser has a slightly larger cavity length in comparison with the second fs laser. This larger cavity length results in a longer round-trip time, and wider spacing of subsequent fs laser pulses. Both lasers can be free-running if they are well manufactured, when passive stabilization is sufficient. Another way is to control one cavity actively by placing one cavity mirror on a fast piezo stage, and locking the round-trip frequency of the detection laser, with a slight offset to the excitation laser. Typically, the excitation laser is stabilized to the detection laser using techniques such as ASOPS, or electronically controlled optical sampling (ECOPS). It is also possible to lock the lasers to separate frequency synthesizers. These methods require a very high electronic stability. During the development of so-called frequency combs used for metrology, the stability has increased significantly because the mode number is large, typically between 10^5

and 10^6 . The mode number follows from the laser round-trip frequency, ranging from 80 MHz to 5 GHz, and the center wavelength of the visible fs-lasers around 385 THz (780 nm).

This configuration leads to a sampling scheme in which the detection pulse is increasingly time delayed with increasing pulse number (Fig. 6.13d). Adjusting the frequency offset to the cavity length difference determines the number of sampling time delay positions, until a zero delay reoccurs between two laser pulses.

Various methods are used for THz generation with fs-lasers which involve charged carriers in semiconductors and photoconductive antennas, or crystals using a nonlinear optical process, i.e., optical rectification or even nonlinear processes in an air plasma [621, 622]. In the following paragraphs, the main features of the techniques are described. Current research attempts to find better materials for THz generation, with regard to power output, and increased efficiency in converting visible to THz light over a broad frequency range. Improvements lead to very short THz pulses containing components at high frequencies with high peak power.

6.6.1 *Simplified Models of THz Pulse Generation and Detection*

The generation of a THz pulse by charged carriers is typically attributed to fast carrier generation (Chap. 4.7). The fast generation of the number of charge carriers on the fs to sub-ps timescale can be considered as a time-varying electrical current in a Hertzian dipole, which in turn leads to an electrical field, that changes on the same timescale as described by the Maxwell equations.

A simple model of a THz TDS can be obtained in the following form: two electrodes deposited on a semiconductor, separated by a few micrometers, form a photoconductive gap, which is illuminated by a visible fs-pulse that is assumed to have a Gaussian intensity distribution (Fig. 6.18). The pulse creates free electron–hole pairs with a generation rate $G(t)$. The initial fs pulse intensity leads to a Gaussian distribution of the free carriers N_C , distributed in time with a width Δt , and is described as:

$$G(t) = N_C \exp(-(t/\Delta t)^2). \quad (6.34)$$

The concentration of free carriers depends on the laser power focused in the gap, and an approximate value of $10^{15} \text{ cm}^{-3}/\text{mW}$ can be assumed [603, 623]. A bias voltage is applied to the electrodes, which are typically part of a deposited antenna structure, to improve free space emission of the THz pulse. The electric bias field induces a force and acceleration of the carriers. This leads to a separation of the hole and electron distributions which in part, where they do not compensate each other, also weaken the bias field E_B , due to screening caused by the polarization P of the space charge. The local field experienced by the carriers is:

$$E = E_B - \frac{P}{\alpha \varepsilon}. \quad (6.35)$$

The parameter α is a geometrical factor of the semiconductor, which is three for an isotropic medium, and ε is the dielectric constant of the substrate. The time dependence of the polarization is connected to the current density j and the recombination time τ_r between a hole and an electron:

$$\frac{dP}{dt} = -\frac{P}{\tau_r} + j. \quad (6.36)$$

The time-dependent current density as a function of velocity $v_{e,h}$ and mobility $\mu_{e,h}$ is:

$$j(t) = q_{e,h} N_C v_{e,h} = q_{e,h} N_C \mu_{e,h} E \quad (6.37)$$

with a direction, depending on the charge $q_{e,h}$ of electrons and holes. The current leads to the emission of an electric field pulse according to Maxwell's equations

$$E_{\text{THz}} \propto \frac{dj}{dt} = q_{e,h} \frac{dN_C}{dt} v_{e,h} + q_{e,h} N_C \frac{dv_{e,h}}{dt}, \quad (6.38)$$

including the change of hole and electron velocities. The electron–hole generation rate depends on the carrier trapping time τ_c . Short trapping times are obtained for semiconductors with a large number of defects, or precipitates such as metal clusters. With the generation rate G , the change of the carrier concentration is obtained:

$$\frac{dN_C}{dt} = -\frac{N_C}{\tau_c} + G(t). \quad (6.39)$$

The carrier acceleration follows as

$$\frac{dv_{e,h}}{dt} = -\frac{v_{e,h}}{\tau_s} + \frac{q_{e,h}}{m_{e,h}} E. \quad (6.40)$$

The momentum relaxation time or carrier collision time τ_s and similarly the electron–hole recombination time τ_r are below 100 fs. The trapping time τ_c often exceeds 1 ps. The dependence of the current density on the carrier mobility $\mu_{e,h}$ and the recombination, and therefore the switching time as a function of defects, indicates the challenges for materials design. A carrier velocity overshoot, and ballistic acceleration of carriers is possible, if the average collision time is much smaller than the momentum relaxation time. The collision time depends on the concentration of the electron–hole pairs n as a consequence of the fs-laser pulse power.

The detector material changes conductivity $\sigma(t)$ as a function of the incident time-delayed fs-laser pulse, with the impinging THz electric field serving as bias

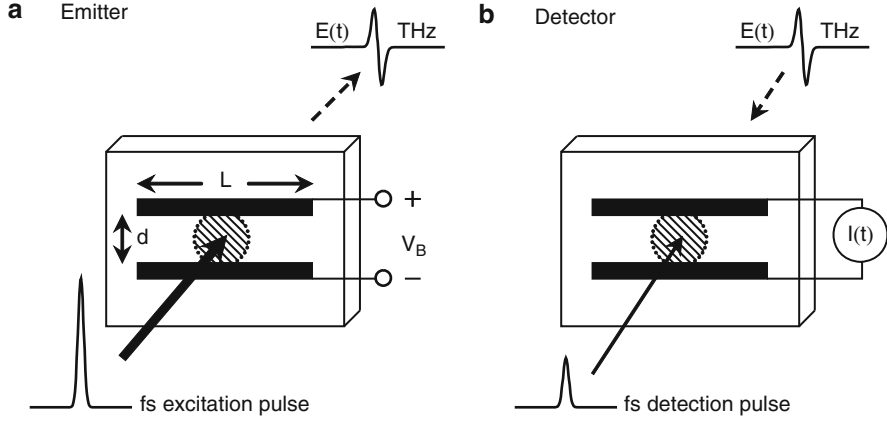


Fig. 6.16 (a) Emitter excited by an fs-laser pulse focused on an approximately $5\ \mu\text{m}$ photoconducting gap of LTG-GaAs substrate. The dipole antenna structure is biased at a voltage V_B to increase the THz pulse emitted from the device (not to scale). (b) Photoconducting detector chip with dipole antenna structure gated by an fs-laser beam. The incident THz pulse and electric field leads to a voltage across the photoconductive gap with varying polarity, which produces a measurable current I during electron-hole pair generation of the fs-laser pulse

voltage for the photo-injected electron-hole pairs, resulting in a time-dependent current density given by

$$j(\tau) = \frac{1}{T} \int_0^T E(t) \sigma(t + \tau) dt. \quad (6.41)$$

If the conductivity $\sigma(t) \propto e\mu N_C$ is a delta function, the detected current has the same pulse shape as the THz electric field. The time integrated field of the THz pulse is obtained if the conductivity is a square function, with a length longer than the THz pulse, but shorter than the repetition rate $1/T$ of the fs-laser. A suitable model assumes that the conductivity rise time is governed by the collision time τ_s , and the carrier recombination and/or trapping time.

The peak field amplitude is proportional to the peak current amplitude ΔJ

$$\begin{aligned} \Delta J &= \sigma \frac{w\delta}{d} V_B T / \tau_c = e\mu \bar{N}_C \frac{w\delta}{d} V_B T / \tau_c \\ &= e\mu T \frac{1-R}{h\nu} \frac{P_{\text{in}}}{d} \frac{V_B}{d}, \end{aligned} \quad (6.42)$$

where δ is the skin depth reached by the excitation laser (Sect. 3.8), $\bar{N}_C$ is the average carrier density, σ is the average conductivity of the semiconductor, and μ the electron mobility, which for SI GaAs exceeds $3,000\ \text{cm}^2/(\text{Vs})$ [623, 624]. $h\nu$ is the photon energy of the excitation laser, R the reflectance of the GaAs substrate,

and P_{in} the average incoming power of the excitation laser. The parameters d , distance, and w , effective width, characterize the surface area between the electrodes which can be illuminated.

The maximum electromagnetic energy which can be emitted by the photoconductor corresponds to the electrostatic energy which is stored in the gap region of the antenna structure. Considering the gap region as a capacitor, with capacitance C on the order of a fF, the energy for an applied bias voltage V_B is obtained

$$E = \frac{1}{2} C V_B^2, \quad (6.43)$$

which is typically a few pJ.

The amplitude of the emitted THz electric field is a linear function of the bias voltage V_B but shows a saturation behavior

$$E_{\text{THz}}^{\text{max}} \propto \frac{P_{\text{in}}}{P_0 + P_{\text{in}}}. \quad (6.44)$$

The saturation behavior (with P_0 corresponding to half of the maximum THz electric field) can be measured by increasing the bias voltage V_B or the input power P_{in} . For example, at 30 V a saturation power of approximately 60–70 mW can be obtained which, however, would destroy a real device.

Ti:Sa lasers emit in the wavelength range around 780 nm, and are typically combined with GaAs materials, due to its appropriate bandgap. In the simplest case, SI GaAs, or As⁺-implanted in GaAs, is used [623], but GaAs array emitters are also available [625]. Fiber lasers around 1,500 nm wavelength require different materials such as InP and InAs [626]. Pure GaAs and Si are transparent around 1,500 nm, which can be useful, for instance in delivering the exciting fs-pulse through an Si substrate to a layer, which contains electrical fields, as in a computer circuit. A typical antenna is a double dipole antenna at a bias voltage V_B of 30–50 V across a gap of 5 μm (Fig. 6.16).

Optical rectification uses a material with nonlinear properties to generate THz radiation (Sect. 2.5). The process down-converts an optical photon into two photons, one in the IR and one in the THz range. The process is comparable to a difference frequency generation process. The efficiency increases with intensity.

Second-order rectification is the most important nonlinear process (Sect. 2.5) for generating THz radiation from $\langle 110 \rangle$ crystals in the zinc-blend structure. ZnTe crystal belongs to the point symmetry group $43m$, which leads to a simplified tensor for the second-order susceptibility:

$$d_{ijk} = d_{ij} = \begin{pmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{36} \end{pmatrix}. \quad (6.45)$$

Further symmetry operations lead to the following identification $d_{14} \equiv d_{123} = d_{213} \equiv d_{25} = d_{312} \equiv d_{36}$. The optical beam needs to be perpendicular to the $\langle 110 \rangle$ ZnTe surface with the electric field described as [627]:

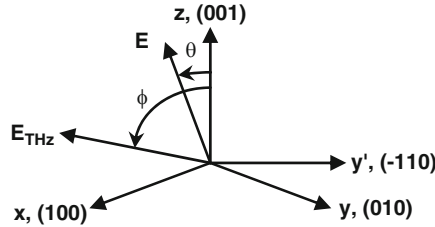


Fig. 6.17 The (110) plane lies in the plane of the paper and the graph illustrates the conversion from the original coordinates xyz to the dashed ones $x'y'z'$: θ is the angle between optical polarization and the z -axis while Φ gives the angle between the THz E -field and z -axis (adapted from [627])

$$E = \frac{E_0}{\sqrt{2}} \begin{pmatrix} -\sin(\theta) \\ \sin(\theta) \\ \sqrt{2} \cos(\theta) \end{pmatrix} \quad (6.46)$$

with $|E| = E_0$ and θ the angle between the E -field vector and the $\langle 001 \rangle$ axis (Fig. 6.17). Using the new coordinates, and rotation by 45° degrees in the $(x - y)$ -plane, the following nonlinear polarization is obtained:

$$P = d_{14} E_0^2 \begin{pmatrix} 0 \\ 2 \sin(\theta) \cos(\theta) \\ \sin^2(\theta) \end{pmatrix}. \quad (6.47)$$

From this, it follows that the maximum THz amplitude is achieved for both θ and Φ equal to 54.7° corresponding to $\sin(\theta) = \sin(\Phi) = \pm \sqrt{2/3}$ with $|P_{\max}| = \sqrt{4/3} d_{14} E_0^2$ [627].

The crystal thickness needs to be long enough, ideally as long as the coherence length, to amplify the THz radiation sufficiently, before destructive interference between the wave packet and the THz radiation reduces efficiency. The refractive index of ZnTe varies, but is the same at 780 nm and 2.28 THz ($n = 3.28$) resulting in a maximum of the coherence length. For THz generation up to 2.5 THz, a crystal length of more than 1.5 mm is ideal while for frequencies above 10 THz, due to the very small coherence length, crystals of $10 \mu\text{m}$ thickness, typically supported on a substrate such as quartz, can be used. At 5.3 THz, the transverse optical (TO) phonon in ZnTe is present, which leads to a suppression of emission around this frequency. Choosing different crystal materials such as GaSe [628] or GaP can extend or complement the emitted frequency range.

The techniques described retrace the THz pulse intensity in time. It is appropriate to highlight applications which retrace the THz pulse intensity in space. In the near-IR spectral range, high-speed systems, such as streak cameras, are available to monitor ps-pulses, by mapping the pulse on a CCD array. In the THz frequency range, it is possible to map the time intervals on an array. Direct mapping requires a THz array, or at least a line array. THz arrays are discussed in detail in Chap. 7 in the context of THz imaging. EOS can be used to map the near-IR detection pulse,

which was modified by the THz pulse, on an array designed for visible light or on a CCD camera imaging array [629]. In principle, this technique can monitor individual pulse traces with a chirped optical detection pulse [630] (Sect. 2.8).

6.6.2 Applications

An illustrative example is the transmission of a THz pulse through a 1 mm thick plane-parallel window material such as ultra-pure Si. Si has very little absorption and dispersion in the THz frequency range. It mainly introduces a time delay due to its refractive index. The pulse recurs approximately every 23 ps, corresponding to an optical length of $2nd$ given by the refractive index n of Si, and the geometrical length d of the window.

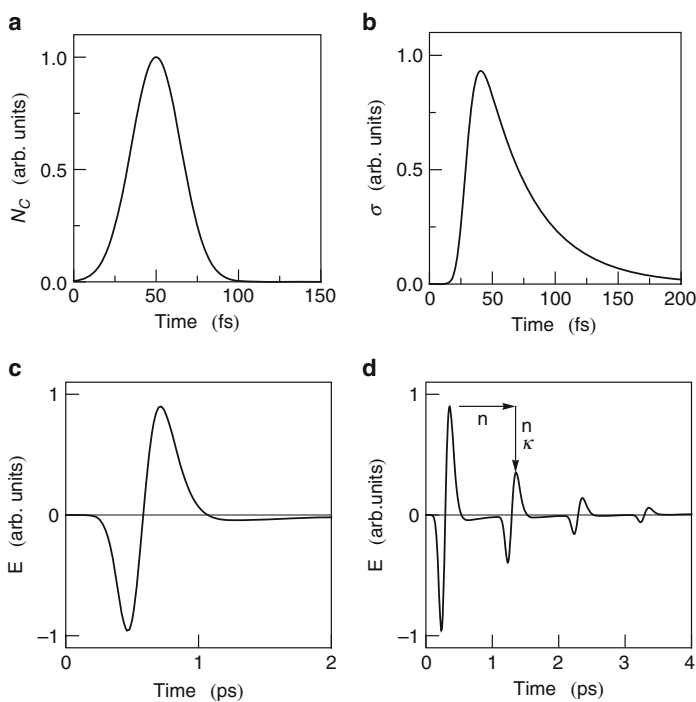


Fig. 6.18 Illustration of (a) Photo-injected carrier concentration N_C (6.39), (b) change of conductivity σ , (c) resulting emitted E-field of the THz pulse, (d) ringdown signal for a THz pulse shown in (c) transmitted through a plane-parallel plate of thickness d . The pulse shift and amplitude reduction of the following pulses is a result of the complex refractive index $\hat{n} = n + i\kappa$. The pulse peak is shifted by $2(\hat{n} - 1)d$ due to the average refractive index $\hat{n}$. The shape of the shifted pulse is very similar to the initial pulse shape, if ultra-pure Si is used, because the refractive index is reasonably constant throughout much of the THz range. The pulse amplitude of the following pulses is governed by the absorption κ and reflection losses at the plate interfaces because of the refractive index n

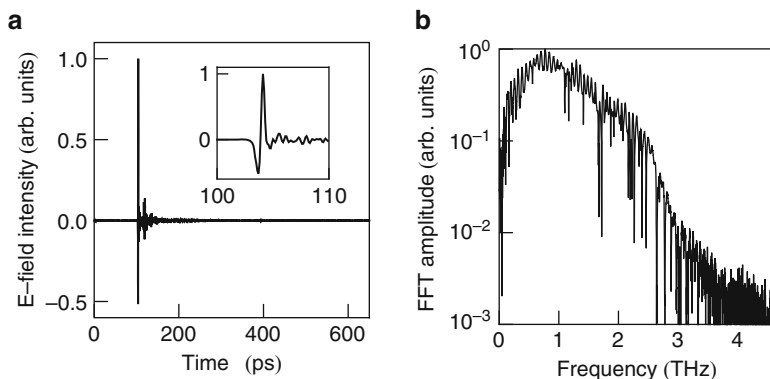


Fig. 6.19 (a) Typical electric field transient of a THz pulse in the time domain and (b) the Fourier-transformed spectrum of field amplitudes using a GaAs emitter and EOS detection giving a dynamic range in intensity of 10^6 . The spectral resolution of 1 GHz (1 ns delay time) allows for resolving residual water vapor absorption lines

The intensity decays due to reflection losses at each material interface, leading to a so-called ringdown signal. The pulse length is much shorter than the round trip time, so that interference is not observed between pulses of the transmitted pulse train. In the time domain, the refractive index, the thickness, and, under some conditions, the absorption can be immediately extracted by using the ringdown signal (Fig. 6.18).

A THz pulse can also be viewed as a wave package consisting of a superposition of a multitude of single frequency waves. The amplitude and phase at each frequency can be extracted by Fourier transformation. The assignment of water vapor lines illustrates the possibility of deriving amplitude and phase, absorption and dispersion, simultaneously at THz frequencies with very high accuracy (Fig. 6.19) [607].

Since THz TDS are commercially available in various forms at a reasonable price, the number of techniques and applications using TDS systems is expanding. For a review, see [631] and references within.

6.7 Coherent Source Spectrometers

In this section, spectrometers which are based on coherent cw tunable, and narrow band, i.e. $\Delta\nu/\nu \approx 10^{-6}$, THz sources are discussed. Naturally, the domain of such spectrometers is high-resolution spectroscopy, specifically molecular spectroscopy. Several approaches to coherent THz sources have been discussed in Chap. 4. From these, harmonic generators and backward wave oscillators (BWOs) can be used more or less directly for THz spectroscopy. It will be shown that gas lasers are also used, although some effort is required in order to achieve tunability.

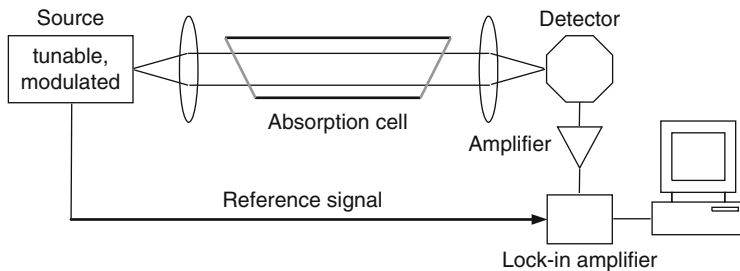


Fig. 6.20 Simplified scheme of a THz spectrometer for molecular spectroscopy. The frequency of the source is tunable and its power or frequency can be modulated by some method. The radiation is transmitted through an absorption cell and focused onto a detector, whose output signal is amplified and phase-sensitively detected with a lock-in amplifier

Spectrometers based on these sources were introduced in the 1970s to 1980s. More recently, spectrometers based on photomixing of either dye lasers [632], or cw diode lasers [633] in a photoconductive switch, have been reported, and a spectrometer based on a quantum cascade laser (QCL) has been demonstrated [331]. However, comparatively little spectroscopic work has been done with these spectrometers and therefore the focus here is on more established approaches. The principles of coherent source spectroscopy can easily be transferred, from the spectrometers described in the following sections, to spectrometers based on other sources.

Before discussing actual spectrometers, it is useful to consider the sensitivity limits. A simple spectroscopic system is shown in Fig. 6.20. Some type of THz source emits radiation with power P_S and frequency ν . The radiation is transmitted through an absorption cell, which is filled with a gas at a particular pressure. The transmitted radiation impinges on a detector, and generates an output voltage. The THz source is modulated by an appropriate method, and the output signal of the detector is measured with reference to the modulation frequency by a lock-in technique. For small absorption, the minimum detectable absorption $\delta_{\min}$ is defined as $\delta_{\min} = \frac{\Delta P}{P_S} \approx \alpha L$ with the absorption coefficient α of the absorbing medium and the length L of the absorption path. Ultimately, it is given by the ratio of the noise power P_N of the detector to the source power P_S . The lock-in amplifier converts an AC signal into a DC signal. This results in a factor of 2 power loss of the signal, while the noise power does not change. The minimum detectable absorbance is further increased because the modulation of the signal has to be taken into account. M is the modulation index of the detection scheme. Its value is between 0 (no modulation) and 1 (full modulation of the absorption feature). This leads to the following expression:

$$\delta_{\min} = \frac{2}{M} \frac{P_N}{P_S}. \quad (6.48)$$

As an example, a detector with an NEP of $10^{-14} \text{ W}/\sqrt{\text{Hz}}$ is considered. The THz source may deliver $10 \mu\text{W}$, the integration time is 1 s and the modulation depth is

0.3. For a frequency modulation scheme, this corresponds to about the half width at half maximum of a molecular absorption line and yields $\delta_{\min} = 6.7 \times 10^{-9}$. The fundamental detection limit of such a system is given by the shot noise limit of the THz source. Note that the power of the THz source is much higher than the power from the background ($P_B \ll P_S$), which can therefore be neglected. For a photovoltaic detector, the shot noise limit of $\delta_{\min}$ is given by (5.10)

$$\delta_{\min} = \frac{2}{M} \sqrt{\frac{h\nu}{\eta P_S \tau}}. \quad (6.49)$$

Here, η is the quantum efficiency of the detector and τ is the integration time. To convert the bandwidth B of the detector into an integration time, a filter with $B = 1/(2\tau)$ has been assumed. If the power is constant, the minimum detectable absorption increases with frequency. This is typical for the statistical nature of the shot noise limit, because increasing the frequency at constant power means fewer photons, and consequently larger relative noise. For a photoconductive detector, (6.49) needs to be multiplied by a factor of $\sqrt{2}$, due to the higher shot noise limit of such a detector (5.47). As an example consider a spectrometer with a THz source delivering $10 \mu\text{W}$ at 2.5 THz , a detector with a quantum efficiency of 0.2 , an integration time of 1 s and, a modulation depth of 0.3 . With these numbers, the minimum detectable absorbance is 4.8×10^{-8} . Obviously, with a state-of-the art detector it should be possible to achieve quantum-limited sensitivity. However, in real spectrometers, this is often not the case because other effects, such as power fluctuations of the source, or standing waves in the spectrometer, are a limitation. It should be noted that the sensitivity of a spectrometer can be increased significantly when using a source with higher output power. Alternatively, the lack of source power can be compensated for by using a more sensitive detector.

In the following three sections, the design and performance of different spectrometers are discussed. Emphasis is placed on those parts of the spectrometer which are specific for the THz range, rather than on particular spectroscopic features such as absorption cells or sample holders.

6.7.1 Spectrometers with Multiplier-Based Sources

As pointed out in the introduction to this chapter, spectrometers based on harmonic generation of THz radiation were the first to be used for molecular spectroscopy in the THz region. Modern multiplier-based spectrometers rely on the same principles as the early instruments, but are more powerful in the sense of frequency coverage, frequency agility, output power, and ease of handling.

As an example, a laboratory spectrometer for molecular spectroscopy is described which is based on the local oscillator technology for the HIFI instrument on the Herschel Space Observatory [552] (Fig. 6.21). The core of the spectrometer is the source, which consists of a commercial sweep synthesizer operating from 11

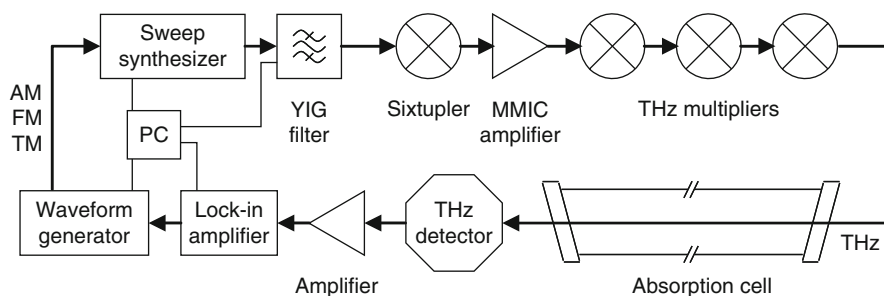


Fig. 6.21 Scheme of the multiplier-based THz spectrometer. The THz multipliers are exchangeable and with a set of these it is possible to provide radiation for spectroscopy up to 2.6 THz (adapted from [552])

to 18 GHz that has a low phase-noise output. It is frequency locked to $1:10^{11}$. It is important that the output of the synthesizer is not turned off while the system is stepping frequency. The synthesizer's output is filtered by a tunable YIG filter in order to provide the necessary spectral purity. The sweep synthesizer and the YIG filter are swept simultaneously. Different sixtuplers are used to reach 70–120 GHz. From this signal, different MMIC amplifiers generate 200–500 mW between 71 and 115 GHz. This amplification is an essential requirement for generating sufficient drive power to the following cascaded multipliers. A variety of cascaded doublers or triplers follows the power amplifiers and generates the THz radiation. With a set of exchangeable multipliers, it is possible to provide radiation for spectroscopy up to 2.6 THz. The THz radiation is passed through an absorption cell and detected with a Schottky diode (below 0.7 THz), an InSb detector (below 1.2 THz), or a composite helium-cooled Si bolometer (above 1.2 THz). The output signal of the detector is amplified and detected with a lock-in amplifier. Three frequency modulation schemes are implemented. For amplitude modulation (AM), the synthesizer is turned on and off with frequency f . For frequency modulation (FM), the source frequency is modulated with an applied sine wave, and $2f$ detection results in a second-derivative Gaussian line shape. Toneburst (TM) modulation is a combination of the two previous methods. In TM, a fast dithering sine wave is switched on and off at a slow rate; $1f$ detection then results in an approximate second-derivative Gaussian line shape. For best (S/N)-ratio, the modulation depth, and the modulation frequency, have to match both the absorption linewidth and the detector speed. The slow Si bolometer requires either AM or TM while the faster Schottky and InSb detectors can be operated with all schemes.

In a path length of 2 m, absorptions on the order of $\Delta P/P_S = 10^{-5}$ can be detected for frequencies less than 1.7 THz. Above 1.7 THz, this threshold increases to $\Delta P/P_S = 10^{-2}$, due to the rapid decrease in available source power P_S . Spectrometers such as the one described here are relatively simple and, especially with detectors operating at room temperature, they are attractive for a number of applications in the laboratory or in the field. A fast and sensitive spectrometer for gas monitoring and quantification operating around 250 GHz has been demonstrated

[634]. Because of its high spectral resolution, the spectrometer provides both specificity and low false alarm rates.

6.7.2 Spectrometers with Backward Wave Oscillators

Ever since their invention backward wave oscillators (BWOs) have been used for spectroscopy. Their attractive features are narrow linewidth, high output power, and tunability. Despite being a narrow band source BWOs are used not only for molecular spectroscopy, but also for solid state transmission or reflection spectroscopy. This is different from the sideband and difference frequency spectrometers discussed in the following sections, which are used exclusively for molecular spectroscopy. The main reason is the broad frequency coverage achievable with a single BWO in combination with >1 mW output power. In addition, the whole THz range up to 1 THz can be covered continuously with a few BWOs, which allows the measurement of broad features typically occurring in solid-state systems. As examples, BWO-based spectrometers, for high-resolution molecular spectroscopy and solid-state spectroscopy, will be described in this section. More details can be found in [372, 554].

Figure 6.22 shows the principal layout of a spectrometer for molecular spectroscopy. The key element is a phase-locked BWO. The reference frequency is provided by a mmW synthesizer. A small amount of the output radiation from the BWO is mixed in a Schottky diode with one of the harmonics generated from the millimeter wave radiation inside the diode. The mixing signal is the reference for the phase-lock loop. When the synthesizer frequency is tuned, the BWO frequency tracks it. Most of the BWO radiation is coupled into an absorption cell and an

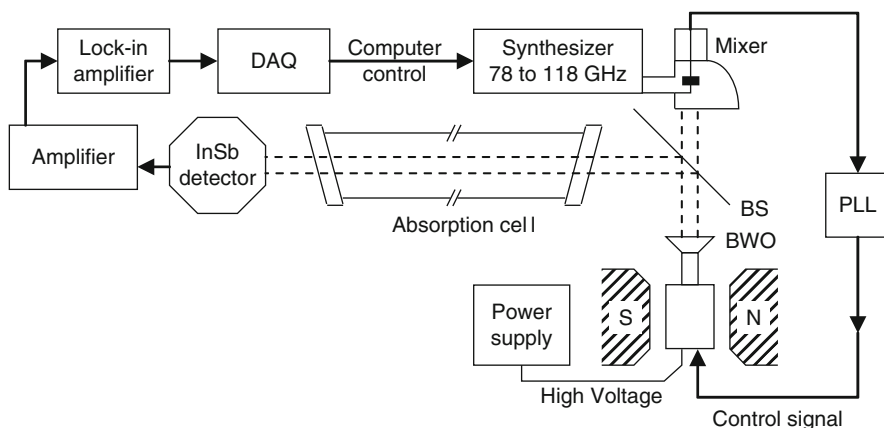


Fig. 6.22 Schematic diagram of an absorption spectrometer with a BWO source. The frequency of the BWO is controlled by a phase-lock loop (PLL). A portion of the power is mixed with the harmonic signal from a microwave synthesizer to provide the lock-in signal (adapted from [635])

InSb detector measures the fractional absorption with the lock-in technique. The linewidth of the phase-locked BWO is ~ 30 kHz, which is much smaller than the typical Doppler limited linewidth of a molecular absorption line at THz frequencies. It enables very precise measurements of line profiles and line shifts. In comparison with the MIM diode-based difference frequency spectrometer (Sect. 6.7.4), which provides a similar linewidth, the BWO spectrometer has the advantage of a much higher (more than three orders of magnitude) output power and, correspondingly, a higher sensitivity. Due to the high spectral purity and output power of the BWO, it is possible to perform saturation spectroscopy, and to resolve the ammonia rotational spectrum with Lamb dips and crossover resonances, which are only ~ 50 kHz wide [636]. The development of a BWO spectrometer without a phase-lock loop [637] uses the inherently small linewidth of a free-running BWO (< 20 kHz). A part of the BWO emission is sent through a Fabry–Pérot interferometer (FPI) and this serves as the frequency standard. The BWO is swept so fast, from one mode of the FPI to the next, that during this time its linewidth is not affected by thermal and power supply induced instabilities. A correspondingly fast data acquisition stores the absorption signal, together with the frequency determined by the FPI. The mode separation of the FPI should be small (in this case ~ 4 MHz) in order to minimize errors induced by interpolation of the frequency between two modes. With this spectrometer, it is possible to scan a large spectral portion (10^5 spectral resolution elements per second) with a linewidth as small as ~ 0.1 MHz.

In solid-state spectroscopy, the parameters of interest are the dielectric properties of the material or, in other words, the refractive index and the absorption coefficient. These can be extracted from a measurement of the transmission coefficient of the sample, the phase shift of the transmitted wave, the reflection coefficient, or the phase shift of the reflected wave. The measurement of two of these parameters is sufficient to determine the dielectric properties unambiguously, and without the need for a Kramers–Kronig transformation [638]. Depending on the sample properties, i.e. its transmission, usually either the transmission or the reflection parameters are measured. The design of a spectrometer for transmission measurements is shown in Fig. 6.23 [372].

Essentially, it is the THz analogue to a Mach–Zehnder interferometer. The spectrometer has two beam paths, one with the sample and one with a phase-shift compensator. Transmission measurements are made by sweeping the frequency of the BWO. This is achieved by ramping its supply voltage up or down. As pointed out above, the unstabilized linewidth of $\Delta\nu/\nu \approx 10^{-5}$ is sufficient and phase-locking is not required. The output beam is shaped by lenses and diaphragms in order to provide a Gaussian beam profile in front of the sample. A wire grid polarizer defines the polarization, and a wire grid step attenuator is used for adjusting the power on the sample under investigation. The transmitted radiation is focused onto a detector. For most measurements, a Golay cell is sufficient because of the high power available from the BWO. The transmission is measured twice, once with the sample in place and once without the sample. By dividing both spectra, the transmission function of the spectrometer is eliminated, and the transmission of the sample is determined. For phase measurements, a beam splitter directs a part of the radiation onto a phase-shift

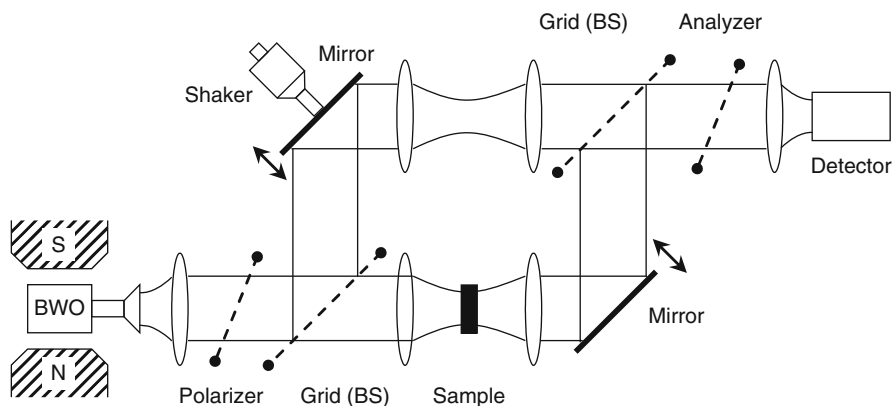


Fig. 6.23 BWO spectrometer for dielectric measurements. The transmission of the sample is measured as a function of the BWO frequency in transmission geometry without the grid beam splitter (BS) and the upper beam path. The phase is measured in a Mach-Zehnder type configuration with a grid beam splitter, where the frequency of the BWO is tuned. One mirror is movable to compensate for phase-shifts, so that the signal at the detector is zero. The second mirror is mounted on a motor (shaker) and modulates the phase for lock-in detection. In this case, the movement of the phase-shift compensator is measured as a function of the BWO frequency (adapted from [639])

compensator. This is a movable mirror. While tuning the frequency of the BWO, the phase shift compensator is moved so that a zero-signal is maintained on the detector. The displacement of the mirror is measured as a function of the frequency of the BWO. As with the transmission measurement, one measurement with the sample and one without the sample is performed. From this the phase shift induced by the sample can be determined. The reproducibility and accuracy of the data is limited by the matching of the sample and reference spectra. Typically, this is 0.1% to 1% in transmission and $\sim 1 \mu\text{m}$ in phase. There are various other possibilities e.g. for reflection or birefringence measurements. These can be found in [372] for example.

6.7.3 Spectroscopy by Tunable Terahertz Sideband Generation

Above 1–1.5 THz, BWOs, or techniques based on harmonic generation of THz radiation, do not provide sufficient power and frequency tunability for high-resolution molecular spectroscopy. Laser sideband spectroscopy, which is based on mixing the radiation of a fixed frequency THz gas laser, and some type of frequency tunable microwave source in a Schottky diode, provides the possibility of sensitive spectroscopy at higher THz frequencies. This technique dates back to the 1970s [347–349]. In the following years, it has been improved significantly and has been successfully used to detect high-resolution spectra of many molecules, free radicals, and ions [350, 351].

The principle of the technique relies on the ability of Schottky diodes to generate sidebands when irradiated with radiation from two sources. One source is a THz gas laser operating at ν_{THz} and emitting more than 10 mW of power. The other is a microwave source operating at ν_{MW} . The Schottky diode generates the sideband frequencies $\nu_{\text{SB}} = \nu_{\text{THz}} \pm m\nu_{\text{MW}}$ with $m = 1, 2, 3, \dots$. The $+$ sign refers to the upper sideband and the $-$ sign to the lower sideband. Physically, the generation process is the same as in the case of a Schottky diode mixer but with two differences. First, the microwave radiation is quite powerful, typically tens of mW, as compared to the signal radiation, and second, the sideband frequency is close to the laser frequency, while for heterodyne detection the frequencies of the signal and local oscillator radiation are close to each other. In some sense, sideband generation is the inverse process to heterodyne detection. The sideband radiation is emitted from the Schottky diode in the opposite direction to the incoming laser radiation. However, laser radiation is emitted as well as reflected from the Schottky diode and most of the power in the emitted beam is at the fundamental laser frequency. This makes it necessary to have sophisticated schemes to separate the sideband radiation from the laser radiation. It should be noted that not only the first-order sideband ($m = 1$) but also higher-order sidebands can be used. However, the power emitted in the sideband decreases markedly for higher-order sidebands [350]. Nevertheless, almost the entire frequency range from 0.5 to 3 THz is accessible by sideband generation [350]. The output power emitted in the first order is up to 10 μ W.

The principal layout of a laser sideband spectrometer is shown in Fig. 6.24a, with a powerful THz gas laser providing the fundamental frequency. Since sideband generation is not a very efficient method of THz generation, and because the power emitted in the sidebands grows linearly with the power from the gas laser, 10 mW or more is needed to generate sufficient sideband power. Also, it should be possible to operate the laser on as many frequencies as possible to provide good frequency coverage, and its power and frequency should be stable. Microwave radiation can be supplied by any kind of commonly available source, such as a microwave synthesizer [350, 351] or a BWO [640]. Which type is used depends mainly on the frequency to be covered. While a typical synthesizer has a frequency coverage of about 0.1–40 GHz, yielding sidebands very close to the laser carrier frequency, a BWO provides a larger tuning range as well as a higher frequency, which leads to a larger offset of the sidebands. In any case, the microwave frequency should be stabilized and controlled by a phase-lock system, which is inherent in synthesizers, but requiring some effort for the BWO. The frequency of the microwave source is automatically swept and controlled. The radiation from the laser is quasi-optically coupled to the Schottky diode via a diplexer. The Schottky diode is in an open structure mount (Sect. 5.5.2), which is essentially the same as for a quasi-optical Schottky mixer, i.e. the Schottky diode is whisker-contacted and equipped with a $\lambda/4$ corner cube antenna. A different method is used for the coupling of a microwave source. Depending on its frequency, a waveguide system is used, replacing the coaxial output for the intermediate frequency in a Schottky mixer, which can be done at all frequencies where waveguides are available, typically up to about 115 GHz. Alternatively, if a higher frequency source, such as a BWO, is used, its radiation is coupled quasi-optically to the Schottky diode [640].

Special care needs to be taken to separate the laser carrier frequency from the sideband frequency. This is especially demanding in the case of a low-frequency microwave source where the two frequencies are only about 1% apart. A Martin–Puplett interferometer as diplexer has been found to be most efficient. If used for sideband separation, one port serves as the input for the laser radiation, while the other port is the output for the sideband radiation. In addition, a monochromator or a Fabry–Pérot interferometer is used for further suppression of the carrier frequency. A common problem is baseline variations, with a period of a few 10 MHz, in the spectrum. Since the linewidth of rotational transitions is a few MHz, this may cause serious degradation of the data. The reasons are manifold. Standing waves can occur between the Schottky diode which, due to its corner cube design, is a very good reflector, and any other optical component, such as the detector. Another reason is leakage of sideband radiation into the laser. These effects can be reduced by putting a wavelength modulator into the beam path, or by using an isolating

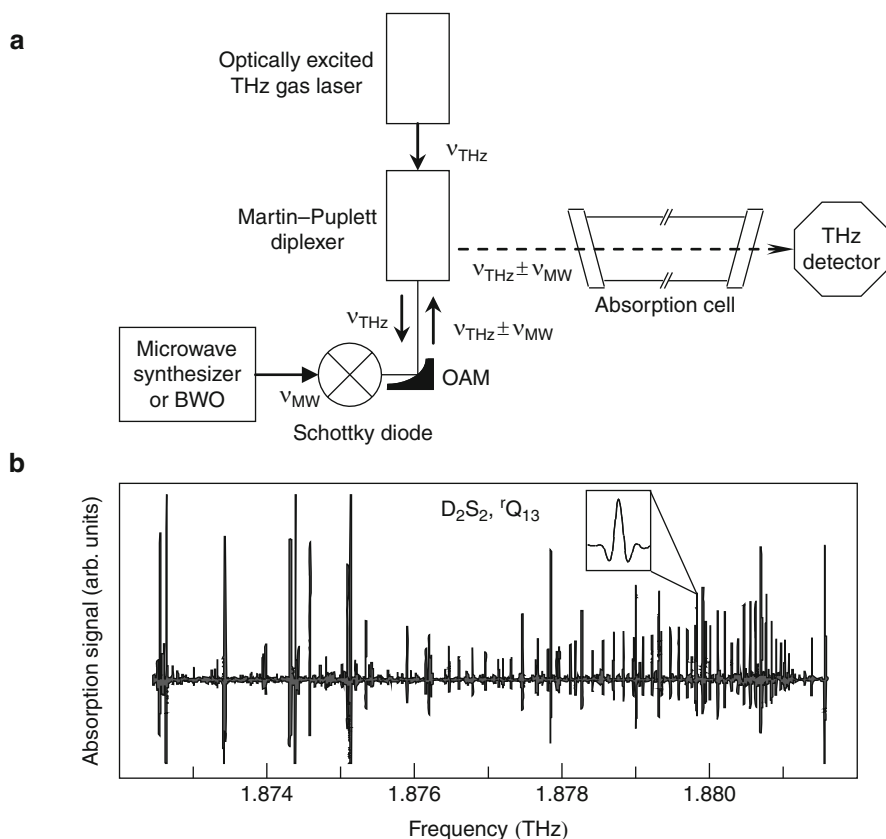


Fig. 6.24 (a) Scheme of a laser sideband spectrometer. (b) Absorption spectrum (second derivative signal) of D_2S_2 measured with a laser sideband spectrometer. The spectrometer provides broad frequency tunability, high resolution, and high accuracy at the same time. The inset displays an expanded view of one line (adapted from [640])

Fabry–Pérot interferometer in front of the laser output. The sidebands can be detected by any sensitive THz detector, such as an InSb hot electron bolometer, a Ge:Ga photoconductive detector, or a Si bolometer. Heterodyne detection has also been used [641]. This has the advantage of an inherent separation of the sidebands and laser radiation. Modulation of the sideband radiation and use of the lock-in detection technique is required to achieve maximum sensitivity. A convenient way of doing this is to modulate the bias voltage of the Schottky diode. Despite sensitive detectors and lock-in detection, the minimum detectable absorption is no better than 10^{-6} [351]. The reasons for this are fluctuations of the laser power and leakage of the carrier frequency, which increase the minimum detectable absorption beyond the shot noise limit according to (6.49).

The linewidth of the sideband radiation is determined by the frequency stability of the gas laser, because the contribution from the microwave source, or from the BWO if it is phase-locked, is negligible. For a free-running laser, the linewidth is about 100 kHz and this can be improved by frequency locking of the laser. However, for Doppler limited spectroscopy, this is not necessary since typical Doppler widths are >1 MHz, and the spectral resolution is determined by the intrinsic linewidth of the absorbing molecules. An example of the capability of this technique is shown in (Fig. 6.24b), which displays the absorption spectrum of D_2S_2 . The broad frequency tunability, high resolution, and high accuracy are evident.

6.7.4 Spectroscopy by Difference Frequency Generation

A somewhat similar technique to sideband generation is difference frequency generation by mixing the radiation of two lasers, operating in the IR or visible spectral region, in a nonlinear element (Sect. 4.5.2). While the concept is quite simple and dates back more than 30 years [354], it was 10 years later that K.M. Evenson and his co-workers implemented this method successfully for spectroscopic applications [355]. They used two CO_2 lasers and a metal–insulator–metal (MIM) diode for generation of the difference frequency. One of the CO_2 lasers was a waveguide laser, which is frequency tunable by $\Delta\nu_{WG} = \pm 120$ MHz and, by tuning its frequency, the THz radiation can be swept (second-order spectrometer, $\nu_{THz} = |\nu_1 - \nu_{WG}| \pm \Delta\nu_{WG}$).

In another configuration (third-order spectrometer), both CO_2 lasers are at fixed frequencies, and frequency tunable microwave radiation ν_{MW} from a synthesizer is coupled into the MIM diode, to allow for frequency tuning of the THz radiation. In this case, the MIM diode generates sidebands at the difference frequency of the two CO_2 lasers, i.e. $\nu_{THz} = |\nu_1 - \nu_2| \pm \nu_{MW}$. Since the third-order spectrometer provides a much better frequency coverage, along with a higher absolute accuracy, it has found more applications, and this type of spectrometer will now be described in more detail (Fig. 6.25a).

The MIM diode (sometimes also called an MOM diode, metal-oxide-metal diode) consists of two metals which are separated by a thin (~ 1 nm) metal oxide

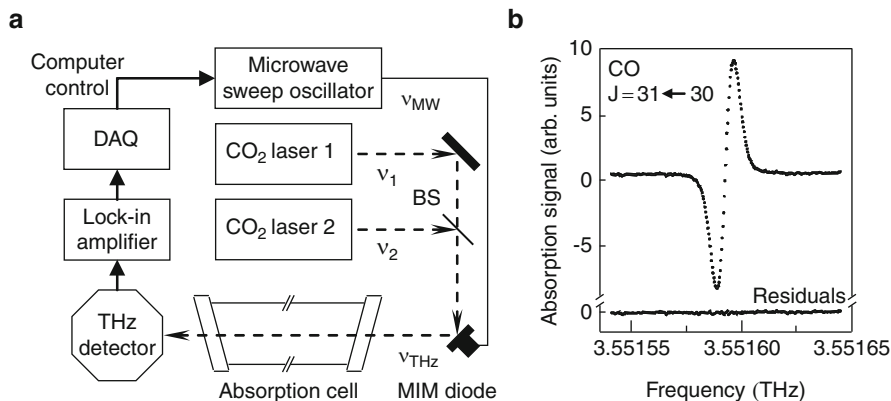


Fig. 6.25 (a) Block diagram of a frequency tunable spectrometer. The radiation of two CO₂ lasers and a microwave source is mixed in a metal–insulator–metal diode. The difference frequency of the two CO₂ lasers is in the THz range and the tunable microwave radiation enables frequency tuning around this difference frequency (adapted from [555]). (b) Absorption spectrum (first derivative signal) of the $J=31 \leftarrow 30$ rotational transition of CO in the ground state (upper trace). The lower trace displays the residuals obtained from a least-squares fit to the measured line shape. The frequency accuracy is better than 10 kHz for 2 s integration time (adapted from [645])

layer. For second-order generation, a tungsten-nickel diode is the most efficient, while for third-order generation a tungsten-cobalt diode is more effective. The MIM diode has a tungsten whisker similar to that of quasi-optical GaAs Schottky diodes. The whisker is electrochemically sharpened to have a conical tip with a radius of ~ 1 nm. The other metal is a post on which the thin oxide is grown. When a bias voltage is applied to the diode, electrons tunnel from one metal to the other. This results in a slightly nonlinear I-V curve. Although the nonlinearity is far weaker than that found in GaAs Schottky diodes, it is sufficient for mixing. Since the current transport is based on tunneling [642], which is a faster process than thermionic emission, a MIM diode is a much faster device than a GaAs Schottky diode, working well for the mixing of IR and visible radiation despite its weak nonlinearity [356]. With about 150 mW from each of the CO₂ lasers and 5–10 mW from the microwave source, about 100 nW of third-order THz radiation can be generated (about three times more for a second-order spectrometer). The laser radiation is focused onto the MIM diode where the conical tip of the whisker serves as an antenna [357]. The microwave radiation is coupled to the MIM diode with a bias tee. The THz radiation is emitted by the whisker, which serves as a long-wire antenna for THz radiation [643]. Both CO₂ lasers are Lamb-dip stabilized. The accuracy of the frequency stabilization of each laser is better than 25 kHz, yielding an overall uncertainty of $\sqrt{2} \times 25 \text{ kHz} = 35 \text{ kHz}$ for the THz radiation [644]. This corresponds to a frequency accuracy of about 10^8 , which is about one order of magnitude better than that of a laser sideband spectrometer. The THz radiation is sent through an absorption cell and the fractional absorption is measured with the lock-in technique using an InSb detector, a Si bolometer, or a Ge:Ga photoconductive detector. Due to

the many laser lines available from CO₂ lasers (note that all the isotopologues of CO₂ and even the weakest laser lines can be used), the spectral coverage of this spectrometer ranges from a few hundred GHz to 6 THz, with an almost continuous coverage up to 4.5 THz [555]. The frequency coverage can be further extended to 9 THz by replacing one of the CO₂ lasers with a ¹⁵NH₃ laser, using a Ge:Be photoconductor as detector [358]. Due to its excellent frequency precision, this spectrometer is especially suited for measurements where very high resolution and accuracy are required, such as measurements of the shape of an absorption line (Fig. 6.25b). It is worth noting that the largest contribution to the uncertainty in the measurement of a transition frequency comes from finding the line center of a Doppler broadened line of the molecule under investigation, which is about 10% of the linewidth. The minimum detectable absorption is roughly 10⁻⁴ [648], which is only slightly worse than that of a laser sideband spectrometer, although the source power is almost 100 times less. Here, the difference frequency method profits from the large separation of the carrier frequencies from the THz frequency, and this is a major difference compared to the sideband method. This spectrometer is called a tunable far-infrared (TuFIR) spectrometer. Note that similar schemes have been realized using other lasers and mixers such as Ti:Sa lasers, cw diode lasers, and LTG-GaAs photomixers (for example, [647, 648]). These approaches have not yet reached the performance of the TuFIR spectrometer in terms of spectral resolution and frequency coverage. However, they are much more compact, robust, and versatile. A particularly interesting approach is a fiber-assembled system operating at 1.5 μm. The mixer is an integrated high speed photodiode and the photoconductive detector is based on LTG-InGaAs/InAlAs multi-layer structures which allow coherent detection (Sect. 7.4.2) [346]

6.8 Heterodyne Spectroscopy

Heterodyne spectroscopy in the THz region is a powerful and important spectroscopic technique for remote sensing. Since heterodyne detection is inherently narrow band, i.e., a bandwidth of less than 1 MHz can be easily obtained, a major application is in molecular spectroscopy. For example, the chemical composition, the evolution, and the dynamical behavior of astronomical objects such as star forming regions, molecular clouds, photo-dominated regions around young and bright stars, or shock processes in the interstellar medium, can all be investigated by high-resolution spectroscopy. Some astronomically relevant emission lines are rotational transitions of CO [94], which are distributed across the whole THz region, for example, the CII fine structure line at 1.6 THz [649], the 1 → 0 transition of HD at 2.7 THz, and the OI fine structure line at 4.7 THz [650]. The HIFI heterodyne spectrometer on the Herschel Space Observatory has provided a wealth of exciting data, which can be found for example in [558]. Furthermore, THz heterodyne receivers play an important role in the investigation of the Earth's atmosphere. For example, trace gases such as OH and ClO have been detected with high spectral

resolution, in order to obtain the height distribution of these molecules in the stratosphere [171, 524, 651]. THz heterodyne receivers are also valuable for the investigation of planetary atmospheres and comets [652].

The important features of a heterodyne spectrometer are shown in Fig. 5.28. For spectroscopic applications, it is not sufficient to detect the intermediate frequency (IF) signal at the output of the heterodyne receiver. Instead, spectral analysis of the signal is required. One way of achieving this is to replace the detector at the end of the heterodyne chain by a spectrum analyzer. However, commercially available spectrum analyzers are scanning devices, and only one frequency or resolution channel is measured at a time. Since heterodyne spectrometers are often employed in space- or airborne missions this is not acceptable, because the expensive measurement time is not used efficiently. Instead of scanning, the whole IF band has to be analyzed simultaneously with a so-called back-end spectrometer. However, before coming to these devices, some system aspects will be discussed. For reviews on THz heterodyne spectroscopy, see for example [653–655].

6.8.1 System Aspects

Ultimately, it is not the performance of single components, but the overall performance of the whole heterodyne spectrometer, which is important. Single components may have an exceptional performance but, if the system design does not take into account the interplay between the various components, the heterodyne receiver will not be optimized. Therefore, a few system aspects will be discussed here. In general, the optimal goal is to obtain as much high quality data, i.e. data with a significant (S/N)-ratio in a given time. This is not only a question of quality but also a question of economy, for example to make the most efficient use of expensive observing time on a satellite.

For a practical heterodyne spectrometer, the minimum detectable temperature ΔT_{MIN} is of prime importance, because it describes the temperature contrast which can be resolved within a certain integration time τ . For a total power receiver, it is given by

$$\Delta T_{\text{MIN}} = \frac{k_{\text{REC}} T_{\text{REC}}}{\sqrt{\tau \Delta \nu + \Delta P/P}}. \quad (6.50)$$

If the term $\Delta P/P$ is neglected, this equation reduces to the radiometer equation (5.63). The basic assumption is that the noise is uncorrelated and integrates with the square root of the integration time. However, for a nonideal receiver or spectrometer, other sources, such as gain instabilities of the IF amplifiers or of the back-end spectrometer, contribute to the sensitivity fluctuations, and raise ΔT_{MIN} . This is taken into account by the additional term $\Delta P/P$.

Radiometric calibration of a heterodyne spectrometer is usually accomplished by the Y-factor method, as described in Sect. 5.5.1. From the noise temperature, the

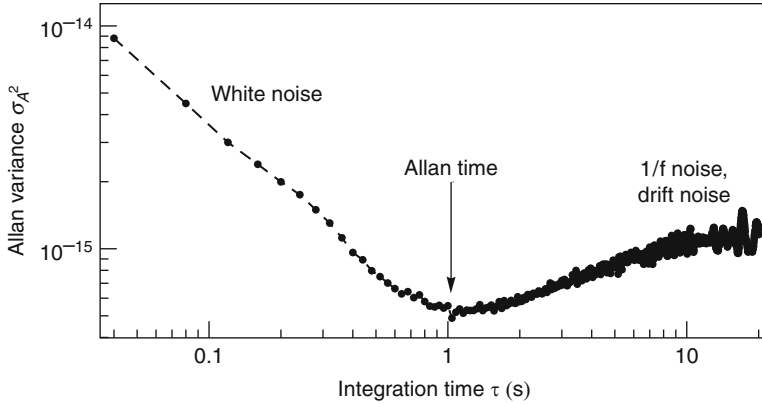


Fig. 6.26 Example of an Allan variance measurement. The Allan time is 1 s. For shorter times, the Allan variance decreases as τ^{-1} as expected for white noise, while for longer times it increases due to $1/f$ noise or drift noise

minimum detectable signal temperature can be calculated, provided that $\Delta P/P$ is negligible or known. For an analysis of the noise stability of a heterodyne spectrometer, it is useful to consider the “Allan” variance σ_A^2 (named after D. W. Allan, who introduced it for the characterization of atomic frequency standards [656]). It can be described by

$$\sigma_A^2(t) = \sigma^2(t)/2 = at^\beta + b/t + c, \quad (6.51)$$

where σ is the standard deviation of the signal (here $\sigma(t) = \Delta T_{\text{MIN}}$). For a receiver with a noise spectrum that contains white noise, $1/f$ noise, and drift noise, the Allan variance takes the right-hand side of (6.51), where a , b , c , and β are constants, which depend on the specific receiver. For example, for short integration times the Allan variance decreases normally as τ^{-1} , as expected for white noise. For longer integration times, the drift will dominate as shown by the term at^β . In this case, the variance starts to increase with a slope β , which is experimentally found to be between 1 and 2. Therefore, an optimum integration time (Allan time t_A) exists, after which observing efficiency is lost (Fig. 6.26) [657, 658]. The Allan time marks the crossover from white noise to $1/f$ noise or drift noise. In a typical measurement, difference data (signal on minus signal off) are taken and integrated during a period τ . If this integration time is longer than the Allan time of the receiver, integration efficiency will be lost. In most THz heterodyne receivers, the Allan time is limited by the front-end. Typical receiver Allan times are 8–9 s (SIS mixer with 100 MHz IF bandwidth [659]) or ~ 0.3 s (HEB mixer with 60 MHz IF bandwidth [660]), while back-ends have Allan times of hundreds of seconds.

Ideally, the spectrum measured with a heterodyne spectrometer is a one-to-one transformation of a THz spectrum into a spectrum, centered around the IF at a few GHz. Effectively, the received THz spectrum is convolved with the line shape of the LO. In order to preserve the absolute frequency of the spectrum, the absolute

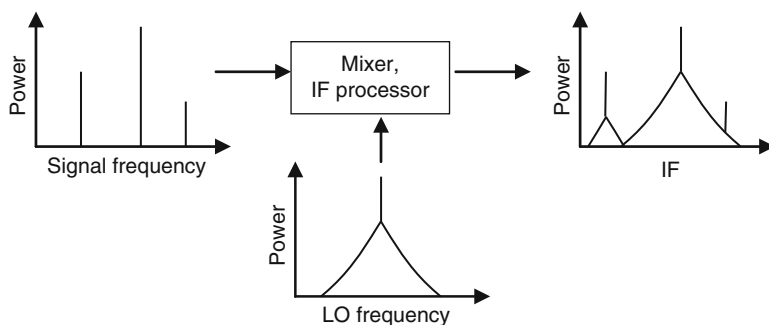


Fig. 6.27 Influence of LO noise on the output spectrum of a heterodyne receiver. The frequency width of the LO is broadened by noise. The signal frequencies which appear at the IF output, are convolved with the LO noise. The central line of these is the strongest and therefore has the strongest noise component. The line on the right is nearly masked by the LO noise

frequency of the LO must be known. Usually, an accuracy ten times better than the resolving power of the heterodyne spectrometer is sufficient. Obviously, the frequency stability should be equally good. In order to be able to measure a spectrum with weak lines, adjacent to strong lines, the LO has to contain all of its power in a very narrow frequency band. In other words, the spectral power density of the LO has to fall rapidly to a very small value at a frequency offset, which corresponds to the spectral resolution of the spectrometer. Noise in the LO broadens the emission of the LO and degrades its performance. This is illustrated in Fig. 6.27. Three emission lines are in the input band of the heterodyne spectrometer. The strongest is in the center and weaker lines appear on each side. The line of the LO is broadened by noise, which decreases with increasing offset frequency. The three emission lines appear at the IF output, and each line is convolved with the LO noise. The central line of these is the strongest and therefore has the strongest noise component. The weaker lines have less noise, but the line on the right is nearly masked by the noise from the strong signal line, and frequency, power, and line shape can be determined only with considerable uncertainty.

An actual LO will exhibit both an amplitude noise modulation $A_n(t)$ and a phase noise modulation $\Phi_n(t)$. Instead of having an ideal sinusoidal waveform, with an amplitude V_0 and a frequency ν_0 , the waveform of the LO is $V(t) = V_0[1 + A_n(t)] \cos[2\pi\nu_0 t + \Phi_n(t)]$, where $A_n(t)$ and $\Phi_n(t)$ are random processes. Both modulation processes produce sidebands on either side of the carrier frequency ν_0 . In the case of amplitude-modulated noise, these are in phase, and in the case of phase modulated noise they are in anti-phase. If the LO is operated in saturation, the amplitude noise is largely suppressed and can often be neglected. The phase noise is specified in a bandwidth of 1 Hz, in decibels relative to the power of the carrier, and the unit is dBc/Hz (note that dBc is not an SI unit). It should be noted that the phase noise power spectral density of a multiplier-based LO increases with the multiplication N as N^2 , and the phase noise level increases as $20 \log(N)$.

6.8.2 Back-End Spectrometers

The output of the IF amplifier is a time-dependent voltage. Any spectroscopic signature needs to be retrieved from this signal. Two different approaches can be followed. Either the spectrum can be transformed directly into the frequency domain, or it is measured in the time-domain and correlation techniques are used to generate the power spectrum (Fig. 6.28).

The filter bank spectrometer belongs to the first group. The IF signal is split into many frequency channels where each has a separate bandpass filter and a power detector [661]. Since they contain many components, filter banks are quite large and power consumptive. They are mainly used for atmospheric observations, where their wide bandwidth and flexibility is advantageous.

Another type of back-end spectrometer is the acousto-optical spectrometer (AOS). Its central component is a Bragg cell, which consists of a piezoelectric transducer in close mechanical contact to a crystal, typically LiNbO_3 . The IF signal is transformed by the transducer into a sound wave, which propagates in the crystal and generates a grating. A red laser diode provides a collimated beam, which illuminates the crystal and is diffracted by the grating. Since the period of the grating depends on the IF, the angle of diffraction is frequency dependent. The diffracted light is detected with a linear CCD array, where each pixel corresponds to a particular angle of diffraction and IF frequency. The frequency band is limited by the properties of the crystal, which impact the attenuation of the acoustic wave. For LiNbO_3 , this band is between 1.6 and 2.6 GHz. The spectral resolution is ~ 1.5 MHz, the dynamic range is >13 dB, and the Allan time, i.e. the maximum useful integration time, is >600 s [662]. Although an AOS contains somewhat delicate optical components, it can be built to be very compact and rugged. In practice, AOSs have been used on the SWAS and ODIN satellites [663, 664]. The

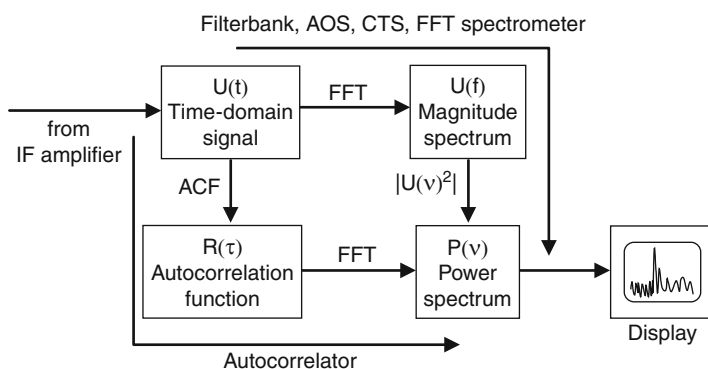


Fig. 6.28 Principal paths to obtain a spectrum, i.e. the power of a signal as a function of frequency, from a time-varying IF signal. The autocorrelator spectrometer performs as a first step an autocorrelation of the time-domain signal, which is followed by a Fourier transformation. All other spectrometers perform a Fourier transformation first

bandwidth of an AOS can be increased by using a LiNbO_3 Bragg cell with four transducers, one on top of the other and each covering an IF band of 1 GHz [662]. For SOFIA and for the HIFI instrument, such wide band 4 GHz AOS are in use.

With an autocorrelator, the spectrum is obtained by using the Fourier-transform relationship between the correlation function of the IF signal and its power spectrum. Autocorrelators can be constructed with either digital or analog electronics. In a digital autocorrelator, a fast analog-to-digital converter (ADC) samples the IF signal. Time shifting, multiplying, and Fourier transformation are subsequently performed with high speed digital electronics. The bandwidth of a digital autocorrelator is limited by the ADC, which must digitize the IF signal with, at least, twice the IF frequency. The subsequent digital electronics do not need to be as fast, because the output of the ADC can be processed in parallel. In an analog autocorrelator, the limitation set by the speed of the ADC is overcome by using high-speed analog electronics. Such a device, covering an IF band from 0.3 to 3.9 GHz with 33 MHz spectral resolution and an Allan time of a few seconds, has been demonstrated [665, 666]. Autocorrelators are especially attractive for use in interferometers, because of their capability of cross correlating signals from different antennas.

The chirp transform spectrometer (CTS) is based on the chirp transform algorithm, a generalized form of the Fourier transformation. In a CTS, a chirp signal is generated by impulsing a dispersive surface acoustic wave filter. This is mixed with the IF signal and convolved with a matched surface acoustic wave filter. As an example, the CTS for the GREAT heterodyne spectrometer on SOFIA has a dynamic range of 30 dB and an Allan time of ~ 550 s. It covers a bandwidth of 215 MHz with a spectral resolution of 41.7 kHz [667].

The development of ADCs with GHz sampling rates, and powerful field programmable gate arrays, has paved the way for digital fast Fourier-transform spectrometers, where the IF signal is directly digitized. A state-of-the-art spectrometer provides an Allan time of $\sim 4,000$ s, a dynamic range of > 30 dB, and a spectral resolution of 100 kHz in an IF band of 2×1 GHz [668]. Such back-end spectrometers are now in use in many THz observatories including ALMA and SOFIA. With the ongoing rapid progress in the performance of ADCs, similar progress for the back-end spectrometers can be expected.

6.8.3 Examples of Heterodyne Spectrometers

As mentioned above, heterodyne spectrometers are predominantly employed in remote sensing applications. The major reason why it is scarcely used for laboratory spectroscopy is the limited spectral coverage and its relative complexity.

A few examples may illustrate the state-of-the art. A suite of mm-wave and THz heterodyne receivers is operated on the Aura satellite [669], which was launched in July 2004. The THz receivers map the concentration of atmospheric species such as HCl, ClO, BrO at 0.64 THz, and OH at 2.5 THz [524]. A major requirement is a mission duration of several years, in order to cover many annual and diurnal

cycles of the atmosphere. Therefore, cryogenic SIS or HEB mixers were not an option. Also, the technical readiness level of HEBs was low at the time of the definition of the mission. Planar Schottky diode mixers with MMIC amplifiers are used in all receivers. A particular challenge was the 2.5 THz LO, which had to deliver an output power of a few mW for a lifetime of more than three years. This was solved with an optically excited THz gas laser. Filter banks, as well as digital autocorrelators, served as back-ends. The receivers deliver a wealth of data regarding global distribution, as well as height distribution, of many trace molecules in the stratosphere and upper troposphere (Fig. 6.29).

Probably, the most complex THz heterodyne receiver ever built is the HIFI on the Herschel Space Observatory. The instrument design was driven by both the

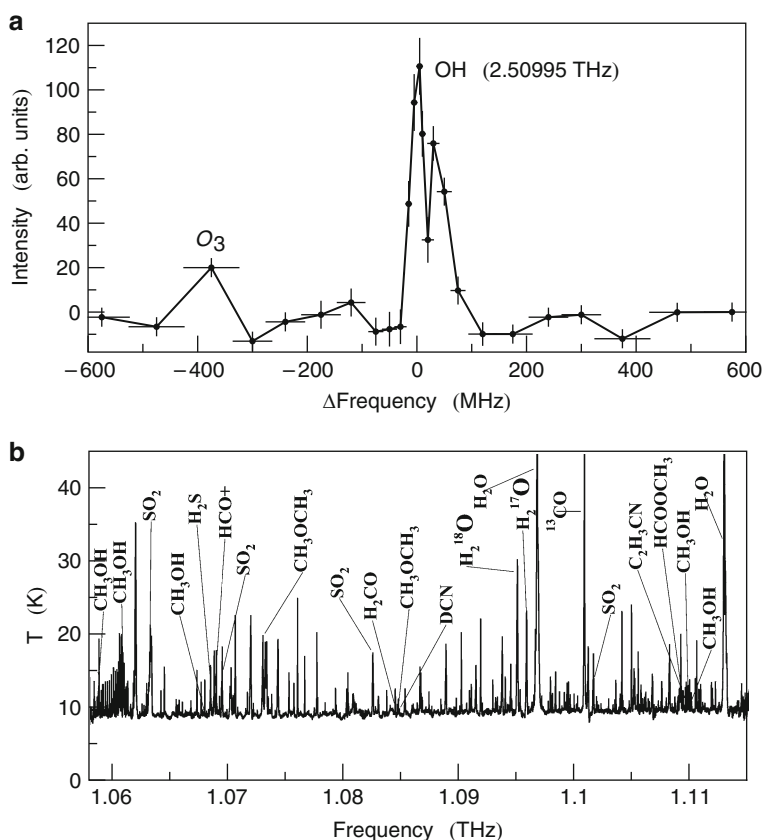


Fig. 6.29 (a) Rotational transition of atmospheric OH measured with the 2.5-THz heterodyne spectrometer on board the Aura satellite. The two fine structure components for OH are easily seen. The frequency scale corresponds to the deviation from the center frequency of the OH transition (adapted from [669]). (b) Spectral scan of Orion KL measured with the HIFI heterodyne spectrometer on board the Herschel Space Observatory. Strong lines are easily identified (adapted from [670])

astronomical requirements and the restrictions regarding power, mass, volume, etc. of a space mission [671]. HIFI covers a frequency range from 0.48 to 1.28 THz and from 1.4 to 1.9 THz. The lower portion is divided into five frequency bands, each equipped with two SIS mixers for orthogonal polarizations, which also provides some redundancy. Waveguide SIS mixers with corrugated horn antennas are used for the four lower bands up to 1.1 THz, while for band five from 1.1 to 1.28 THz quasi-optical coupling is employed. The band from 1.4 to 1.9 THz is equipped with four quasi-optical HEB mixers covering two polarizations for two subbands at 1.4–1.7 and 1.7–1.9 THz. The LOs are multiplier based. In bands 1 and 2 up to 0.8 THz, the output power is sufficient for beam splitter coupling. Above that, signal and LO radiation are coupled to the mixers with Martin–Puplett-type diplexers. Spectroscopy of planetary atmospheres or galactic objects requires a resolution of ~ 0.1 and ~ 1 MHz, respectively, while for extragalactic objects a large bandwidth of a few GHz is most important. Therefore, a CTS and a wideband array AOS are installed on board to comply with the spectral resolution and bandwidth requirements.

Another astronomical heterodyne receiver is the German Receiver for Astronomy at THz Frequencies (GREAT) [672], which is in operation on SOFIA. GREAT is a modular heterodyne receiver covering three frequency bands from 1.3 to 1.9, 2.5 to 2.7, and at 4.7 THz. It consists of a front-end with two cryostats, each carrying phonon-cooled HEB mixers, cryogenic amplifiers, and some optical elements for one frequency channel. The LOs are multiplier sources below 3 THz. At 4.7 THz, it will be a QCL. The back-end includes an AOS array spectrometer, a CTS, and a digital fast Fourier-transform spectrometer.

As in any other field, a constant improvement of the existing heterodyne technology with respect to performance, mass, power consumption, ease-of-use, and cost effectiveness is required. This is not at all trivial, as illustrated by the example of the heterodyne receivers for the Atacama Large Millimeter Array (ALMA). It is composed of 66 telescopes covering the frequency range from 84 to 950 GHz, which is subdivided into eight frequency bands. This project requires hundreds of heterodyne receivers.

Chapter 7

Terahertz Imaging

7.1 Introduction

THz imaging is a common task in radio astronomy, where the term mapping is used instead of imaging. The first astronomical THz images date back to the 1960s. THz imaging experiments for other than astronomical applications have been reported by D.H. Barker et al. as early as 1975. Already at that time the authors pointed out that “an FIR imaging system could be developed for industrial, military, law enforcement, and medical applications in the next few years” [84]. Their imaging system was based on an HCN laser operating at $337\text{ }\mu\text{m}$ and an LED-based THz detector. With this system, they obtained THz images of concealed objects which were raster scanned at the focus of the HCN laser beam. In addition, they determined the transmittance of common packaging materials, clothing, and biological samples at several frequencies between 0.3 and 0.9 THz.

Barker and his colleagues anticipated the potential of THz imaging, which is now being explored by many research groups. In the field of security, THz imaging is promising because of the ability of THz radiation to penetrate many substances, especially clothing and packaging materials. Research efforts are therefore being devoted to the development of THz imaging systems that are capable of detecting hidden threats, such as weapons or explosives, at stand-off distances up to several tens of meters. Compared with passive mmW scanners, which are already deployed for test and evaluation purposes at security checkpoints in several airports, THz imaging systems have the advantage that the entrance aperture can be much smaller than that of mmW cameras. Commercial mmW scanners operate at around 30 GHz and provide a spatial resolution of about 1 mm at a distance of less than 1 m. These systems are person-sized and can be integrated into a security checkpoint with metal and X-ray detectors. In contrast, a 0.35 THz imaging system can provide a spatial resolution of about 1.5 cm at 8 m distance with a 0.5-m diameter aperture [673] (Fig. 7.1). This allows the use of such an imager for both security checkpoints and in video surveillance-type applications. Besides the fact that THz technology is less mature than mmW technology, a further drawback is that THz radiation does not

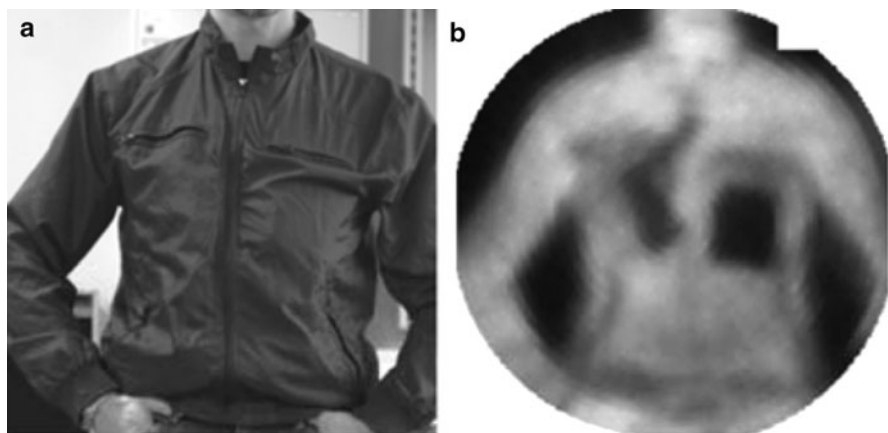


Fig. 7.1 (a) Visible image of a person with hidden objects under the jacket. (b) THz image of the same person. Two objects are clearly visible. The THz image was acquired with a passive camera equipped with a superconducting array of transition edge sensors operating at 0.35 THz. The image was obtained at 5 m distance (adapted from [674])

penetrate clothing as well as mm-waves. An increasingly important issue is the privacy implications, because both mmW and THz imaging systems are capable of exposing the naked body which, for privacy and ethical reasons, might not be acceptable to many people.

Industrial applications of THz imaging are, for example, in nondestructive testing (NDT) or process control. Again, these applications are based on the penetration capabilities of THz radiation into many plastics and composite materials. A first step toward NDT has been made by using an optically excited THz gas laser, which illuminates a high voltage cable. By measuring the scattered laser radiation voids, as small as $100\text{ }\mu\text{m}$ in diameter within the polyethylene insulation of the cable, were detected during cable production. By setting a scanning mechanism above the extrusion direction of the cable, some degree of imaging was achieved [249]. A particularly impressive example of NDT was the detection of defects in the thermal insulation of the space shuttle foam, for which THz imaging has proved to be superior compared with other established testing methods such as X-ray inspection [675]. Another example is shown in Fig. 7.2, which is a composite visible and THz image of an airbag cover [676].

In biomedicine, hopes have been raised that skin cancer and other diseases might be detected by THz imaging [677]. Comparison of THz images with histological analysis shows a coincidence between features in the THz image and cancer cells (Fig. 7.3). However, to establish this as a reliable and competitive diagnosis method requires a thorough understanding of the contrast mechanism, and how it is related to the disease. In addition, the diagnosis might require images with a much better resolution than the size of a typical cell (about several tens of μm), which calls for sub-wavelength spatial resolution. This might be achieved with near-field imaging techniques.

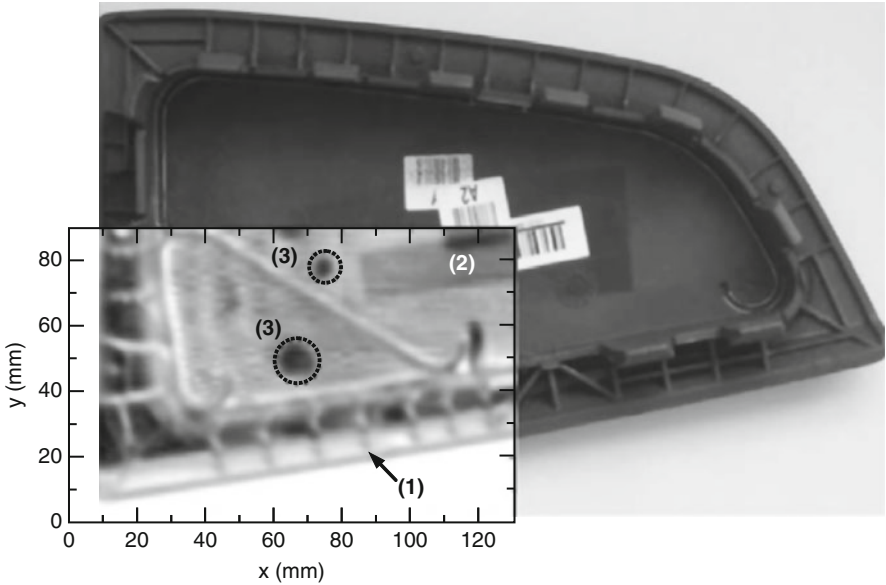


Fig. 7.2 Composite visible and THz image of an airbag cover. The THz image was measured at 0.3 THz in transmission with a cw source and detection of the transmitted amplitude. The THz image reveals reinforcing ribs (1), paper labels (2), and slightly thicker regions (3) (adapted from [676])

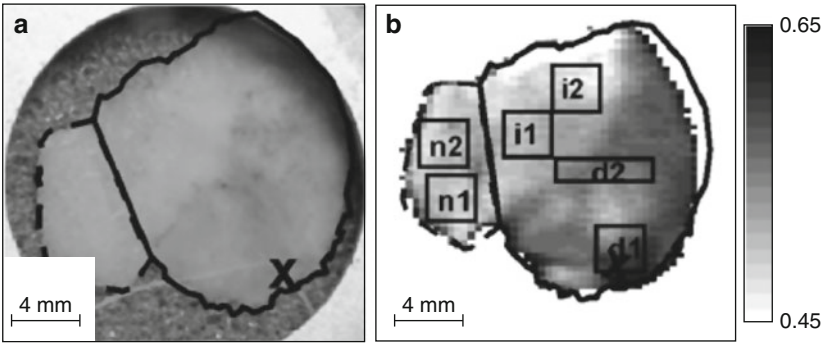


Fig. 7.3 (a) Visible image of tissue with basal cell carcinoma (skin cancer). The *solid line* marks the region of diseased tissue and the *dashed line* marks the region of normal tissue. (b) THz pulse image measured with a time-domain spectrometer in reflection mode (n1, n2: normal tissue; i1, i2: inflammation; d1, d2: diseased tissue). There are significant differences between the diseased tissue and the other tissue (adapted from [677])

From an instrumentation point of view, THz imaging systems can be roughly divided into passive and active imagers. Passive imagers are similar to IR cameras in that they detect the natural THz radiation emitted by an object, or the THz radiation from the surroundings which is reflected or scattered by the object. Active imaging

systems use some type of THz source for illumination of the object. Although this source can, in principal, be a broadband thermal source, active imagers normally employ a coherent source. Depending on the type of source, it is convenient to distinguish between cw and pulsed imaging systems. The latter are usually based on time-domain THz generation and detection. In principle, all THz detectors and sources could be used, and transmission and reflection geometries are possible for active imaging.

Another method of classification is based on spatial resolution. In a conventional imaging system, the spatial resolution is limited by the ratio of wavelength and aperture diameter. This holds good for camera-like imaging systems working at stand-off distances of several meters as well as for short-range imaging, where the sample is only a few centimeters or millimeters from the source or detector. To overcome these limitations, near-field imaging or microscopy techniques need to be used. At THz frequencies these allow sub-wavelength spatial resolution down to 1/10,000th of the wavelength. This is particularly relevant for biological applications, for example, imaging of inner parts of a cell or the inspection of semiconductor devices, which require sub- μm resolution.

One of the major challenges is the acquisition of real-time images with a good (S/N)-ratio (20 dB is usually sufficient), and an adequate spatial resolution. This is generally not the case for existing THz imaging systems, which either lack speed, contrast, or spatial resolution. In the main, practical considerations such as cost, complexity, and ease of handling limit the use of THz imaging systems to specialized applications in science. Major breakthroughs in the development of large format detector arrays and compact sources, both preferably operating at 300 K, or at cryogenic temperatures that are easily accessible with simple coolers, are necessary to achieve the full potential of the above-mentioned applications.

7.2 Some Fundamentals of Imaging

Before discussing THz imaging systems, it is helpful to review some basics relevant to image evaluation. For the purpose of this book, it is sufficient to restrict the consideration to the questions of spatial resolution and contrast.

Due to diffraction, a point source is imaged as an Airy disc with encircling rings. For a circular aperture with diameter d , the minimum of the first dark ring occurs at $1.22\lambda L/d$ where L is the distance between the aperture and the image. Note that if the aperture is a lens or mirror, and L the focal length, L/d can be replaced by the f -number $f^\#$. If two equally bright point sources are imaged, their diffraction patterns will overlap. When the overlap is such that it is possible to distinguish the two points, they are said to be resolved. There are several criteria for the definition of the limiting resolution of an optical system. The most common is the Rayleigh criterion. This states that there is a clear indication of two maxima in the combined diffraction pattern if the maximum of one diffraction pattern coincides with the first dark ring of the other diffraction pattern. In this case, the difference between the

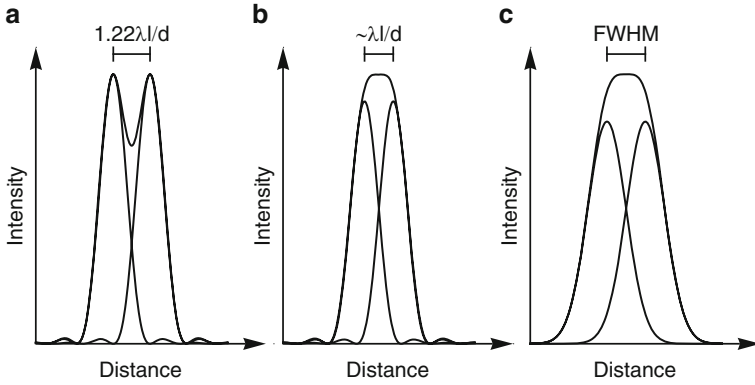


Fig. 7.4 Illustration of different criteria for spatial resolution. (a) two Airy patterns with the maximum of one pattern at the position of the first minimum of the other one (Airy criterion) (b) two Airy patterns laterally offset to achieve a flat top (Sparrow's criterion), and (c) two Gaussian beam profiles laterally offset by one FWHM

maxima and the central minimum is 26% (Fig. 7.4a), and the angular separation of the two object points is given by

$$\alpha = 1.22\lambda/d, \quad (7.1)$$

where d is the diameter of the limiting aperture. Note that for a differently shaped aperture the position of the first minimum is not the same. For example, for a slit aperture of width w it occurs at $\lambda L/w$ and (7.1) becomes $\alpha = \lambda/w$ with w the width of the slit. Somewhat less strict is Sparrow's criterion. This states that two point sources are detectable when the minimum between the two maxima just disappears (Fig. 7.4b). For a circular aperture, this is the case if $\alpha \approx \lambda/d$.

As discussed in Chap. 2, it is essential to take into account the Gaussian amplitude distribution of THz beams. If the aperture of an optical system is sufficiently large ($4.6w$ for a beam waist radius w), diffraction effects due to beam truncation are negligible, and a Gaussian beam keeps its shape while passing through an optical system. The image of a point source with Gaussian amplitude distribution is described by a Gaussian profile. In this case, two point sources are said to be resolved if the maxima of their images are separated by one FWHM. In this case, the difference between the maxima disappears (Fig. 7.4c).

A straightforward way to determine the spatial resolution is the knife-edge method. A sharp edge, for example a razor blade, is moved perpendicularly to the optical axis of the THz beam and covers it progressively (Fig. 7.5a). The intensity arriving at the detector placed behind the knife edge depends on the position of the knife edge in the following way

$$I(x) = \frac{I_0}{2} \left(1 + \operatorname{erf} \left(2\sqrt{\ln(2)} \left(\frac{x - x_c}{d} \right) \right) \right). \quad (7.2)$$

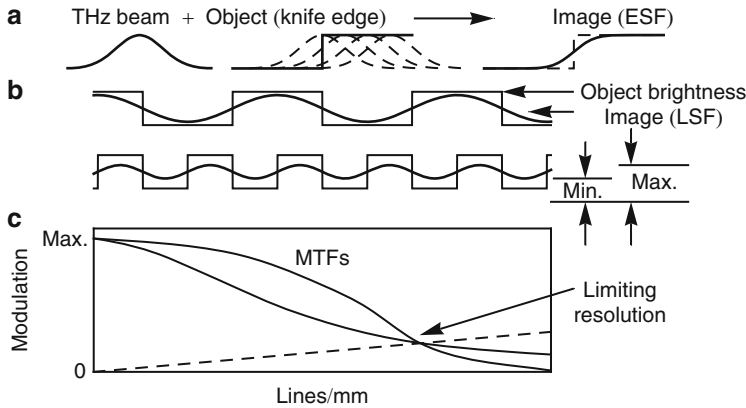


Fig. 7.5 Imagery of a bar target. (a) A knife edge is moved perpendicularly to the optical axis of a THz beam and covers it progressively. The blurred response of an imaging system is known as the edge spread function (ESF). (b) A bar target occurs as a blurred line with rounded corners. The blurred response is known as the line spread function (LSF). (c) Modulation transfer functions (MTFs) of two imaging systems (*solid lines*). The *dashed line* represents the smallest amount of modulation which the imaging system can detect. The intersection between this line and the MTFs gives the limiting resolution of the system. Although both MTFs have the same limiting resolution, the system with larger modulation at lower frequency is superior (adapted from [678])

Here, erf is the error function, I_0 is the total intensity of the beam, x is the position of the knife edge, x_c is the position of the intensity maximum of the beam, and d is the FWHM. The blurred response of an imaging system to a sharp edge is also known as the edge spread function (ESF). If, for example, the beam has a Gaussian shape, the derivation of the ESF yields the Gaussian profile of the beam. Another commonly used test to evaluate the resolution of an optical system is to image a series of alternating bright and dark bars. The limit of resolution is given by the pattern with the finest spacing that can be resolved. The imagery of a bar target is shown in Fig. 7.5b. In the image, each bar occurs as a blurred line with rounded corners. The blurred response of an imaging system to a bar is known as the line spread function (LSF). The LSF is the first derivative of the ESF. The finer the pattern the smaller is the contrast between the maximum and minimum intensity. If the noise of the imaging detector is larger than the difference between the maximum and minimum intensity, the pattern is no longer resolved. The contrast in the image is expressed as modulation M

$$M = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (7.3)$$

with $I_{\max}$ and $I_{\min}$ as the maximum and minimum intensity on the detector. In Fig. 7.5c, the modulation is plotted as a function of lines per millimeter. This is the modulation transfer function (MTF). The MTF is the Fourier-transform of the LSF. The dashed line in Fig. 7.5c represents the smallest amount of modulation, which the

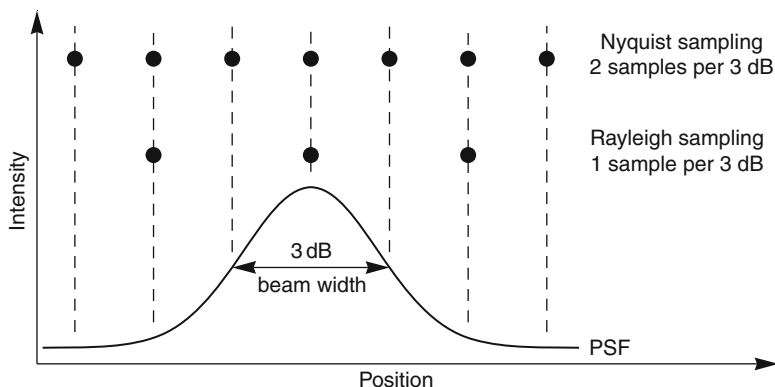


Fig. 7.6 Sampling with a beam of finite width. If two samples per 3-dB beam width are taken, it is called Nyquist sampling. If only one sample per 3-dB width is taken, it is called Rayleigh sampling. The complete spatial information of an object can only be obtained with Nyquist sampling

imaging system can detect. The intersection between this line and the MTF gives the limiting resolution of the system. For example, the limiting resolution of an aberration-free system is $\nu_0 = 1/(L f^\#)$, where ν_0 is in cycles per millimeter. Note that the shape of the MTF depends on many factors in the optical system and that the limiting resolution does not fully characterize the optical system. In Fig. 7.5c, two MTFs with the same limiting resolution are shown. Obviously, the system with an MTF with larger modulation at lower frequency is superior. An important feature of the MTF is that each component of an optical system can be described by an individual MTF, and that by multiplying them together the total MTF of the system is obtained.

Besides image blur, sampling is an important issue in achieving a well-resolved image. Nyquist sampling theory states that a signal must be sampled at a rate twice that of its frequency. This means that in order to record all spatial information the scene has to be sampled at a rate of two sampling points for an FWHM of the beamwidth of the optical system (Nyquist sampling). This is illustrated in Fig. 7.6. Sampling at half the frequency (Rayleigh sampling) yields an image which contains significantly less information, while sampling at twice the Nyquist rate does not yield significantly more information.

7.3 Camera-Like Imaging

In this section, imaging systems are described which are similar to cameras widely used in the IR and visible regions. Their characteristic features are that the object which is imaged is at some distance from the camera, and that only radiation which is reflected or emitted from the object is detected. If this is only natural background

radiation, it is called passive imaging. In analogy with flash lights used for illumination in photography, the object can be illuminated with radiation from some type of THz source. This is called active imaging. Camera-like imaging is different from the imaging approaches described in Sect. 7.4 where the object is usually raster-scanned in the focus of a THz beam and the transmitted radiation is detected.

7.3.1 Contrast Formation

Before going into the details of some camera-like THz imaging systems, it is worth considering briefly how the contrast in a THz image arises. For a small solid angle Ω and a small frequency interval $\Delta\nu$, the power that reaches a detector with an area A is

$$P_B = \varepsilon(\nu) A \Omega \frac{2h\nu^3 \Delta\nu}{c^2} \left(\exp\left(\frac{h\nu}{k_B T}\right) - 1 \right)^{-1} \approx \frac{2A\Omega \Delta\nu k_B T}{\lambda^2}. \quad (7.4)$$

On the right-hand side, the Rayleigh–Jeans approximation with the emissivity $\varepsilon(\nu) = 1$ is used. For simplicity, these assumptions are used in the following discussion. As in IR imagery, it is convenient to quote the sensitivity of an imaging system as the minimum temperature change that can be detected in a given integration time. In analogy with the NEP, this is called the noise equivalent delta temperature (NEDT). It is given by

$$\text{NEDT} = \frac{\text{NEP}}{dP_B/dT}. \quad (7.5)$$

From this one obtains for a direct detector

$$\text{NEDT}_{\text{Dir}} = \text{NEP}_{\text{Dir}} \frac{\lambda^2}{A\Omega} \frac{1}{2k_B \Delta\nu} \approx \frac{\text{NEP}_{\text{Dir}}}{2k_B \Delta\nu}. \quad (7.6)$$

The approximation on the right-hand side holds for a diffraction-limited detector when only the fundamental mode is detected, and the throughput is $A\Omega = \lambda^2$. Note that the factor 2 in the denominator arises from the two polarizations of each mode. In the case of a heterodyne receiver, the input bandwidth $\Delta\nu$ is equal to the intermediate frequency (IF) bandwidth $\Delta\nu_{\text{IF}}$ for an SSB receiver, and it is $2\Delta\nu_{\text{IF}}$ for a DSB receiver. In contrast to a direct detector, a heterodyne receiver detects a single mode and only one polarization. Taking this into account, for a DSB receiver one obtains

$$\text{NEDT}_{\text{Het}} = \frac{\text{NEP}_{\text{Het}}}{2k_B \Delta\nu_{\text{IF}}}, \quad (7.7)$$

which has a similar form to that for a diffraction-limited direct detector. Comparison of both equations shows that the difference in NEDT between heterodyne and direct detection reduces to $\text{NEP}_{\text{Het}}/\text{NEP}_{\text{Dir}} = \Delta\nu/\Delta\nu_{\text{IF}}$. Therefore, direct detection is advantageous for broadband signals for which $\Delta\nu \gg \Delta\nu_{\text{IF}}$, while heterodyne

detection is advantageous when the signal bandwidth is in the order of $\Delta\nu_{\text{IF}}$. This is because there are no bandpass filters available that are as narrow as the IF bandwidth of a heterodyne receiver, and the direct detector receives more background noise. It should be noted that the considerations above are for an output bandwidth of 1 Hz. For a different bandwidth, (7.6) and (7.7) need to be multiplied by the square root of the output bandwidth B_{out} . The NEP of a heterodyne receiver can be expressed as $\text{NEP}_{\text{HET}} = 2k_B T_{\text{REC}} \sqrt{\Delta\nu_{\text{IF}}}$ [679]. Using this and $B_{\text{out}} = 1/\tau$, (7.7) takes the well-known form of the radiometer equation (5.63) with $k_{\text{REC}} = 1$

$$\text{NEDT}_{\text{Het}} = \Delta T = \frac{T_{\text{REC}}}{\sqrt{\tau \Delta\nu_{\text{IF}}}}. \quad (7.8)$$

A simple scheme of an imaging situation is depicted in Fig. 7.7: a person with an object under their clothing is imaged by some type of THz camera. This is a typical scenario for security applications. The person, the object, the clothing, the atmosphere, and the surroundings are described by their temperature T , emissivity ϵ , transmission t , and reflectivity r [679]. Using the Rayleigh–Jeans approximation, the signal temperature at the camera received from the body is

$$T_{\text{REC,B}} = t_C(r_B \epsilon_C T_C + t_C r_B T_{\text{BG}} + \epsilon_B T_B) + \epsilon_C T_C + r_C T_{\text{BG}}, \quad (7.9)$$

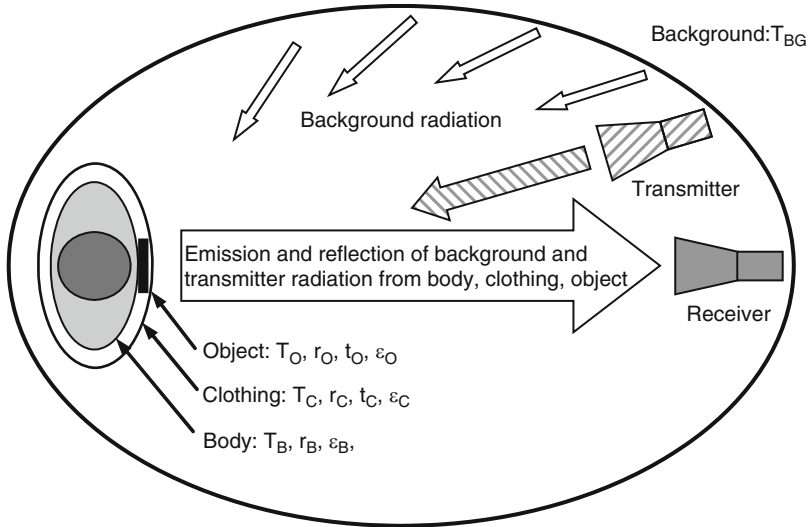


Fig. 7.7 Imaging scenario: a person carries a hidden object under the clothing. In the case of passive imaging, the receiver detects the emitted radiation from the body, the object, and the clothing as well as the reflected radiation from the surrounding background. In the case of active imaging, the person is additionally illuminated by THz radiation from a transmitter (hatched symbol). The symbols refer to the temperature (T), reflection (r), transmission (t), and emissivity (ϵ) of the object (O), clothing (C), and body (B)

Table 7.1 Typical values of emissivity ε , reflectivity r , and transmission t of skin, jeans (Denim) and metal (adapted from [680])

Frequency	0.5 THz	1 THz	0.5 THz	1 THz	0.5 THz	1 THz
	ε	ε	r	r	t	t
Skin	0.91	0.93	0.09	0.07	0	0
Denim	0.49	0.85	0.01	0.05	0.5	0.1
Metal	0	0	1	1	0	0

and the signal temperature received from the covered object is

$$T_{\text{REC,O}} = t_{\text{C}}(r_{\text{O}}\varepsilon_{\text{C}}T_{\text{C}} + t_{\text{C}}r_{\text{O}}T_{\text{BG}} + \varepsilon_{\text{O}}T_{\text{O}}) + \varepsilon_{\text{C}}T_{\text{C}} + r_{\text{C}}T_{\text{BG}}. \quad (7.10)$$

Here, the subscript C refers to clothing, B refers to body, and BG refers to background. The temperature contrast between the two signals is $\Delta T = T_{\text{REC,B}} - T_{\text{REC,O}}$. With the values listed in Table 7.1, and assuming $T_{\text{BG}} = 300$ K, $T_{\text{B}} = 310$ K, $T_{\text{C}} = 300$ K, this yields for a metal object covered by denim $\Delta T = 5.9$ K at 0.5 THz and $\Delta T = 2.3$ K at 1 THz.

It should be kept in mind that this is a rather simplified scenario. For example, atmospheric absorption or scattering by skin, clothing, and the atmosphere are neglected. These effects worsen the temperature contrast. Suppose a contrast of 1 K needs to be imaged. In order to obtain a good image, an (S/N) -ratio of 20 dB is required. Therefore, the NEDT should be about 0.01 K for an integration time of 1 s. If the THz camera has a 0.5-m diameter aperture which creates a 1 cm spot size at the object (corresponding to an $A\Omega$ of 2×10^7) and the input bandwidth is 100 GHz at 0.35 THz, this corresponds to an NEP of 5×10^{-13} W/ $\sqrt{\text{Hz}}$. This can only be achieved with state-of-the-art liquid helium-cooled detectors. For heterodyne detection with 50 GHz IF bandwidth, it corresponds to $T_{\text{sys}} = 2200$ K ($k_{\text{REC}} = 1$, integration time = 1 s), which can be achieved with uncooled Schottky diode mixers.

However, many applications require real-time or video-rate imaging. If the THz camera has only one or a few detecting elements, the requirement regarding detector NEP increases dramatically. For example, 10^5 pixels with a 10 Hz frame rate require a detector response time $< 1 \mu\text{s}$ or an NEP $\approx 5 \times 10^{-16}$ W/ $\sqrt{\text{Hz}}$. This can only be achieved by a very few direct detectors such as a superconducting transition edge sensor (TES). A similar consideration leads to a T_{sys} of about 2 K for heterodyne detection, which is clearly not feasible. Here, the solution is additional illumination of the scene with a narrow-band source to take full advantage of the narrow-band heterodyne detection.

The contrast obtained with a THz camera in astronomy, often referred to as a photometer or imaging spectrometer, is determined by the signal flux from the object under investigation, relative to a known calibration source. A convenient calibration source is the sky at a position where no signal except the 2.7 K background radiation is observed. In addition, a dedicated cooled blackbody calibration source is part of the photometer. In order to prevent masking of the astronomical signals by background radiation, the optics need to have a low emissivity and low temperature. In practice, this means that only mirrors should be used and the optics

should perhaps be cooled. As an example, typical photon fluxes which should be detectable by a state-of-the art THz photometer are a few mJy. A “Jansky” is not an SI unit but is widely used in astronomy. It is given as $1 \text{ Jy} = 10^{-26} \text{ W}/(\text{Hz m}^2)$ within one hour integration time, and for a (S/N) -ratio five times larger than the rms noise.

7.3.2 *Imaging with Direct Detectors*

Camera-like imaging with direct detectors is in some sense the extension of IR imaging towards longer wavelengths. While for IR cameras large format (for example 260×340 pixel) detector arrays with good sensitivity for imaging are available, this is not the case at THz frequencies. IR detector arrays such as VO_2 or Si microbolometers do not require cooling, although for ultimate performance moderately cooled (to about 250 K) InSb arrays are used. At THz frequencies, the situation is different. Uncooled detector arrays based, for example, on microbolometers or pyroelectric detectors, do not provide an NEP which is sufficient for passive imaging. These arrays are used in conjunction with powerful THz sources for image acquisition and such arrangements will be discussed in Sect. 7.4. To achieve good performance for imaging, THz detectors need to be cooled. If bolometers are to be used for detection, this immediately requires an improvement of 100 in the NEP, which can only be achieved by cooling. Besides which, the most sensitive detectors are based on superconductors or photoconductors and these, in any case, require cooling. This cooling requirement makes the development of passive THz imaging systems with direct detectors quite a challenging task. Another challenge is the fabrication of the array itself. If the array is larger than a few tens of pixels, individual readout and cabling of each pixel is an almost impossible task, because of the number of cables and the cryogenic requirements.

For astronomical applications, a wide variety of THz cameras have been developed in the past years or are currently under development. These are too numerous to be discussed here and the reader is referred to the SPIE proceedings series “Millimeter and Submillimeter Detectors and Instrumentation for Astronomy”. These give a nearly complete overview of ongoing developments of THz cameras for astronomy. Only a few examples of the various systems and their main features will be described here.

The Herschel Space Observatory, launched in May 2009, is equipped with three instruments with imaging capabilities. PACS, the Photoconductor Array Camera and Spectrometer, and SPIRE, the Spectral and Photometric Imaging Receiver, are both equipped with arrays of direct detectors. PACS was designed and built by a consortium of partners across Europe led by the Max-Planck Institute for Extraterrestrial Physics. It operates either as an imaging photometer (camera) or as an integral field spectrometer over the spectral band from 57 to $210 \mu\text{m}$. The imaging photometer allows simultaneous observation (with the same field-of-view) at 60–85 μm , or at 85–130 μm , and at 130–210 μm . It has two bolometer arrays,

one with 64×32 pixels for bands $<130 \mu\text{m}$ and one with 32×16 pixels for the $130\text{--}210 \mu\text{m}$ band. Each of these arrays is made from 16×16 sub-arrays, which are cooled to 0.3 K . The NEP is $\sim 10^{-16} \text{ W}/\sqrt{\text{Hz}}$. The selection between the two short-wavelength bands is achieved with filters. The integral field spectrometer is essentially a grating spectrometer. The focal plane has an unstressed and a stressed Ge:Ga detector array, with 16×25 pixels for $57\text{--}105 \mu\text{m}$ and $105\text{--}210 \mu\text{m}$, respectively. In order to perform imaging and spectroscopy at the same time, each array provides 5×5 spatial pixels of the sky, with each of them providing 16 spectral channels. In order to comply with the optics, governed by the entrance slit of the grating spectrometer, an image slicer rearranges the 5×5 pixels into a 1×25 line to match the entrance slit. Imaging is achieved by scanning the telescope across the sky. Each telescope position provides 25 image points. The instantaneous spectral coverage is insufficient for spectroscopy and, in order to increase the spectral coverage, the grating needs to be rotated. An example of the capability of the PACS photometer is the “Whirlpool Galaxy” M51 shown in Fig. 7.8. This is one of the first images taken with PACS after the launch of Herschel.

SPIRE is an imaging photometer and an imaging Fourier-transform spectrometer. It has been built by an international consortium led by Cardiff University in the UK. The camera operates in three wavelength bands centered on 250 , 360 , and $520 \mu\text{m}$. The bands are defined by mesh-filters giving a resolution of $\lambda/\Delta\lambda \approx 3$. SPIRE can make images of the sky simultaneously in the three bands, with a field of view of $4 \times 8 \text{ arcmin}$. The detectors are arrays of spider-web bolometers using neutron transmutation doped Ge thermistors cooled to 0.3 K . Each band has one array containing 43 ($520 \mu\text{m}$), 88 ($360 \mu\text{m}$), and 139 ($250 \mu\text{m}$) detectors. The radiation is coupled to each pixel via single-mode conical feedhorns giving diffraction-limited beams. The background-limited NEP is $10^{-15}\text{--}10^{-16} \text{ W}/\sqrt{\text{Hz}}$. The imaging Fourier spectrometer covers the range $200\text{--}670 \mu\text{m}$ with an approximately 2.6 arcmin diameter circular field of view. It employs a Mach–Zehnder-type dual-beam configuration, with broadband beam splitters to provide high efficiency, and separated output and input ports. It has two output ports with overlapping bands of

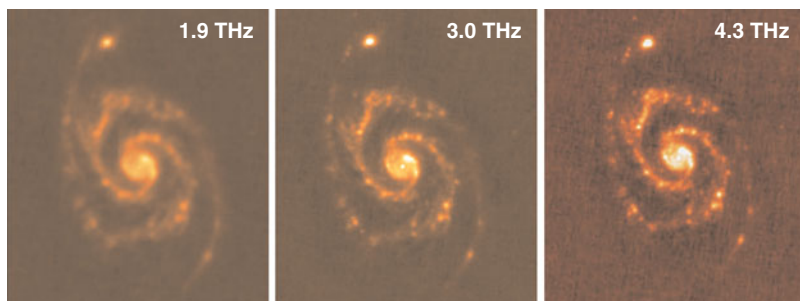


Fig. 7.8 THz images of M51 (“Whirlpool galaxy”) obtained with the PACS instrument on board the European Space Agency’s Herschel Space Observatory (copyright: ESA and the PACS consortium)

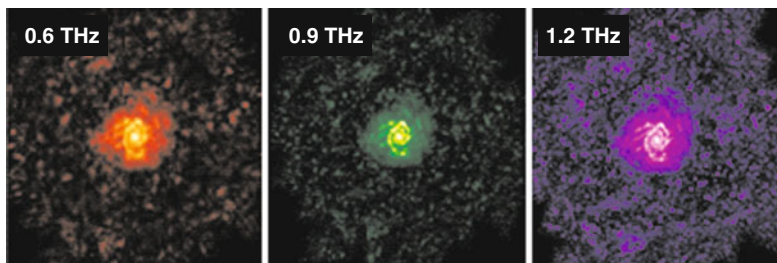


Fig. 7.9 THz images of M74 obtained with the SPIRE instrument on board the European Space Agency's Herschel Space Observatory (copyright: ESA and the SPIRE consortium)

200–325 and 315–670 μm . Each is equipped with a detector array similar to that of the photometer (19 detectors for the long wavelength band and 37 detectors for the short wavelength band). The spectral resolution can be adjusted between 0.04 and 2 cm^{-1} . Impressive images taken with the SPIRE photometer are shown in Fig. 7.9.

Probably the most complex ground-based THz camera ever built is SCUBA-2, the Submillimeter Common User Bolometer Array at the James Clerk Maxwell telescope in Hawaii. It is capable of imaging at 450 and 850 μm simultaneously. The two wavelengths are separated by a dichroic beam splitter. For each wavelength, one optimized focal plane array is used, which is made from four sub-arrays with 1280 TES each in a 32×40 format. The TES are made from Mo-Cu bi-layers. They are voltage biased and operate at 100 mK. The NEP is $\sim 10^{-16} \text{ W}/\sqrt{\text{Hz}}$. The read-out is achieved with SQUIDS operated in a time division multiplexing scheme. The TES array and the SQUIDS are on two separate silicon wafers, which are bump-bonded together [681]. The arrays are cooled to 100 mK by a dilution refrigerator. In order to achieve background-limited performance with the TES, the last mirrors in front of the arrays need to be cooled to below 10 K. In addition, the large field-of-view (50 arcmin^2) requires rather large optics (up to 1.2 m across). This results in a very large cryostat of $2.3 \times 1.7 \times 2.1 \text{ m}^3$ (height $\times$ width $\times$ length) cooled by a powerful pulse tube cooler.

Somewhat similar to SCUBA is SABOCA, the Submillimeter APEX Bolometer Camera for the APEX (Atacama Pathfinder Experiment) telescope in the Atacama desert in Chile. SABOCA is designed to function in the atmospheric window at 0.8 THz. Its detectors are composite bolometers with superconducting TES thermistors made from Mo-Au/Pd bilayers on an SiN membrane. The operating temperature of the detectors is 300 mK, and the NEP is $\sim 10^{-16} \text{ W}/\sqrt{\text{Hz}}$, with a background at 300 K. The detector array consists of 37 TES in a hexagonal layout. It is read out in the time domain with four SQUID amplifiers/multiplexer chains providing 10 channels each. A monolithic array of conical horn antennas is placed in front of the TES wafer. The bandpass is defined by filters. The array and the read-out are operated in a liquid helium cryostat with a closed-cycle helium-3 sorption cooler. What makes the SABOCA detector array particularly noteworthy is that a similar array is used in a spin-off application for THz security imaging. This THz

camera utilizes similar TES detectors, but the operational frequency is 0.35 THz with a bandwidth of 40 GHz because, at this frequency, the transmission of clothing and the atmosphere is better than in the SABOCA band at 0.85 THz. The frequency adaptation is done by exchanging the filters and the conical horns, optimized for 0.35 THz. A prototype of such a THz security camera has been built [673, 674]. The detector array is integrated in a He^3/He^4 sorption cooler, which in turn is mounted on the cold plate of a pulse tube cooler. A Cassegrain-type telescope (diameter: 0.5 m) with a scanning secondary mirror yields an on-axis spatial resolution of better than 2 cm at 8 m distance. The frame rate is 10 Hz and the temperature resolution is below 0.1 K. An image obtained with this camera is shown in Fig. 7.1.

A similar approach for passive THz imaging for security applications is PEAT-CAM (passive European-American THz camera). This camera utilizes NbN or Nb superconducting bolometers. They are embedded in a planar logarithmic-spiral antenna, which limits the signal bandwidth to 0.2–1 THz. Optical coupling is achieved with a 2-mm diameter hemispherical Si lens-antenna. These bolometers are somewhat similar to the HEB mixers described in Sect. 5.5.4. However, in order to reduce the response time and to increase the responsivity for direct detection, the superconducting bridges (8 μm long, 2 μm wide, 350 nm thick) are suspended in vacuum. The detector is operated as a voltage-biased TES with an NEP of $\sim 10^{-14} \text{ W}/\sqrt{\text{Hz}}$. Readout is done with a JFET amplifier at room temperature. Images with a single detector mounted in a liquid helium cryostat show encouraging results, with an NEDT of 125 mK within a 30 ms integration time. The camera employs a 68 element linear array of these detectors and allows for near video-rate imaging [682].

7.3.3 Heterodyne Imaging

Heterodyne imaging has attracted significant interest because of applications in the field of security, namely the stand-off detection of concealed objects, for example a weapon carried by a person under their clothing. Most of these imaging systems are active systems with coherent illumination. The principal scheme is similar to the one shown in Fig. 7.7, with the addition of a radiation source, the transmitter, which acts as a source for illumination of the object. Radiation, which is scattered or reflected from the object, is detected by a heterodyne receiver. Using a narrow-band source allows full advantage of the restricted detection bandwidth to be taken, thus achieving the high (S/N)-ratio, which is attainable with a heterodyne receiver, even at 300 K. However, coherent illumination tends to produce speckles, or very strong specular reflection signals in the image. An important part of such an imaging system is therefore the optics and the way that the object is illuminated. Either the optical paths of the transmitter and the receiver are collinear or they are not. In the latter case, the transmitter may illuminate the whole object rather than a single point, as with the collinear alignment. Since large heterodyne arrays are not yet available, the field-of-view of the single pixel of the heterodyne receiver needs to be scanned

across the object. Raster-scanning and conical-scanning schemes are employed, with the latter being faster. Typical frequencies for THz security imagers are in the atmospheric windows at 0.35, 0.6, and 0.8 THz. The size of the optics is determined by the stand-off distance and operating frequency. With a 0.7-m diameter mirror, and at a frequency of 0.8 THz, imaging with a spatial resolution of 2 cm at a distance of 20 m has been demonstrated [683]. This imager used a Gregorian-type telescope. An image taken with a 0.64 THz active imager based on a multiplier source and a Schottky diode mixer is shown in Fig. 7.10 [684]. The object was placed 3.2 m from a scanning mirror. The image consists of 216×181 pixels.

The first work on heterodyne arrays at frequencies above 0.3 THz dates back to the early 1980s. This development was driven by applications in plasma diagnostics. These arrays were linear Schottky diode mixers on the rear side of a single lens [204, 685]. About 10 years later, the development of SIS mixer arrays for applications in radioastronomy began. These became operational in the late 1990s and operate in the two atmospheric windows around 0.35 and 0.8 THz. All major ground-based submm/THz telescopes are now equipped with heterodyne receiver arrays. Their design is complex although, in essence, it is a rather straightforward extension of single pixel receivers. This is because each mixer is fabricated on a separate chip and the chips are assembled to individual pixels, each with its own optical coupling (either waveguides with horn antennas, or quasi-optical coupling with a lens-antenna). The individual mixers are assembled in a common mixer block. Basically, this is an extension of single pixel receivers but with special, compact mixer blocks in order to keep the array dimensions small. However, the fill-factor is not optimal. The IF readout is accomplished using low-noise amplifiers. Each mixer/pixel is equipped with one IF amplification chain. No multiplexing of the IF signal is done. The distribution of the LO power is achieved, either by a series of beam splitters, or by phase gratings (Fourier or Dammann type) [686]. The grating approach is more efficient and compact. If the optical coupling to the mixer is carried out with horn antennas, the size of the horn antennas ultimately limits the packaging density of the mixers. Micromachined silicon horn antennas may open a way for

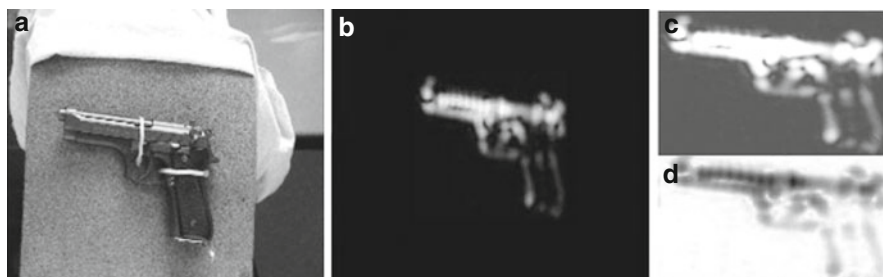


Fig. 7.10 (a) Visible image of a pistol, (b) THz image of the same pistol underneath a cotton T-shirt taken at a distance of 3.2 m. The image was acquired with 0.64 THz illumination from a multiplier-based source and heterodyne detection of the returned radiation. The intensity scale is linear. In (c), the same image is shown with a logarithmic intensity scale and (d) is the same as (c) but with an inverted intensity scale (adapted from [684])

compact heterodyne mixer arrays at THz frequencies [687]. Similarly, the size of the lens limits the packaging, if individual lens-antennas are used for radiation coupling, a so-called fly's-eye configuration [207]. However, with lens-antenna systems this can be overcome to some extent by using a single lens with several mixers on the rear side. The price to pay is that only one mixer is on the optical axis. All other mixers are off-axis, which results in distorted beam profiles, i.e. less Gaussian-shaped beams, with side lobes on one side of the antenna pattern that are higher than those on the other side. Furthermore, the reflection loss at the lens-air interface increases with increasing off-axis position [688, 689]. Which one of the two approaches is appropriate for a particular application depends on whether efficient coupling to each mixer, or efficient filling of the focal plane has the higher priority.

7.3.4 Three-Dimensional Heterodyne Imaging

Depth information can be obtained by combining a heterodyne imaging system with frequency modulated continuous wave (FMCW) radar. An FMCW radar is based on a transmitter which emits at a frequency ν_{TX} . Its frequency is modulated, for example with a sawtooth or triangular function (Fig. 7.11). The linear frequency change is $d\nu_{TX}/dt$. The return signal has a time delay $\Delta t = 2d/c$, where d is the distance to the object. A heterodyne receiver which operates at a frequency ν_{TX} detects an IF signal at $\nu_{IF} = (d\nu_{TX}/dt)\Delta t$. The distance to the object is proportional to the IF according to

$$d = \frac{c\nu_{IF}}{2d\nu_{TX}/dt}. \quad (7.11)$$

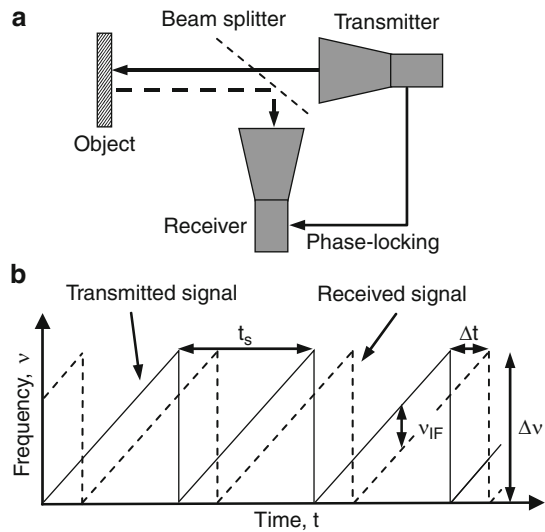


Fig. 7.11 (a) Simplified scheme of an FMCW radar and (b) the operating principle

The minimum range resolution r_0 , which is the minimum depth distance between two objects that can be resolved, is given by

$$r_0 = \frac{c}{2\Delta\nu}, \quad (7.12)$$

where $\Delta\nu$ is the chirp bandwidth, i.e. the maximum frequency change of the transmitter. For a 10 GHz chirp bandwidth, the minimum range resolution is 1.5 cm. While for a conventional mmW FMCW radar this represents quite a large bandwidth, it is only 2% bandwidth at 0.5 THz. In fact, a 10–20% bandwidth is readily achievable for a THz heterodyne receiver. Therefore, FMCW radar has the potential for small range resolution.

In Fig. 7.11 a simplified block diagram of a THz FMCW radar is shown. It consists of a transmitter and a heterodyne receiver. The transmitter and the LO are phase-locked. The received signal is downconverted by the receiver, and Fourier analysis of the IF signal yields the power spectrum as a function of frequency, which can then be directly transformed into a distance according to (7.11). Scanning the beam of such a single pixel FMCW radar enables imaging in three dimensions. With a high-resolution imaging radar operating around 0.6 THz weapons concealed by clothing were detected at ranges of 4 to 25 m. A range resolution of about 1 cm at 4 m distance has been demonstrated (Fig. 7.12) [690].

7.4 Short-Range Imaging

The principal scheme of a typical setup for a short-range imaging experiment is shown in Fig. 7.13. The term “short-range” is understood as an imaging modality, where the components of the imaging system are relatively close to the object i.e. typically below one meter. Usually, in this configuration, a source illuminates the object and the transmitted or reflected radiation is detected. An appropriate optical system focuses the radiation from the source onto the object, which is step-scanned in a plane orthogonal to the optical axis. The spatial resolution of such a system is defined by the optics and the step size of the scanning motion. Ultimately, it is limited by the wavelength of the source, and diffraction introduced by the optical elements, unless near-field techniques are employed (Sect. 7.5). In principle any detector and any source described in Chaps. 4 and 5 could be used for such an imaging setup. In practice the performance of a newly developed source or detector is often demonstrated by an imaging experiment. However, it transpires that certain detector-source combinations have found more widespread use than others. In general, many imaging techniques explored, and frequently used in the visible range, have also been demonstrated in a short-range configuration at THz frequencies. Examples are tomography [691, 692], synthetic aperture imaging [693], reciprocal imaging [694], or compressed sensing [695]. In

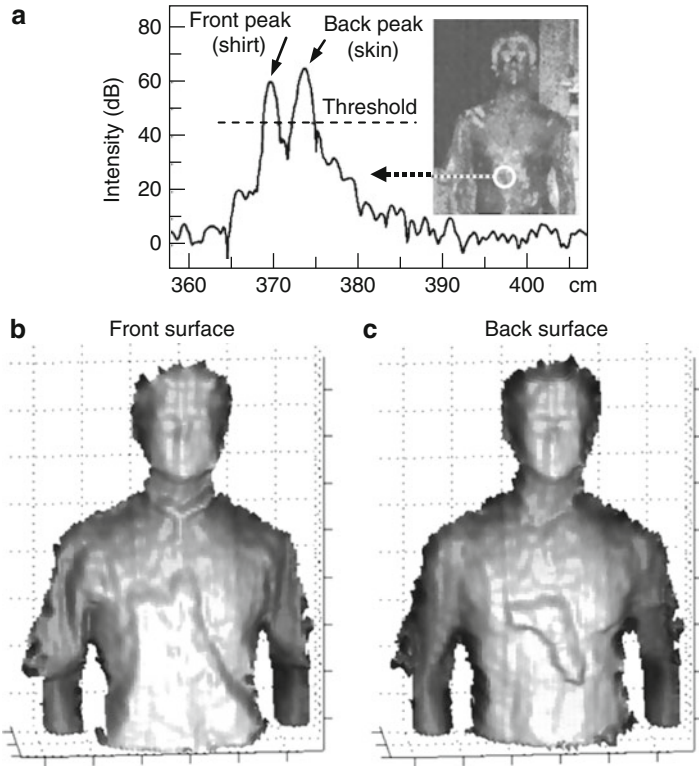


Fig. 7.12 (a) FMCW power spectrum of an individual pixel. The pixel location is indicated by a circle in the inset total-power image. The two peaks exceeding the threshold come from the subject's shirt and body surface. (b) and (c) Three-dimensional reconstructions of the front and back surface ranges consisting of the identified front and back peaks for each pixel. The concealed handgun is visible as a bulge in the back surface (adapted from [690])

this section, a few of the more widely employed imaging setups and techniques will be described.

7.4.1 Imaging with cw and Long-Pulse Sources

7.4.1.1 Imaging with Backward Wave Oscillators

This imaging system relies on a backward wave oscillator (BWO), which has the particular advantage of being a high output power source. In addition, its narrow bandwidth and frequency tunability are helpful for imaging. Such a system has been described by Dobroiu et al. [696]. The output from a BWO operating in the 0.52 THz

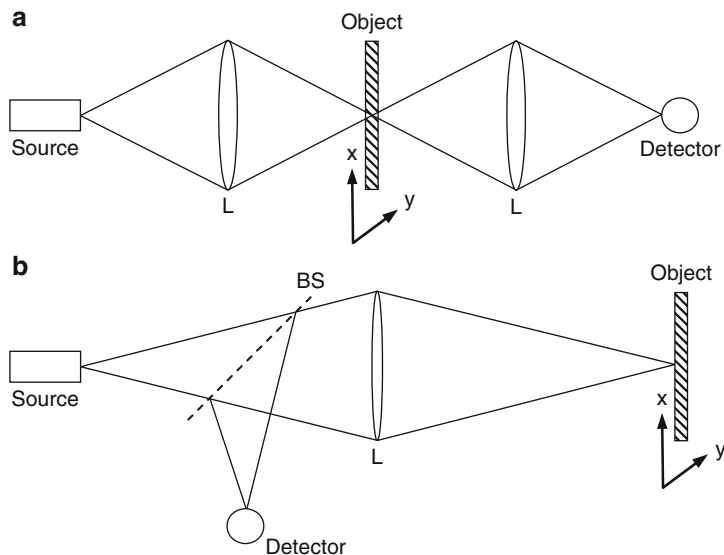


Fig. 7.13 Simple arrangements for (a) transmission and (b) reflection imaging in a short-range configuration (L: lens, BS: beam splitter)

to 0.71 THz region is focused onto the sample, which is then raster-scanned. The transmitted radiation is collected by a pair of off-axis paraboloidal mirrors and detected with a pyroelectric element. At 0.59 THz, an FWHM of $550\text{ }\mu\text{m}$ was obtained, which is 15% larger than the theoretical diffraction limit. A particular problem that occurs, when imaging with a coherent source, is standing waves, due to partial reflection at optical components of the setup which are perpendicular to the optical beam. These manifest themselves as interference fringes in the image. With the BWO system, these can be reduced efficiently by modulating the BWO frequency. The imaging speed is limited by the sensitivity of the pyroelectric detector, which has an output noise power that is larger than the power instabilities of the BWO. When imaging a 1 mm diameter hole in a thin piece of aluminum foil, a (S/N) -ratio of 27 dB was achieved with a DLATGS pyroelectric detector. Replacing this detector by a superconducting tunnel junction detector improved the (S/N) -ratio to 46 and 64 dB, when the detector was cooled to 4.2 and 0.3 K, respectively [697].

7.4.2 Photomixer-Based Imaging

In 2002 a cw photomixer imaging system was demonstrated by Siebert et al. [698]. It consists of a two-color Ti:Sa laser with independently tunable wavelengths centered

at 800 nm. The emitter, as well as the detector, is an LTG-GaAs photomixer with double dipole antennas mounted on the rear side of a hyperhemispherical Si lens. Detection is achieved by homodyne mixing of the transmitted radiation with the difference signal created by the two-color laser in the detector. The detected signal V_{out} varies with the time delay t' between the THz and the IR beams according to the autocorrelation function

$$V_{\text{out}}(t') \propto \int_{-\infty}^{+\infty} \cos(\omega t) \cos(\omega(t + t') + \phi) dt. \quad (7.13)$$

With a delay line integrated into the optical path between the laser and the detector, the sinusoidal variation of the THz signal is registered, and the amplitude and phase ϕ can be recovered for each pixel. The spatial resolution is then determined by the knife-edge method. At 1.1 THz, it was $\sim 320 \mu\text{m}$, which is about 18% larger than one wavelength. A comparison of different imaging modalities with such a system is shown in Fig. 7.14.

Another imaging system with cw photomixers has been developed by Gregory et al. [699]. This system works in reflection. The generation of THz radiation is based on two free-running diode lasers operating at 830 nm. The frequency changes introduced by the drift and the linewidths of the diode lasers are small compared to the target difference frequency of 0.5 THz. Detection is achieved by homodyne mixing and lock-in detection of the reflected radiation with the mixing signal of the two diode lasers in the photomixer detector, in a similar way to the system described in [698]. For a 30 ms time constant, per pixel a (S/N) -ratio of ~ 60 dB was obtained.

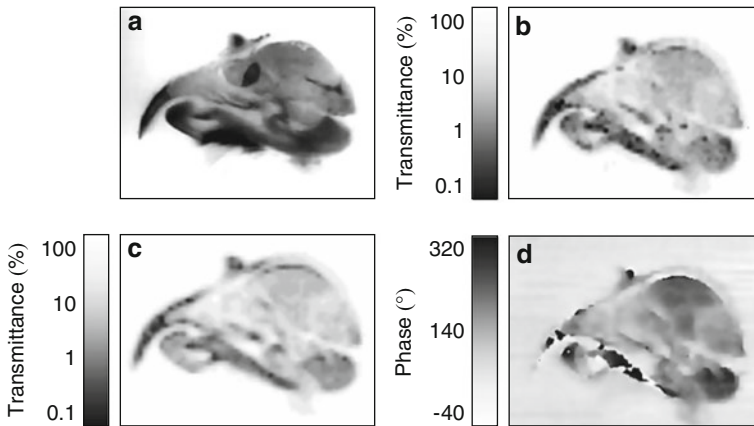


Fig. 7.14 Images of a wax-mounted thin-slice of tissue from a canary's head. (a) photograph, (b) power transmission image measured with a cw photomixer system at approximately 1 THz, (c) pulsed THz image obtained from the pulse amplitude at 1 THz, and (d) cw THz image obtained from the phase information at 1 THz (adapted from [698])

7.4.3 *Imaging with Quantum Cascade Lasers*

While BWOs and photomixer-based systems have been used for imaging below about 1 THz, imaging systems based on quantum cascade lasers (QCLs) are employed at higher frequencies, typically above 2 THz. Analogous to BWOs they deliver several mW, or even several tens of mW, and this compensates for absorption losses, which are generally higher in the upper THz range. Shortly after the invention of the THz-QCL, imaging experiments were performed [700]. This system consisted of a 3.43 THz QCL which was operated at 5 K and ran in a pulsed mode at a repetition frequency of 400 Hz, with a pulse width of 100 ns. The peak power was about 20 mW. The QCL emission was focused onto the sample, a section of a rat brain, which was raster-scanned at the focus. The reflected radiation was focused onto a liquid helium-cooled Si bolometer. With this system, images with a spatial resolution better than $250\text{ }\mu\text{m}$ ($\approx 3\lambda$), and a dynamic range of about 1,000, were obtained [700]. A similar system, with a 2.9-THz QCL operated in cw at 10 K, used a room temperature Schottky diode mixer for detection of the reflected radiation. The Schottky diode detected the beating of two longitudinal modes of the QCL separated by 13 GHz. A (S/N) -ratio of ~ 60 dB was obtained with an acquisition time of 100 ms/pixel, limited only by the acquisition software. In principle, this might be reduced by several orders of magnitude [701].

7.4.4 *Real-Time Imaging with cw Sources*

The drawback of the imaging methods described so far is that they rely on raster scanning of the object under test. This results in relatively long image acquisition times, because just the measurement of a single pixel requires several tens of ms. Pyroelectric detectors (Sect. 5.3.2), as well as microbolometers of the type described in Sect. 5.3.7, are available in large format arrays. If the imaging source provides enough power at a frequency where the detector array has a good sensitivity, and if the transmission of the imaged object is not too small, a combination of an array detector with a mW-range source allows real-time imaging. The combination of a THz gas laser or QCL, with a microbolometer array camera, fulfills this condition, and real-time imaging has been demonstrated by Lee et al. with a THz gas laser and a 160×120 pixel microbolometer array [471]. The gas laser delivered 10 mW of power at 2.5 THz. Approximately 60 frames per second were obtained and the best (S/N) -ratio was 13 dB for a 16 ms frame time. A particularly impressive example of this imaging technique is the real-time imaging of a dried seedpod placed at 22.5 m from the QCL [330]. For this experiment, a 320×240 pixel microbolometer array was used. The QCL was mounted in a pulse-tube cryo-cooler. It was operated in a narrow atmospheric window at 4.9 THz, in a pulsed mode with a 27% duty cycle and ~ 17 mW peak power, from which $\sim 150\text{ }\mu\text{W}$ reached the object. The pulsing was necessary for synchronization with the readout of the microbolometer camera. This

permits differential imaging where the strong IR background can be subtracted from the THz signal. The beam from the QCL was collimated with a paraboloidal mirror close to the exit window of the cooler. The dried seedpod was placed at 22.5 m from the QCL and the transmitted radiation was collected by a spherical mirror 2 m away from the seedpod, and imaged with a silicon lens onto the microbolometer array (Fig. 7.15). In another configuration, an off-axis paraboloidal mirror, placed after the spherical mirror, created an object plane in front of the Si lens. This provided a smaller field-of-view and a higher spatial resolution (Fig. 7.15). The (S/N) -ratio at the focal plane was 2.5 for a single frame (50 ms integration time) and 10 for a 20-frame average (1 s integration time).

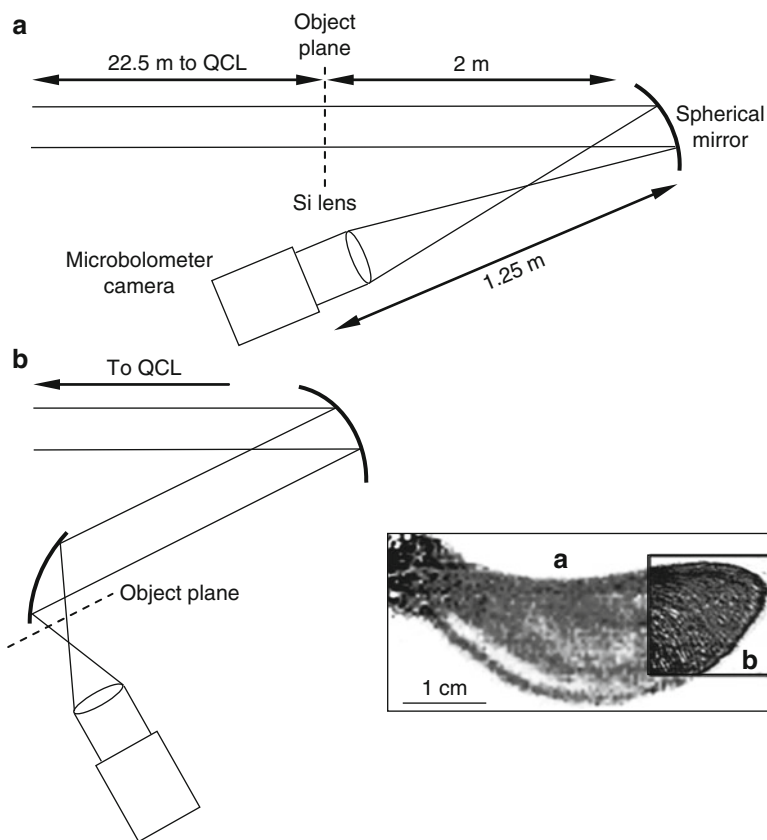


Fig. 7.15 Experimental setup for real-time imaging with a QCL and a microbolometer camera. In setup (a) the object plane is at a distance of 22.5 m from the QCL and 2 and 3.25 m from a focusing mirror and the microbolometer camera, respectively. In setup (b) the object plane is several centimeters from an off-axis paraboloidal mirror and the microbolometer camera. The inset shows the THz transmission image of a dried seedpod measured with setup (a) and measured with setup (b). Both images were taken with 1 s integration time (20 frames) (adapted from [330])

7.4.5 *Spectroscopic Imaging*

Spectroscopic imaging requires a source which covers a much broader frequency range than the spectral signatures and their distribution. Except for relatively narrow (<100 MHz) molecular transitions, typical signatures which are of interest for THz spectroscopic imaging, such as absorption bands of biomolecules or explosives, have a width in the order of 0.1 THz. In addition, such signatures are usually distributed across a spectral range in the order of 1 THz or more. Obviously, TDS imaging systems are very suitable for this task unless very high resolution below 1 GHz is required, and this will be discussed in the following section.

For other source-detector combinations, the situation is more complicated. While broadband detectors are readily available, sources which can cope with the spectral requirements are sparse, and so are imaging systems which provide spectral and spatial information at the same time. A step in this direction is a THz-wave parametric oscillator (TPO)-based spectroscopic imaging system [702]. Although this is, strictly speaking, not a cw imaging system, it will be described here, because the pulse is relatively long (ns) and its phase is not used to create image information as, for example, in time-domain imaging systems (Sect. 7.5). The emission frequency of the TPO was changed from 1.0 to 2.5 THz by rotating the nonlinear optical crystal. The emission is focused and the object under test is raster-scanned at the focus over 20×38 mm², which corresponds to 40×76 pixels. The transmitted radiation is detected with a pyroelectric detector, or an Si bolometer using a lock-in scheme, which is synchronized to the pulse frequency. The measurement time for one image was about 10 minutes. This was repeated for seven frequencies between 1.3 and 2.0 THz. By using a component spatial pattern analysis, it was possible to distinguish between three different drugs which were hidden in a closed airmail envelope (Fig. 7.16). It should be noted that the imaging procedure is a sequential one in space as well as for frequency, and that a similar experiment might be performed with QCLs operating at several frequencies, or across a wide frequency range. However, the experiment demonstrates in principal the capability of THz spectroscopic imaging.

7.4.6 *Three-Dimensional Imaging with a Quantum Cascade Laser*

A widely used method to generate three-dimensional images is tomography. As the word, which is derived from the Greek word “*tomos*,” meaning “section” or “slice,” suggests, tomography is imaging by sections. The three-dimensional image is reconstructed from slices by a mathematical procedure. At THz frequencies, this technique has been pioneered with TDS imaging systems (for a review, see [692]). More recently, a QCL was used for three-dimensional imaging [703]. The setup resembles X-ray computed tomography (CT). The emission from a pulsed 2.9-THz

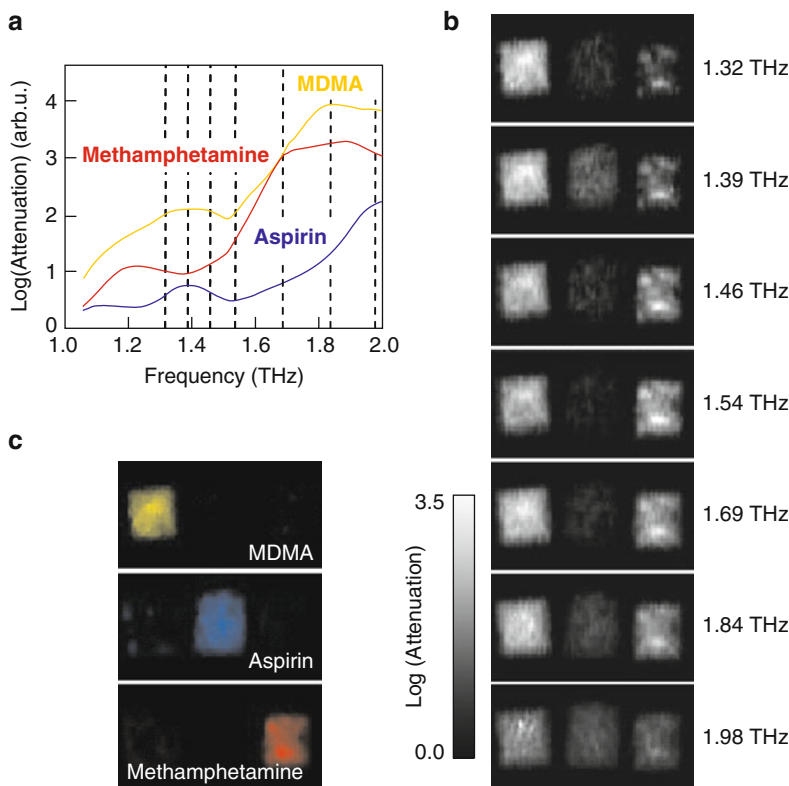


Fig. 7.16 (a) Absorption spectra of MDMA (3,4-methylenedioxymethamphetamine, commonly known as ecstasy), aspirin, and methamphetamine. The *dashed lines* indicate seven frequencies, which were used for the component spatial analysis. (b) Images of the three drugs in an envelope at seven frequencies (spatial arrangement of the drugs: MDMA, aspirin, and methamphetamine, from left to right placed in an envelope). (c) Spatial patterns extracted from the spectroscopic images. The drugs are identified on the basis of a component spatial analysis (adapted from [702])

QCL (250 ns pulse length, 80 kHz repetition rate, 70 mW peak power) was focused onto the sample. The FWHM of the beam was about 1 mm, which is rather large but is a compromise caused by the required depth of focus. For imaging, the sample was rotated in both directions orthogonal to the optical axis scanned through the focus. By this means, a series of parallel beams is transmitted through the sample and the intensity of the beams is detected with a Golay cell. Then the sample is rotated by a certain angle and the procedure is repeated. Finally, a slice through the sample is imaged. This is then repeated at different sample heights (Fig. 7.17). Data analysis is based on the Fourier slice theorem, which states that the one-dimensional Fourier-transform of a parallel beam of an object is equal to its two-dimensional Fourier-transform [704]. There are several algorithms for image reconstruction based on this theorem. In this particular case, the filtered back-projection algorithm was used,

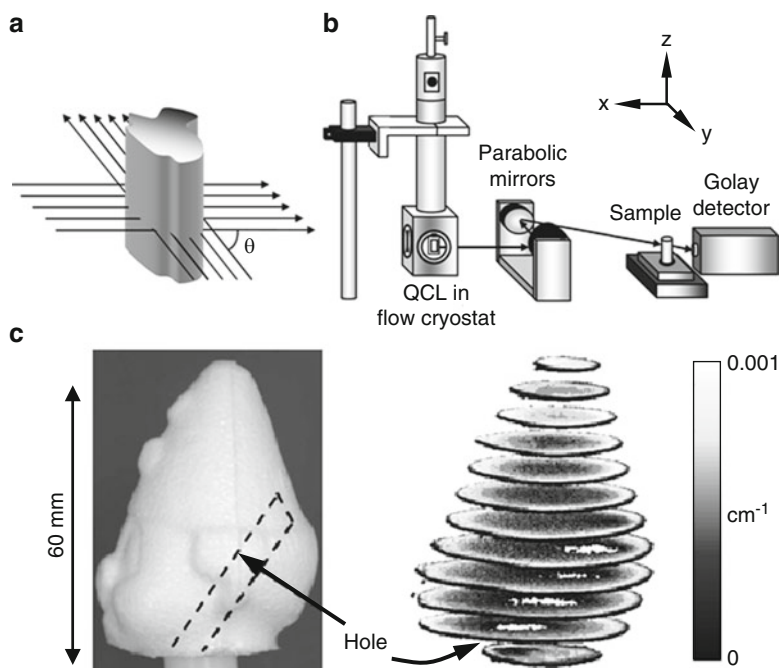


Fig. 7.17 Principle of a cw tomography approach (a) and experimental setup (b). The radiation from a QCL is transformed by two paraboloidal mirrors into a parallel beam, which is passed through the sample in the plane to be imaged. The intensity of the transmitted beam is detected with a Golay cell. This is done for several angles θ until a range of 180° is covered. The procedure is repeated for another plane by moving the sample in the z-direction. (c) Photograph of a polystyrene phantom with a 10-mm diameter hole (indicated by the dashed line) and reconstructed three-dimensional THz image obtained with the setup shown in (b). The hole is clearly visible (adapted from [703])

and a result of that procedure is shown in Fig. 7.17c. The image of a polystyrene phantom was taken with linear steps of 0.2 mm and the rotation angle was 10° . The total image acquisition time was about three hours. The reconstructed image reveals the shape and some surface details of the phantom, as well as the 10-mm diameter hole in its center.

7.4.7 Imaging with Time-Domain Spectrometers

THz time-domain spectrometers (TDS) were employed for imaging shortly after their invention [107] and numerous applications and imaging setups have been reported since (for a review, see [705]). There are two features which are unique to THz TDS systems: firstly, image information can be obtained from the amplitude

of the THz pulse as well as from its phase. Secondly, spectroscopic imaging is quite straightforward because THz TDS systems, emit in a broad spectral range, typically from 0.1 to 2 THz. However, realizing this potential makes data processing more demanding. The low power of the THz beam, (typically in the order of tens of μW average power), is a serious limitation. This can be compensated for because of the large dynamic range of a THz TDS system, due to the coherent detection scheme. However, because of the low power, most TDS imaging systems rely on raster-scanning of either the object or the THz beam. Each THz waveform obtained from one pixel contains a large amount of information which has to be processed. In order to record the waveform of the THz pulse, an optical delay line needs to be scanned during the measurement of one pixel. With a standard step-scan optical delay line, it can easily take several seconds to minutes to measure a 1024-point THz waveform. This time has to be multiplied by the number of pixels required for the image. Obviously, for real-time imaging, this poses a major challenge.

Figure 7.18 illustrates the operating modes of a TDS imaging system [706]. The two gray scale images were taken by raster-scanning a nylon plate in steps through the focus of a THz TDS beam. The left image is generated by plotting the peak-to-peak amplitude of the transmitted beam. The steps are visible because of scattering effects, which reduce the power arriving on the detector. If the phase of the THz pulse is used for image reconstruction, the steps are very well resolved (Fig. 7.18b). This illustrates the capability of this imaging mode. In general, if phase detection is employed, either the change in the arrival time of the peak or the first zero crossing of the pulse is detected. Changes Δt in the arrival time are given by

$$\Delta t = \frac{1}{c} \int n(z) dz, \quad (7.14)$$

where z is the direction of the THz beam and $n(z)$ is the refractive index along this direction. Changes in the arrival time of the THz pulse, as a function of position on

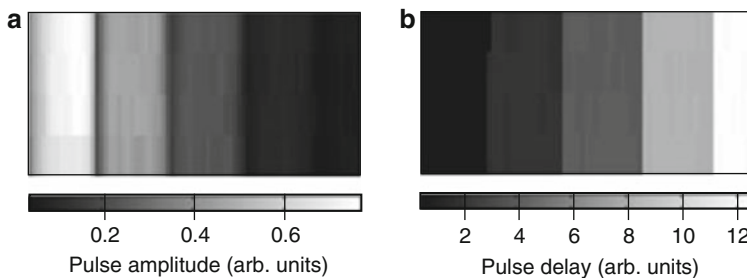


Fig. 7.18 Gray scale images of a nylon plate obtained by raster-scanning the object through the focus of a TDS-THz beam. Image (a) is generated by plotting the peak-to-peak amplitude of the transmitted beam. The steps are visible because of scattering effects which reduce the power arriving on the detector. If the phase of the THz pulse is used for image reconstruction, the steps are very well resolved (b) (adapted from [706])

the sample, indicate changes in the optical path length which are caused by changes in the thickness of the sample, or changes of the refractive index, or both. The phase or pulse delay can be determined with greater accuracy than the amplitude, because noise appears predominantly as amplitude fluctuations of the THz pulse, rather than affecting the pulse delay [707].

The spatial resolution of a TDS imaging system is affected by its broadband emission. The diameter of the focus or the minimum waist is dependent on the frequency. If a particular frequency interval is used to generate the image, the spatial resolution is roughly proportional to the average wavelength of this interval. If the whole spectrum is used, the spatial resolution is determined by the wavelength of the frequency component with the largest amplitude.

7.4.8 Spectroscopic Imaging with Time-Domain Spectrometers

Spectroscopic imaging is a capability which is inherent to TDS imaging systems. For each pixel of an image, the THz waveform can be detected and the spectrum obtained. As pointed out above, this is at the expense of measurement time. An example of TDS spectroscopic imaging is shown in Fig. 7.19 [708]. The test sample consists of a lactose pellet, a sucrose pellet, and a sheet of the explosive RDX (1,3,5-trinitro-1,3,5-triacyclohexane). The images were obtained by raster-scanning the THz beam across the sample, which was mounted at the focal position of the THz beam. The scanned area was $8 \times 24 \text{ mm}^2$ with a $100 \mu\text{m}$ pixel spacing. The total imaging time was 16 min. The standard nonspectroscopic image is shown in Fig. 7.19b. There is some structure and contrast in the sucrose and RDX images. However, this is hardly sufficient for discrimination or identification of the

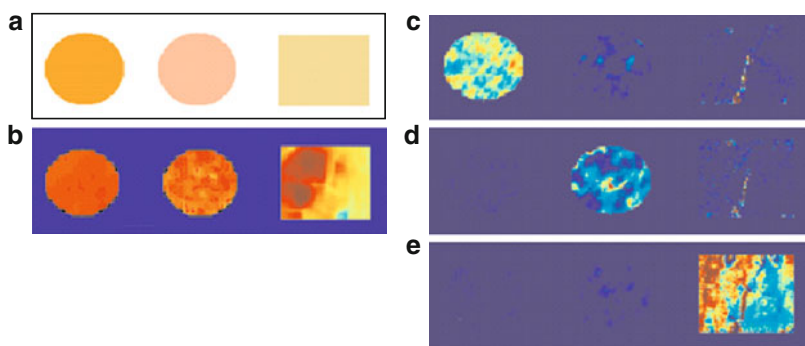


Fig. 7.19 (a) Scheme of the sample used for THz spectroscopic imaging. The left circle is a lactose pellet, the right circle is a fructose pellet, and the square is an RDX explosive sample. (b) TDS image using the pulse amplitude. (c)–(e) Spectral images obtained by a component spatial analysis similar to that applied in (Fig. 7.16). The different components are clearly identified (adapted from [708])

substances. By calculating the spectrum for all pixels at frequencies from 0.3 to 2.4 THz a three-dimensional data cube was obtained. Applying a component spatial pattern analysis, similar to the procedure described by Kawase et al. [702], it was possible to determine the spatial distribution of the chemicals (Fig. 7.19c–e).

7.4.9 *Three-Dimensional Imaging with Time-Domain Spectrometers*

The first imaging experiment with a TDS system providing three-dimensional information was reported in 1997 [691]. Although initially called “T-ray tomography,” it might be more appropriately described as “time-of-flight imaging.” A typical setup consists of a TDS emitter illuminating the object and a receiver detecting the reflected pulses. Figure 7.20 shows the imaging principle for the example of a floppy disc, which is essentially a layered structure with each layer having a different refractive index. These layers are a series of step discontinuities. When illuminated with the THz pulse, each step generates a reflected THz pulse and the signal measured in reflection consists of a series of such pulses, each with a unique time delay (Fig. 7.20a). The time delay between two pulses provides information about the optical thickness of the layer in between (cf. (7.14)), while the amplitude of each pulse contains information about the magnitude of the refractive index change at the layer interface. It should be noted that time-of-flight imaging is also possible in transmission geometry. In this case, the number of reflections of each pulse is even rather than odd.

Obviously, the depth resolution in time-of-flight imaging is determined by the ability to resolve two closely spaced reflections, which in turn is determined by the temporal duration of the THz pulse. Shorter pulses allow for better depth resolution. A figure-of-merit is the coherence length L_c of the radiation, which in terms of spectral bandwidth $\Delta\omega$ is given by $L_c = c/\Delta\omega$. The depth resolution is half the coherence length, because in reflection geometry each pulse must travel twice through the medium. The depth resolution can be improved by several techniques, for example by the time-domain analogue of optical coherence tomography. A depth resolution of less than 2% of the coherence length has been demonstrated with this technique [709].

As pointed out earlier, THz tomography was pioneered with TDS systems and many different modalities have been investigated. A review on pulsed THz tomography is given in [692]. Here, only a few results of THz tomography with TDS systems will be described. One example is the THz analog to X-ray CT. This is similar to the procedure described in Sect. 7.4.6. The sample is placed at the focus of a THz emitter and the signal is measured in transmission mode. Generating a three-dimensional image requires scanning four dimensions: two directions orthogonal to the optical axis, the rotational dimension for the slice, and time. With an optical delay line, it would take several hours for an image with a moderate number of a few thousand pixels. This can be speeded up by using a chirped probe beam

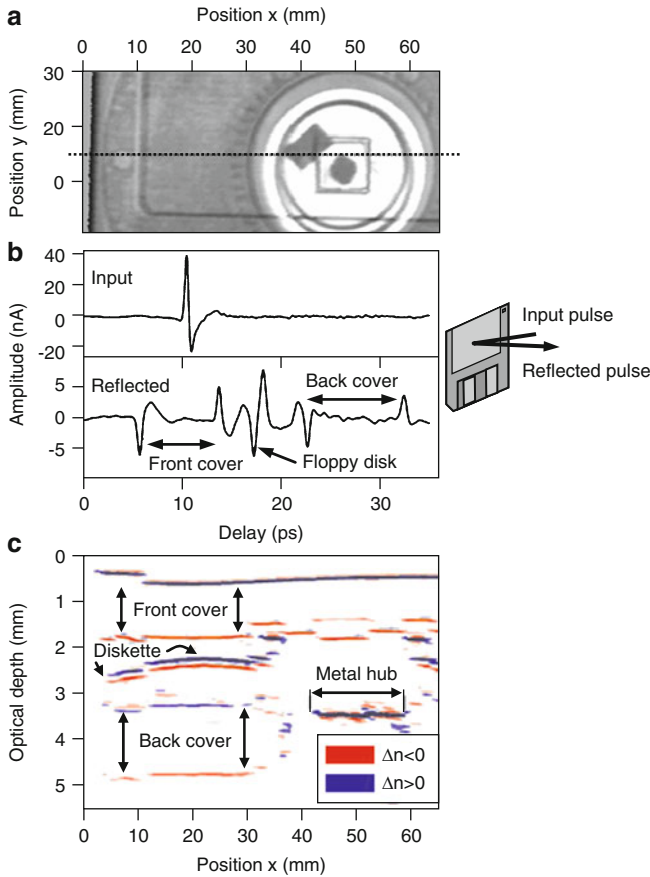


Fig. 7.20 (a) THz transmission image of a part of a floppy disk. The image was generated from the pulse amplitude. In (b) the pulse reflected from a floppy disk is shown in the lower trace. Reflections from each dielectric interface can be seen. For comparison, the upper trace shows the input THz pulse reflected from a mirror. Diagram (c) shows a time-of-flight image obtained from a slice along $y = 15$ mm (dotted line in (a)). The different components of the floppy disk such as its front and back cover, the diskette itself, and the metal hub are well resolved (adapted from [691])

technique, which enables the measurement of the full THz waveform using a single THz probe pulse. This hastens the imaging speed significantly [629, 710]. An example of a THz CT image measured with this setup is shown in Fig. 7.21. The three-dimensional image has been reconstructed using the filtered back-projection algorithm. In addition, the frequency-dependent optical properties were determined. The fine structure of the bone below about 0.5 mm could not be resolved, due to limitations of the reconstruction algorithm. It should be noted that there is a wide variety of pulsed THz tomography approaches, many of them inspired by tomographic techniques, developed for other applications and other parts of the electromagnetic spectrum. For these, the reader is referred to [692, 705].

Fig. 7.21 (a) Photograph of a piece of a turkey leg bone and (b) three-dimensional CT image of the bone. The fine structure of the bone is not resolved due to limitations of the reconstruction algorithm (adapted from [692])

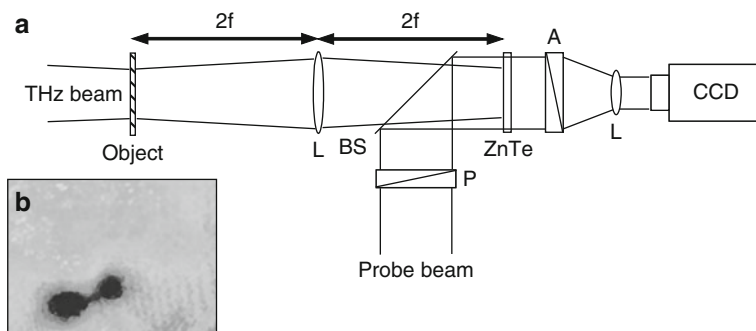
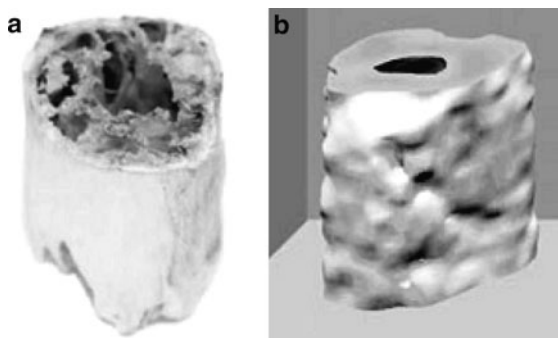


Fig. 7.22 (a) Two-dimensional real-time THz imaging system based on electro-optic sampling and a CCD camera and (b) THz transmission image of an ant obtained with this system (L: lens, BS: beam splitter, P: polarizer, A: analyzer; adapted from [711])

7.4.10 Real-Time Imaging with Time-Domain Spectrometers

Real-time or video-rate imaging with TDS systems faces two challenges. Firstly, the raster-scanning needs to be eliminated and, secondly, the waveform acquisition must be very fast. As with cw techniques, the first challenge can be tackled by using a detector array. As described in Sect. 6.6, electro-optic sampling relies on polarization rotation of a near-IR femtosecond probe pulse, with the degree of polarization change depending on the amplitude and phase of the THz field. This transfers the THz image into the near-IR, where large format CCD detector arrays are available. In contrast to raster-scan imaging, a large $\sim\text{cm}^2$ -sized electrooptic crystal, and a correspondingly large-area THz beam, need to be used. The polarization change across the area of the electro-optic crystal can be detected by using crossed polarizers and a CCD camera [711] (Fig. 7.22). However, this technique needs considerably more power in the femtosecond pulse beam. Therefore, an amplified femtosecond laser system is needed. This system has real-time imaging capability, but it can only obtain one image per time delay. For acquisition of the THz pulse waveform, a delay line or the chirped probe beam technique is required.

7.5 Near-Field Microscopy

In far-field imaging the typical maximum resolution is in the order of the wavelength λ (Sect. 7.1). At THz frequencies with wavelengths λ between 1 mm and 30 μm , this is not always sufficient since many objects, such as cells or nano-structured semiconductor components, have dimensions well below 30 μm . For example, imaging of the inner parts of a cell requires sub- μm resolution and the inspection of semiconductor devices requires nm-scale resolution. At THz frequencies, this corresponds to a spatial resolution scale down to $\lambda/10,000$. To overcome the far-field limitation, information has to be obtained from the near-field of the sample. Every sample surface consists of atoms with an electric field associated with each atom, leading to an electric charge distribution, and to the so-called near-field distribution. If an electromagnetic wave impinges on a surface, it reacts by re-radiating waves from each local resonator on and in the surface. At a distance of λ all near-fields have sufficiently interfered, so that the resulting propagating far-field wave contains only information with length scales of the order of λ . Detecting the near-field distribution allows information to be obtained on a length scale of much less than λ . Near-field microscopy describes the mapping of such local near-fields of a sample, ultimately with atomic resolution. In 1928, E.H. Synge proposed that sub-wavelength resolution might be obtained by scanning a sub-wavelength sized hole in a metal film illuminated from the rear and scanned very close to the surface of the sample [712]. Later on, H. A. Bethe [713] and C.J. Bouwkamp [714] developed a theory of diffraction by small holes.

While these concepts were readily adapted and developed in the visible part of the spectrum, it was not until 1985 that THz near-field imaging experiments were performed [715]. Generally, two approaches can be distinguished. One follows the original idea of Synge by using a small aperture, for example in a tapered cone similar to those described in Sect. 3.10. This apertures-based microscopy allows a resolution in the order of $\lambda/10$. Major limitations are the manufacture of small holes, and the power reduction which scales with the aperture size d at a rate of d^{-4} leading to a loss of power of up to $\sim 10^4$ [716]. Alternatively, the metallized tip of an atomic force microscope (AFM) could be used. According to Babinet's principle, this can be viewed as an inverted structure of a small hole in a metal plate. The tip apex is irradiated by a THz source and serves as a scattering source. In early 2000, this method was applied at THz frequencies [717, 718].

7.5.1 Aperture-Based Scanning Near-Field Microscopy

The demonstration of sub-wavelength resolution at THz frequencies was reported in 1985. Massey et al. used a gas laser operating on the very strong 2.5 THz (118.8 μm) line of methanol [715]. They achieved a spatial resolution down to 30 μm by scanning a 10 μm wide and 2 mm long slit against another slit with

the same dimensions. The separation of the slits was approximately $10\text{ }\mu\text{m}$. The laser radiation was focused onto one slit, while a pyroelectric detector was placed behind the scanning slit. Similar experiments have been performed with round apertures down to $50\text{ }\mu\text{m}$ diameter [715]. This initial experiment illustrates the limitations of this type of microscopy. Firstly there is a strong power reduction, as the nonevanescence power, which is transmitted through an aperture with diameter d varies as $(d/\lambda)^4$. To achieve a resolution of one tenth of the wavelength, the loss in transmitted power is 10^4 . The second challenge is to keep the distance between the object and the hole constant and below at least $\lambda/2$.

A first step to improve the power transmittance was the introduction of a cone with a small pinhole at the exit (Fig. 7.23a). Essentially, this is a straight cone as described in Sect. 3.10, which can be viewed as a tapered hollow metal waveguide with a specific cutoff frequency. Assuming an infinite conductivity, and neglecting input and output coupling losses, the power transmittance in the cutoff region is [719]

$$T = \exp\left(-4\pi L \sqrt{\sigma_c^2 - \sigma^2}\right), \quad (7.15)$$

where σ is the wavenumber of the incident radiation, σ_c is the cutoff wavenumber, and L is the length of the waveguide. As with a hole in a foil, there is a rather steep decrease of the transmitted power beyond the cutoff frequency. A coaxial waveguide is known to have no cutoff effect and can be employed to improve the power transmittance. F. Keilmann inserted an electrochemically sharpened $25\text{-}\mu\text{m}$ diameter tungsten wire into a metal cone. To ensure electrical insulation between the wire and the pinhole, the wire was sputter-coated with SiO_2 . The transmitted power at a wavelength of $392\text{ }\mu\text{m}$ ($\approx 0.8\text{ THz}$) for cones, with pinholes of different diameters, is shown in Fig. 7.23. Without the central wire, there is a sharp cutoff at about $\lambda/2$, while with the wire the transmittance is independent of the wavelength from $\lambda/3$ to $\lambda/20$.

Tapered cones have been applied for THz microscopy experiments with a variety of sources. Merz et al. [720] used a THz gas laser operating at $392\text{ }\mu\text{m}$ to resolve spatially the photoconductivity of a two-dimensional electron gas. They achieved a spatial resolution of $\lambda/2$. Near-field imaging with a time-domain system and broadband THz pulses has been reported by Hunsche et al. [721]. The emission pulse had a maximum at 0.5 THz . Using a tapered cone with a $100\text{-}\mu\text{m}$ diameter pinhole, the power was strongly attenuated, and the low-frequency part was cutoff, resulting in a peak pulse power at 1.4 THz . With this setup, a spatial resolution of $\lambda/4$ was obtained. Sensitivity and spatial resolution of such a system can be improved by placing the aperture in front of the detector. Mitrofanov et al. observed a spatial resolution of better than $40\text{ }\mu\text{m}$ over a broad spectrum from 300 to $600\text{ }\mu\text{m}$, which corresponds to $\lambda/15$ at the longest wavelength [722]. Schade et al. [405] used broadband coherent THz synchrotron radiation. With a conical cone and a $100\text{-}\mu\text{m}$ diameter pinhole, they obtained a spatial resolution of $\lambda/12$ at 0.36 THz , which increased to $\lambda/40$ at 60 GHz . Examples of the capability of THz microscopy with tapered cones are shown in Fig. 7.24.

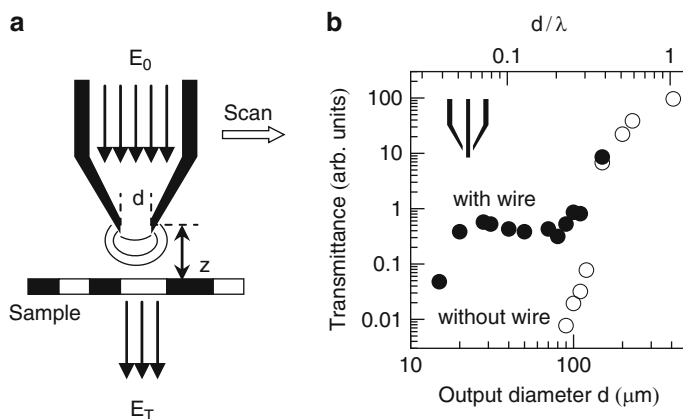


Fig. 7.23 (a) Scanning near-field microscopy using an aperture of diameter d with incident electric field E_0 in transmission. The evanescent electric field is scanned along a sample with structural variation in the μm to nm range at a sufficiently small distance $z \ll d$ (not to scale in the drawing). The transmitted field E_T coupled out of the aperture through the sample is measured with a detector in the far-field (b) Relative transmittance of focusing cones with different output apertures at $\lambda = 392 \mu\text{m}$ (open circle: cone without central wire; closed circle: cone with central wire). Note that for the cone with a central wire (closed circles) the transmittance is independent of the wavelength from approximately $\lambda/3$ to $\lambda/20$ (adapted from [719])

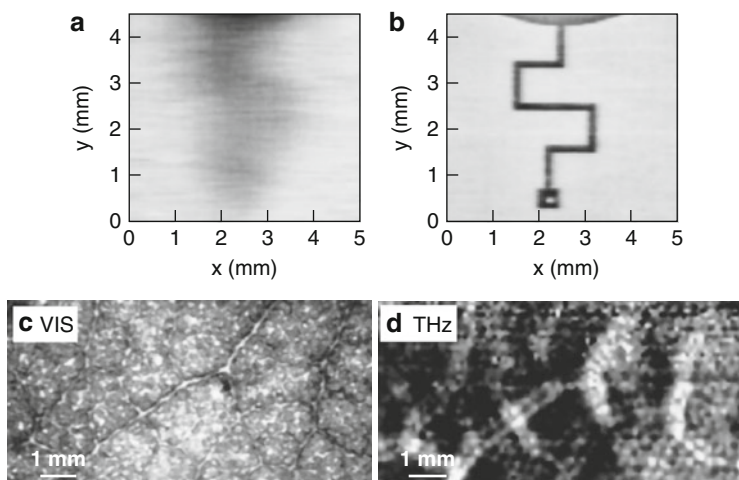


Fig. 7.24 Examples of THz near-field images. (a) Image of a metal circuit deposited on a GaAs wafer obtained by conventional THz imaging. (b) The same circuit imaged with the dynamic aperture technique (adapted from [723]). (c) Visible image of a leaf. (d) THz image of the same portion of the leaf as in (c) obtained with a cone with an aperture diameter of $200 \mu\text{m}$ and coherent THz synchrotron radiation as the source. Note that less absorption is indicated by a darker region in the THz image (adapted from [405])

Another development is a BWO-based THz micro-spectrometer [724, 725]. High spatial resolution is obtained with small pinholes (diameter $\lambda/100$ to $\lambda/10$) in a thin copper foil. The radiation from the BWO is focused onto the pinhole. The focusing is accomplished with an aberration-free hyperbolic lens and a hemispherical Si immersion lens. The sample is placed immediately behind the pinhole. The hyperbolic lens gives a diffraction limited spot. The immersion lens reduces this spot by an additional factor of n , where n is the refractive index of the lens material. The small pinhole in close contact to the sample defines the spatial resolution. The transmitted radiation is detected with a Golay cell. By tuning the frequency of the BWO spectral information is obtained and spatial information is extracted, by raster-scanning the object under investigation. The instrument is of interest for the study of biological samples with high spatial and frequency resolution [725].

An alternative approach is the so-called dynamic aperture. This method was developed for the IR by Palanker and co-workers [726], and was adapted by Chen et al. [723] for THz frequencies. The method is based on a modulated optical gating beam from a laser, which is focused to a micrometer-sized spot on a thin silicon wafer (Fig. 7.25). The energy of the laser photons is larger than the band gap energy of Si, thus creating an electron-hole plasma in the wafer. The plasma acts as a small mirror while the rest of the Si wafer is nearly transparent at THz frequencies. The sample is placed on the rear side of the Si wafer and the change of transmitted THz power is detected with a lock-in amplifier. In general, the method requires Si, or possibly other semiconductors, for generation of the aperture and keeping the sample in place. Chen et al. [723] implemented this method in combination with a TDS, which was set around 0.9 THz. This achieved a spatial resolution of 50 μm . Gompf et al. [727] developed a cw near-field spectrometer for 30 GHz to 1.4 THz. Their spectrometer was based on a series of BWOs, for coverage of the frequency range, and a dynamic aperture created by a green laser. The laser beam was focused to 35 μm while the THz beam was focused to 1 mm. With this setup, it was possible to obtain a spatial resolution of 100 μm at 0.45 THz, which corresponds to $\lambda/7$.

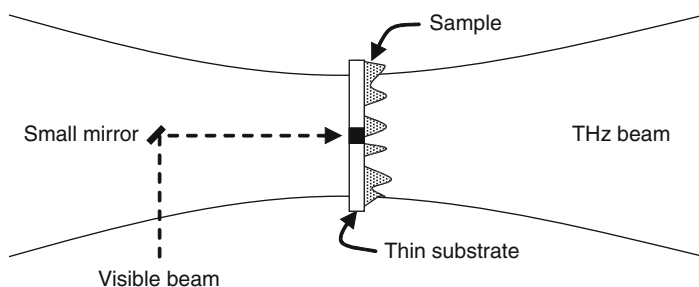


Fig. 7.25 Principle of a dynamic aperture (black region), which is formed by an optical beam from a laser illuminating a thin semiconductor substrate creating free carriers. The THz beam penetrates the substrate and the sample is placed on the backside of the substrate. The spatial resolution is determined by the size of the aperture

Due to the polarized radiation from the BWO, anisotropic sample properties could be investigated on spot sizes much smaller than the wavelength. The resolution was limited by diffusion of the photo-induced free carriers in the Si wafer rather than by the THz beam.

7.5.2 Aperture-Less Scanning Near-Field Microscopy

A metallized atomic force microscope (AFM) nano-tip can be viewed as an inverted hole. Since AFMs are widely used these tips can be obtained commercially, with high quality and nano-scale sizes ranging from 5 to 40 nm. The tip apex serves as a scattering source of the light. The IR and THz wavelength is much larger than the nano-tip so that in a first approximation a dipole is induced in the tip. The dipole now samples the material with a resolution comparable to the size of the nano-tip [728, 729]. Such an instrument is called a scattering scanning near-field optical microscope (s-SNOM). Several other abbreviations are used in the literature which use the same concept, although the spectral range or specifics of the tip may vary, for example, aperture-less near-field scanning optical microscopy named ANSOM or aSNOM.

In analogy with chemical microscopes, a chemical nanoscope can map the chemical composition of a material, with ultimately atomic resolution, by using the spectral fingerprint of chemicals originating from the types of bonds and forces which build up the material. s-SNOM in the THz region has been demonstrated by using an optically excited gas laser for illumination of an AFM nano-tip and an HEB detector [730], by employing THz TDS systems [608, 717, 718] and FELs [731, 732]. In general, any source and detector combination can be used if the (S/N) -ratio is sufficient. The illuminated area of the sample is still governed by far-field limitations, and the focus corresponds to a diameter of $\sim \lambda$, in which the nano-tip only samples an area which is up to 10^{-8} smaller. This signal reduction is partially compensated by the local field enhancement of the metal nano-tip, which has a function that can be crudely described as being analogous to a lightning rod. The nano-tip is typically elongated, so that approaches using antenna theory [733–735] or surface plasmons (Sect. 3.17) to describe the near-field interaction are also used. For biological samples, the signal power has to be low to avoid damage, which demands sophisticated modulation techniques and sensitive detectors. The THz region has been found to be especially useful for mapping charge distribution in semiconductors, where the THz power impinging on the sample is also less critical [730, 736].

The contrast in near-field microscopy is determined by the complex refractive index $\hat{n} = n + i\kappa$ of a material, so it is a combination of amplitude and phase information. Using special modulation and demodulation techniques, it is possible to extract phase and amplitude contrast independently [737]. For comparison, the complex refractive index of samples can be measured by far-field methods such as THz TDS, DFTS, heterodyne spectroscopy, or ellipsometry. Often, it is possible

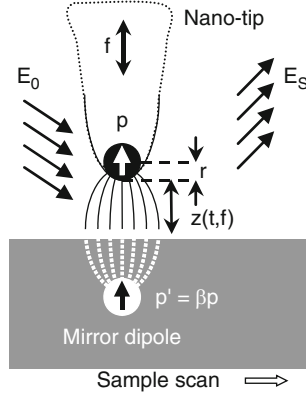


Fig. 7.26 Aperture-free scanning near-field microscopy at a distance z using a nano-tip (approximated by a tip dipole of radius r) in a medium inducing a mirror dipole in the sample. The scattered field E_s is recorded with a detector in the far-field. The indicated field lines are perpendicular to the tip and sample surface if both are ideal conductors with infinite conductivity

to obtain near-field contrast by modulating the signal, due to the AFM tip and cantilever oscillation f , and demodulation at higher harmonics $2f, 3f, \dots, nf$ of the cantilever eigen-frequency f , while recording the scattered light. In this case, the scattered light from the cantilever interferes with the light scattered from the sample–tip apex interaction zone. The contrast is then not separated into amplitude and phase but, due to the strong interference in this mode, very small amounts, for example a few hundred molecules [738], can be measured.

The measured signal follows from scattering of the incident electric field E_0 at the tip apex-sample system, described by the effective polarizability (Fig. 7.26). The tip of the AFM cantilever, in the simplest case, can be approximated by a point dipole. The polarizability α_T of such a point dipole is the product of the polarizability of a dense material, which can be derived from the Clausius-Mosotti equation, and the volume of the sphere. In a medium with the complex permittivity ε_M , it is given as

$$\alpha_T = 4\pi r^3 \frac{\varepsilon_T - \varepsilon_M}{\varepsilon_T + 2\varepsilon_M}. \quad (7.16)$$

For highly conductive materials, like Au or Pt coated tips, and in air ($\varepsilon_M \approx 1$), the polarizability approaches $4\pi r^3$. For a tip made from very pure Si, the permittivity may be expressed by the refractive index $\varepsilon_T = n^2$, and the tip polarizability is still 78% of the value for metals. If ε_T approaches $-2\varepsilon_M$ (in air the real and imaginary parts of ε_T approach -2 and 0 , respectively), a resonance occurs and very high enhancement factors can be obtained.

In the mirror-dipole model [739, 740], which is typically employed for extended samples or surfaces, the effective polarizability is derived as

$$\alpha_{\text{eff}} = \varepsilon_M \alpha_T (1 + \beta) \left(1 - \frac{\alpha_T \beta}{16\pi(z+r)^3} \right)^{-1}. \quad (7.17)$$

Concerning the resonant mode the same reasoning holds for β with

$$\beta = \frac{\varepsilon - \varepsilon_M}{\varepsilon + \varepsilon_M}. \quad (7.18)$$

So-called polariton-resonant materials have a real part of ε close to -1 (in air) and a small imaginary part (in air). If the tip and the sample are sufficiently separated on the order of one wavelength ($z \gg \lambda$), then the effective polarizability in (7.17) mathematically simplifies to the sum of the tip dipole and its mirror dipole.

The scattered electric field E_S is proportional to the effective polarizability α_{eff} and the incident electric field E_0 , leading to

$$E_S \propto \alpha_{\text{eff}} E_0. \quad (7.19)$$

Equation (7.17) has been modified for samples with flat surfaces [741]. With the Fresnel reflection coefficient R of the flat surface, the scattered field is expressed as

$$E_S = \varepsilon_M \alpha_T (1 + R)^2 \left(1 - \frac{\alpha_T \beta}{16\pi(z+r)^3} \right)^{-1} E_0. \quad (7.20)$$

The factor $(1 + \beta)$, derived from electrostatic reflection, leads to inconsistencies because in the far-field the mirror dipole can radiate more strongly than the tip dipole, so it is replaced by $(1 + R)$. Another factor $(1 + R)$ follows from additional illumination of the tip apex by reflected light from the sample interface [741]. This simple model describes a variety of experiments very well. Sophisticated models include the tip geometry and sample nano-scale geometry, which may consist of multilayers, three-dimensional structures, or of naturally very heterogeneous materials like biological membranes [742]. The z -dependent enhancement factor in (7.20) approaches 1 for large values of z (Fig. 7.27) and $4/3$ for $z = 0$. More precisely, the value deviates only slightly from one for distances of $2r$ to $3r$. This leads to a maximum spatial resolution of tip-based scanning near-field microscopy close to the tip curvature radius unless resonance effects occur, which additionally enhance the tip-sample interaction. The scattering and absorption cross-sections σ_S and σ_A , according to classical electrostatics, are

$$\sigma_S = \frac{k^4}{6\pi} |\alpha_{\text{eff}}|^2, \quad (7.21)$$

$$\sigma_A = k \Im(\alpha_{\text{eff}}), \quad (7.22)$$

and connect absorption and scattering to the effective polarizability. For a sphere with a polarizability of negligible imaginary part and absorption ($\sigma_A \approx 0$) and a real part approaching $4\pi r^3$, the product of the geometrical cross-section and the Rayleigh factor in the fourth power of the wavevector $k = 2\pi/\lambda$ is obtained: $\sigma_S = (8\pi r^2/3)(kr)^4$. For an AFM tip curvature radius r of 30 nm and at 1 THz, a scattering cross-section of only 10^{-9} nm^2 results because at THz frequencies the

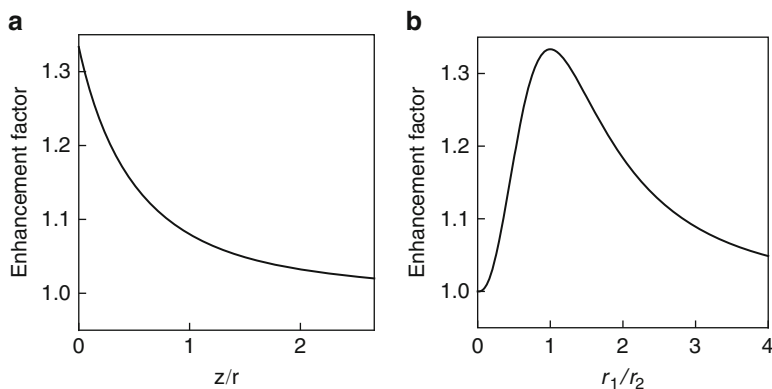


Fig. 7.27 (a) Dipole interaction enhancement factor for an infinitely conductive sample and tip (curvature radius of $r = 30$ nm and $\beta \approx 1$). The z -dependent near-field enhancement and therefore the maximum spatial resolution corresponds approximately to r . (b) Enhancement due to interaction of two dipoles of radius r_1 and r_2

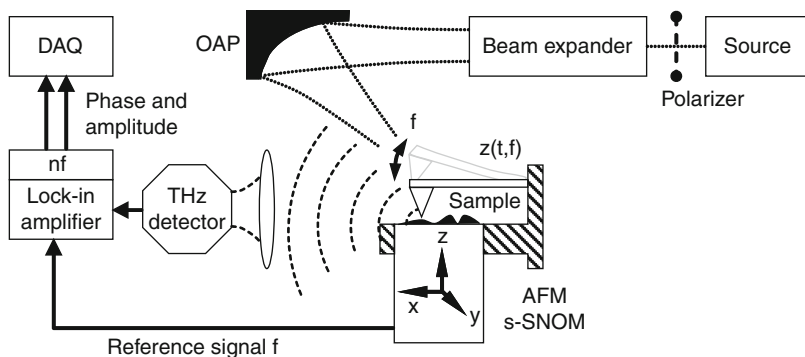


Fig. 7.28 Experimental setup with optical modulation due to the cantilever tip motion. The cantilever and the tip apex are modulated with the frequency f , while the detected signal is demodulated at higher harmonics $n \times f$. The laser or light source is shaped by optical components to achieve optimal far-field focusing close to the AFM tip. The polarizer is aligned to allow perpendicular polarization along the elongation of the metalized AFM tip for efficient coupling

wavevector is very small. In an experiment (Fig. 7.28), the cantilever oscillates with a frequency f , close to its eigen-frequency, determined by the geometry of the cantilever and the material stiffness. The frequency can range from a few kHz to MHz. The oscillation leads to a tip displacement from the surface of $z(t) = \Delta z(1 + \cos(2\pi f t))/2$. This displacement modulates the near-field interaction (7.20) and the effective polarizability α_{eff} as a function of time. Fourier decomposition of the time-dependent scattered field $E = E(z(t))$, at higher harmonic demodulation n , allows the n -th component of the polarizability α_{eff} to be extracted and the following is obtained

$$E_n = s_n \exp(i\Phi_n) \propto (1 + r)^2 \alpha_{\text{eff},n} E_0, \quad (7.23)$$

in which s_n and Φ_n describe the near-field amplitude and phase, respectively. The contrast for two materials A and B in amplitude and phase is then given as the ratio $s_{A,n}/s_{B,n} = |\alpha_{\text{eff},A,n}/\alpha_{\text{eff},B,n}|$ and the argument $\Phi_{A,n}/\Phi_{B,n} = \arg(\alpha_{\text{eff},A,n}/\alpha_{\text{eff},B,n})$. This is valid under the assumption that the Fresnel coefficient r does not change at the junction of the two materials. Using demodulation at higher harmonics reduces the influence of scattered light components from the cantilever surface, because simple reflection is mainly modulated at the frequency f , if the cantilever is operated in such a way that it mainly oscillates harmonically. This is not easy to achieve because a high (S/N)-ratio is only reached if the tip oscillates very close to the surface for which the z -dependent near-field enhancement is sufficient. If the tip interacts with the surface, nonlinear frequencies can be excited in the cantilever arm, which translate in contributions of the reflected light at the cantilever arm at higher harmonics of f . However, the nonlinear near-field interaction has frequency components with sufficient (S/N)-ratio even at higher orders of f at $2f$, $3f$, or $4f$, and good tip-sample distance controls with sub-nm resolution are commercially available to minimize nonlinear effects from sample-tip mechanical interactions. Heterodyne [743] or interferometric detection with a reference beam (Fig. 7.29a) is used in experimental setups, which allows the separation of amplitude and phase contrast. Near-field images, for example of semiconductor circuits (Fig. 7.29b), can be obtained which are similar to those obtained with far-field microscopes, but with a spatial resolution far beyond the diffraction limit. With THz near-

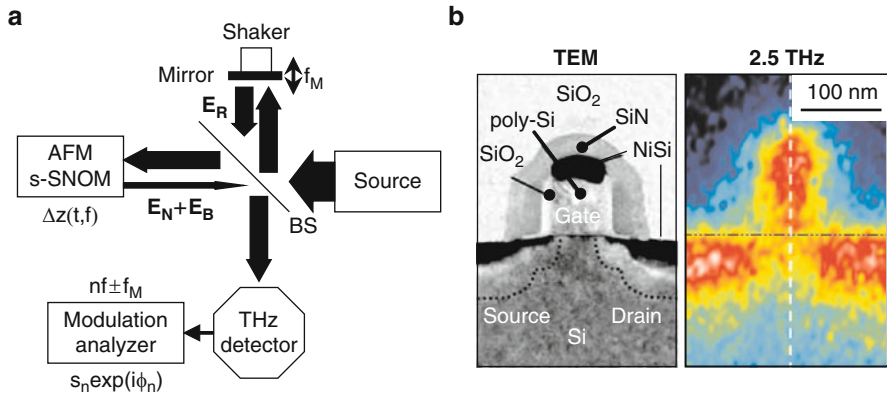


Fig. 7.29 (a) Experimental setup for interferometric detection using a reference beam with an electric field of E_R . The near-field E_N and unavoidable background field E_B reflected from the AFM tip and the rest of cantilever, respectively, is demodulated at the n -th harmonic of the tip oscillation frequency. Amplitude s_n and phase Φ_n information is extracted and can be used to plot images with pure amplitude or pure phase contrast (adapted from [737]). (b) Near-field imaging of a transistor at 2.5 THz ($\lambda \approx 118 \mu\text{m}$) with a spatial resolution of 40 nm, several orders of magnitude below the diffraction limit, in comparison with a transmission electron microscope (TEM) image (adapted from [730])

field techniques, it is possible to distinguish very small numbers of molecules in chemistry and biology [744], device structures in current nanoelectronics, or to map the electron density differences due to a few hundred electrons [730]. The availability of commercial near-field microscopes [745], and the wealth of potential applications in nanotechnology, nanooptics, and life sciences is resulting in very active research in near-field techniques.

References

1. T.G. Blaney, *Infrared and Millimeter Waves* (Academic Press, New York, London, Toronto, Sydney, San Francisco, 1980), *Submillimeter Techniques*, vol. 3, chap. 1 – Detection Techniques at Short Millimeter and Submillimeter Wavelengths: An Overview, p. 2
2. J.W. Fleming, High resolution submillimeter-wave Fourier-transform spectrometry of gases, *IEEE Trans. Microw. Theor. Tech.* **22**, 1023 (1974)
3. S. Paine, R. Blundell, D.C. Papa, J.W. Barrett, S.J.E. Radford, A Fourier transform spectrometer for measurement of atmospheric transmission at submillimeter wavelengths, *Publ. Astron. Soc. Pac. (PASP)* **112**(767), 108 (2000)
4. H. Yang, C.A. Kulesa, C.K. Walker, N.F.H. Tothill, J. Yang, M.C.B. Ashley, X. Cui, L. Feng, J.S. Lawrence, D.M. Luong-Van, M.J. McCaughrean, J.W.V. Storey, L. Wang, X. Zhou, Z. Zhu, Exceptional terahertz transparency and stability above dome A, Antarctica, *Publ. Astron. Soc. Pac. (PASP)* **122**, 490 (2010)
5. D.P. Marrone, R. Blundell, E. Tong, S.N. Paine, D. Loudkov, J.H. Kawamura, D. Lühr, C. Barrientos, in *Proc. 16th International Symposium on Space Terahertz Technology* (2005), Observations in the 1.3 and 1.5 THz atmospheric windows with the Receiver Lab Telescope, pp. 64–67
6. F.C.D. Lucia, Science and technology in the submillimeter region, *Opt. Photon. News* **14**, 45 (2003)
7. A. Crocker, H.A. Gebbie, M.F. Kimmitt, L.E.S. Mathias, Stimulated emission in the far infrared, *Nature* **201**, 250 (1964)
8. J.C. Bose, On a self-recovering coherer and the study of the cohering action of different metals, *Proc. Roy. Soc.* **65**, 166 (1899)
9. H. Rubens, B.W. Snow, On the refraction of rays of great wavelength in rock salt, sylvine, and fluorite, *Phil. Mag.* **35**, 35 (1893)
10. E.F. Nichols, A method for energy measurements in the infrared spectrum and the properties of the ordinary ray in quartz for waves of great wavelength, *Phys. Rev.* **4**, 297 (1897)
11. H. Rubens, E.F. Nichols, Heat rays of great wave length, *Phys. Rev.* **4**, 314 (1897)
12. H. Rubens, K. Kurlbaum, Über die Emission langwelliger Wärmestrahlen durch den schwarzen Körper bei verschiedenen Temperaturen, *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin* **II**, 929 (1900)
13. M. Planck, Gedächtnisrede auf Heinrich Rubens, *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, pp. CVIII–CXII (1923)
14. H. Rubens, R.W. Wood, XXVII. Focal isolation of long heat-waves, *Phil. Mag.* **21**, 249 (1911)
15. R.W. Wood, The echelette grating for the infrared, *Phil. Mag.* **20**, 770 (1910)
16. L. Rayleigh, On the manufacture and theory of diffraction gratings, *Phil. Mag.* **47**(4), 193 (1874)

17. E. Hagen, H. Rubens, On some relations between the optical and the electrical qualities of metals, *Phil. Mag.* **7**, 157 (1904)
18. H. Rubens, Gittermessungen im langwelligen Spektrum, *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 8–27 (1921)
19. E.F. Nichols, J.D. Tear, Joining the infra-red and electric-wave spectra, *Astrophys. J.* **61**, 17 (1925)
20. A.G. Arkadiewa, Short electromagnetic waves of wave-length up to 92μ , *Nature* **113**, 640 (1924)
21. R.M. Badger, Absolute intensities in the hydrogen-chloride rotation spectrum, *Proc. Natl. Acad. Sci. USA* **13**(6), 408 (1927)
22. C.H. Cartwright, Black bodies in the extreme infra-red, *Phys. Rev.* **35**, 415 (1930)
23. C.H. Cartwright, Radiation thermopiles for use at liquid-air temperatures, *Rev. Sci. Instr.* **4**, 382 (1933)
24. H.M. Randall, J. Strong, A self recording spectrometer, *Rev. Sci. Instrum.* **2**(10), 585 (1931)
25. J. Strong, S.C. Woo, Far infrared spectra of gases, *Phys. Rev.* **42**(2), 267 (1932)
26. J. Strong, G.A. Vanasse, Interferometric spectroscopy in the far infrared, *J. Opt. Soc. Am.* **49**, 844 (1959)
27. A.H. Pfund, The electric Welsbach lamp, *J. Opt. Soc. Am.* **26**, 439 (1936)
28. V. Cleeton, N.H. Williams, Electromagnetic waves of 1.1 cm wave-length and the absorption spectrum of ammonia, *Phys. Rev.* **45**, 234 (1934)
29. E.K. Plyler, Prism spectrometry from 24 to 37 microns., *J. Chem. Phys.* **15**, 885 (1947)
30. E.K. Plyler, F.P. Phelps, Growth and infrared transmission of cesium iodide crystals, *J. Opt. Soc. Am.* **42**, 432–1 (1952)
31. E.K. Plyler, N. Acquista, Infrared spectrometry with a cesium iodide prism, *J. Opt. Soc. Am.* **43**, 212–1 (1953)
32. M.J.E. Golay, A pneumatic infra-red detector, *Rev. Sci. Instr.* **18**, 347 (1947)
33. H.A. Zahl, M.J.E. Golay, Pneumatic heat detector, *Rev. Sci. Instr.* **17**, 511 (1946)
34. D.H. Andrews, W.F. Bruksch, W.T. Zeigler, E.R. Blanchard, Attenuated superconductors I. For measuring infra-red radiation, *Rev. Sci. Instrum.* **13**, 281 (1942)
35. D.H. Andrews, R.M. Milton, W. DeSorbo, A fast superconducting bolometer, *J. Opt. Soc. Am.* **36**(9), 518 (1946)
36. H.M. Randall, D.M. Dennison, N. Ginsburg, L.R. Weber, The far infrared spectrum of water vapor, *Phys. Rev.* **52**, 160 (1937)
37. F.A. Firestone, A periodic radiometer for eliminating drifts, *Rev. Sci. Instrum.* **3**, 163 (1932)
38. R. Beringer, The absorption of one-half centimeter electromagnetic waves in oxygen, *Phys. Rev.* **70**, 53 (1946)
39. W. Gordy, Early events and some later developments in microwave spectroscopy, *J. Mol. Struct.* **97**, 17 (1983)
40. J.H.N. Loubser, C.H. Townes, Spectroscopy between 1.5 and 2 mm wave-length using magnetron harmonics, *Phys. Rev.* **76**, 178 (1949)
41. J.A. Klein, J.H.N. Loubser, A.H. Nethercot, C.H. Townes, Magnetron harmonics at millimeter wavelengths, *Rev. Sci. Instrum.* **23**, 78 (1952)
42. C.A. Burrus, W. Gordy, Submillimeter wave spectroscopy, *Phys. Rev.* **93**, 897 (1954)
43. A. Hadni, A short history of 50 years of research in the far infrared: 1952–2002, *Int. J. Infrared Millimeter Waves* **24**(2), 91 (2003)
44. A.A. Michelson, S.W. Stratton, A new harmonic analyzer, *Am J. Sci.* **s4–5**, 1 (1898)
45. M.J.E. Golay, Multi-slit spectrometry, *J. Opt. Soc. Am.* **39**, 437 (1949)
46. P. Fellgett, On numerical Fourier transformation, with special reference to Lipson-Beevers strips, *J. Sci. Instrum.* **35**, 257 (1958)
47. P. Jacquinot, How the search for a throughput advantage led to Fourier transform spectroscopy, *Infrared Phys.* **24**(2–3), 99 (1984)
48. J. Strong, Fourier transform spectroscopy reminiscences, *Infrared Phys.* **24**(2–3), 103 (1984)
49. P.B. Fellgett, Three concepts make a million points, *Infrared Phys.* **24**(2–3), 95 (1984)

50. H.A. Gebbie, Fourier transform spectroscopy—Recollections of the period 1955–1960, *Infrared Phys.* **24**(2-3), 105 (1984)
51. P. Connes, Early history of Fourier transform Spectroscopy, *Infrared Phys.* **24**, 69 (1984)
52. H. Rubens, H. Hollnagel, Measurements in the extreme infra-red spectrum, *Phil. Mag.* **19**, 761 (1910)
53. W.S. Boyle, K.F. Rogers, Performance characteristics of a new low-temperature bolometer, *J. Opt. Soc. Am.* **49**, 66 (1959)
54. S.J. Fray, J.F.C. Oliver, Photoconductive detector of radiation of wavelength greater than 50 μ , *J. Sci. Instr.* **36**, 195 (1959)
55. P. Guenard, O. Doehler, B. Epszstein, R. Warnecke, New UHF tubes with wide electronic tuning range, *C. R. Acad. Sci. (Paris)* **235**, 235 (1952)
56. R. Kompfner, N.T. Williams, Backward-wave tubes, *Proc. IRE* **41**, 1602 (1953)
57. E. Burstein, G.S. Picus, H.A. Gebbie, Cyclotron resonance at infrared frequencies in InSb at room temperature, *Phys. Rev.* **103**, 825 (1956)
58. W.M. Sinton, Observation of solar and lunar radiation at 1.5 millimeters, *J. Opt. Soc. Am.* **45**, 975 (1955)
59. P.L. Richards, M. Tinkham, Far infrared energy gap measurements in bulk superconducting In, Sn, Hg, Ta, V, Pb and Nb, *Phys. Rev.* **119**, 575 (1960)
60. M.F. Kimmitt, A.C. Prior, V. Roberts, in *Plasma Diagnostic Techniques*, vol. 21, ed. by R.H. Huddleston, S.L. Leonard (Academic Press, New York, 1965), chap. 9 - Far-infrared techniques, pp. 399–430
61. R.Q. Twiss, Radiation transfer and the possibility of negative absorption in radio astronomy, *Austral. J. Phys.* **11**, 564 (1958)
62. A.V. Gaponov, Interaction between electron fluxes and electromagnetic waves in waveguides, *Izv. Vyssh. Uchebn. Zaved. Radiofiz.* **2**, 450 (1959)
63. R.H. Pantell, The design and characteristics of a megawatt space-harmonic traveling-wave tube, *Proc. IRE* **48–53**, 1146 (1959)
64. Y. Ta, Effects des radiations sur les cristaux pyro electriques, *Compt. Rend.* **207**, 1042 (1938)
65. E.H. Putley, Impurity photoconductivity in n-type InSb, *Proc. Phys. Soc.* **76**, 802 (1960)
66. F.J. Low, Low temperature germanium bolometer, *J. Opt. Soc. Am.* **51**, 1300 (1961)
67. M.A.C.S. Brown, M. Kimmitt, Far infrared resonant photoconductivity in indium antimonide, *Infrared Phys.* **5**, 93 (1965)
68. W.J. Moore, H. Shenker, A high-detectivity gallium-doped germanium detector for the 40–120 μ region, *Infrared Phys.* **5**, 99 (1965)
69. A.G. Kazanskii, P.L. Richards, E.E. Haller, Far-infrared photoconductivity of uniaxially stressed germanium, *Appl. Phys. Lett.* **31**, 496 (1977)
70. N.S. Nishioka, P.L. Richards, D. Woody, Composite bolometers for submillimeter wavelengths, *Appl. Opt.* **17**, 1562 (1978)
71. C.C. Grimes, P.L. Richards, S. Shapiro, Josephson-effect far-infrared detector, *J. Appl. Phys.* **39**, 3905 (1968)
72. D.T. Young, J.C. Irvin, Millimeter frequency conversion using Au-n-type GaAs Schottky barrier epitaxial diodes with a novel contacting technique, *Proc. IEEE* **53**, 2130 (1965)
73. H.A. Gebbie, N.W.B. Stone, F.D. Findlay, A stimulated emission source at 0.34 millimetre wave-length, *Nature* **202**, 685 (1964)
74. J.E. Chamberlain, J.E. Gibbs, H.A. Gebbie, Refractometry in the far infra-red using a two-beam interferometer, *Nature* **198**, 874 (1963)
75. E.E. Bell, Measurement of spectral transmittance and reflectance with a far infrared michelson interferometer, *Japan J. Appl. Phys.* **4**, 412 (1964). Supplement I
76. D.H. Martin, E. Puplett, Polarised interferometric spectrometry for the millimeter and submillimetre spectrum, *Infrared Phys.* **10**, 105 (1970)
77. J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Math. Comput.* **19**, 297 (1965)
78. R. Ulrich, Far-infrared properties of metallic mesh and its complementary structure, *Infrared Phys.* **7**, 37 (1967)

79. G.W. Chantry, H.M. Evans, J.W. Fleming, H.A. Gebbie, TPX, a new material for optical components in the far infra-red spectral region, *Infrared Phys.* **9**, 31 (1969)
80. A.M. Nicolson, Broad-band microwave transmission characteristics from a single measurement of the transient response, *IEEE Trans. Instrum. Meas.* **17**, 395 (1968)
81. T.Y. Chang, T.J. Bridges, Laser action at 452, 496 and 541 μm in optically pumped CH_3F , *Opt. Comm.* **1**, 423 (1970)
82. R.W. Wilson, K.B. Jefferts, A.A. Penzias, Carbon monoxide in the Orion nebula, *Astrophys. J.* **161**, L43 (1970)
83. T.G. Phillips, P.J. Huggins, G. Neugebauer, M.W. Werner, Detection of submillimeter (870 μm) CO emission from the Orion molecular cloud, *Astrophys. J.* **217**, L161 (1977)
84. D.H. Barker, D.T. Hodges, T.S. Hartwick, Far infrared imagery, *Proc. SPIE* **67**, 27 (1975)
85. K.H. Yang, P.L. Richards, Y.R. Shen, Generation of far-infrared radiation by picosecond light pulses in LiNbO_3 , *Appl. Phys. Lett.* **19**, 320 (1971)
86. T. Yajima, N. Takeuchi, Far-infrared difference-frequency generation by picosecond laser pulses, *Jpn. J. Appl. Phys.* **9**, 1361 (1970)
87. W.M. Kelly, G.T. Wrixon, *Infrared and Millimeter Waves*, (Academic Press, New York, London, Toronto, Sydney, San Francisco, 1980), Submillimeter Techniques, vol. 3, chap. 2 – Optimization of Schottky-Barrier Diodes for Low-Noise, Low-Conversion Loss Operation at Near-Millimeter Wavelengths, pp. 75–110
88. E.E. Haller, M.R. Hueschen, P.L. Richards, Ge:Ga photoconductors in low infrared backgrounds, *Appl. Phys. Lett.* **34**, 495 (1979)
89. G.J. Dolan, T.G. Phillips, D.P. Woody, Low-noise 115-GHz mixing in superconducting oxide-barrier tunnel junctions, *Appl. Phys. Lett.* **34**, 347 (1979)
90. P.L. Richards, T.M. Shen, R.E. Harris, F. Lloyd, Quasiparticle heterodyne mixing in SIS tunnel junctions, *Appl. Phys. Lett.* **34**, 345 (1979)
91. B. Carli, F. Mencaraglia, A. Bonetti, Submillimeter high-resolution FT spectrometer for atmospheric studies, *Appl. Opt.* **23**, 2594 (1984)
92. D.M. Watson, J.W.V. Storey, Far infrared fine structure lines in the interstellar medium, *Int. J. Infrared Millimeter Waves* **1**(4), 609 (1980)
93. H.R. Fetterman, G.A. Koepf, P.F. Goldsmith, B.J. Clifton, D. Buhl, N.R. Erickson, D.D. Peck, N. Mcavoy, P.E. Tannenwald, Submillimeter heterodyne detection of interstellar carbon monoxide at 434 micrometers, *Science* **211**(4482), 580 (1981)
94. H. Roeser, R. Wattenbach, E. Durwen, G. Schultz, A high resolution heterodyne spectrometer from 100 μm to 1000 μm and the detection of CO ($J = 7 - 6$), CO ($J = 6 - 5$) and (^{13}CO) ($J = 3 - 2$), *Astron. Astrophys.* **165**, 287 (1986)
95. J. Zmuidzinas, A.L. Betz, D.M. Goldhaber, Observations of neutral atomic carbon at 809 GHz, *Astrophys. J. Lett.* **307**, L75 (1986)
96. M.D. Petroff, M.G. Stapelbroek, Blocked impurity band detectors, U.S. Patent 4,568,960, filed 1980 (1986)
97. S. Komiyama, Far-infrared emission from population-inverted hot-carrier system in p-Ge, *Phys. Rev. Lett.* **48**, 271 (1982)
98. A.A. Andronov, I. Zverev, V.A. Kozlov, Y.N. Nozdrin, S.A. Pavlov, V.N. Shastin, Stimulated emission in the long-wavelength IR region from hot holes in Ge in crossed electric and magnetic fields, *Pis'ma Zh. Eksp. Teor. Fiz.* **40**, 69 (1984)
99. E. Bründermann, D.R. Chamberlin, E.E. Haller, High duty cycle and continuous terahertz emission from germanium, *Appl. Phys. Lett.* **76**(21), 2991 (2000)
100. E. Bründermann, H.P. Röser, First operation of a far-infrared p-germanium laser in a standard closed-cycle machine at 15 K, *Infrared Phys. Technol.* **38**(4), 201 (1997)
101. L.R. Elias, J. Hu, G. Ramian, The UCSB electrostatic accelerator free electron laser: First operation, *Nucl. Instr. Meth.* **A237**, 203 (1984)
102. W.D. Duncan, G.P. Williams, Infrared synchrotron radiation from electron storage rings, *Appl. Opt.* **22**, 2914 (1983)

103. M. Abo-Bakr, J. Feikes, K. Holldack, G. Wüstefeld, H.-W. Hübers, Steady-state far-infrared coherent synchrotron radiation detected at BESSY II, *Phys. Rev. Lett.* **88**, 254801 (2002)
104. D.H. Auston, K.P. Cheung, P.R. Smith, Picosecond photoconducting Hertzian dipoles, *Appl. Phys. Lett.* **45**, 284 (1984)
105. E.M. Gershenzon, M.E. Gershenzon, G.N. Goltsman, A.M. Lyulkin, A.D. Semenov, A.V. Sergeev, On the limiting characteristics of high-speed superconducting bolometers, *Sov. Phys. Tech. Phys.* **34**, 195 (1989)
106. T.S. Hartwick, D.H. Barker, D.T. Hodges, B.F. Foote, Far infrared imagery, *Appl. Opt.* **67**, 1919 (1976)
107. B.B. Hu, M.C. Nuss, Imaging with terahertz waves, *Opt. Lett.* **20**(16), 1716 (1995)
108. R. Köhler, A. Tredicucci, F. Beltram, H.E. Beere, E.H. Linfield, A.G. Davies, D.A. Ritchie, R.C. Iotti, F. Rossi, Terahertz semiconductor-heterostructure laser, *Nature* **417**, 156 (2002)
109. D. Abbott, X.C. Zhang (eds.), Special issue on *T-ray imaging, sensing, and retection*, vol. 95 (Proc. IEEE, 2007)
110. K. Fukunaga, Y. Ogawa, S. Hayashi, I. Hosako, Terahertz spectroscopy for art conservation, *IEICE Electron. Express* **4**(8), 258 (2007)
111. E.P. Ippen, C.V. Shank, A. Dienes, Passive mode locking of the cw dye laser, *Appl. Phys. Lett.* **21**(8), 348 (1972)
112. P.F. Moulton, Spectroscopic and laser characteristics of $\text{Ti:Al}_2\text{O}_3$, *J. Opt. Soc. Am. B* **3**(1), 125 (1986)
113. A.S. Müller, I. Birkel, S. Casalbuoni, B. Gasharova, E. Huttel, Y.L. Mathis, D. Moss, N. Smale, P. Wesolowski, E. Bründermann, T. Bückle, M. Klein, in *Proc. European Particle Accelerator Conference* (European Particle Accelerator Conference, 2008), Characterizing THz coherent synchrotron radiation at the ANKA storage ring, WEPC046, pp. 2091–2093
114. V. Blackmore, G. Doucas, C. Perry, B. Ottewell, M.F. Kimmitt, M. Woods, S. Molloy, R. Arnold, First measurements of the longitudinal bunch profile of a 28.5 GeV beam using coherent Smith-Purcell radiation, *Phys. Rev. ST Accel. Beams* **12**(3), 032803 (2009)
115. Y. Neo, H. Shimawaki, T. Matsumoto, H. Mimura, Smith-Purcell radiation from ultraviolet to infrared using a Si field emitter, *J. Vac. Sci. Technol. B* **24**(2), 924 (2006)
116. P.F. Goldsmith, *Quasioptical systems-Gaussian beam quasioptical propagation and applications* (IEEE Press, New York, 1998)
117. A.E. Siegman, *Lasers* (University Science Books, Mill Valley, California, 1986)
118. K. Halbach, Matrix representation of gaussian optics, *Am. J. Phys.* **32**(2), 90 (1964)
119. A. Yariv, *Optical Electronics*, 4th int. edn. (Harcourt Brace Jovanovich College Publishers, Orlando, 1991)
120. J.D. Kraus, *Radio Astronomy*, 2nd edn. (Cygnus-Quasar Books, Powell, 1986)
121. V.G. Veselago, The electrodynamics of substances with simultaneously negative values of ϵ and μ , *Sov. Phys. Usp.* **10**, 509 (1968), translated from V.G. Veselago, *Usp. Fiz. Nauk* **92**, 517 (1967)
122. K.S. Cole, R.H. Cole, Dispersion and absorption in dielectrics, *J. Chem. Phys.* **9**, 341 (1941)
123. D.W. Davidson, R.H. Cole, Dielectric relaxation in glycerol, propylene glycol and n-propanol, *J. Chem. Phys.* **19**, 1484 (1951)
124. H. Zhang, P. Guo, P. Chen, S. Chang, J. Yuan, Liquid-crystal-filled photonic crystal for terahertz switch and filter, *J. Opt. Soc. Am. B* **26**, 101 (2009)
125. R. Wilk, N. Vieweg, O. Kopschinski, T. Hasek, M. Koch, THz spectroscopy of liquid crystals from the CB family, *J. Infrared Milli. Terahz. Waves* **30**(11), 1139 (2009)
126. C.F. Hsieh, R.P. Pan, T.T. Tang, H.L. Chen, C.L. Pan, Voltage-controlled liquid-crystal terahertz phase shifter and quarter-wave plate, *Opt. Lett.* **31**(8), 1112 (2006)
127. C.Y. Chen, C.F. Hsieh, Y.F. Lin, R.P. Pan, C.L. Pan, Magnetically tunable room-temperature 2π liquid crystal terahertz phase shifter, *Opt. Express* **12**(12), 2625 (2004)
128. T.R. Tsai, C.Y. Chen, R.P. Pan, C.L. Pan, X.C. Zhang, Electrically controlled room temperature terahertz phase shifter with liquid crystal, *IEEE Microw. Wireless Compon. Lett.* **14**, 2 (2004)

129. C.Y. Chen, C.L. Pan, C.F. Hsieh, Y.F. Lin, R.P. Pan, Liquid-crystal-based terahertz tunable Lyot filter, *Appl. Phys. Lett.* **88**, 101107 (2006)
130. T. Kleine-Ostmann, K. Pierz, G. Hein, P. Dawson, M. Koch, Audio signal transmission over THz communication channel using semiconductor modulator, *Electron. Lett.* **40**, 124 (2004)
131. J.-ichi Nishizawa, T. Yamada, T. Sasaki, T. Tanabe, T. Wadayama, T. Tanno, K. Suto, Terahertz dichroism of MBBA liquid crystal on rubbed substrate, *Appl. Surf. Sci.* **252**, 4226 (2006)
132. G.G. Stokes, On the composition and resolution of streams of polarized light from different sources, *Trans. Cambridge Phil. Soc.* **9**, 399 (1852)
133. A.J. Gatesman, R.H. Giles, J. Waldman, High-precision reflectometer for submillimeter wavelengths, *J. Opt. Soc. Am. B* **12**, 212 (1995)
134. J. Houghton, J.D. Smith, *Infra-red Physics* (Clarendon Press, United Kingdom, 1966)
135. M. Naftaly, R.E. Miles, in *T-ray imaging, sensing, and detection*, ed. by D. Abbott, X.C. Zhang (2007), Terahertz time-domain spectroscopy for material characterization, *Proc. IEEE* **95**, pp. 1658–1665
136. G.W. Chantry, J.W. Fleming, G.W. F. Pardoe, W. Reddish, H.A. Willis, Absorption spectra of polypropylene in the millimeter and submillimeter regions, *Infrared Phys.* **11**, 109 (1971)
137. J.R. Birch, The far-infrared optical constants of polypropylene, PTFE and polystyrene, *Infrared Phys.* **33**, 33 (1992)
138. D. Larie, I. Booth, M.L.W. Thewalt, B.P. Clayman, Use of polypropylene film for infrared cryostat windows, *Appl. Opt.* **25**, 171 (1986)
139. A. Podzorov, G. Gallot, Low-loss polymers for terahertz applications, *Appl. Opt.* **47**, 3254 (2008)
140. E.V. Loewenstein, D. R. Smith, Optical constants of far-infrared materials. 1: Analysis of channelled spectra and application to Mylar, *Appl. Opt.* **10**, 577 (1971)
141. H.-W. Hübers, J. Schubert, A. Krabbe, M. Birk, G. Wagner, A. Semenov, G. Goltsman, B. Voronov, E. Gershenzon, Parylene anti-reflection coating of a quasi-optical hot-electron bolometric mixer at terahertz frequencies, *Infrared Phys. Technol.* **42**, 41 (2001)
142. J.R. Birch, F.P. Kong, Birefringence and dichroism in fluorogold at near-millimetre wavelengths, *Infrared Phys.* **26**, 131 (1986)
143. A. Blanco, S. Fonti, M. Bancarella, V.D. Cosimo, Polarization properties of some materials at near millimeter wavelengths, *Int. J. Infrared Millimet. Waves* **4**, 751 (1983)
144. S. Sato, S. Hayakawa, T. Matsumoto, H. Matsuo, H. Murakami, K. Sakai, A.E. Lange, P.L. Richards, Submillimeter wave low pass filters made of glass beads, *Appl. Opt.* **28**, 4478 (1989)
145. D.J. Benford, M.C. Gaidis, J.W. Kooi, Optical properties of Zitex in the infrared to submillimeter, *Appl. Opt.* **42**, 5118 (2003)
146. E.V. Loewenstein, D. R. Smith, R.L. Morgan, Optical constants of far infrared materials 2: Crystalline solids, *Appl. Opt.* **12**, 398 (1973)
147. J.R. Birch, C.C. Bradley, M.F. Kimmitt, Absorption and refraction in germanium at 293°K in the range 12–50 cm⁻¹, *Infrared Phys.* **14**, 189 (1974)
148. T.S. Moss, *Optical properties of semiconductors, a semiconductor monograph* (Butterworths Scientific Publications and Academic Press, London and New York, 1959)
149. D. Grischkowsky, S. Keiding, M. van Exter, C. Fattinger, Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors, *J. Opt. Soc. Am. B* **7**, 2006 (1990)
150. K. Kawase, N. Hiromoto, Terahertz-wave antireflection coating on Ge and GaAs with fused quartz, *Appl. Opt.* **37**(10), 1862 (1998)
151. I. Hosako, Antireflection coating formed by plasma-enhanced chemical-vapor deposition for terahertz-frequency germanium optics, *Appl. Opt.* **42**(19), 4045 (2003)
152. A. Hadni, J. Claudel, X. Gerbaux, G. Morlot, J.M. Munier, Sur le comportement différent des cristaux et des verres dans l'absorption de l'infrarouge lointain (40-1500μ) à la température de l'hélium liquide, *Appl. Opt.* **4**, 487 (1965)

153. J.W. Lamb, Miscellaneous data on materials for millimeter and submillimetre optics, *Int. J. Infrared Millimet. Waves* **17**, 1997 (1996)
154. J.R. Birch, G.J. Simonis, M.N. Afsar, R.N. Clarke, J.M. Dutta, H.M. Frost, X. Gerbaux, A. Hadni, W.F. Hall, R. Heidinger, W.W. Ho, C.R. Jones, F. Königer, R.L. Moore, H. Matsuo, T. Nakano, W. Richter, K. Sakai, M.R. Stead, U. Stumper, R.S. Vigil, T.N. Wells, An intercomparison of measurement techniques for the determination of the dielectric properties of solids at near millimeter wavelengths, *IEEE Trans. Microw. Theory Tech.* **42**, 956 (1994)
155. D.E. McCarthy, Black polyethylene as a far-infrared filter, *J. Opt. Soc. Am.* **57**(5), 699-1 (1967)
156. Y. Yamada, A. Mitsuishi, H. Yoshinaga, Transmission filters in the far-infrared region, *J. Opt. Soc. Am.* **52**(1), 17 (1962)
157. T.R. Manley, D.A. Williams, Scattering filters in the far infrared, *Spectrochim. Acta* **21**, 737 (1965)
158. C. Christiansen, Untersuchungen über die optischen Eigenschaften von fein vertheilten Körpern, *Ann. Phys. Chem.* **23**, 298 (1884)
159. K.D. Möller, R.V. McKnight, Measurements on transmission-filter gratings in the far infrared, *J. Opt. Soc. Am.* **55**, 1075 (1965)
160. M.F. Kimmitt, *Far-infrared techniques* (Pion, 1970)
161. K. Sakai, T. Yoshida, Single mesh narrow bandpass filters from the infrared to the submillimeter region, *Infrared Phys.* **18**, 137 (1978)
162. R. Ulrich, Interference filters for the far infrared, *Appl. Opt.* **7**, 1987 (1968)
163. D.W. Porterfield, J.L. Hesler, R. Densing, E.R. Mueller, T.W. Crowe, R.M. Weikle, Resonant metal-mesh bandpass filters for the far infrared, *Appl. Opt.* **33**, 6046 (1994)
164. F. Keilmann, Infrared high-pass filter with high contrast, *Int. J. Infrared Millimeter Waves* **2**, 259 (1981)
165. T. Timusk, P.L. Richards, Near millimeter wave bandpass filters, *Appl. Opt.* **20**, 1355 (1981)
166. P.G. Huggard, G. Schneider, W. Prettl, W. Blau, A simple method of producing far-infrared high-pass filters, *Meas. Sci. Technol.* **2**, 203 (1991)
167. P.G. Huggard, M. Meyringer, A. Schilz, K. Goller, W. Prettl, Far-infrared bandpass filters from perforated metal screens, *Appl. Opt.* **33**, 39 (1994)
168. J. Zmuidzinis, N.G. Ugras, D. Miller, M.C. Gaidis, H.G. LeDuc, Low noise slot antenna SIS mixers, *IEEE Trans. Appl. Superconductivity* **5**, 3053 (1995)
169. J. Lau, J. Fowler, R. Marriage, L. Page, J. Leong, E. Wishnow, R. Henry, E. Wollack, M. Halpern, D. Marsden, G. Marsden, Millimeter-wave antireflection coating for cryogenic silicon lenses, *Appl. Opt.* **45**, 3746 (2006)
170. S. Cherednichenko, V. Drakinskiy, T. Berg, P. Khosropanah, E. Kollberg, Hot-electron bolometer terahertz mixers for the Herschel Space Observatory, *Rev. Sci. Instrum.* **79**, 034501 (2008)
171. C.R. Englert, B. Schimpf, M. Birk, F. Schreier, M. Krocka, R. Nitsche, R. Titz, M. Summers, The 2.5 THz heterodyne spectrometer THOMAS: measurement of OH in the middle atmosphere and comparison with photochemical model results, *J. Geophys. Res. D* **105**, 22211 (2000)
172. J. Xu, J.M. Hensley, D.B. Fenner, R.P. Green, L. Mahler, A. Tredicucci, M.G. Allen, F. Beltram, H.E. Beere, D.A. Ritchie, Tunable terahertz quantum cascade lasers with an external cavity resonator, *Appl. Phys. Lett.* **91**, 121104 (2007)
173. S.B. Cohn, in *Antenna Engineering Handbook* (McGraw-Hill, New York, 1961), chap. 14 - Lens type radiators
174. C. Brückner, T. Käsebier, B. Pradarutti, S. Riehemann, G. Notni, E.B. Kley, A. Tünnermann, Broadband antireflective structures applied to high resistive float zone silicon in the THz spectral range, *Opt. Express* **17**(5), 3063 (2009)
175. J. Ruze, Antenna tolerance theory: A review, *Proc. IEEE* **54**, 633 (1966)
176. I. Anderson, The effect of small phase errors upon transmission between confocal apertures, *Bell Syst. Tech. J.* **54**, 783 (1975)

177. H. Davies, The reflection of electromagnetic waves from a rough surface, *Proc. Inst. Electr. Eng.* **101**, 209 (1954)
178. M. Ortolani, J.S. Lee, U. Schade, H.-W. Hübers, Surface roughness effects on the terahertz reflectance of pure explosive materials, *Appl. Phys. Lett.* **93**, 081906 (2008)
179. J.A. Murphy, Distortion of a simple Gaussian beam on reflection from off-axis ellipsoidal mirrors, *Int. J. Infrared Millimeter Waves* **8**, 1165 (1987)
180. J.A. Murphy, S. Withington, Perturbation analysis of Gaussian-beam-mode scattering at off-axis ellipsoidal mirrors, *Infrared Phys. Technol.* **37**, 205 (1996)
181. R.C. Ohlmann, P.L. Richards, M. Tinkham, Far-infrared transmission through metal light pipes, *J. Opt. Soc. Am.* **48**, 531 (1958)
182. R.E. Harris, R.L. Cappelletti, D.M. Ginsberg, Far infrared transmission through metal light pipes with low thermal conductance, *Appl. Opt.* **5**(6), 1083 (1966)
183. E. Fu, Transmission of submillimeter waves through metal light pipes, *J. Opt. Soc. Am. B* **13**, 702 (1996)
184. E.V. Loewenstein, D.C. Newell, Ray traces through hollow metal light-pipe elements, *J. Opt. Soc. Am.* **59**, 407 (1969)
185. D.G. Hawthorn, T. Timusk, Transmittance of skew rays through metal light pipes, *Appl. Opt.* **38**(13), 2787 (1999)
186. R. Winston, Light collection within the framework of geometrical optics, *J. Opt. Soc. Am.* **60**, 245 (1970)
187. M. Born, E. Wolf, *Principles of optics* (Pergamon Press Ltd., Oxford, 1965)
188. D.A. Harper, R.H. Hildebrand, R. Stiening, R. Winston, Heat trap: an optimized far infrared field optics system, *Appl. Opt.* **15**, 53 (1976)
189. E. Bründermann, M. Havenith, G. Scaleri, M. Giovannini, J. Faist, J. Kunsch, L. Mechold, M. Abraham, Turn-key, compact high temperature terahertz quantum cascade lasers: imaging and room temperature detection, *Opt. Express* **14**(5), 1829 (2006)
190. J. Keene, R.H. Hildebrand, S.E. Whitcomb, R. Winston, Compact infrared heat trap field optics, *Appl. Opt.* **17**, 1107 (1978)
191. D.E. Williamson, Cone channel condenser optics, *J. Opt. Soc. Am.* **42**, 712 (1952)
192. W. Witte, Cone channel optics, *Infrared Phys.* **5**, 179 (1965)
193. H. Lamb, On the reflection and transmission of electric waves by a metallic grating, *Proc. Math. Soc.* **XIX**, 523 (1898)
194. J.P. Casey, E.A. Lewis, Interferometer action of a parallel pair of wire gratings, *J. Opt. Soc. Am.* **42**, 971 (1952)
195. K.F. Renk, L. Genzel, Interference filters and Fabry-Perot interferometers for the far infrared, *Appl. Opt.* **1**, 643 (1962)
196. J.P. Autou, Infrared transmission polarizers by photolithography, *Appl. Opt.* **6**, 1023 (1967)
197. A. Mitsuishi, Y. Yamada, S. Fujita, H. Yoshinaga, Polarizer for the far-infrared region, *J. Opt. Soc. Am.* **50**, 433 (1960)
198. C.C. Homes, G.L. Carr, R.P.S.M. Lobo, J.D. LaVeigne, D.B. Tanner, Si beam splitter for far-infrared and terahertz spectroscopy, *Appl. Opt.* **46**, 7884 (2007)
199. N.L. Rowell, E.A. Wang, Bilayer free-standing beam splitter for Fourier transform infrared spectrometry, *Appl. Opt.* **35**, 2927 (1996)
200. J.A. Dobrolowski, W.A. Traub, New designs for far-infrared beam splitters, *Appl. Opt.* **35**, 2934 (1996)
201. T.O. Klaassen, J.H. Blok, N.J. Hovenier, G. Jakob, D. Rosenthal, K.J. Wildeman, in *IR Space Telescopes and Instruments*, ed. by J.C. Mather (2003), Scattering of sub-millimeter radiation from rough surfaces: absorbers and diffuse reflectors for HIFI and PACS, *Proc. IEEE* **4850**, pp. 788–796
202. H. Hemmati, J.C. Mather, W.L. Eichhorn, Submillimeter and millimeter wave characterization of absorbing materials, *Appl. Opt.* **24**, 4489 (1985)
203. J. Lehman, A. Sanders, L. Hanssen, B. Wilthan, J. Zeng, C. Jensen, Very black infrared detector from vertically aligned carbon nanotubes and electric-field poling of lithium tantalate, *Nano Lett.* **10**, 3261 (2010)

204. D.B. Rutledge, D.P. Neikirk, D.P. Kasilingam, in *Infrared and Millimeter Waves*, K. Button (ed.) (Academic Press, New York, 1983), vol. 10, Integrated-circuit antennas, pp. 1–90
205. G.M. Rebeiz, Millimeter-wave and terahertz integrated circuit antennas, *Proc. IEEE* **80**, 1748 (1992)
206. D.F. Filipovic, S.S. Gearhart, G.M. Rebeiz, Double-slot antennas on extended hemispherical and elliptical Si dielectric lenses, *IEEE Trans. Microw. Theory Tech.* **41**, 1738 (1993)
207. T.H. Büttgenbach, An improved solution for integrated array optics in quasi-optical mm and submm receivers: the hybrid antenna, *IEEE Trans. Microw. Theory Tech.* **41**, 1750 (1993)
208. J.V. Rudd, D.M. Mittleman, Influence of substrate-lens design in terahertz time-domain spectroscopy, *J. Opt. Soc. Am. B* **19**, 319 (2002)
209. D.M. Pozar, *Microwave engineering*, 2nd edn. (John Wiley and Sons, New York, 1998)
210. C. Balanis, *Antenna Theory* (John Wiley and Sons, New York, 1982)
211. S.A. Ramakrishna, Physics of negative refractive index materials, *Rep. Prog. Phys.* **68**, 449 (2005)
212. R. Schiwon, G.W. Schwaab, E. Bründermann, M. Havenith, FIR multilayer mirrors, *Appl. Phys. Lett.* **83**, 4119 (2003)
213. J.D. Joannopoulos, R.D. Meade, J.N. Winn, *Photonic Crystals* (Princeton University Press, Princeton, 1995)
214. H. Kroemer, Nano-whatever: Do we really know where we are heading?, *Phys. Status Solidi A* **202**(6), 957 (2005)
215. J.B. Pendry, Negative refraction makes a perfect lens, *Phys. Rev. Lett.* **85**, 3966 (2000)
216. V.G. Veselago, The electrodynamics of substances with simultaneously negative values of ϵ and μ , *Sov. Phys. Usp.* **10**, 509 (1968), translated from V.G. Veselago, *Usp. Fiz. Nauk* **92**, 517 (1967)
217. A. Schuster, *An Introduction to the Theory of Optics* (Edward Arnold, London, 1904)
218. H. Lamb, On group - velocity, *Proc. Math. Soc.* **1**, 473 (1904)
219. H.C. Poklington, Growth of a wave-group when the group velocity is negative, *Nature* **71**, 607 (1905)
220. P. Vogel, L. Genzel, Transmission and reflection of metallic mesh in the far infrared, *Infrared Phys.* **4**, 257 (1964)
221. H.T. Chen, W.J. Padilla, J.M.O. Zide, S.R. Bank, A.C. Gossard, A.J. Taylor, R.D. Averitt, Ultrafast optical switching of terahertz metamaterials fabricated on ErAs/GaAs nanoisland superlattices, *Opt. Lett.* **32**(12), 1620 (2007)
222. H.T. Chen, H. Lu, A.K. Azad, R.D. Averitt, A.C. Gossard, S.A. Trugman, J.F. O'Hara, A.J. Taylor, Electronic control of extraordinary terahertz transmission through subwavelength metal hole arrays, *Opt. Express* **16**(11), 7641 (2008)
223. O. Paul, C. Imhof, B. Lägler, S. Wolff, J. Heinrich, S. Höfling, A. Forchel, R. Zengerle, R. Beigang, M. Rahm, Polarization-independent active metamaterial for high-frequency terahertz modulation, *Opt. Express* **17**, 819 (2009)
224. J. Han, A. Lakhtakia, Semiconductor split-ring resonators for thermally tunable terahertz metamaterials, *J. Mod. Opt.* **56**(4), 554 (2009)
225. J. Han, A. Lakhtakia, C.W. Qiu, Terahertz metamaterials with semiconductor split-ring resonators for magnetostatic tunability, *Opt. Express* **16**(19), 14390 (2008)
226. J. Brown, Artificial dielectrics having refractive indices less than unity, *Proc. IEEE* **100**(62R), 51 (1953)
227. L. Tonks, I. Langmuir, Oscillations in ionised gases, *Phys. Rev.* **33**, 195 (1929)
228. J.D. Jackson, *Classical electrodynamics*, 2nd edn. (John Wiley and Sons, New York, 1975)
229. J.B. Pendry, A.J. Holden, W.J. Stewart, I. Youngs, Extremely low frequency plasmons in metallic mesostructures, *Phys. Rev. Lett.* **76**, 4773 (1996)
230. D.R. Smith, W.J. Padilla, D.C. Vier, S.C. Nemat-Nasser, S. Schultz, Composite medium with simultaneously negative permeability and permittivity, *Phys. Rev. Lett.* **84**, 4184 (2000)

231. R.A. Shelby, D.R. Smith, S. Schultz, Experimental verification of a negative index of refraction, *Science* **292**, 77 (2001)
232. T.J. Yen, W.J. Padilla, N. Fang, D.C. Vier, D.R. Smith, J.B. Pendry, D.N. Basov, X. Zhang, Terahertz magnetic response from artificial materials, *Science* **303**(5), 1494 (2004)
233. R.W. Boyd, *Radiometry and the Detection of Optical Radiation* (John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1983), chap. 3 - Theory of Blackbody Radiation, pp. 28–50
234. E.K. Plyler, D.J.C. Yates, H.A. Gebbie, Radiant energy from sources in the far infrared, *J. Opt. Soc. Am.* **52**, 859 (1962)
235. M.F. Kimmitt, K. Miller, C.L. Platt, J.E. Walsh, Infrared output from a compact high pressure arc source, *Infrared Phys. Technol.* **37**, 471 (1996)
236. G.W. Chantry, in *Submillimetre Spectroscopy: a guide to the theoretical and experimental physics of the far infrared*, (Academic Press, London and New York, 1971), chap. 5 – Submillimeter Physics, pp. 217–210
237. R.J. Emery, H.A. Gebbie, An improved far infrared continuum source, *Infrared Phys.* **17**, 231 (1977)
238. J.P. Kotthaus, High power output from a submillimeter cw gas laser, *Appl. Opt.* **7**(12), 2422-1 (1968)
239. L.E. Sharp, A.T. Wetherell, High power pulsed HCN laser, *Appl. Opt.* **11**(8), 1737 (1972)
240. Y.X. Jie, X. Gao, Y.F. Cheng, K. Yang, X.D. Tong, Multi-channel FIR HCN laser interferometer on HT-7 Tokamak, *Int. J. Infrared Millimet. Waves* **21**, 1375 (2000)
241. C. Sturzenegger, H. Vetsch, F. Kneubühl, Transversely excited double-discharge HCN laser, *Infrared Phys.* **19**, 277 (1979)
242. G. Dodel, On the history of far-infrared (FIR) gas lasers: Thirty-five years of research and application, *Infrared Phys. Technol.* **40**, 127 (1999)
243. K.J. Button, M. Inguscio, F. Strumia, in *Reviews of infrared and millimetre waves*, ed. by K.J. Button, M. Inguscio, F. Strumia (Plenum Press, New York, 1984), vol. 2 - Optically pumped far infrared lasers
244. P. K. Cheo (ed.), *Handbook of molecular gas lasers* (Marcel Dekker Inc., New York, 1987)
245. N.G. Douglas, *Millimetre and submillimetre wavelength lasers* (Springer, Berlin, Heidelberg, 1989)
246. D.T. Hodges, A review of advances in optically pumped far-infrared lasers, *Infrared Phys.* **18**(5-6), 375 (1978)
247. E.R. Mueller, R. Henschke, W.E. Robotham, L.A. Newman, L.M. Laughman, R.A. Hart, J. Kennedy, Terahertz local oscillator for the Microwave Limb Sounder on the Aura satellite, *Appl. Opt.* **46**, 4907 (2007)
248. A.G. Adam, T.E. Gough, N.R. Isenor, CO₂ laser stabilization using an external cavity locked to a reference HeNe laser, *Rev. Sci. Instrum.* **57**, 6 (1986)
249. A.J. Cantor, P.K. Cheo, M.C. Foster, L.A. Newman, Application of submillimeter wave lasers to high voltage cable inspection, *IEEE J. Quantum Electron.* **17**, 477 (1981)
250. A.J. Beaulieu, Transversely excited atmospheric pressure CO₂ lasers, *Appl. Phys. Lett.* **16**, 504 (1970)
251. P.R. Pearson, H.M. Lamberton, Atmospheric pressure CO₂ lasers giving high output energy per unit volume, *IEEE J. of Quantum Electron.* **8**, 145 (1972)
252. D.E. Evans, L.E. Sharp, B.W. James, W.A. Peebles, Far-infrared superradiant laser action in methyl fluoride, *Appl. Phys. Lett.* **26**, 630 (1975)
253. R. Behn, D. Dicken, J. Hackmann, S.A. Salito, M.R. Siegrist, P.A. Krug, I. Kjelberg, B. Duval, B. Joye, A. Pochelon, Ion temperature measurement of tokamak plasmas by collective Thomson scattering of D₂O laser radiation, *Phys. Rev. Lett.* **62**, 2833 (1989)
254. J. Wiggins, Z. Drozdowicz, R. Temkin, Two-photon transitions in optically pumped submillimeter lasers, *IEEE J. Quantum Electron.* **14**, 23 (1978)
255. A.J. Alcock, K. Leopold, M.C. Richardson, Continuously tunable high-pressure CO₂ laser with uv photopreionization, *Appl. Phys. Lett.* **23**, 562 (1973)

256. D.G. Biron, R.J. Temkin, B. Lax, B.G. Danly, High-intensity CO₂ laser pumping of a CH₃F Raman FIR laser, *Opt. Lett.* **4**, 381 (1979)
257. P. Mathieu, J.R. Izatt, Continuously tunable CH₃F raman far-infrared laser, *Opt. Lett.* **6**, 369 (1981)
258. U.P. Schießl, J. John, P.J. McCann, in *Long-wavelength Infrared Semiconductor Lasers* (John Wiley and Sons, New York, 2004), chap. 4 - Lead-Chalcogenide-based Mid-Infrared Diode Lasers, pp. 145–216
259. H. Krömer, Proposed negative mass microwave amplifier, *Phys. Rev.* **109**, 1856 (1958)
260. W. Shockley, Hot electrons in germanium and Ohm's law, *Bell Syst. Tech. J.* **30**, 990 (1951)
261. S. Komiyama, S. Kuroda, I. Hosako, Y. Akasaka, N. Iizuka, Germanium lasers in the range from far-infrared to millimetre waves, *Optical & Quantum Electron.* **23**(2), S133 (1991)
262. I. Hosako, N. Hiromoto, in *Proc. 7th Int. Conf. Terahertz Electronics* (IEEE, 25–26 November 1999, Nara, Japan, 1999), p-type germanium sub-terahertz maser oscillation in the Voigt configuration, pp. 193–194
263. H. Maeda, T. Kurosawa, Hot electron population inversion in crossed electric and magnetic fields, *J. Phys. Soc. Jpn.* **33**, 562 (1972)
264. A.A. Andronov, V.A. Kozlov, L.S. Mazov, V.N. Shastin, Amplification of far IR radiation in Ge on 'hot' hole population inversion, *Pis'ma Zh. Eksp. Teor. Fiz.* **30**, 585 (1979)
265. S. Komiyama, R. Spies, Hot-carrier population inversion in p-Ge, *Phys. Rev. B* **23**, 6839 (1981)
266. L.E. Vorob'ev, F.I. Osokin, V.I. Stafeev, V.N. Tulupenko, Discovery of hot hole population inversion in Ge, *Pis'ma Zh. Eksp. Teor. Fiz.* **34**(81), 125 (1981)
267. S. Komiyama, T. Masumi, K. Kajita, Streaming motion and population inversion of hot electrons in silver halides at crossed electric and magnetic fields, *Phys. Rev. B* **20**, 5192 (1979)
268. E. Bründermann, E.E. Haller, A.V. Muravjov, Terahertz emission of population-inverted hot-holes in single-crystalline silicon, *Appl. Phys. Lett.* **73**, 723 (1998)
269. Y.L. Ivanov, Generation of cyclotron resonance radiation by light holes in germanium, *Opt. Quantum Electron.* **23**, S253 (1991)
270. L.A. Reichertz, O.D. Dubon, G. Sirmain, E. Bründermann, W.L. Hansen, D.R. Chamberlin, A.M. Linhart, H.P. Röser, E.E. Haller, Stimulated far-infrared emission from combined cyclotron resonances in germanium, *Phys. Rev. B* **56**(19), 12069 (1997)
271. E. Bründermann, in *Long-wavelength Infrared Semiconductor Lasers* (John Wiley and Sons, New York, 2004), chap. 6 - Widely Tunable Far Infrared Hot Hole Semiconductor Lasers, pp. 279–350
272. E. Bründermann, D.R. Chamberlin, E.E. Haller, Novel design concepts of widely tunable germanium terahertz lasers, *Infrared Phys. Technol.* **40**(3), 141 (1999)
273. A. Andronov, V. Kozlov, S. Pavlov, S. Pavlov, Bragg selection in hot hole FIR lasers, *Opt. Quantum Electron.* **23**, S205 (1991)
274. D.R. Chamberlin, E. Bründermann, E.E. Haller, Planar contact geometry for far-infrared germanium lasers, *Appl. Phys. Lett.* **74**(25), 3761 (1999)
275. S. Komiyama, S. Kuroda, Far-infrared laser oscillation in p-Ge using external reflectors, *Jpn. J. Appl. Phys.* **26**, L71 (1987)
276. K. Unterrainer, M. Helm, E. Gornik, E.E. Haller, J. Leotin, Mode structure of the p-germanium far-infrared laser with and without external mirrors: single line operation, *Appl. Phys. Lett.* **52**, 564 (1988)
277. S. Komiyama, H. Morita, I. Hosako, Continuous wavelength tuning of inter-valence-band laser oscillation in p-type germanium over range of 80–120 μm , *Jpn. J. Appl. Phys.* **32**, 4987 (1993)
278. E. Bründermann, H.P. Röser, A.V. Muravjov, S.G. Pavlov, V.N. Shastin, Mode fine structure of the FIR p-Ge intervalenceband laser measured by heterodyne mixing spectroscopy with an optically pumped ring gas laser, *Infrared Phys. Technol.* **36**(1), 59 (1995)
279. A.V. Muravjov, S.H. Withers, H. Weidner, R.C. Strijbos, S.G. Pavlov, V.N. Shastin, R.E. Peale, Single axial-mode selection in a far-infrared p-Ge laser, *Appl. Phys. Lett.* **76**, 1996 (2000)

280. K. Park, R.E. Peale, H. Weidner, J.J. Kim, Bulk semiconductor lasers at submillimeter/far infrared wavelengths using a regular permanent magnet, *IEEE J. of Quantum Electron.* **32**, 1203 (1996)
281. F. Keilmann, H. Zuckermann, Transient gain of the germanium hot hole laser, *Opt. Communications* **109**, 296 (1994)
282. A.V. Muravjov, H. Saxena, R.E. Peale, C.J. Fredricksen, O. Edwards, V.N. Shastin, Injection-seeded internal-reflection-mode p-Ge laser exceeds 10 W peak terahertz power, *J. Appl. Phys.* **103**, 083112 (2008)
283. J.N. Hovenier, A.V. Muravjov, S.G. Pavlov, V.N. Shastin, R.C. Stribos, W.T. Wenckebach, Active mode locking of a p-Ge hot hole laser, *Appl. Phys. Lett.* **71**, 443 (1997)
284. A.V. Muravjov, S.H. Withers, R.C. Stribos, S.G. Pavlov, V.N. Shastin, R.E. Peale, Actively mode-locked p-Ge laser in Faraday configuration, *Appl. Phys. Lett.* **75**, 2882 (1999)
285. J.N. Hovenier, R.M. de Kleijn, T.O. Klaassen, W.T. Wenckebach, D.R. Chamberlin, E. Bründermann, E.E. Haller, Mode-locked operation of the copper-doped germanium terahertz laser, *Appl. Phys. Lett.* **77**(20), 3155 (2000)
286. S. Ebbinghaus, S.J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D.M. Leitner, M. Havenith, An extended dynamical solvation shell around proteins, *Proceedings of The National Academy of Sciences of the USA* **104**(52), 20749 (2007)
287. E. Bründermann, A.M. Linhart, L. Reichertz, H.P. Röser, O.D. Dubon, W.L. Hansen, G. Sirmain, E.E. Haller, Double acceptor doped Ge: a new medium for inter-valence-band lasers., *Appl. Phys. Lett.* **68**(22), 3075 (1996)
288. G. Sirmain, L.A. Reichertz, O.D. Dubon, E.E. Haller, W.L. Hansen, E. Bründermann, A.M. Linhart, H.P. Röser, Stimulated far-infrared emission from copper-doped germanium crystals, *Appl. Phys. Lett.* **70**(13), 1659 (1997)
289. N. Hiromoto, I. Hosako, M. Fujiwara, Far-infrared laser oscillation from a very small p-Ge crystal under uniaxial stress, *Appl. Phys. Lett.* **74**, 3432 (1999)
290. R.C. Stribos, J.G.S. Lok, W.T. Wenckebach, A Monte Carlo simulation of mode-locked hot-hole laser operation, *J. Phys.: Condens. Matter* **6**, 7461 (1994)
291. A.V. Muravjov, R.C. Stribos, C.J. Fredricksen, H. Weidner, W. Trimble, S.H. Withers, S.G. Pavlov, V.N. Shastin, R.E. Peale, Evidence for self-mode-locking in p-Ge laser emission, *Appl. Phys. Lett.* **73**, 3037 (1998)
292. S.V. Demihovsky, A.V. Murav'ev, S.G. Pavlov, V.N. Shastin, Stimulated emission using the transitions of shallow acceptor states in germanium, *Semicond. Sci. Technol.* **7**, B622 (1992)
293. E.E. Orlova, V.N. Shastin, in *Proc. 21th Int. Conf. on Infrared and Millimeter Waves*, eds. M. von Ortenberg, H.-U. Mueller (Humboldt-Universität zu Berlin, 1996), p. CTh4, ISBN 3-00-000800-4
294. S.G. Pavlov, R.K. Zhukavin, E.E. Orlova, V.N. Shastin, A.V. Kirsanov, H.-W. Hübers, K. Auen, H. Riemann, Stimulated emission from donor transitions in silicon, *Phys. Rev. Lett.* **84**, 5220 (2000)
295. H.-W. Hübers, S.G. Pavlov, V.N. Shastin, Terahertz lasers based on germanium and silicon, *Semicond. Sci. Technol.* **20**, 211 (2005)
296. S.G. Pavlov, R. Eichholz, N.V. Abrosimov, B. Redlich, H.-W. Hübers, Multifrequency terahertz lasing from codoped silicon crystals, *Appl. Phys. Lett.* **98**, 061102 (2011)
297. S.G. Pavlov, H.-W. Hübers, J.N. Hovenier, T.O. Klaassen, D.A. Carder, P.J. Phillips, B. Redlich, H. Riemann, R. Zhukavin, V.N. Shastin, Stimulated terahertz stokes emission of silicon crystals doped with antimony donors, *Phys. Rev. Lett.* **96**, 037404 (2006)
298. S.A. Lynch, R. Bates, D.J. Paul, D.J. Norris, A.G. Cullis, Z. Ikonic, R.W. Kelsall, P. Harrison, D.D. Arnone, C.R. Pidgeon, Intersubband electroluminescence from Si/SiGe cascade emitters at terahertz frequencies, *Appl. Phys. Lett.* **81**, 1543 (2002)
299. A. Borak, Toward bridging the terahertz gap with silicon-based lasers, *Science* **308**, 638 (2005)
300. R.F. Kazarinov, R.A. Suris, Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice, *Sov. Phys. Semicond.* **5**, 707 (1971)

301. M. Helm, P. England, E. Colas, F. DeRosa, S. Allen, Intersubband emission from semiconductor superlattices excited by sequential resonant tunneling, *Phys. Rev. Lett.* **63**, 74 (1989)
302. J. Faist, F. Capasso, D.L. Sivco, C. Sirtorio, A.L. Hutchinson, A.Y. Cho, Quantum cascade laser, *Science* **264**, 553 (1994)
303. J. Faist, C. Sirtori, in *Long-wavelength infrared semiconductor lasers* (John Wiley and Sons, New York, 2004), chap. 4 - InP and GaAs-based quantum cascade lasers, pp. 217–278
304. A. Tredicucci, R. Köhler, in *Intersubband Transitions in Quantum Structures* (McGraw-Hill, New York, 2006), chap. 2 - Terahertz quantum cascade lasers, pp. 45–105
305. B. Williams, Terahertz quantum cascade lasers, *Nature Photonics* **1**, 517 (2007)
306. G. Scalari, L. Ajili, J. Faist, H. Beere, E. Linfield, D. Ritchie, G. Davies, Far-infrared $\lambda \sim 87 \mu\text{m}$ bound-to-continuum quantum-cascade lasers operating up to 90 K, *Appl. Phys. Lett.* **82**, 3165 (2003)
307. B.S. Williams, H. Callebaut, S. Kumar, Q. Hu, J.L. Reno, 3.4-THz quantum cascade laser based on longitudinal-optical-phonon scattering for depopulation, *Appl. Phys. Lett.* **82**, 1015 (2003)
308. G. Fasching, V. Tamosiunas, A. Benz, A.M. Andrews, K. Unterrainer, R. Zobl, T. Roch, W. Schrenk, G. Strasser, Subwavelength microdisk and microring terahertz quantum-cascade lasers, *IEEE J. of Quantum Electron.* **43**, 687 (2007)
309. L.A. Dunbar, R. Houdr, G. Scalari, L. Sirigu, M. Giovannini, J. Faist, Small optical volume terahertz emitting microdisk quantum cascade lasers, *Appl. Phys. Lett.* **90**, 141114 (2007)
310. Y. Chassagneux, R. Colombelli, W. Maineult, S. Barbieri, H.E. Beere, D.A. Ritchie, S.P. Khanna, E.H. Linfield, A.G. Davies, Electrically pumped photonic-crystal terahertz lasers controlled by boundary conditions, *Nature* **457**, 174 (2009)
311. E. Mujagic, C. Deutsch, H. Detz, P. Klang, M. Nobile, A.M. Andrews, W. Schrenk, K. Unterrainer, G. Strasser, Vertically emitting terahertz quantum cascade ring lasers, *Appl. Phys. Lett.* **95**, 011120 (2009)
312. L. Mahler, R. Köhler, A. Tredicucci, F. Beltram, H.E. Beere, E.H. Linfield, D.A. Ritchie, A.G. Davies, Single-mode operation of terahertz quantum cascade lasers with distributed feedback resonators, *Appl. Phys. Lett.* **84**, 5546 (2004)
313. H.-W. Hübers, S.G. Pavlov, A.D. Semenov, R. Köhler, L. Mahler, A. Tredicucci, H.E. Beere, D.A. Ritchie, E.H. Linfield, Terahertz quantum cascade laser as local oscillator in a heterodyne receiver, *Opt. Express* **13**, 5890 (2005)
314. H. Richter, A.D. Semenov, S. Pavlov, L. Mahler, A. Tredicucci, K. Ilin, M. Siegel, H.-W. Hübers, Terahertz heterodyne receiver with quantum cascade laser and hot electron bolometer mixer in a pulse tube cooler, *Appl. Phys. Lett.* **93**, 141108 (2008)
315. R. Köhler, A. Tredicucci, F. Beltram, H.E. Beere, E.H. Linfield, A.G. Davies, D.A. Ritchie, S.S. Dhillon, C. Sirtori, High-performance continuous-wave operation of superlattice terahertz quantum-cascade lasers, *Appl. Phys. Lett.* **82**(10), 1518 (2003)
316. H.-W. Hübers, S.G. Pavlov, H. Richter, A.D. Semenov, L. Mahler, A. Tredicucci, H.E. Beere, D.A. Ritchie, Molecular spectroscopy with terahertz quantum cascade lasers, *J. Nanoelectronics Optoelectronics* **2**, 101 (2007)
317. A.J.L. Adam, I. Kašalynas, J.N. Hovenier, T.O. Klaassen, J.R. Gao, E.E. Orlova, B.S. Williams, S. Kumar, Q. Hu, J.L. Reno, Beam patterns of terahertz quantum cascade lasers with subwavelength cavity dimensions, *Appl. Phys. Lett.* **88**, 151105 (2006)
318. E.E. Orlova, J.N. Hovenier, T. Klaassen, I. Kašalynas, A.J.L. Adam, J.R. Gao, T. M. Klapwijk, B. S. Williams, S. Kumar, Q. Hu, J.L. Reno, Antenna model for wire lasers, *Phys. Rev. Lett.* **96**, 173904 (2006)
319. M.L. Wei, Q.Q. Alan, S. Kumar, B.S. Williams, Q. Hu, J.L. Reno, High-power and high-temperature THz quantum-cascade lasers based on lens-coupled metal-metal waveguides, *Opt. Lett.* **32**, 2840 (2007)
320. M.I. Amanti, M. Fisher, C. Walther, G. Scalari, J. Faist, Horn antennas for terahertz quantum cascade lasers, *Electron. Lett.* **43**(10), 573 (2007)

321. S. Kumar, B.S. Williams, Q. Qin, A.W. Lee, Q. Hu, J.L. Reno, Surface-emitting distributed feedback terahertz quantum-cascade lasers in metal-metal waveguides, *Opt. Express* **15**, 113 (2007)
322. A. Barkan, F.K. Tittel, D.M. Mittleman, R. Dengler, P.H. Siegel, G. Scalari, L. Ajili, J. Faist, H.E. Beere, E.H. Linfield, A.G. Davies, D.A. Ritchie, Linewidth and tuning characteristics of terahertz quantum cascade lasers, *Opt. Lett.* **295**, 575 (2004)
323. A. Baryshev, J.N. Hovenier, A.J.L. Adam, I. Kaalynas, J.R. Gao, T.O. Klaassen, B.S. Williams, S. Kumar, Q. Hu, J. Reno, Phase-locking and spectral linewidth of a two-mode terahertz quantum cascade laser, *Appl. Phys. Lett.* **89**, 031115 (2006)
324. A.L. Betz, R.T. Boreiko, B.S. Williams, S. Kumar, Q. Hu, J.L. Reno, Frequency and phase-lock control of a 3 THz quantum cascade laser, *Opt. Lett.* **30**, 1837 (2005)
325. D. Rabanus, U.U. Graf, M. Philipp, O. Ricken, J. Stutzki, B. Vowinkel, M. C. Wiedner, C. Walther, M. Fischer, J. Faist, Phase locking of a 1.5 terahertz quantum cascade laser and use as a local oscillator in a heterodyne HEB receiver, *Opt. Express* **17**, 1159 (2009)
326. H. Richter, S.G. Pavlov, A.D. Semenov, L. Mahler, A. Tredicucci, H.E. Beere, D.A. Ritchie, H.-W. Hübers, Sub-megahertz frequency stabilization of a terahertz quantum cascade laser to a molecular absorption line, *Appl. Phys. Lett.* **96**, 071112 (2010)
327. S. Barbieri, P. Gellie, G. Santarelli, L. Ding, W. Maineult, C. Sirtori, R. Colombelli, H. Beere, D. Ritchie, Phase-locking of a 2.7-THz quantum cascade laser to a mode-locked erbium-doped fibre laser, *Nature Photonics* **4**, 636 (2010)
328. L. Ajili, G. Scalari, D. Hofstetter, M. Beck, J. Faist, H. Beere, G. Davies, E. Linfield, D. Ritchie, Continuous-wave operation of far-infrared quantum cascade lasers, *Electron. Lett.* **38**, 1675 (2002)
329. Q. Qin, B.S. Williams, S. Kumar, J.L. Reno, Q. Hu, Tuning a terahertz wire laser, *Nature Photonics* **3**, 732 (2009)
330. A.W.M. Lee, Q. Qin, S. Kumar, B.S. Williams, Q. Hu, J.L. Reno, Real-time terahertz imaging over a standoff distance (>25 meters), *Appl. Phys. Lett.* **89**, 141125 (2006)
331. H.-W. Hübers, S.G. Pavlov, H. Richter, A.D. Semenov, L. Mahler, A. Tredicucci, H.E. Beere, D.A. Ritchie, High resolution gas phase spectroscopy with a distributed feedback terahertz quantum cascade laser, *Appl. Phys. Lett.* **89**, 061115 (2006)
332. J.R. Gao, J.N. Hovenier, Z.Q. Yang, J.J.A. Baselmans, A. Baryshev, M. Hajenius, T.M. Klapwijk, A.J.L. Adam, T.O. Klaassen, B.S. Williams, S. Kumar, Q. Hu, J.L. Reno, Terahertz heterodyne receiver based on a quantum cascade laser and a superconducting bolometer, *Appl. Phys. Lett.* **86**, 244104 (2005)
333. H. Richter, M. Greiner-Bär, S.G. Pavlov, A.D. Semenov, M. Wienold, L. Schrottke, M. Giehler, R. Hey, H. T. Grahn, H.-W. Hübers, A compact, continuous-wave terahertz source based on a quantum-cascade laser and a miniature cryocooler, *Opt. Express* **18**, 10177 (2010)
334. M.C. Wanke, E.W. Young, C.D. Nordquist, M.J. Cich, A.D. Grine, C.T. Fuller, J.L. Reno, M. Lee, Monolithically integrated solid-state terahertz transceivers, *Nature Photonics* **4**, 565 (2010)
335. E.R. Brown, F.W. Smith, K.A. McIntosh, Coherent millimeter-wave generation by heterodyne conversion in low-temperature-grown GaAs photoconductors, *J. Appl. Phys.* **73**, 1480 (1993)
336. K.A. McIntosh, K.B. Nichols, S. Verghese, E.R. Brown, Investigation of ultrashort photocarrier relaxation times in low-temperature-grown GaAs, *Appl. Phys. Lett.* **70**, 354 (1997)
337. W. Shi, Y.J. Dinga, A monochromatic and high-power terahertz source tunable in the ranges of 2.7–38.4 and 58.2–3540 μm for variety of potential applications, *Appl. Phys. Lett.* **84**(10), 1635 (2004)
338. S. Matsuura, M. Tani, K. Sakai, Generation of coherent terahertz radiation by photomixing in dipole photoconductive antennas, *Appl. Phys. Lett.* **70**(5), 559 (1997)
339. M. Tani, P. Gu, M. Hyodo, K. Sakai, T. Hidaka, Generation of coherent terahertz radiation by photomixing of dual-mode lasers, *Opt. Quantum Electron.* **32**, 503 (2000)
340. T. Kleine-Ostmann, P. Knobloch, M. Koch, S. Hoffmann, M. Breede, M. Hofmann, G. Hein, K. Pierz, M. Sperling, K. Donhuijsen, Continuous-wave THz imaging, *Electron. Lett.* **37**, 1461 (2001)

341. S. Hoffmann, M. Hofmann, E. Bründermann, M. Havenith, M. Matus, J.V. Moloney, A.S. Moskalenko, M. Kira, S.W. Koch, S. Saito, K. Sakai, Four-wave mixing and direct terahertz emission with two-color semiconductor lasers, *Appl. Phys. Lett.* **84**, 3585 (2004)
342. M. Scheller, J.M. Yarbrough, J.V. Moloney, M. Fallahi, M. Koch, S.W. Koch, Room temperature continuous wave milliwatt terahertz source, *Opt. Express* **18**, 27112 (2010)
343. M.A. Belkin, F. Capasso, A. Belyanin, D.L. Sivco, A.Y. Cho, D.C. Oakley, C.J. Vineis, G.W. Turner, Terahertz quantum-cascade-laser source based on intracavity difference-frequency generation, *Nature Photonics* **1**, 288 (2007)
344. M.A. Belkin, F. Capasso, F. Xie, A. Belyanin, M. Fischer, A. Wittmann, J. Faist, Room temperature terahertz quantum cascade laser source based on intracavity difference-frequency generation, *Appl. Phys. Lett.* **92**, 201101 (2008)
345. S.M. Duffy, S. Verghese, K.A. McIntosh, A. Jackson, A.C. Gossard, S. Matsuura, Accurate modeling of dual dipole and slot elements used with photomixers for coherent terahertz output power, *IEEE Trans. Microw. Theory Tech.* **49**(6), 1032 (2001)
346. B. Sartorius, M. Schlak, D. Stanze, H. Roehle, H. Künzel, D. Schmidt, H.-G. Bach, R. Kunkel, M. Schell, Continuous wave terahertz systems exploiting 1.5 μm telecom technologies, *Opt. Express* **17**, 15001 (2009)
347. D.D. Bicanic, B.F.J. Zuidberg, A. Dymanus, Generation of continuously tunable laser sidebands in the submillimeter region, *Appl. Phys. Lett.* **32**, 367 (1978)
348. W.A.M. Blumberg, H.R. Fetterman, D.D. Peck, P.F. Goldsmith, Tunable submillimeter sources applied to the excited state rotational spectroscopy and kinetics of CH_3F , *Appl. Phys. Lett.* **35**, 582 (1979)
349. J. Farhoomand, G.A. Blake, M.A. Frerking, H.M. Pickett, Generation of tunable laser sidebands in the far-infrared region, *J. Appl. Phys.* **57**, 1763 (1984)
350. P. Verhoeve, E. Zwart, M. Versluis, M. Drabbels, J.J. ter Meulen, W.L. Meerts, D.B. McLay, A far infrared laser sideband spectrometer in the frequency region 550-2700 GHz, *Rev. Sci. Instrum.* **61**, 1612 (1990)
351. G.A. Blake, K.B. Laughlin, R.C. Cohen, K.L. Busarow, D.H. Gwo, C.A. Schmittenmaer, D.W. Steyert, R. Saykally, The Berkeley tunable far infrared spectrometers, *Rev. Sci. Instrum.* **62**, 1701 (1991)
352. E.R. Mueller, J. Waldman, Power and spatial mode measurements of sideband generated, spatially filtered, submillimeter radiation, *IEEE Trans. Microw. Theory Tech.* **42**, 1891 (1994)
353. D.S. Kurtz, J.L. Hesler, T.W. Crowe, R.M. Weikle, Submillimeter-wave sideband generation using varactor Schottky diodes, *IEEE Trans. Microw. Theory Tech.* **50**(11), 2610 (2002)
354. R.L. Aggarwal, B. Lax, H.R. Fetterman, P.E. Tannenwald, B.J. Clifton, CW generation of tunable narrow-band far-infrared radiation, *J. Appl. Phys.* **45**, 3972 (1974)
355. K.M. Evenson, D.A. Jennings, F. R. Petersen, Tunable far-infrared spectroscopy, *Appl. Phys. Lett.* **44**, 576 (1984)
356. H.-W. Hübers, G.W. Schwaab, H.P. Röser, Video detection and mixing performance of GaAs Schottky-barrier diodes at 30 THz and comparison with metal-insulator-metal diodes, *J. Appl. Phys.* **75**, 4243 (1994)
357. L.M. Matarese, K.M. Evenson, Improved coupling to infrared whisker diodes by use of antenna theory, *Appl. Phys. Lett.* **17**, 8 (1970)
358. H. Odashima, L.R. Zink, K.M. Evenson, Tunable far-infrared spectroscopy extended to 9.1 THz, *Opt. Lett.* **24**, 406 (1999)
359. C. Fumeaux, W. Herrmann, F.K. Kneubühl, H. Rothuizen, Nanometer thin film Ni-NiO-Ni diodes for detection and mixing of 30 THz radiation, *Infrared Phys. Technol.* **39**, 123 (1998)
360. A. Stöhr, A. Malloci, A. Sauerwald, I.C. Mayorga, R. Güsten, D.S. Jäger, Ultra-wide-band traveling-wave photodetectors for photonic local oscillators, *IEEE/OSA J. of Lightwave Technol.* **21**(12), 3062 (2003)
361. H. Ito, F. Nakajima, T. Furuta, T. Ishibashi, Continuous THz-wave generation using antenna-integrated uni-travelling-carrier photodiodes, *Semicond. Sci. Technol.* **20**, S191 (2005)

362. A. Madjar, N. Koka, J. Bloch, P. Yu, A. Stöhr, D. Jäger, A novel analytical model as a design tool for uni-traveling-carrier traveling wave photo detectors approaching 1 THz, *IEEE Trans. Microw. Theory Tech.* **57**(1), 223 (2009)
363. I. Mehdi, in *Proc. Conf. on Millimeter and Submillimeter Detectors for Astronomy II* (SPIE, Glasgow, 2004), THz local oscillator technology, *Proc. SPIE* **5498**, pp. 103–112
364. S.A. Maas, *Nonlinear microwave circuits* (IEEE Press, New York, 1997)
365. G. Chattopadhyay, E. Schlecht, J.S. Ward, J.J. Gill, H.H.S. Javadi, F. Maiwald, I. Mehdi, An all-solid-state broad-band frequency multiplier chain at 1500GHz, *IEEE Trans. Microw. Theory Tech.* **52**, 1538 (2004)
366. A. Maestrini, J. Ward, J. Gill, H. Javadi, E. Schlecht, G. Chattopadhyay, F. Maiwald, N. Erickson, I. Mehdi, A 1.7-1.9 THz local oscillator source, *IEEE Microw. and Wireless Components Lett.* **14**, 253 (2004)
367. Y. Fu, J. Stake, L. Dillner, M. Willander, E.L. Kollberg, AlGaAs/GaAs and InAlAs/InGaAs heterostructure barrier varactors, *J. Appl. Phys.* **82**, 5568 (1997)
368. J. Stake, T. Bryllert, T.A. Emadi, J. Vukusic, in *Terahertz Frequency Detection and Identification of Materials and Objects* (Springer, 2007), chap. 1 - Terahertz generation by multiplication, pp. 17–30
369. R. Scheurer, M. Haeussler, K.F. Renk, E. Schomburg, Y.I. Koschurinov, D.G. Pavelev, N. Mleev, V. Ustinov, A. Zhukov, Frequency multiplication of microwave radiation by propagating space-charge domains in a semiconductor superlattice, *Appl. Phys. Lett.* **82**, 2826 (2003)
370. S. Winnerl, E. Schomburg, S. Brandl, K.F. Renk, M.C. Wanke, S.J. Allen, A.A. Ignatov, V. Ustinov, A. Zhukov, P.S. Kopeev, Frequency doubling and tripling of terahertz radiation in a GaAs/AlAs superlattice due to frequency modulation of bloch oscillations, *Appl. Phys. Lett.* **77**, 1259 (2000)
371. C.P. Endress, F. Lewen, T.F. Giesen, S. Schlemmer, D.G. Pavelev, Y.I. Koschurinov, V.M. Ustinov, A.E. Zhucov, Application of superlattice multipliers for high-resolution terahertz spectroscopy, *Rev. Sci. Instrum.* **78**, 043106 (2007)
372. G. Kozlov, A. Volkov, in *Millimeter and submillimeter wave spectroscopy of solids* (Springer-Verlag, Berlin, Heidelberg, 1998), chap. 3 - Coherent source submillimeter wave spectroscopy, pp. 51–109
373. M. Philipp, U.U. Graf, A. Wagner-Gentner, D. Rabanus, F. Lewen, Compact 1.9 THz BWO local-oscillator for the GREAT heterodyne receiver, *Infrared Phys. Technol.* **51**(1), 54 (2007)
374. S.J. Smith, E.M. Purcell, Visible light from localised surface charges moving across a grating, *Phys. Rev.* **92**, 10693 (1953)
375. F.S. Rusin, G.D. Bogomolov, Orotron – an electronic oscillator with an open resonator and reflecting grating, *Proc. IEEE* **57**, 720 (1969)
376. K. Mizuno and S. Ono, *Infrared and Millimeter Waves* (Academic Press, New York, San Francisco, London, 1979), Sources of Radiation, Vol. 1, chap. 5 - The Ledatron, pp. 213–233
377. D.E. Wortman, R.P. Leavitt, H. Dropkin, Improved orotron performance in the 50- to 75-GHz frequency region, *IEEE Trans. Electron. Dev.* **29**, 1639 (1982)
378. V.L. Bratman, B.S. Dumesh, A.E. Fedotov, Y.A. Grishin, F.S. Rusin, Broadband orotron operation at millimeter and submillimeter waves, *Int. J. Infrared Millimeter Waves* **23**, 1595 (2002)
379. J. Urata, M. Goldstein, M.F. Kimmitt, A. Naumov, C. Platt, J.E. Walsh, Superradiant Smith-Purcell emission, *Phys. Rev. Lett.* **80**, 516 (1998)
380. H.L. Andrews, C.H. Boulware, C.A. Brau, J.D. Jarvis, Dispersion and attenuation in a Smith-Purcell free electron laser, *Phys. Rev. ST Accel. Beams* **8**(5), 050703 (2005)
381. H.L. Andrews, C.A. Brau, J.D. Jarvis, C.F. Guertin, A. O'Donnell, B. Durant, T.H. Lowell, M.R. Mross, Observation of THz evanescent waves in a Smith-Purcell free-electron laser, *Phys. Rev. ST Accel. Beams* **12**(8), 080703 (2009)
382. M. Mross, T. Lowell, R. Durant, M. Kimmitt, Performance characteristics of a Smith-Purcell tunable terahertz source, *J. Biol. Phys.* **29**, 295 (2003)

383. V.L. Bratman, M. Yu. Glyavin, Yu. K. Kalynov, A. G. Litvak, A. G. Luchinin, A. V. Savilov and V. E. Zapevalov, Terahertz Gyrotrons at IAP RAS: Status and New Designs, *J. Infrared Milli. Terahz. Waves* **32**, 371–379 (2011)
384. V.L. Bratman, A.E. Fedotov, P.B. Makhalov, Experimental demonstration of Smith-Purcell radiation enhancement by frequency multiplication in open cavity, *Appl. Phys. Lett.* **98**, 061503 (2011)
385. M.K. Hornstein, V.S. Bajaj, R.G. Griffin, R.J. Temkin, Continuous-wave operation of a 460-GHz second harmonic gyrotron oscillator, *IEEE Trans. Plasma Science* **34**, 524 (2006)
386. N.I. Zaytsev, T.B. Pankratova, M.I. Petelin, V.A. Flyagin, Millimeter- and submillimeter-wave gyrotrons, *Radio Eng. Electron Phys.* **19**, 103 (1974)
387. J. Schneider, Stimulated emission of radiation by relativistic electrons in a magnetic field, *Phys. Rev. Lett.* **2**, 504 (1959)
388. T. Idehara, H. Tsuchiya, L. Agusu, S. Mitsudo, H. Murase, H. Mori, T. Kanemaki, T. Saito, Development of a THz gyrotron with 20 T pulsed magnet, *J. Phys.: Conference Series* **51**, 553 (2006)
389. L. Agusu, T. Idehara, H. Mori, T. Saito, I. Ogawa, S. Mitsudo, Design of a CW THz gyrotron (Gyrotron FU CW III) using a 20 T superconducting magnet, *Int. J. Infrared Millimeter Waves* **28**, 315 (2007)
390. V.A. Flyagin, A.G. Luchinin, G.S. Nusinovich, Submillimeter-wave gyrotrons: Theory and experiment, *Int. J. Infrared Millimeter Waves* **4**, 629 (1983)
391. S.N. Vlasov, L.I. Zagryadskaya, M.I. Petelin, Transformation of a whispering gallery mode, propagating on a circular waveguide, into a beam of waves, *Radio Eng. Electron Phys.* **20**, 14 (1975)
392. E.M. Choi, M.A. Shapiro, J.R. Sirigiri, R.J. Temkin, Calculation of radiation from a helically cut waveguide for a gyrotron mode converter in the quasi-optical approximation, *Int. J. Infrared Millimeter Waves* **30**, 8 (2009)
393. M.K. Hornstein, V.S. Bajaj, R.G. Griffin, K.E. Kreischer, I. Mastovsky, M.A. Shapiro, J.R. Sirigiri, R.J. Temkin, Second harmonic operation at 460 GHz and broadband continuous frequency tuning of a gyrotron oscillator, *IEEE Trans. Electron Devices* **52**, 798 (2005)
394. K.D. Hong, G.F. Brand, T. Idehara, A 150–600 GHz step-tunable gyrotron, *J. Appl. Phys.* **74**, 5250 (1993)
395. F.R. Elder, A.M. Gurewitsch, R.V. Langmuir, H.C. Pollock, Radiation from electrons in a synchrotron, *Phys. Rev.* **71**, 829 (1947)
396. G.P. Williams, Filling the THz gap - high power sources and applications, *Rep. Prog. Phys.* **69**, 301 (2006)
397. H. Wiedemann, *Synchrotron radiation* (Springer, Berlin, Heidelberg, 2007)
398. T. Nakazato, M. Oyamada, N. Niimura, S. Urasawa, O. Konno, A. Kagaya, T. Kamiyama, Y. Torizuka, T. Nanba, Y. Kondo, Y. Shibata, K. Ishi, T. Ohsaka, M. Ikezawa, Observation of coherent synchrotron radiation, *Phys. Rev. Lett.* **63**, 1245 (1989)
399. G.L. Carr, M.C. Martin, W.R. McKinney, K. Jordan, G.R. Neil, G.P. Williams, High-power terahertz radiation from relativistic electrons, *Nature* **420**, 153 (2002)
400. H.-W. Hübers, A. Semenov, K. Holldack, U. Schade, G. Wüstefeld, G. Goltsman, Time domain analysis of coherent terahertz synchrotron radiation, *Appl. Phys. Lett.* **87**, 184103 (2005)
401. J.S. Nodvick, D. Saxon, Suppression of coherent radiation by electrons in a synchrotron, *Phys. Rev.* **96**, 180 (1954)
402. M. Abo-Bakr, J. Feikes, K. Holldack, P. Kuske, W.B. Peatman, U. Schade, G. Wüstefeld, H.-W. Hübers, Brilliant, coherent far infrared (THz) synchrotron radiation, *Phys. Rev. Lett.* **90**, 094801 (2003)
403. J. Feikes, M. von Hartrott, M. Ries, P. Schmid, G. Wüstefeld, A. Hoehl, R. Klein, R. Müller, G. Ulm, Metrology Light Source: The first electron storage ring optimized for generating coherent THz radiation, *Phys. Rev. ST Accel. Beams* **14**, 030705 (2011)
404. K. Holldack, S. Khan, R. Mitzner, T. Quast, Femtosecond terahertz radiation from femtoslicing at BESSY, *Phys. Rev. Lett.* **96**, 054801 (2006)

405. U. Schade, K. Holldack, P. Kuske, G. Wüstefeld, H.-W. Hübers, THz near-field imaging employing synchrotron radiation, *Appl. Phys. Lett.* **84**, 1422 (2004)
406. E.J. Singley, M. Abo-Bakr, D.N. Basov, J. Feikes, P. Guptasarma, K. Holldack, H.-W. Hübers, P. Kuske, M.C. Martin, W.B. Peatman, U. Schade, G. Wüstefeld, New scientific opportunities with intense coherent THz synchrotron radiation: measuring the Josephson plasma resonance in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$, *Phys. Rev. B* **69**, 092512 (2004)
407. F. Sannibale, J.M. Byrd, A. Loftsdottir, M. Venturini, M. Abo-Bakr, J. Feikes, K. Holldack, P. Kuske, G. Wüstefeld, H.-W. Hübers, R. Warnock, A model for producing stable, broadband terahertz coherent synchrotron radiation in storage rings, *Phys. Rev. Lett.* **93**, 094801 (2004)
408. B.N. Murdin, Far-infrared free-electron lasers and their applications, *Contemp. Phys.* **50**(2), 391 (2009)
409. H. Motz, Applications of the radiation from fast electron beams, *J. Appl. Phys.* **22**, 527 (1951)
410. R.M. Phillips, The Ubitron, a high-power traveling-wave tube based on a periodic beam interaction in unloaded waveguide, *IRE Trans. Electron Dev.* **7**, 231 (1960)
411. J.M.J. Madey, H.A. Schwettman, W.M. Fairbank, A free electron laser, *IEEE Trans. Nuclear Sci.* **20**, 980 (1973)
412. E. Giovenale, M. D'Arienzo, A. Doria, G. Gallerano, A. Lai, G. Messina, D. Piccinelli, Absorption and diffusion measurements of biological samples using a THz free electron laser, *J. Biol. Phys.* **29**, 159 (2003)
413. S. Lynch, G. Matmon, S.G. Pavlov, K.L. Litvinenko, B. Redlich, A.F.G. van der Meer, N.V. Abrosimov, and H.-W. Hübers, Inhomogeneous broadening of phosphorus donor lines in the infrared spectra of a SiGe single crystal, *Phys. Rev.* **82**, 245206 (2010)
414. B.C. Johnson, H.E. Puthoff, J. SooHoo, S.S. Sussman, Power and linewidth of tunable stimulated far-infrared emission in LiNbO_3 , *Appl. Phys. Lett.* **18**, 181 (1971)
415. M.A. Piestrup, R.N. Fleming, R.H. Pantell, Continuously tunable submillimeter wave source, *Appl. Phys. Lett.* **26**, 418 (1975)
416. J. Nishizawa, K. Suto, Semiconductor Raman laser, *J. Appl. Phys.* **51**, 2429 (1980)
417. K. Kawase, J.-ichi Shikata, H. Ito, Terahertz wave parametric source, *J. Phys. D: Appl. Phys.* **35**, R1 (2002)
418. K. Kawase, Y. Ogawa, H. Minamide, H. Ito, Terahertz parametric sources and imaging applications, *Semicond. Sci. Technol.* **20**, S258 (2005)
419. T.J. Edwards, D. Walsh, M.B. Spurr, C.F. Rae, M.H. Dunn, P.G. Browne, Compact source of continuously and widely-tunable terahertz radiation, *Opt. Express* **14**(4), 1582 (2006)
420. G. Torosyan, K. Nerkararyan, Y. Avetisyan, R. Beigang, Generation of narrowband tunable THz-radiation via optical rectification in periodically poled materials, *J. Biol. Phys.* **29**, 287 (2003)
421. R.C. Jones, Phenomenological description of the response and detecting ability of radiation detectors, *Proc. IRE* **47**, 1495 (1959)
422. R.A. Smith, F.E. Jones, R.P. Chasmar, in *Detection and Measurement of Infra-red Radiation* (Clarendon, Oxford, 1968), chap. 7 - The Ultimate Sensitivity of Infra-Red Detectors, p. 221
423. K.M. van Vliet, Irreversible thermodynamics and carrier density fluctuations in semiconductors, *Phys. Rev.* **110**, 50 (1958)
424. T.W. Kenny, W.J. Kaiser, S.B. Waltman, J.K. Reynolds, Novel infrared detector based on a tunneling displacement transducer, *Appl. Phys. Lett.* **59**, 1820 (1991)
425. T.W. Kenny, J.K. Reynolds, J.A. Podosek, E.C. Vote, L.M. Miller, H.K. Rockstad, W.J. Kaiser, Micromachined infrared sensors using tunneling displacement transducers, *Rev. Sci. Instr.* **67**, 112 (1996)
426. J.B. Chevrier, K. Baert, T. Slater, An infrared pneumatic detector made by micromachining technology, *J. Micromech. Microeng.* **5**, 193 (1995)
427. O. Ajakaiye, J. Grade, C. Shin, T. Kenny, Wafer-scale fabrication of infrared detectors based on tunneling displacement transducers, *Sensor Actuator A* **134**(2), 575 (2007)
428. C. Jiang, M.E. McConney, S. Singamaneni, E. Merrick, Y. Chen, J. Zhao, L. Zhang, V.V. Tsukruk, Thermo-optical arrays of flexible nanoscale nanomembranes freely suspended over microfabricated cavities as IR microimagers, *Chem. Mater.* **18**, 2632 (2006)

429. A.G. Chynoweth, Dynamic method for measuring the pyroelectric effect with special reference to barium titanate, *J. Appl. Phys.* **27**, 78 (1956)
430. J. Cooper, A fast -response pyroelectric thermal detector, *J. Sci. Instr.* **39**, 462 (1962)
431. A.G. Chynoweth, Pyroelectricity, internal domains, and interface charges in triglycine sulfate, *Phys. Rev.* **117**, 1235 (1960)
432. A.L. Stanford, Detection of electromagnetic radiation using the pyroelectric effect, *Solid State Electron.* **8**, 747 (1965)
433. P.J. Lock, Doped triglycine sulfate for pyroelectric applications, *Appl. Phys. Lett.* **19**, 390 (1971)
434. R.J. Phelan, R.J. Mahler, A.R. Cook, High D* pyroelectric polyvinylfluoride detectors, *Appl. Phys. Lett.* **19**, 337 (1971)
435. A.M. Glass, J.H. McFee, J.G. Bergman, Pyroelectric properties of polyvinylidene fluoride and its use for infrared detection, *J. Appl. Phys.* **42**, 5219 (1971)
436. W.-s. Zhu, J.R. Izatt, B.K. Deka, Pyroelectric detection of submicrosecond laser pulses between 230 and 530 μm , *Appl. Opt.* **28**(17), 3647 (1989)
437. F.B. Foote, D.T. Hodges, H.B. Dyson, Calibration of power and energy meters for the far infrared/near millimeter wave region, *Int. J. Infrared Millimeter Waves* **2**, 773 (1981)
438. B. Vowinkel, Broad-band calorimeter for precision measurement of millimeter and submillimeter-wave power, *IEEE Trans. Instrum. Meas.* **29**, 183 (1980)
439. A.C. Macpherson, D.M. Kerns, A microwave microcalorimeter, *Rev. Sci. Instr.* **26**, 27 (1955)
440. N. Erickson, in *Proc. of the 13th Int. Symp. on Space Terahertz Technology* (Harvard, 2002), A fast, very sensitive calorimetric power meter for millimetre to submillimeter wavelengths, pp. 301–307
441. B. Vowinkel, H.P. Röser, Precision measurement of power at millimeter and sub-millimeter wavelengths using a waveguide calorimeter, *Int. J. Infrared Millimeter Waves* **3**, 471 (1982)
442. R.J. Wyld, Installation and operating instructions for the TK TeraHertz Absolute Power Meter System. Tech. rep., Thomas Keating Ltd., Station Mills, Billingshurst, West Sussex, RH14 9SH, England (2002)
443. A.M. Danishevskii, A.A. Kastal'skii, S.M. Ryvkin, I.D. Yaroshetskii, Photon drag of free carriers in direct interband transitions in semiconductors, *Zh. Eksp. Teor. Fiz.* **58**, 544 (1970)
444. A.F. Gibson, M.F. Kimmitt, A.C. Walker, Photon drag in germanium., *Appl. Phys. Lett.* **17**, 75 (1970)
445. C. Baxter, R. Loudon, Radiation pressure and the photon momentum in dielectrics, *J. Mod. Opt.* **57**, 830 (2010)
446. M.F. Kimmitt, Recent development of infrared detectors, *Infrared Phys.* **17**, 459 (1977)
447. S.D. Ganichev, Y.V. Terent'ev, I.D. Yaroshetskii, Photon-drag photodetectors for the far-IR and submillimeter regions, *Sov. Tech. Phys. Lett.* **11**(1), 20 (1985)
448. A.F. Gibson, M.F. Kimmitt, in *Infrared Millimeter Waves* (Academic Press, New York, London, Toronto, Sydney, San Francisco, 1980), Submillimeter Techniques, vol. 3, chap. 2 - Photon Drag Detection, pp. 181–217 (Calibration Standards for 100–2,000 μm , pp. 212–213)
449. M.F. Kimmitt, A.A. Serafetinides, H.P. Roser, D.A. Huckridge, Submillimeter performance of photon-drag detectors, *Infrared Phys.* **18**, 675 (1978)
450. M.A. Kinch, Compensated silicon-impurity conduction bolometer, *J. Appl. Phys.* **42**, 5861 (1971)
451. E.E. Haller, Physics and design of advanced IR bolometers and photoconductors, *Infrared Phys.* **25**, 257 (1985)
452. P.L. Richards, Bolometers for infrared and millimeter waves, *J. Appl. Phys.* **76**, 1 (1994)
453. J. Clarke, G.I. Hoffer, P.L. Richards, N.H. Yeh, Superconductive bolometers for submillimeter wavelengths, *J. Appl. Phys.* **48**, 4865 (1977)
454. P.D. Mauskopf, J.J. Bock, H. DelCastillo, W.H. Holzappel, A.E. Lange, Composite infrared bolometers with Si_3N_4 micromesh absorbers, *Appl. Opt.* **36**, 765 (1997)
455. H.T. Nguyen, P. Ringold, P.A.R. Ade, J. Battle, J.W. Beeman, J.J. Bock, S.C. Elliott, P.C. Hargrave, B. Schulz, A.D. Turner, V.V. Hristov, M. Weilert, and L. Zhang, in *Millimeter and*

- submillimeter detectors and instrumentation for astronomy III*, ed. by J. Zmuidzinas, W.S. Holland, S. Withington, W.D. Duncan (2006), A report on laboratory performance of the bolometric detector arrays for SPIRE/HSO (paper II), Proc. SPIE **6275**, pp. 627518
456. N. Jethava, E. Kreysa, G. Siringo, W. Esch, H.P. Gemünd, T. May, S. Anders, L. Fritzsche, R. Boucher, V. Zakosarenko, H.G. Meyer, in *Millimeter and Submillimeter Detectors and Instrumentation for Astronomy III*, ed. by J. Zmuidzinas, W.S. Holland, S. Withington, W.D. Duncan (2006), A superconducting bolometer camera for APEX, Proc. SPIE **6275**, p. 62751A
 457. S.F. Lee, J.M. Gildemeister, W. Holmes, A.T. Lee, P.L. Richards, Voltage-biased superconducting transition-edge bolometer with strong electrothermal feedback operated at 370 mK, Appl. Opt. **37**(16), 3391 (1998)
 458. A.T. Lee, P.L. Richards, S.W. Nam, B. Cabrera, K.D. Irwin, A superconducting bolometer with strong electrothermal feedback, Appl. Phys. Lett. **69**, 1801 (1996)
 459. J.M. Gildemeister, A.T. Lee, P.L. Richards, A fully lithographed voltage-biased superconducting spiderweb bolometer, Appl. Phys. Lett. **74**, 868 (1999)
 460. J.M. Gildemeister, A.T. Lee, P.L. Richards, Monolithic arrays of absorber-coupled voltage-biased superconducting bolometers, Appl. Phys. Lett. **77**, 4040 (2000)
 461. D.J. Benford, J.G. Staguhn, T.J. Ames, C.A. Allen, J.A. Chervenak, C.R. Kennedy, S. Lefranc, S.F. Maher, S.H. Moseley, F. Pajot, C. Rioux, R. Shafer, G.M. Voellmer, in *Millimeter and Submillimeter detectors and instrumentation for astronomy III*, ed. by J. Zmuidzinas, W.S. Holland, S. Withington, W. Duncan (2006), First astronomy images with a multiplexed superconducting bolometer array, Proc. SPIE **6275**, pp. 62,751C–2
 462. W. Holland, M. MacIntosh, A. Fairley, D. Kelly, D. Montgomery, D. Gostick, E. Atad-Ettinger, M. Ellis, I. Robinson, M. Hollister, A. Woodcraft, P. Ade, I. Walker, K. Irwin, G. Hilton, W. Duncan, C. Reintsema, A. Walton, W. Parkes, C. Dunars, M. Fich, J. Kycia, M. Halpern, D. Scott, A. Gibb, J. Molnar, E. Chapin, D. Bintley, S. Craig, T. Chylek, T. Jenness, F. Economou, G. Davis, in *Millimeter and Submillimeter detectors and instrumentation for astronomy III*, ed. by J. Zmuidzinas, W.S. Holland, S. Withington, W. Duncan (2006), SCUBA-2: a 10,000 pixel submillimeter camera for the James Clerk Maxwell Telescope, Proc. SPIE **6275**, pp. 62,751E–1
 463. P.A.J. de Korte, J. Beyer, S. Deiker, G.C. Hilton, K.D. Irwin, M. MacIntosh, S.W. Nam, C.D. Reintsema, L.R. Vale, M.E. Huber, Time-division superconducting quantum interference device multiplexer for transition-edge sensors, Rev. Sci. Instr. **74**, 3807 (2003)
 464. T.M. Lanting, H.M. Cho, J. Clarke, W.L. Holzapfel, A.T. Lee, M. Lueker, P.L. Richards, M.A. Dobbs, H. Spieler, A. Smith, Frequency-domain multiplexed readout of transition-edge sensor arrays with a superconducting quantum interference device, Appl. Phys. Lett. **86**, 112511 (2005)
 465. J.T. Skidmore, J. Gildemeister, A.T. Lee, M.J. Meyers, P.L. Richards, Superconducting bolometer for far-infrared Fourier transform spectroscopy, Appl. Phys. Lett. **82**, 469 (2003)
 466. K. I'lin, M. Lindgren, M. Currie, A.D. Semenov, G.N. Goltsman, S.I. Cherednichenko, E.M. Gershenzon, Picosecond hot-electron energy relaxation in NbN superconducting photodetectors, Appl. Phys. Lett. **76**, 2752 (2000)
 467. S.E. Schwarz, B.T. Ulrich, Antenna-coupled infrared detectors, J. Appl. Phys. **48**, 1870 (1977)
 468. T.L. Hwang, S.E. Schwarz, D.B. Rutledge, Microbolometers for infrared detection, Appl. Phys. Lett. **34**, 773 (1979)
 469. D. Rutledge, S. Schwarz, Planar multimode detector arrays for infrared and millimeter-wave applications, IEEE J. Quant. Electron. **17**, 407 (1981)
 470. A.J. Steckl, R.D. Nelson, B.T. French, R.A. Gudmundsen, D. Schechter, Application of charge-coupled devices to infrared detection and imaging, Proc. IEEE **63**, 67 (1975)
 471. A.W. Lee, Q. Hu, Real-time, continuous-wave terahertz imaging by use of a microbolometer focal-plane array, Opt. Lett. **30**, 2563 (2005)
 472. B. Fieque, A. Crastes, O. Legras, J.L. Tissot, in *Infrared Technology and Applications XXXI*, ed. by B.F. Andresen, G.F. Fulop (2005), MWIR uncooled microbolometer, a way to increase the number of applications, Proc. SPIE **5783**, p. 531

473. N. Oda, M. Sano, K. Sonoda, H. Yoneyama, S. Kurashina, M. Miyoshi, T. Sasaki, I. Hosako, N. Sekine, T. Sudou, S. Ohkubo, in *Infrared Technology and Applications XXXVII*, ed. by Bjørn F. Andresen, Gabor F. Fulop, Paul R. Norton (2011), Development of Terahertz Focal Plane Arrays and Handy Camera, Proc. SPIE **8012**, 80121B
474. B.N. Behnken, G. Karunasiri, D.R. Chamberlin, P.R. Robrish, J. Faist, Real-time imaging using a 2.8 THz quantum cascade laser and uncooled infrared microbolometer camera, Opt. Lett. **33**, 440 (2008)
475. E. Burstein, J.W. Davisson, E.E. Bell, W.J. Turner, H.G. Lipson, Infrared photoconductivity due to neutral impurities in germanium, Phys. Rev. **93**(1), 65 (1954)
476. A. Baldereschi, N.O. Lipari, Spherical model of shallow acceptor states in semiconductors, Phys. Rev. B **8**, 2697 (1973)
477. A. Baldereschi, N.O. Lipari, Binding energy of shallow acceptors in group IV elements and III-V compounds, J. Luminescence **12–13**, 489 (1976)
478. E.L. Dereniak, D.G. Crowe, *Optical Radiation Detectors* (John Wiley, New York, 1984)
479. R.H. Kingston, *Detection of Optical and Infrared Radiation, Springer Series in Optical Sciences*, vol. 10 (Springer Verlag, Berlin, 1978)
480. P.R. Bratt, in *Semiconductors and Semimetals* (Academic Press, New York, San Francisco, London, 1977), Infrared Detectors II, vol. 12, chap. 2 - Impurity Germanium and Silicon Infrared Detectors, p. 52
481. N. Hiromoto, M. Saito, H. Okuda, Ge:Ga far-infrared photoconductor with low compensation, Jpn. J. Appl. Phys. **29**, 1739 (1990)
482. F.A. Hegmann, J.B. Williams, B. Cole, M.S. Sherwin, J.W. Beeman, E.E. Haller, Time-resolved photoresponse of a gallium-doped germanium photoconductor using a variable pulse-width terahertz source, Appl. Phys. Lett. **76**(3), 262 (2000)
483. M. Fujiwara, N. Hiromoto, Ge:Ga far-infrared photoconductor with a low Ga concentration of $1 \times 10^{14} \text{ cm}^{-3}$, Jpn. J. Appl. Phys. **36**, 4262 (1997)
484. E.E. Haller, Advanced far-infrared detectors, Infrared Phys. Technol. **25**, 127 (1994)
485. N.M. Haegel, E.E. Haller, P.N. Luke, Performance and materials aspects of Ge:Be photoconductors, Int. J. Infrared Millimeter Waves **4**, 945 (1983)
486. J.W. Beeman, W.L. Hansen, O.D. Dubon, E.E. Haller, High performance antimony-doped germanium photoconductors, Infrared Phys. **37**(7), 715 (1996)
487. G.D. Peskett, B.V. Rollin, Energy exchange between hot carriers and the lattice in indium antimonide, Proc. Phys. Soc. **82**, 467 (1963)
488. G. Strasser, K. Bochter, M. Witzany, E. Gornik, Improved tunable InSb FIR detectors, Infrared Phys. **32**, 439 (1991)
489. M.A. Kinch, B.V. Rollin, Detection of millimetre and sub-millimetre wave radiation by free carrier absorption in a semiconductor, Brit. J. Appl. Phys. **14**, 672 (1963)
490. E.H. Putley, Indium antimonide submillimeter photoconductive detectors, Appl. Opt. **4**(6), 649 (1965)
491. E.R. Brown, M.J. Wengler, T.G. Phillips, Absolute response and noise equivalent power of cyclotron resonance-assisted InSb detectors at submillimeter wavelengths, J. Appl. Phys. **58**, 2051 (1985)
492. E. Gornik, V. Roskopf, W. Heiss, Tunable lasers and detectors in the FIR, Infrared Phys. Technol. **36**, 113 (1995)
493. M.F. Kimmitt, G.B.F. Niblett, Infra-red emission from the theta pinch, Proc. Phys. Soc. **82**, 938 (1963)
494. QMC Instruments Ltd., School of Physics and Astronomy, Cardiff University, Cardiff, UK
495. E.R. Brown, J. Keene, T.G. Phillips, A heterodyne receiver for the submillimeter wavelength region based on cyclotron resonance in indium antimonide at low temperatures, Int. J. Infrared Millimeter Waves **6**, 1121 (1985)
496. J.J. Whalen, C.R. Westgate, Temperature dependence of the conversion loss and response time of InSb mixers, IEEE Trans. Electron Devices **17**, 310 (1970)
497. A. Isayama, N. Isei, S. Ishida, M. Sato, A 20-channel electron cyclotron emission detection system for a grating polychromator in JT-60U, Rev. Sci. Instr. **73**, 1165 (2002)

498. G.E. Stillman, C.M. Wolfe, J.O. Dimmock, in *Semiconductors and Semimetals* (Academic Press, New York, San Francisco, London, 1977), Infrared Detectors II, vol. 12, chap. 4 – Far-Infrared Photoconductivity in High Purity GaAs, pp. 169–290
499. G.E. Stillman, C.M. Wolfe, J.O. Dimmock, in *Semiconductors and Semimetals* (Academic Press, New York, San Francisco, London, 1977), *Infrared Detectors II*, vol. 12, chap. 4 – Far-Infrared Photoconductivity in High Purity GaAs, p. 215
500. J. Leotin, in *Far-infrared science and technology* (1986), Far-infrared photoconductive detectors, Proc. SPIE **666**, p. 81
501. M.F. Kimmitt, G.C. Lopez, J.C. Giles, M. Takai, H.P. Röser, B.T. McGuckin, A. Black, Far-infrared detection with $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, Infrared Phys. **25**, 767 (1985)
502. J. Bandaru, J.W. Beeman, E.E. Haller, S. Samperi, N.M. Haegel, Influence of the Sb dopant distribution on far infrared photoconductivity in Ge:Sb blocked impurity band detectors, Infrared Phys. Technol. **43**(6), 353 (2002)
503. N.M. Haegel, S.A. Samperi, A.M. White, Electric field and responsivity modeling for far-infrared blocked impurity band detectors, J. Appl. Phys. **93**, 1305 (2003)
504. J.W. Beeman, S. Goyal, L.A. Reichertz, E.E. Haller, Ion-implanted Ge:B far-infrared blocked-impurity-band detectors, Infrared Phys. Technol. **51**(1), 60 (2007)
505. J.C. Garcia, N.M. Haegel, E.A. Zagorski, Alternate operating mode for long wavelength blocked impurity band detectors, Appl. Phys. Lett. **87**, 043502 (2005)
506. R. Fessenden, Wireless signalling, U. S. Patent 706,740 (1902)
507. W. Schottky, Über den Ursprung des Superheterodyn-Gedankens, E.N.T., 454 (1925)
508. J.S. Belrose, Reginald Aubrey Fessenden and the birth of wireless telephony, IEEE Antennas and Propagation Magazine **44**, 38 (2002)
509. A. Semenov, O. Cojocari, H.-W. Hübers, F. Song, A. Klushin, A.S. Müller, Application of zero-bias quasi-optical Schottky-diode detectors for monitoring short-pulse and weak terahertz radiation, IEEE Electron. Device Lett. **31**, 674 (2010)
510. S.A. Maas, *Microwave Mixers*, 2nd edn. (Artech House, Boston, 1993). ISBN 0-89006-605-1
511. A.R. Kerr, Suggestions for revised definitions of noise quantities, including quantum effects, IEEE Trans. Microw. Theor. Tech. **47**, 325 (1999)
512. E.H. Rhoderick, R.H. Williams, *Metal-Semiconductor contacts* (Clarendon Press, Oxford, 1988)
513. H.-W. Hübers, H.P. Röser, Temperature dependence of the barrier height of Pt/n-GaAs Schottky diodes, J. Appl. Phys. **84**, 5326 (1998)
514. J.H. Werner, H.H. Güttler, Barrier inhomogeneities at Schottky contacts, J. Appl. Phys. **69**, 1522 (1991)
515. A. van der Ziel, Infrared detection and mixing in heavily doped Schottky barrier diodes, J. Appl. Phys. **47**, 2059 (1976)
516. H.P. Röser, R.U. Titz, G.W. Schwaab, M.F. Kimmitt, Current-frequency characteristics of submicron GaAs Schottky barrier diodes with femtofarad capacitances, J. Appl. Phys. **72**, 3194 (1992)
517. T. Nozokido, J.J. Chang, C.M. Mann, T. Suzuki, K. Mizuno, Optimization of a Schottky barrier mixer diode in the submillimeter wave region, Int. J. Infrared Millimeter Waves **15**, 1851 (1994)
518. K.S. Champlin, G. Eisenstein, Cutoff frequency of submillimeter Schottky-barrier diodes, IEEE Trans. Microw. Theory Tech. **26**, 31 (1978)
519. W.C.B. Peatman, T.W. Crowe, Design and fabrication of 0.5 micron GaAs Schottky barrier diodes for low-noise terahertz receiver applications, Int. J. Infrared Millimeter Waves **11**, 355 (1990)
520. A. Aydinli, R.J. Mattauch, The effects of surface treatments on the Pt/n-GaAs Schottky interface, Solid State Electron. **25**, 551 (1982)
521. H.P. Röser, H.-W. Hübers, T.W. Crowe, W.C.B. Peatman, Nanostructure GaAs Schottky diodes for far-infrared heterodyne receivers, Infrared Phys. Technol. **35**, 451 (1994)
522. S.M. Marazita, W.L. Bishop, J.L. Hesler, K. Hui, W.E. Bowen, T.W. Crowe, Integrated GaAs Schottky mixers by spin-on-dielectric wafer bonding, IEEE Trans. Electron Device **47**(6), 1152 (2000)

523. P. Siegel, R.P. Smith, M.C. Gaidis, S. Martin, 2.5-THz GaAs monolithic membrane-diode mixer, *IEEE Trans. Microw. Theory Tech.* **47**(5), 596 (1999)
524. H.M. Pickett, The Microwave Limb Sounder THz module on Aura, *IEEE Trans. Geoscience Remote Sensing* **44**, 1122 (2006)
525. A.H. Dayem, R.J. Martin, Quantum interaction of microwave radiation with tunnelling between superconductors, *Phys. Rev. Lett.* **8**, 246 (1992)
526. P.K. Tien, J.P. Gordon, Multiphoton process observed in the interaction of microwave fields with the tunnelling between superconductor films, *Phys. Rev.* **129**, 647 (1963)
527. J.R. Tucker, M.J. Feldman, Quantum detection at millimeter wavelengths, *Rev. Mod. Phys.* **57**, 1055 (1985)
528. B.B. Hu, X.C. Zhang, D.H. Auston, P.R. Smith, Free-space radiation from electro-optic crystals, *Appl. Phys. Lett.* **56**, 506 (1990)
529. M. Bin, M.C. Gaidis, J. Zmuidzinas, T.G. Phillips, H.G. LeDuc, Low-noise 1 THz niobium superconducting tunnel junction mixer with normal metal tuning circuit, *Appl. Phys. Lett.* **68**, 1714 (1996)
530. B.D. Jackson, A.M. Baryshev, G. de Lange, J.R. Gao, S. Shitov, N.N. Iossad, T.M. Klapwijk, Low-noise 1 THz superconductor-insulator-superconductor mixer incorporating a NbTiN/SiO₂/Al tuning circuit, *Appl. Phys. Lett.* **79**, 436 (2001)
531. A.R. Kerr, Some fundamental and practical limits on broadband matching to capacitive devices, and the implications for SIS mixer design, *IEEE Trans. Microw. Theory Tech.* **43**, 2 (1995)
532. M.C. Gaidis, H.G. LeDuc, M. Bin, D. Miller, J.A. Stern, J. Zmuidzinas, Characterization of low-noise quasi-optical SIS mixers for the submillimeter band, *IEEE Trans. Microw. Theor. Tech.* **44**, 1130 (1996)
533. J. Kawamura, D. Miller, J. Chen, J. Zmuidzinas, B. Bumble, H.G. LeDuc, J.A. Stern, Very high-current-density Nb/AlN/Nb tunnel junctions for low-noise submillimeter mixers, *Appl. Phys. Lett.* **76**(15), 2119 (2000)
534. G.N. Goltsman, B.S. Karasik, O. Okunev, A. Dzadanov, E.M. Gershenzon, H. Ekström, S. Jacobsson, E.L. Kollberg, NbN hot electron superconducting mixers for 100 GHz operation, *IEEE Trans. Appl. Superconductivity* **5**, 3065 (1995)
535. R. Nebosis, A.D. Semenov, Y. Gousev, K.F. Renk, in *Proc. of the 7th Int. Symp. on Space Terahertz Technology* (Charlottesville, 1996), Rigorous analysis of a superconducting hot-electron bolometer mixer: theory and comparison with experiment, p. 601
536. A.D. Semenov, G.N. Goltsman, R. Sobolewski, Hot-electron effect in superconductors and its applications for radiation sensors, *Semicond. Sci. Technol.* **15**, R1 (2002)
537. W. Skocpol, M.R. Beasley, M. Tinkham, Self-heating hotspots in superconducting thin-film microbridges, *J. Appl. Phys.* **45**, 4054 (1974)
538. D.W. Floet, E. Miedema, T.M. Klapwijk, J.R. Gao, Hotspot mixing: A framework for heterodyne mixing in superconducting hot-electron bolometers, *Appl. Phys. Lett.* **74**, 433 (1999)
539. H. Merkel, P. Khosropanah, P. Yagubov, E. Kollberg, A hot-spot mixer model for phonon-cooled NbN hot electron bolometric mixers, *IEEE Trans. Appl. Superconductivity* **9**, 4201 (1999)
540. A.D. Semenov, H.-W. Hübers, Bandwidth of a hot-electron bolometer mixer according to the hotspot model, *IEEE Trans. Appl. Superconductivity* **11**, 196 (2001)
541. E.K. Hollmann, A.G. Zaitsev, Optimal magnetron sputtering parameters for superconducting NbN thin film deposition, *Vacuum* **44**, 847 (1993)
542. N. Perrin, C. Vanneste, Response of superconducting films to a periodic optical radiation, *Phys. Rev. B* **28**, 5150 (1983)
543. B. Voronov, G. Goltsman, E. Gershenzon, L. Seidman, T. Gubkina, V. Siomash, Superconductive properties of NbN ultrathin films on different substrates, *Fiz. Cim. Tekh.* **7**, 1097 (1994)
544. K. Ilin, R. Schneider, D. Gerthsen, A. Engel, H. Bartolf, A. Schilling, A. Semenov, H.-W. Huebers, B. Freitag, M. Siegel, Ultra-thin NbN films on Si: crystalline and superconducting properties, *J. Phys.: Conf. Series* **97**, 012045 (2008)

545. A.D. Semenov, H.-W. Hübers, J. Schubert, G.N. Goltsman, A.I. Elantiev, B.M. Voronov, E.M. Gershenzon, Design and performance of the lattice-cooled hot-electron terahertz mixer, *J. Appl. Phys.* **88**, 6758 (2000)
546. A.D. Semenov, H.-W. Hübers, H. Richter, M. Birk, M. Krocka, U. Mair, K. Smirnov, G. Goltsman, B.V. Voronov, 2.5 THz heterodyne receiver with NbN hot-electron bolometer, *Phys. C* **372–376**, 448 (2002)
547. H.A. Gebbie, Fourier transform versus grating spectroscopy, *Appl. Opt.* **8**(3), 501 (1969)
548. P.B. Fellgett, Ph.D. thesis, University of Cambridge, UK (1951). See also: *Infrared Phys.* **24**, 95–98 (1984)
549. P. Jacquinot, New developments in interference spectroscopy, *Rep. Prog. Phys.* **23**, 267 (1960)
550. C.A. Burrus, W. Gordy, Submillimeter wave spectroscopy, *Phys. Rev.* **93**(4), 897 (1954)
551. P. Helminger, F.C.D. Lucia, W. Gordy, Extension of microwave absorption spectroscopy to 0.37-mm wavelength, *Phys. Rev. Lett.* **25**, 1397 (1970)
552. B.J. Drouin, F.W. Maiwald, J.C. Pearson, Application of cascaded frequency multiplication to molecular spectroscopy, *Rev. Sci. Instr.* **76**, 093113 (2005)
553. A.F. Krupnov, in *Modern Aspects of Microwave Spectroscopy*, ed. G.W. Chantry (Academic Press, New York, 1979), chap. 4 - Modern submillimetre microwave scanning spectroscopy, p. 217–256
554. B. Gorshunov, A. Volkov, I. Spektor, A. Prokhorov, A. Mukhin, M. Dressel, S. Uchida, A. Loidl, Terahertz BWO-spectroscopy, *Int. J. Infrared Millim. Waves* **26**, 1217 (2005)
555. K.M. Evenson, D.A. Jennings, M.D. Vanek, in *Frontiers of laser spectroscopy of gases* (Kluwer Academic Publishers, Dordrecht, 1988), chap. Tunable far infrared laser spectroscopy, pp. 43–51
556. H.I. Ewen, E.M. Purcell, Radiation from galactic hydrogen at 1,420 Mc/sec, *Nature* **168**, 356 (1951)
557. C.A. Muller, J.H. Oort, The interstellar hydrogen line at 1,420 Mc/sec, and an estimate of galactic rotation, *Nature* **168**, 357 (1951)
558. C.M. Walmsley, C. Bertout, F. Combes, A. Ferrara, T. Forveille, T. Guillot, A. Jones, S. Shore (eds.), *Astronomy & Astrophysics special feature: Herschel: the first science highlights*, A&A **518** (2010)
559. R.P. Madden, J. Strong, *Appendix P* in: *Concepts in Classical Optics* (Freeman and Co., San Francisco and London, 1958)
560. S. Jovicevic, S. Sesnic, Diffraction of a parallel- and perpendicular-polarized wave from an echelette grating, *J. Opt. Soc. Am.* **62**, 865 (1972)
561. G.W. Stroke, Attainment of high efficiencies in blazed optical gratings by avoiding polarisation in the diffracted light, *Phys. Lett.* **5**, 45 (1963)
562. H. Czerny, A.F. Turner, Über den Astigmatismus bei Spiegelspektrometern, *Z. Physik* **61**, 792 (1930)
563. H. Ebert, G.H. Wiedemann ed., Zwei Formen von Spectrographen, *Annalen der Physik und Chemie (Wied. Ann.)* **38**, 489 (1889)
564. W.G. Fastie, A small plane grating monochromator, *J. Opt. Soc. Am.* **42**, 641 (1952)
565. A. Girard, Possibilités offertes par un spectrometre a modulation sélective, *J. Phys. Colloques* **28**, C2 (1967)
566. M. Harwit, P.G. Phillips, T. Fine, N.J.A. Sloane, Doubly multiplexed dispersive spectrometers, *Appl. Opt.* **9**(5), 1149 (1970)
567. M.E. Gehm, S.T. McCain, N.P. Pitsianis, D.J. Brady, P. Potuluri, M.E. Sullivan, Static two-dimensional aperture coding for multimodal, multiplex spectroscopy, *Appl. Opt.* **45**(13), 2965 (2006)
568. A.A. Michelson, XXX. On the application of interference methods to spectroscopic measurements.–II, *Phil. Mag. Series 5* **34**, 280 (1892)
569. L. Rayleigh, XLVII. On the interference bands of approximately homogeneous light; in a letter to Prof. A. Michelson, *Phil. Mag.* **34**, 407 (1892)
570. H.F. Talbot, LXXXVI. Facts relating to optical science. No. IV, *Phil. Mag.* **9**, 401 (1836)

571. Q. Wu, C.A. Werley, K.H. Lin, A. Dorn, M.G. Bawendi, K.A. Nelson, Quantitative phase contrast imaging of THz electric fields in a dielectric waveguide, *Opt. Express* **17**(11), 9219 (2009)
572. H.L. Kung, A. Bhatnagar, D.A.B. Miller, Transform spectrometer based on measuring the periodicity of Talbot self-images, *Opt. Lett.* **26**(21), 1645 (2001)
573. D.W. van der Weide, J. Murakowski, F. Keilmann, Gas-absorption spectroscopy with electronic terahertz techniques, *IEEE Trans. Microw. Theory Tech.* **48**(4), 740 (2000)
574. B.I. Greene, J.F. Federici, D.R. Dykaar, R.R. Jones, P.H. Bucksbaum, Interferometric characterization of 160 fs far-infrared light pulses, *Appl. Phys. Lett.* **59**, 893 (1991)
575. S.E. Ralph, D. Grischkowsky, THz spectroscopy and source characterization by optoelectronic interferometry, *Appl. Phys. Lett.* **60**, 1070 (1992)
576. C. Karadi, S. Jauhar, L.P. Kouwenhoven, K. Wald, J. Orenstein, P.L. McEuen, Dynamic response of a quantum point contact, *J. Opt. Soc. Am. B* **11**, 2566 (1994)
577. A. Schliesser, M. Brehm, F. Keilmann, D.W. van der Weide, Frequency-comb infrared spectrometer for rapid, remote chemical sensing, *Opt. Express* **13**(22), 9029 (2005)
578. T. Ganz, H.G. von Ribbeck, M. Brehm, F. Keilmann, Compact frequency-comb Fourier-transform infrared spectrometer, *Opt. Comm.* **281**, 3827 (2008)
579. S. Yokoyama, R. Nakamura, M. Nose, T. Araki, T. Yasui, Terahertz spectrum analyzer based on a terahertz frequency comb, *Opt. Express* **16**(17), 13052 (2008)
580. D.S. Yee, Y. Jang, Y. Kim, D.C. Seo, Terahertz spectrum analyzer based on frequency and power measurement, *Opt. Lett.* **35**(15), 2532 (2010)
581. M.N. Afsar, K.J. Button, Millimeter and submillimeter wave measurements of complex optical and dielectric parameters of materials, *Int. J. Infrared Millim. Waves* **2**(5), 1029 (1981)
582. H. Jamshidi, T.J. Parker, A two-beam interferometer for dispersive reflection spectroscopy of solids in the far infrared at temperatures between 4 and 300K, *Int. J. Infrared Millim. Waves* **4**(4), 681 (1983)
583. J.R. Birch, Dispersive Fourier transform spectroscopy, *Mikrochim. Acta* **III**, 105 (1987)
584. D.H. Martin, E. Puplett, Polarised interferometric spectrometry for the millimetre and submillimetre spectrum, *Infrared Phys.* **10**, 105 (1969)
585. D.H. Martin, K.J. Button (ed.), in *Infrared and Millimeter Waves* (Academic Press, New York, London, Paris, San Diego, San Francisco, Sao Paulo, Sydney, Tokyo, Toronto, 1982), vol. 6: Systems and Components, chap. 2 - Polarizing (Martin-Puplett) interferometric spectrometers for the near- and submillimeter spectra pp. 65–148
586. D.K. Lambert, P.L. Richards, Martin-puplett interferometer: an analysis, *Appl. Opt.* **17**, 1595 (1978)
587. D.H. Martin, J.C.G. Lesurf, Submillimetre-wave optics, *Infrared Phys.* **18**, 405 (1978)
588. J.C.G. Lesurf, Gaussian optics and the design of Martin-Puplett diplexers, *Infrared Phys.* **21**, 383 (1981)
589. J.C.G. Lesurf, Gaussian beam-mode optics and the design of millimeter-wave Martin-Puplett instruments, *Infrared Phys.* **28**, 129 (1988)
590. R. Ulrich, K.F. Renk, L. Genzel, Tunable submillimeter interferometers of the Fabry-Perot type, *IEEE Trans. Microw. Theor. Tech.* **11**, 363 (1963)
591. K.F. Renk, J. Betz, A. Prückle, B. Brunner, H. Langfeller, Use of high T_c superconductors for far-infrared Fabry-Perot resonators, *Appl. Phys. Lett.* **57**, 2148 (1990)
592. E.V. Pechen, S. Vent, B. Brunner, A. Prückle, S. Lipp, G. Lindner, O. Alexandrov, J. Schützmann, K.F. Renk, Far-infrared Fabry-Perot resonator with high T_c $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films on silicon plates, *Appl. Phys. Lett.* **61**, 1980 (1992)
593. A. Poglitsch, J.W. Beeman, N. Geis, R. Genzel, M. Haggerty, E.E. Haller, J. Jackson, M. Rumitz, G.J. Stacey, C.H. Townes, The MPE/UCB far-infrared imaging Fabry-Perot interferometer (FIFI), *Int. J. Infrared Millim. Waves* **12**, 859 (1991)
594. T.D. Graauw, in *Proc. of the conference "The universe as seen by ISO"*, P. Cox and M.F. Kessler (eds.), (ESA, Paris, France, 1999), Summary of ISO SWS performance and science highlights, vol. ESA SP-427, pp. 31–38

595. P.E. Clegg on behalf of the LWS consortium, in *Proc. of the conference "The universe as seen by ISO"*, P. Cox and M.F. Kessler (eds.) (ESA, Paris, France, 1999), The ISO Long-Wavelength Spectrometer: description, performance and highlights, vol. ESA SP-427, pp. 39–43
596. R. Ulrich, T.J. Bridges, M.A. Pollack, Variable metal mesh coupler for far infrared lasers, *Appl. Opt.* **9**, 2511 (1970)
597. B.M. Oliver, Time domain spectroscopy, *Hewlett-Packard J.* **15**, 1 (1964)
598. H. Fellner-Feldegg, Measurement of dielectrics in the time domain, *J. Phys. Chem.* **73**, 616 (1969)
599. D.E. Spence, P.N. Kean, W. Sibbett, 60-fsec pulse generation from a self-mode-locked Ti:sapphire laser, *Opt. Lett.* **16**(1), 42 (1991)
600. A. Stingl, M. Lenzner, C. Spielmann, F. Krausz, R. Szipöcs, Sub-10-fs mirror-dispersion-controlled Ti:sapphire laser, *Opt. Lett.* **20**(6), 602 (1995)
601. J.N. Eckstein, A.I. Ferguson, T.W. Hänsch, High-resolution two-photon spectroscopy with picosecond light pulses, *Phys. Rev. Lett.* **40**(13), 847 (1978)
602. D.H. Auston, K.P. Cheung, J.A. Valdmanis, D.A. Kleinman, Cherenkov radiation from femtosecond optical pulses in electro-optic media, *Phys. Rev. Lett.* **53**(16), 1555 (1984)
603. P.U. Jepsen, R.H. Jacobsen, S.R. Keiding, Generation and detection of terahertz pulses from biased semiconductor antennas, *J. Opt. Soc. Am. B* **13**(11), 2424 (1996)
604. Q. Wu, F.G. Sun, P. Campbell, X.C. Zhang, Dynamic range of an electro-optic field sensor and its imaging applications, *Appl. Phys. Lett.* **68**, 3224 (1996)
605. C. Fattinger, D. Grischkowsky, Point source terahertz optics, *Appl. Phys. Lett.* **53**, 1480 (1988)
606. C. Fattinger, D. Grischkowsky, Terahertz beams, *Appl. Phys. Lett.* **54**, 490 (1989)
607. M. van Exter, C. Fattinger, D. Grischkowsky, Terahertz time-domain spectroscopy of water vapor, *Opt. Lett.* **14**(20), 1128 (1989)
608. P.C.M. Planken, C.E.W.M. van Rijmenam, R.N. Schouten, Opto-electronic pulsed THz systems, *Semicond. Sci. Technol.* **20**(7), S121 (2005)
609. A. Leitenstorfer, S. Hunsche, J. Shah, M.C. Nuss, W.H. Knox, Femtosecond charge transport in polar semiconductors, *Phys. Rev. Lett.* **82**(25), 5140 (1999)
610. R. Huber, A. Brodschelm, F. Tauser, A. Leitenstorfer, Generation and field-resolved detection of femtosecond electromagnetic pulses tunable up to 41 THz, *Appl. Phys. Lett.* **76**(22), 3191 (2000)
611. S. Kono, M. Tani, P. Gu, K. Sakai, Detection of up to 20 THz with a low-temperature-grown GaAs photoconductive antenna gated with 15 fs light pulses, *Appl. Phys. Lett.* **77**, 4104 (2000)
612. S. Kono, M. Tani, K. Sakai, Ultrabroadband photoconductive detection: Comparison with free-space electro-optic sampling, *Appl. Phys. Lett.* **79**(7), 898 (2001)
613. S. Kono, M. Tani, K. Sakai, Coherent detection of mid-infrared radiation up to 60 THz with an LT-GaAs photoconductive antenna, *IEEE Proc. Optoelectron.* **149**, 105 (2002)
614. I. Hosako, N. Sekine, M. Patrashin, S. Saito, K. Fukunaga, Y. Kasai, P. Baron, T. Seta, J. Mendrok, S. Ochiai, H. Yasuda, in *T-ray imaging, sensing, and detection*, ed. by D. Abbott, X.C. Zhang (2007), At the dawn of a new era in terahertz technology, *Proc. IEEE* **95**, pp. 1611–1623
615. K. Reimann, R.P. Smith, A.M. Weiner, T. Elsaesser, M. Woerner, Direct field-resolved detection of terahertz transients with amplitudes of megavolts per centimeter, *Opt. Lett.* **28**, 471 (2003)
616. F. Junginger, A. Sell, O. Schubert, B. Mayer, D. Brida, M. Marangoni, G. Cerullo, A. Leitenstorfer, R. Huber, Single-cycle multiterahertz transients with peak fields above 10 MV/cm, *Opt. Lett.* **35**, 2645 (2010)
617. C. Janke, M. Först, M. Nagel, H. Kurz, A. Bartels, Asynchronous optical sampling for high-speed characterization of integrated resonant terahertz sensors, *Opt. Lett.* **30**(11), 1405 (2005)
618. T. Yasui, E. Saneyoshi, T. Araki, Asynchronous optical sampling terahertz time-domain spectroscopy for ultrahigh spectral resolution and rapid data acquisition, *Appl. Phys. Lett.* **87**, 061101 (2005)

619. A. Bartels, F. Hudert, C. Janke, T. Dekorsy, K. Köhler, Femtosecond time-resolved optical pump-probe spectroscopy at kilohertz-scan-rates over nanosecond-time-delays without mechanical delay line, *Appl. Phys. Lett.* **88**, 041117 (2006)
620. A. Bartels, R. Cerna, C. Kistner, A. Thoma, F. Hudert, C. Janke, T. Dekorsy, Ultrafast time-domain spectroscopy based on high-speed asynchronous optical sampling, *Rev. Sci. Instrum.* **78**, 035107 (2007)
621. T. Bartel, P. Gaal, K. Reimann, M. Woerner, T. Elsaesser, Generation of single-cycle THz transients with high electric-field amplitudes, *Opt. Lett.* **30**(20), 2805 (2005)
622. I.C. Ho, X. Guo, X.C. Zhang, Design and performance of reflective terahertz air-based-coherent-detection for time-domain spectroscopy, *Opt. Express* **18**, 2872 (2010)
623. T.A. Liu, M. Tani, M. Nakajima, M. Hangyo, C.L. Pan, Ultrabroadband terahertz field detection by photoconductive antennas based on multi-energy arsenic-ion-implanted GaAs and semi-insulating GaAs, *Appl. Phys. Lett.* **83**(7), 1322 (2003)
624. M. Tani, M. Herrmann, K. Sakai, Generation and detection of terahertz pulsed radiation with photoconductive antennas and its application to imaging, *Meas. Sci. Technol.* **13**, 1739–1745 (2002)
625. A. Dreyhaupt, S. Winnerl, M. Helm, T. Dekorsy, Optimum excitation conditions for the generation of high-electric-field terahertz radiation from an oscillator-driven photoconductive device, *Opt. Lett.* **31**, 1546 (2006)
626. B. Sartorius, H. Roehle, H. Künzel, J. Böttcher, M. Schlak, D. Stanze, H. Venghaus, M. Schell, All-fiber terahertz time-domain spectrometer operating at 1.5 μm telecom wavelengths, *Opt. Express* **16**, 9565 (2008)
627. Q. Chen, M. Tani, Z. Jiang, X.-C. Zhang, in *Nonlinear Optics: Materials, Fundamentals and Applications*, ed. by T. Li, vol. 46, OSA Trends in Optics and Photonics (2000), Electro-optic THz transceiver, p. MC2
628. K. Liu, J. Xu, X.C. Zhang, GaSe crystals for broadband terahertz wave detection, *Appl. Phys. Lett.* **85**(6), 863 (2004)
629. Z. Jiang, X.C. Zhang, Single-shot spatiotemporal terahertz field imaging, *Opt. Lett.* **23**(14), 1114 (1998)
630. Z. Jiang, X.C. Zhang, Electro-optic measurement of THz field pulses with a chirped optical beam, *Appl. Phys. Lett.* **72**(16), 1945 (1998)
631. P.U. Jepsen, D.G. Cooke, M. Koch, Terahertz spectroscopy and imaging - modern techniques and applications, *Laser Photonics Rev.* **5**, 124 (2010)
632. A.S. Pine, R.D. Suenram, E.R. Brown, K.A. McIntosh, A terahertz photomixing spectrometer: application to SO_2 self broadening, *J. Mol. Spectrosc.* **175**, 37 (1996)
633. S. Matsuura, P. Chen, G. Blake, J. Pearson, H.M. Pickett, A tunable cavity-locked diode laser source for terahertz photomixing, *IEEE Trans. Microw. Theory Techn.* **48**, 380 (2000)
634. I.R. Medvedev, M. Behnke, F.C.D. Lucia, Fast analysis of gases in the submillimeter/terahertz with “absolute” specificity, *Appl. Phys. Lett.* **86**, 154105 (2005)
635. G. Winnewisser, A.F. Krupnov, M.Y. Tretyakov, M. Liedtke, F. Lewen, A.H. Saleck, R. Schieder, A.K. Shkarev, S.V. Volokhov, Precision broadband spectroscopy in the terahertz region, *J. Mol. Spectrosc.* **169**, 294 (1994)
636. S.P. Belov, S. Urban, G. Winnewisser, Hyperfine structure of rotation-inversion levels in the excited ν_2 state of ammonia, *J. Mol. Spectrosc.* **189**(1), 1 (1998)
637. D.T. Petkie, T. M. Goyette, R.P.A. Bettens, S.P. Belov, S. Albert, P. Helminger, F.C. DeLucia, A fast scan submillimeter spectroscopic technique, *Rev. Sci. Instrum.* **68**, 1675 (1997)
638. M. Born, E. Wolf, *Principles of Optics* (Pergamon Press, Oxford, 1986)
639. A.A. Volkov, Y.G. Goncharov, G.V. Kozlov, S.P. Lebedev, A.M. Prokhorov, Dielectric measurements in the submillimeter wavelength region, *Infrared Phys.* **25**(1/2), 369 (1985)
640. R. Gendriesch, F. Lewen, G. Winnewisser, J. Hahn, Precision broadband spectroscopy near 2 THz: frequency-stabilized laser sideband spectrometer with backward-wave oscillators, *J. Mol. Spectrosc.* **203**, 205 (2000)
641. D. Boucher, R. Bocquet, J. Burie, W. Chen, A far-infrared heterodyne sidebands spectrometer, *J. Phys. III* **4**, 1467 (1994)

642. M. Heiblum, S. Wang, J.R. Whinnery, T.K. Gustafson, Characteristics of integrated MOM junctions at dc and optical frequencies, *IEEE J. Quantum Electron.* **14**, 159 (1978)
643. K.M. Evenson, M. Inguscio, D.A. Jennings, Point contact diode at laser frequencies, *J. Appl. Phys.* **57**, 956 (1985)
644. T.D. Varberg, K.M. Evenson, Laser spectroscopy of carbon monoxide: a frequency reference for the far infrared, *IEEE Trans. Instrumentation Measurement* **42**, 412 (1993)
645. T.D. Varberg, K.M. Evenson, Accurate far-infrared rotational frequencies of carbon monoxide, *Astrophys. J.* **385**, 763 (1992)
646. P. de Natale, L. Lorini, M. Inguscio, G.D. Lonardo, L. Fusina, P.A.R. Ade, A.G. Murray, Improved sensitivity of tunable far-infrared spectroscopy: application to the detection of HBr in the $\nu = 1$ state, *Appl. Opt.* **36**, 5822 (1997)
647. G. Mouret, S. Matton, R. Bocquet, F. Hindle, E. Peytavit, J.F. Lampin, D. Lippens, Far-infrared cw difference-frequency generation using vertically integrated and planar low temperature grown GaAs photomixers: application to H_2S rotational spectrum up to 3 THz, *Appl. Phys. B* **79**, 725 (2004)
648. S. Matsuura, M. Tani, H. Abe, K. Sakai, H. Ozeki, S. Saito, High-resolution terahertz spectroscopy by a compact radiation source based on photomixing with diode lasers in a photoconductive antenna, *J. Mol. Spectrosc.* **187**, 97 (1998)
649. R.T. Boreiko, A.L. Betz, J. Zmuidzinas, Heterodyne spectroscopy of the 158 micron CII line in M42, *Astrophys. J.* **325**, L47 (1988)
650. R.T. Boreiko, A.L. Betz, Heterodyne spectroscopy of the $63\ \mu\text{m}$ OI line in M42, *Astrophys. J.* **464**, L83 (1996)
651. J. Mees, S. Crewell, H. Nett, G. de Lange, H.V. de Stadt, J. Kuipers, R.A. Panhuyzen, ASUR—an airborne SIS receiver for atmospheric measurements of trace gases at 625 to 760 GHz, *IEEE Trans. Microw. Theory Techn.* **43**, 2543 (1995)
652. S. Gulkis, M. Allen, C. Backus, G. Beaudin, N. Biver, D. Bockelee-Morvan, J. Crovisier, D. Despois, P. Encrenaz, M. Frerking, M. Hofstadter, P. Hartogh, W. Ip, M. Janssen, L. Kamp, T. Koch, E. Lellouch, I. Mann, D. Muhleman, H. Rauer, P. Schloerb, T. Spilker, Remote sensing of a comet at millimeter and submillimeter wavelengths from an orbiting spacecraft, *Planet. Space Sci.* **55**, 1050 (2007)
653. R. Blundell, C.Y.E. Tong, Submillimeter receivers for radio astronomy, *Proc. IEEE* **80**, 1702 (1992)
654. J.E. Carlstrom, J. Zmuidzinas, in *Reviews of Radio Science 1993-1996*, ed. by W.R. Stone (Oxford University Press, Oxford U.K., 1996), chap. Millimeter and submillimeter techniques, pp. 839–882
655. H.-W. Hübers, Terahertz heterodyne receivers, *IEEE Sel. Top. Quant. Electron.* **14**, 378 (2008)
656. D.W. Allan, Statistics of atomic frequency standards, *Proc. IEEE* **54**(2), 221 (1966)
657. R. Schieder, G. Rau, B. Vowinkel, in *Instrumentation for Submillimeter Spectroscopy*, Characterization and measurement of system stability, *Proc. SPIE* **598**, 189 (1985)
658. R. Schieder, C. Kramer, Optimization of heterodyne observations using Allan variance measurements, *Astron. Astrophys.* **373/2**, 746 (2001)
659. J.W. Kooi, G. Chattopadhyay, M. Thielmann, T.G. Phillips, R. Schieder, Noise stability of SIS receivers noise stability, **21**, 689 (2000)
660. J.W. Kooi, J.J.A. Baselmans, A. Baryshev, R. Schieder, M. Hajenius, J.R. Gao, T.M. Klapwijk, B. Voronov, G. Goltsman, Stability of heterodyne terahertz receivers, *J. Appl. Phys.* **100**, 064904 (2006)
661. H. Penfield, in *Methods in Experimental Physics*, ed. by M.L. Meeks (Academic Press, New York, 1976), vol. 12, part. B, chap. 3.4. - Multichannel-Filter Spectrometers, pp. 266–279
662. J. Horn, O. Siebertz, F. Schmülling, C. Kunz, R. Schieder, G. Winnewisser, A 4x1 GHz array acousto-optical spectrometer, *Exp. Astron.* **9**, 17 (1999)
663. G.J. Melnick, J.R. Stauffer, M.L.N. Ashby, E.A. Bergin, G. Chin, N.R. Erickson, P.F. Goldsmith, M. Harwit, J.E. Howe, S.C. Kleiner, D.G. Koch, D.A. Neufeld, B.M. Patten, R. Plume, R. Schieder, R.L. Snell, V. Tolls, Z. Wang, G. Winnewisser, Y.F. Zhang, The Submillimeter Wave Astronomy Satellite: Science objectives and instrument description, *Astrophys. J.* **539**, L77 (2000)

664. U. Frisk, M. Hagström, J. Ala-Laurinaho, S. Andersson, J.C. Berges, J.P. Chabaud, M. Dahlgren, A. Emrich, H.G. Florn, G. Florin, M. Fredrixon, T. Gaier, R. Haas, T. Hirvonen, A. Hjalmarsson, B. Jakobsson, P. Jukkala, P.S. Kildal, E. Kollberg, J. Lassing, A. Lecacheux, P. Lehtikoinen, A. Lehto, J. Mallat, C. Marty, D. Michet, J. Narbonne, M. Nexon, M. Olberg, A.O.H. Olofsson, G. Olofsson, A. Origine, M. Petersson, P. Piironen, R. Pons, D. Pouliquen, I. Ristorcelli, C. Rosolen, G. Rouaix, A.V. Räisänen, G. Serra, F. Sjöberg, L. Stenmark, S. Torchinsky, J. Tuovinen, C. Ullberg, E. Vinterhav, N. Wadefalk, H. Zirath, P. Zimmermann, R. Zimmermann, The Odin satellite I. Radiometer design and test, *Astron. Astrophys.* **402**(3), L27 (2003)
665. A. Emrich, in *Proc. ESA Symp. The Far InfraRed and Submillimetre Universe* (ESA, Grenoble, 1997), vol. ESA SP401, Autocorrelation spectrometers for space borne (sub)millimeter spectroscopy, pp. 361–364
666. A.I. Harris, in *Proc. Conf. on Millimeter and Submillimeter Detectors for Astronomy* (Waikoloa, 2002), Heterodyne spectrometers with very wide bandwidths, *Proc. SPIE* **4855**, pp. 279–289
667. G. Villanueva, P. Hartogh, The high resolution chirp transform spectrometer for the SOFIA-GREAT instrument, *Exp. Astron.* **18**, 77 (2004)
668. B. Klein, S.D. Philipp, I. Krämer, C. Kasemann, R. Güsten, K.M. Menten, The APEX digital Fast Fourier Transform spectrometer, *Astron. Astrophys.* **454**(2), L29 (2006)
669. J.W. Waters, L. Froidevaux, R.S. Harwood, R.F. Jarrot, H.M. Pickett, W.G. Read, P.H. Siegel, R.E. Cofield, M.J. Filipiak, D.A. Flower, J.R. Holden, G.K. Lau, N.J. Livesey, G.L. Manney, H.C. Pumphrey, M.L. Santee, D.L. Wu, D.T. Cuddy, R.R. Lay, M.S. Loo, V.S. Perun, M.J. Schwartz, P.C. Stek, R.P. Thurstans, M.A. Boyles, K.M. Chandra, M.C. Chavez, G.S. Chen, B.V. Chudasama, R. Dodge, R.A. Fuller, M.A. Girard, J.H. Jiang, Y. Jiang, B.W. Knosp, R.C. LaBelle, J.C. Lam, K.A. Lee, D. Miller, J.E. Oswald, N.C. Patel, D.M. Pukala, O. Quintero, D.M. Scaff, W.V. Snyder, M.C. Tope, P.A. Wagner, M.J. Walch, The Earth Observing Microwave Limb Sounder (EOS MLS) on the Aura satellite, *IEEE Trans. Geosci. Rem. Sens.* **44**(5), 1075 (2006)
670. E.A. Bergin, T.G. Phillips, C. Comito, N.R. Crockett, D.C. Lis, P. Schilke, S. Wang, T.A. Bell, G.A. Blake, B. Bumble, E. Caux, S. Cabrit, C. Ceccarelli, J. Cernicharo, F. Daniel, T. de Graauw, M.L. Dubernet, M. Emprechtinger, P. Encrenaz, E. Falgarone, M. Gerin, T.F. Giesen, J.R. Goicoechea, P.F. Goldsmith, H. Gupta, P. Hartogh, F.P. Helmich, E. Herbst, C. Joblin, D. Johnstone, J.H. Kawamura, W.D. Langer, W.B. Latter, S.D. Lord, S. Maret, P.G. Martin, G.J. Melnick, K.M. Menten, P. Morris, H.S.P. Müller, J.A. Murphy, D.A. Neufeld, V. Ossenkopf, L. Pagani, J.C. Pearson, M. Pérou, R. Plume, P. Roelfsema, S.L. Qin, M. Salez, S. Schlemmer, J. Stutzki, A.G.G.M. Tielens, N. Trappe, F.F.S. van der Tak, C. Vastel, H.W. Yorke, S. Yu, J. Zmuidzinas, Herschel observations of EXtra-Ordinary Sources (HEXOS): the present and future of spectral surveys with Herschel/HIFI, *A&A* **521**, L20 (2010)
671. T. de Graauw, F.P. Helmich, T.G. Phillips, J. Stutzki, E. Caux, N.D. Whyborn, P. Dieleman, P.R. Roelfsema, H. Aarts, R. Assendorp, R. Bachiller, W. Baechtold, A. Barcia, D.A. Beintema, V. Belitsky, A.O. Benz, R. Bieber, A. Boogert, C. Borys, B. Bumble, P. Cas, M. Caris, P. Cerulli-Irelli, G. Chattopadhyay, S. Cherednichenko, M. Ciechanowicz, O. Coeur-Joly, C. Comito, A. Cros, A. de Jonge, G. de Lange, B. Delforges, Y. Delorme, T. den Boggende, J.M. Desbat, C. Diez-González, A.M. di Giorgio, L. Dubbeldam, K. Edwards, M. Eggens, N. Erickson, J. Evers, M. Fich, T. Finn, B. Franke, T. Gaier, C. Gal, J.R. Gao, J.D. Gallego, S. Gauffre, J.J. Gill, S. Glenz, H. Golstein, H. Gouloozee, T. Gunsing, R. Güsten, P. Hartogh, W.A. Hatch, R. Higgins, E.C. Honingh, R. Huisman, B.D. Jackson, H. Jacobs, K. Jacobs, C. Jarchow, H. Javadi, W. Jellema, M. Justen, A. Karpov, C. Kasemann, J. Kawamura, G. Keizer, D. Kester, T.M. Klapwijk, T. Klein, E. Kollberg, J. Kooi, P.P. Kooiman, B. Kopf, M. Krause, J.M. Krieg, C. Kramer, B. Kruizenga, T. Kuhn, W. Laauwen, R. Lai, B. Larsson, H.G. Leduc, C. Leinz, R.H. Lin, R. Liseau, G.S. Liu, A. Loose, I. Lopez-Fernandez, S. Lord, W. Luinge, A. Marston, J. Martin-Pintado, A. Maestrini, F.W. Maiwald, C. McCoe, I. Mehdi, A. Megej, M. Melchior, L. Meisma, H. Merkel, M. Michalska, C. Monstein, D. Moratschke, P. Morris, H. Muller,

- J.A. Murphy, A. Naber, E. Natale, W. Nowosielski, F. Nuzzolo, M. Olberg, M. Olbrich, R. Orfei, P. Orleanski, V. Ossenkopf, T. Peacock, J.C. Pearson, I. Peron, S. Phillip-May, L. Piazzo, P. Planesas, M. Rataj, L. Ravera, C. Risacher, M. Salez, L.A. Samoska, P. Saraceno, R. Schieder, E. Schlecht, F. Schlöder, F. Schmülling, M. Schultz, K. Schuster, O. Siebertz, H. Smit, R. Szczerba, R. Shipman, E. Steinmetz, J.A. Stern, M. Stokroos, R. Teipen, D. Teyssier, T. Tils, N. Trappe, C. van Baaren, B.J. van Leeuwen, H. van de Stadt, H. Visser, K.J. Wildeman, C.K. Wafelbakker, J.S. Ward, P. Wesselius, W. Wild, S. Wulff, H.J. Wunsch, X. Tielens, P. Zaal, H. Zirath, J. Zmuidzinas, F. Zwart, The Herschel-Heterodyne Instrument for the Far-Infrared (HIFI), *Astron. Astrophys.* **518**, L6 (2010)
672. U.U. Graf, S. Heyminck, R. Güsten, P. Hartogh, H.-W. Hübers, K. Jacobs, M. Philipp, D. Rabanus, H.P. Röser, J. Stutzki, P. van der Wal, A. Wagner-Gentner, in *Proc Conf. on Millimeter and Submillimeter Detectors for Astronomy III*, ed. by J. Zmuidzinas, W.S. Holland, S. Withington, W. Duncan (2006), GREAT: the German first light heterodyne instrument for SOFIA, *Proc. SPIE* **6275**, pp. 62750K
673. E. Heinz, D. Born, G. Zieger, T. May, T. Krause, A. Krüger, M. Schulz, S. Anders, V. Zakosarenko, H.G. Meyer, M. Starkloff, M. Rössler, G. Thorwirth, U. Krause, Progress report on Safe VISITOR: approaching a practical instrument for terahertz security screening, *Proc. SPIE* **7670**, 767005 (2010)
674. T. May, G. Zieger, S. Anders, V. Zakosarenko, H.G. Meyer, M. Schubert, M. Starkloff, M. Rößler, G. Thorwirth, U. Krause, in *Passive Millimeter-Wave Imaging Technology XII*, ed. by R. Appleby, D.A. Wikner (2009), Safe VISITOR: VISible, Infrared and Terahertz Object Recognition for security screening applications, *Proc. SPIE* **7309**, 73090E
675. N. Karpowicz, H. Zhong, C. Zhang, K. Lin, J.S. Hwang, J. Xu, X.C. Zhang, Compact continuous-wave subterahertz system for inspection applications, *Appl. Phys. Lett.* **86**, 054105 (2005)
676. C. Jördens, S. Wietzke, M. Salhi, R. Wilk, M. Koch, Potenziale der bildgebenden Terahertz-Spektroskopie, *Technisches Messen* **75**, 71 (2008)
677. R.M. Woodward, B.E. Cole, V.P. Walker, R.J. Pye, D.D. Arnone, E.H. Linfield, M. Pepper, Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue, *Phys. Med. Biol.* **47**, 3853 (2002)
678. W.J. Smith, *Modern optical engineering*, 2nd edn. (McGraw-Hill, 1990)
679. T.G. Blaney, Radiation detection at submillimetre wavelengths, *J. Phys. E: Sci. Instrum.* **11**, 856 (1978)
680. R. Appleby, H.B. Wallace, Standoff detection of weapons and contraband in the 100 GHz to 1 THz region, *IEEE Trans. Antennas Propagation* **55**, 2944 (2007)
681. A.L. Woodcraft, M.I. Hollister, D. Bintley, M.A. Ellis, X. Gao, W.S. Holland, M.J. MacIntosh, P.A.R. Ade, J.S. House, C.L. Hunt, R.V. Sudiwala, W.D. Duncan, G.C. Hilton, K.D. Irwin, C.D. Reintsema, C.C. Dunare, W. Parkes, A.J. Walton, J.B. Kycia, M. Amiri, B. Burger, M. Halpern, in *Millimeter and Submillimeter Detectors and Instrumentation for Astronomy III*, ed. by J. Zmuidzinas, W.S. Holland, S. Withington, W.D. Duncan (2006), Characterization of a prototype SCUBA-2 1280 pixel submillimetre superconducting bolometer array, *Proc. SPIE* **6275**, 62751F
682. A. Luukanen, L. Grönberg, M. Grönholm, P. Lappalainen, M. Leivo, A. Rautiainen, A. Tamminen, J. Ala-Laurinaho, C.R. Dietlein, E.N. Grossman, Real-time passive terahertz imaging system for standoff concealed weapons imaging, *Proc. SPIE* **7670**, 767004 (2010)
683. H.-W. Hübers, A.D. Semenov, H. Richter, U. Böttger, in *Terahertz for Military and Security Applications V*, ed. by J.O. Jensen, H.L. Cui, Terahertz imaging system for stand-off detection of threats, *Proc. SPIE* **6549**, 65490A (2007)
684. C.C. Franck, D. Lee, R.L. Espinola, S.R. Murrill, E.L. Jacobs, S.T. Griffin, D.T. Petkie, J. Reynolds, in *Terahertz for Military and Security Applications V*, ed. by J.O. Jensen, H.L. Cui, Terahertz standoff imaging testbed design and performance for concealed weapon and device identification model development, *Proc. SPIE* **6549**, 654908 (2007)
685. K.S. Yngvesson, in *Infrared and Millimeter Waves*, ed. by K.J. Button (Academic Press, New York, 1983), vol. 10, chap. 1 - Near-millimeter imaging with integrated planar receptors: General requirements and constraints, pp. 91

686. S. Heyminck, U.U. Graf, Reflection gratings as THz local oscillator multiplexers, in airborne telescope systems, Proc. SPIE **4014**, 164 (2000)
687. C. Walker, C. Kulesa, J. Kloosterman, D. Lesser, T. Cottam, C. Groppi, J. Zmuidzinas, M. Edgar, S. Radford, P. Goldsmith, W. Langer, H. Yorke, J. Kawamura, I. Mehdi, D. Hollenbach, J. Stutzki, H.-W. Hübers, J. Gao, C. Martin, in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy V*, ed. by W.S. Holland, J. Zmuidzinas, Large Format Heterodyne Arrays for Observing Far-Infrared Lines with SOFIA, Proc. SPIE, **7741**, 77,410Z (2010)
688. D.F. Filipovic, G.P. Gauthier, S. Raman, G.M. Rebeiz, Off-axis properties of silicon and quartz dielectric lens antennas, IEEE Trans. Antennas Propagation **45**, 760 (1997)
689. A.D. Semenov, H. Richter, H.-W. Hübers, B. Günther, A. Smirnov, K. Ilin, M. Siegel, J. Karamarkovic, Terahertz performance of integrated lens antennas with a hot electron bolometer, IEEE Trans. Microw. Theory Techn. **55**(2), 239 (2007)
690. K.B. Cooper, R.J. dengler, N. Llombart, T. Bryllert, G. Chattopadhyay, E. Schlecht, J. Gill, C. Lee, A. Skalare, I. Mehdi, P.H. Siegel, Penetrating 3-D imaging at 4- and 25-m range using submillimeter-wave radiation, IEEE Trans. Microw. Theory Techn. **56**, 2771 (2008)
691. D.M. Mittleman, S. Hunsche, L. Boivin, M.C. Nuss, T-ray tomography, Opt. Lett. **22**, 904 (1997)
692. S. Wang, X.C. Zhang, Pulsed terahertz tomography, J. Phys. D: Appl. Phys. **37**, R1 (2004)
693. K. McClatchey, M.T. Reiten, R.A. Cheville, Time resolved synthetic aperture terahertz impulse imaging, Appl. Phys. Lett. **79**, 4485 (2001)
694. J. Xu, X.C. Zhang, Terahertz wave reciprocal imaging, Appl. Phys. Lett. **88**, 151107 (2006)
695. W.L. Chan, K. Charan, D. Takhar, K.F. Kelly, R.G. Baraniuk, D.M. Mittleman, A single-pixel terahertz imaging system based on compressed sensing, Appl. Phys. Lett. **93**, 121105 (2008)
696. A. Dobroui, M. Yamashita, Y.N. Ohshima, Y. Morita, C. Otani, K. Kawase, Terahertz imaging system based on a backward-wave oscillator, Appl. Opt. **43**, 5637 (2004)
697. S. Ariyoshi, C. Otani, H. Sato, K. Kawase, H.M. Shimizu, T. Taino, H. Matsuo, Terahertz imaging with a direct detector based on superconducting tunnel junctions, Appl. Phys. Lett. **88**, 203503 (2006)
698. K.J. Siebert, H. Quast, R. Leonhardt, T. Löffler, M. Thomson, S. Czasch, Continuous-wave all-optoelectronic terahertz imaging, Appl. Phys. Lett. **80**, 3003 (2002)
699. I.S. Gregory, W.R. Tribe, C. Baker, B.E. Cole, M.J. Evans, L. Spencer, M. Pepper, M. Missous, Continuous-wave terahertz system with a 60 dB dynamic range, Appl. Phys. Lett. **86**, 204104 (2005)
700. J. Darro, V. Tamosiunas, G. Fasching, J. Kröll, K. Unterrainer, M. Beckk, M. Giovannini, J. Faist, C. Kremser, P. Debagge, Imaging with a THz quantum cascade laser, Opt. Express **12**, 1879 (2004)
701. S. Barbieri, J. Alton, C. Baker, T. Lo, H.E. Beere, D. Ritchie, Imaging with THz quantum cascade lasers using a Schottky diode mixer, Opt. Express **13**, 6497 (2005)
702. K. Kawase, Y. Ogawa, Y. Watanabe, H. Inoue, Non-destructive terahertz imaging of illicit drugs using spectral fingerprints, Opt. Express **11**, 2549 (2003)
703. K.L. Nguyen, M.L. Johns, L.F. Gladden, C.H. Worral, P. Alexander, H.E. Beere, M. Pepper, D.A. Ritchie, J. Alton, S. Barbieri, E.H. Linfield, Three-dimensional imaging with terahertz quantum cascade laser, Opt. Express **14**, 2123 (2006)
704. G.T. Herman, Advances in Computer Vision and Pattern Recognition in *Fundamentals of Computerized Tomography*, ed. by S. Singh, 2nd edn. (Springer, London, 2010, originally published by Academic Press, New York, 1980)
705. W.L. Chan, J. Deibel, D.M. Mittleman, Imaging with terahertz radiation, Rep. Prog. Phys. **70**, 1325 (2007)
706. A.J. Fitzgerald, E. Berry, R.E. Miles, N.N. Zinovev, M.A. Smith, J.M. Chamberlain, Evaluation of image quality in terahertz pulsed imaging using test objects, Phys. Med. Biol. **47**, 3865 (2002)
707. L. Duvillaret, F. Garet, J.L. Coutaz, Influence of noise on the characterization of materials by terahertz time-domain spectroscopy, J. Opt. Soc. Am. B **17**, 452 (2000)

708. Y.C. Shen, T. Lo, P.F. Taday, B.E. Cole, W.R. Tribe, M.C. Kemp, Detection and identification of explosives using terahertz pulsed spectroscopic imaging, *Appl. Phys. Lett.* **86**, 241116 (2005)
709. J.L. Johnson, T.D. Dorney, D.M. Mittleman, Enhanced depth resolution in terahertz imaging using phase-shift interferometry, *Appl. Phys. Lett.* **78**, 835 (2001)
710. Y. Kawada, T. Yasuda, A. Nakanishi, H. Takahashi, S.-ichiro Aoshima, Single-shot measurement of terahertz temporal waveform using pulse-front tilting by a direct vision dispersion prism, *Rev. Sci. Instrum.* **80**, 113703 (2009)
711. Z. Jiang, X.C. Zhang, Terahertz imaging via electrooptic effect, *IEEE Trans. Microw. Theory Techn.* **47**, 2644 (1999)
712. E.H. Syngé, A suggested method for extending the microscopic resolution into the ultramicroscopic region, *Phil. Mag.* **6**, 356 (1928)
713. H.A. Bethe, Theory of diffraction by small holes, *Phys. Rev.* **66**, 163 (1944)
714. C.J. Bouwkamp, On Bethe's theory of diffraction by small holes, *Philips Res. Rep.* **5**, 321 (1950)
715. G.A. Massey, J.A. Davis, S.M. Katnik, E. Omon, Subwavelength resolution far-infrared microscopy, *Appl. Opt.* **24**, 1498 (1985)
716. O. Mitrofanov, M. Lee, J.W.P. Hsu, L.N. Pfeiffer, K.W. West, J.D. Wynn, J. Frederici, Terahertz pulse propagation through small apertures, *Appl. Phys. Lett.* **79**, 907 (2001)
717. N.C.J. van der Valk, P.C.M. Planken, Electro-optic detection of subwavelength terahertz spot sizes in the near field of a metal tip, *Appl. Phys. Lett.* **81**(9), 1558 (2002)
718. H.T. Chen, R. Kersting, G.C. Cho, Terahertz imaging with nanometer resolution, *Appl. Phys. Lett.* **83**(15), 3009 (2003)
719. F. Keilmann, FIR microscopy, *Infrared Phys. Technol.* **36**(1), 217 (1995)
720. R. Merz, F. Keilmann, R.J. Haug, K. Ploog, Non-equilibrium edge state transport resolved by far-infrared microscopy, *Phys. Rev. Lett.* **70**, 651 (1993)
721. S. Hunsche, M. Koch, I. Brener, M.C. Nuss, THz near-field imaging, *Opt. Comm.* **150**, 22 (1998)
722. O. Mitrofanov, I. Brener, R. Harel, J.D. Wynn, L.N. Pfeiffer, K.W. West, J. Frederici, Terahertz near-field microscopy based on a collection mode detector, *Appl. Phys. Lett.* **77**, 3496 (2000)
723. Q. Chen, Z. Jiang, G.X. Xu, X.C. Zhang, Near field terahertz imaging with a dynamical aperture, *Opt. Lett.* **25**, 1122 (2000)
724. S. Mair, B. Gompf, M. Dressel, Spatial and spectral behaviour of the optical near field studied by a terahertz near field spectrometer, *Appl. Phys. Lett.* **84**, 1219 (2004)
725. B. Gompf, M. Gerull, T. Müller, M. Dressel, THz-micro-spectroscopy with backward wave oscillators, *Infrared Phys. Technol.* **49**, 128 (2006)
726. D.V. Palanker, G.H.M. Knippels, T.I. Smith, H.A. Schwetman, Fast IR imaging with sub-wavelength resolution using a transient near field probe, *Nucl. Instrum. Meth. Phys. Res. B* **144**, 240 (1998)
727. B. Gompf, N. Gebert, H. Heer, M. Dressel, Polarization contrast terahertz-near-field imaging of anisotropic conductors, *Appl. Phys. Lett.* **90**, 082104 (2007)
728. E. Bründermann, M. Havenith, SNIM - Scanning near-field infrared microscopy in *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.*, vol. 104 (The Royal Society of Chemistry, 2008), pp. 235–255
729. F. Keilmann, R. Hillenbrand, in *Nano-Optics and Near-Field Optical Microscopy*, ed. by A. Zayats, D. Richards (ArtechHouse, 2009), chap. 11 - Near-field nanoscopy by elastic light scattering from a tip, pp. 235–265
730. A.J. Huber, F. Keilmann, J. Wittborn, J. Aizpurua, R. Hillenbrand, Terahertz near-field nanoscopy of mobile carriers in single semiconductor nanodevices, *Nano Lett.* **8**(11), 3766 (2008)
731. S.C. Kehr, M. Cebula, O. Mieth, T. Härtling, J. Seidel, S. Grafström, L.M. Eng, S. Winnerl, D. Stehr, M. Helm, Anisotropy contrast in phonon-enhanced apertureless near-field microscopy using a free-electron laser, *Phys. Rev. Lett.* **100**(25), 256403 (2008)

- 732. H.G. von Ribbeck, M.T. Wenzel, R. Jacob, L.M. Eng, Scattering near-field microscopy in the THz region using a free-electron laser in *35th International Conference on Infrared Millimeter and Terahertz Waves (IRMMW-THz, 2010)*, Tu-E2.4
- 733. K. Wang, D.M. Mittleman, N.C.J.V. der Valk, P.C.M. Planken, Antenna effects in terahertz apertureless near-field optical microscopy, *Appl. Phys. Lett.* **85**(14), 2715 (2004)
- 734. K. Wang, A. Barkan, M. Mittleman, Propagation effects in apertureless near-field optical antennas, *Appl. Phys. Lett.* **84**(2), 305 (2004)
- 735. G.C. Cho, H.T. Chen, S. Kraatz, N. Karpowicz, R. Kersting, Apertureless terahertz near-field microscopy, *Semicond. Sci. Technol.* **20**(7), S286 (2005)
- 736. F. Buersegens, R. Kersting, H.T. Chen, Terahertz microscopy of charge carriers in semiconductors, *Appl. Phys. Lett.* **88**(11), 112115 (2006)
- 737. N. Ocelic, A. Huber, R. Hillenbrand, Pseudoheterodyne detection for background-free near-field spectroscopy, *Appl. Phys. Lett.* **89**, 101124 (2006)
- 738. I. Kopf, C. Grunwald, E. Bründermann, L. Casalis, G. Scoles, M. Havenith, Detection of hybridization on nanografted oligonucleotides using scanning near-field infrared microscopy, *J. Phys. Chem. C* **114**(2), 1306 (2010)
- 739. T. Yamaguchi, S. Yoshida, A. Kinbara, Optical effect of the substrate on the anomalous absorption of aggregated silver films, *Thin Solid Films* **21**, 173 (1974)
- 740. B. Knoll, F. Keilmann, Near-field probing of vibrational absorption for chemical microscopy, *Nature* **399**, 134 (1999)
- 741. A. Cvitkovic, N. Ocelic, R. Hillenbrand, Analytical model for quantitative prediction of material contrasts in scattering-type near-field optical microscopy, *Opt. Express* **15**(14), 8550 (2007)
- 742. G. Wollny, E. Bründermann, Z. Arsov, L. Quaroni, M. Havenith, Nanoscale depth resolution in scanning near-field infrared microscopy, *Opt. Express* **16**, 7453 (2008)
- 743. R. Hillenbrand, F. Keilmann, Complex optical constants on a subwavelength scale, *Phys. Rev. Lett.* **85**(14), 3029 (2000)
- 744. E. Bründermann, D.A. Schmidt, I. Kopf, M. Havenith, Nano-spectroscopy and chemical nanoscopy of biomaterials, *Am. Inst. Phys. Conf. Proc.* **1214**, 7 (2010)
- 745. S. Amarie, T. Ganz, F. Keilmann, Mid-infrared near-field spectroscopy, *Opt. Express* **17**, 21794 (2009)

Index

Symbols

M^2 -value 28
 f -number 28

A

Absorber 85
Absorption coefficient 52
Acceptance solid angle 173
Acousto-optical spectrometer 297
Airy disc 304
Akari (name of Japanese satellite, meaning:
bright, light) 19
Alkali halide 60
Allan time 295
Allan variance 295
Amplified spontaneous emission 118
Amplitude noise 296
Antenna
double-dipole 89
double-slot 89
Gaussian beam 88
horn 91
hybrid 86
integrated 86
logarithmic-periodic 89
logarithmic-spiral 89
planar 87, 89
quasi-optical 89
Antireflection coating 71
grooves 72
materials 72
Apodization 260

Astronomy 293
imaging 311
spectroscopy 299
Asynchronous optical sampling 275
Atacama Large Millimeter Array (ALMA)
300
Atmosphere
absorption 1, 2, 8
transmission 1–3
Attenuator 84
materials 84
wire grid 84
Auston switch 18, 272
Autocorrelator 298

B

Back-end spectrometer 297
Backward wave oscillator 13, 144, 286, 318
frequency tuning 147
linewidth 147
power 147
Band of response 173
Beam propagation factor 28
Beam splitter 83
Bidirectional reflectance distribution function
86
Biomedicine 302
Blackbody
calibration 86
BLIP 176
Blocked impurity band detector 218
Bolometer 181, 312
array 201
composite 196

- electrothermal feedback 199
- Ge 14, 195
- hot electron 157, 213, 239, 335
- InSb 14, 212, 239
- room temperature 202
- semiconducting 195
- Si 14, 195
- superconducting 11, 198
- voltage biased superconducting 199

C

- Calorimeter 192
- Camera 311
 - security 314
- Carcinotron 13, 144
- Chemical microscope 335
- Chemical nanoscope 335
- Chirp transform spectrometer 298
- CO₂ laser 136, 291
- Coherence
 - spatial 258
 - temporal 258
- Coherent synchrotron radiation 18, 155
- Cone 332
- Contrast 306, 308
- Cryocooler 172
- Cryostat 171

D

- Dee-star 175
- Depth of focus 28
- Detectivity 175
- Detector
 - blocked impurity band 218
 - cooling 171
 - direct 308
 - GaAs 217
 - Ge 209
 - Golay 183
 - HgCdTe 217
 - InSb 212
 - photoconductive 203
 - pyroelectric 185
 - response time 181, 182
 - thermal 169, 180
- Diamond 59
- Difference frequency generation 136, 291
- Diffraction limit 308

- Diode
 - heterostructure barrier varactor 142
 - metal-insulator-metal 136, 139, 291
 - Schottky 136, 138, 142
- Diplexer 266, 289
- Direct detector 174, 308, 311
- Dynamic range 173

E

- Edge spread function 306
- Electro-optic sampling 330
- Electron storage ring 156
- Extinction coefficient 52

F

- Fabry–Pérot interferometer 259, 267, 290
 - finesse 268
 - free spectral range 268
- Felgett advantage 249
- Filter 63
 - absorption 64
 - alkali halide 65
 - band-pass 64, 68
 - black polyethylene 64
 - cold 71
 - long-pass 64
 - mesh 67
 - neutral density 84
 - reststrahlen 60, 67
 - scattering 65
 - short-pass 63
 - thick metal plate 70
 - Yoshinaga 65
- Filter bank spectrometer 297
- Fluorogold 57
- Fourier-transform spectrometer 12, 17, 257
 - digital 298
 - dispersive 14, 262
 - multiplex advantage 257
- Free electron laser 17, 160
 - frequency tuning 161
 - gain 161
- Fresnel lens 73

G

- GaAs
 - detector 217
 - low temperature grown 137

Gas laser 269, 321
 efficiency 114
 electrically excited 109
 frequencies 115
 gain 114
 local oscillator 118
 optically excited 112
 pulse 118
 resonator 117
 resonator interferogram 116
 Ge 59
 Ge detector 209
 Ge laser 121
 Globar 106, 107
 Golay detector 11, 183
 Grating 8
 blazed 8, 253
 Grating spectrometer 251
 Czerny-Turner 255
 Ebert-Fastie 255
 Gyrotron 151

H

Hadamard matrix 256
 Half-wave plate 81
 Harmonic generation 141
 Heat trap 79
 Herschel Space Observatory 172, 198,
 239, 245, 299, 311–313
 Heterodyne array 315
 Heterodyne imaging 314, 316
 Heterodyne receiver 15, 17
 Heterodyne spectrometer 221, 266,
 293, 298
 back-end 222
 front-end 221
 mixing 222
 theory 222
 HIFI (Heterodyne Instrument for the
 Far-Infrared) 239, 245
 Hot electron bolometer
 InSb 15, 17, 213
 NbN 18, 234
 superconducting 18, 239
 Hot electron bolometer mixer 239
 diffusion-cooled 241
 hot spot 242
 NbN 243
 noise 241
 noise temperature 234, 244
 phonon-cooled 241, 243

I

Imaging 15
 active 303, 309
 astronomy 311
 backward wave oscillator 318
 bolometer 311
 camera-like 307
 direct detector 311
 heterodyne 314, 316
 passive 303, 309
 photodetector 312
 photomixer 319
 quantum cascade laser 321, 323
 real time 321, 330
 security 315
 short-range 317
 spectrometer 310
 spectroscopic 323, 327
 three-dimensional 316, 323, 328
 time domain spectrometer 325, 327, 328,
 335
 tomography 323
 video-rate 310, 321, 330
 Impurity
 excited states 205
 hydrogenic 205
 shallow 205
 InSb detector 212
 frequency tunable 214
 mixer 216
 NEP 214
 Integrating cavity 79
 Interferogram 265
 Interferometer 247
 Martin-Puplett 263
 Michelson 248
 polarizing 263
 Intermediate frequency 223, 294, 297
 Interstellar medium 15
 IRAS (Infrared Astronomical Satellite) 17

J

Jacquinot advantage 249

K

K-mirror 83
 KAO (Kuiper Airborne Observatory) 17

Knife edge method 305, 320
KRS-5 61

L

Laser

CO₂ 255
femtosecond 269
gas 269
HCN 14
optically excited 15
p-type Ge 121
quantum cascade 128
semiconductor 120
Si 126
TEA 119

Lens

absorption loss 73
antireflection coating 89
elliptical 88
hemispherical 88
hyperhemispherical 89
materials 73
zoned 73

Lens-antenna-system 86

Light concentrator 78

linear 79
parabolic 79

Light pipe 76

Linac 159

Line spread function 306

Linear accelerator 159

Local oscillator 227, 266, 295

backward wave oscillator 147
gas laser 118, 299
multiplier based 228, 296, 300
quantum cascade laser 136

M

Mach-Zehnder interferometer 259, 287

Martin-Puplett interferometer 290

Mercury arc lamp 107

Mesh 67

band-pass 68, 70
capacitive 67–69
equivalent circuit 69
filter 15, 67
inductive 67–69
outcoupler 118

Metal

reflectance 52
reflectivity 8

Metal halide lamp 109

Michelson interferometer 248

Microbolometer 202, 321

Microscope

chemical 335

Microscopy

aperture 331
near-field 331

Mirror 74

distortion 75
ellipsoidal 75
off-axis paraboloid 75
paraboloidal 75
rooftop 82
surface accuracy 74
surface roughness 75

Mixer 170

conversion loss 224
double-sideband 225
hot electron bolometer 239
noise temperature 224
Schottky diode 228
single-sideband 223
square-law 225
superconductor-insulator-superconductor 234

Mixing

three-wave 164

mmW scanner 301

Modulation transfer function 306

Multiplex advantage 249, 256, 257

Multiplier 141, 284

superlattice 143

Mylar 56

N

Near-field microscopy 331, 332, 335

aperture-less 335
dynamic aperture 334

NEDT 308

NEP 173, 175, 308

Noise 174

amplifier 179
current 179
excess 178
generation-recombination 179
Johnson 178, 241
shot 179, 230

thermal 230
 thermal fluctuation 179, 241
 Noise equivalent delta temperature 308
 Noise equivalent power 173, 175, 308
 Noise temperature 224–226, 294
 Non-destructive testing 302
 Nyquist sampling 307
 Nyquist-frequency 261

O

Orotron 149
 Oscillator
 Hertzian 9

P

Parametric oscillator 164, 323
 frequency tuning 165
 phase-matching 166
 quasi-phase-matching 168
 Parylene 56
 Phase noise 296
 Photoacoustic effect 112, 193
 Photoconductor 203
 detectivity 212
 extrinsic 14, 204
 GaAs 217
 gain 206
 Ge:Ga 14, 16, 209
 NEP 208
 responsivity 207
 stressed 211
 Photodetector 170, 177
 background limited infrared 176
 Photodiode 140
 pin 140
 uni-travelling-carrier 140
 Photometer 310, 311
 Photomixer 136, 319
 Picarin 56
 Planck radiation law 7, 104
 Planck's equation 104
 Plasma source 107
 Polarization rotator 82
 Polarization transformers 79
 Polarizer 79
 Brewster angle plate 80
 Poly-4-methylpentene-1 55
 Polyethylene 53, 64

 high-density 53
 low-density 53
 Polyethylene terephthalate 56
 Polypropylene (PP) 55
 Polystyrene (PS) 55
 Polytetrafluorethylene (PTFE) 55
 Power meter 191
 photoacoustic 193
 Pyroelectric detector 14, 185

Q

Quantum cascade laser 128, 137,
 321, 323
 active medium 129, 130
 beam profile 132
 bound-to-continuum 129
 chirped superlattice 129
 frequency stabilization 134
 frequency tuning 135
 imaging 321
 linewidth 134
 resonant phonon 129
 waveguide 131
 Quantum efficiency 177
 Quarter-wave plate 81
 Quartz
 crystalline 57
 fused 58
 z-cut 58
 Quasiparticle 236, 241

R

Radar 316, 317
 Radiometer equation 226, 309
 Rayleigh criterion 304
 Rayleigh sampling 307
 Rayleigh-Jeans approximation 104
 Rectifier 170, 224
 Reflectance 52
 metal 52
 Reflection 52
 reststrahlen 53
 Refractive index 52
 Responsivity 173
 Reststrahlen 6, 53, 67
 alkali halide 60
 KRS-5 61
 quartz 58

S

Sapphire 58
 Schottky barrier 228
 Schottky diode 16, 136, 138, 142, 286, 288, 292
 equivalent circuit 230
 honeycomb 14, 232
 mixer 228, 315
 noise 230
 noise temperature 234
 planar 233
 Security 301, 314
 Semiconductor laser 120
 Si 59
 Si laser 126
 Sideband generation 138, 288
 SIS mixer 234
 noise temperature 234, 239
 Skin depth 74
 Smith-Purcell source 148
 SOFIA 245, 300
 Sparrow's criterion 305
 Spatial resolution 304
 Spectrometer
 coherent source 282
 difference frequency generation 291
 grating 12
 microwave 10, 11
 molecular spectroscopy 283, 284, 286
 sideband generation 288
 solid-state 287
 Speed of response 173
 SQUID 200, 313
 Superconductor-insulator-superconductor 16
 Synchrotron 154
 coherent radiation 154
 CSR 154

T

TEA laser 119
 Teflon 55
 TES 200, 310, 313
 Thermal emission 104

Thermal source 106
 Thermionic emission 228
 Thermopile 9, 190
 Throughput 308
 Throughput advantage 249, 256
 Time domain spectrometer 15, 16, 269, 325, 327, 328, 335
 imaging 325
 tomography 328
 Tomography 323, 328
 TPX 55
 Transition edge sensor 200, 310, 313
 Travelling wave amplifier 144
 Tsurupica 56

U

Undulator 156, 160
 Uni-travelling-carrier photodiode 140

V

Varactor 142
 Varistor 142

W

Wave plate 81
 Waveguide 91
 Wien approximation 104
 Window 62
 materials 62
 Winston cone 78, 79
 Wire grid 80
 reflectance 80
 transmittance 80, 84

Z

Zitex 57

About the Authors

Erik Bründermann and Heinz-Wilhelm Hübers met in 1986 at the Universität Bonn while studying physics, mathematics, and astronomy; and physics and medicine, respectively. In 1991, they both received diploma degrees in different areas of THz technology while working at the Max-Planck-Institut für Radioastronomie (MPIfR) within the group of Hans-Peter Röser. Since Maurice Kimmitt became associated with their research efforts at the MPIfR in 1990, they have shared a longstanding appreciation of each other and become the firmest of friends both inside and outside the laboratory.

Erik Bründermann received a Dr. rer. nat. degree in physics and mineralogy in 1994 while working at the MPIfR, and then he joined the Deutsches Zentrum für Luft- und Raumfahrt (DLR, German Aerospace Center) in Berlin. In 1997, he was awarded a Feodor Lynen-fellowship of the Alexander von Humboldt-foundation, which took him to the Lawrence Berkeley National Laboratory and the Center for Particle Astrophysics. There, together with his host, Eugene E. Haller, he filed a patent on Ge and Si THz lasers. After 2 years in the USA as a visiting researcher, he became a member of the Ruhr-Universität Bochum (RUB), currently as permanent staff scientist serving also as technical advisor to the Applied Competence Cluster (ACC) THz at RUB. His research interests are in physics, mathematics, life science, and technology, spanning the entire spectral range from the THz region to the visible. Within the faculty of chemistry and biochemistry, he developed THz techniques for applications, especially for liquid samples in living conditions. Due to his interest in hyperspectral, multimodal, and multidimensional imaging, he initiated the development of chemical nanoscopes at RUB, which are used for nanoscale objects, molecules, and living cells. For his contribution to the development of a combined microscopy method, he received an Innovation Prize of RUB in 2007. Teaching is an integral part of his daily life, having taught countless students at all levels of education. He designed hands-on experiments ranging from those suitable for second year students up to large-scale research experiments such as a chemical nanoscope at the synchrotron ANKA. He has received several Japanese Center of Excellence awards at the National Institute of Communication and Technology (NICT, Tokyo) and was elected as Honorable

Guest Professor at Shizuoka University in 2009. He serves on a yearly basis as a visiting guest professor at the Department of Nanovision Technology and the Graduate School of Science and Technology, hosted by Norihisa Hiromoto at the Hamamatsu Campus. He is a lifetime member of the Optical Society of America. In 2010, he initiated the regional group “Ruhrgebiet” of the German Humboldt-Club supported by the Alexander von Humboldt-foundation and, as a speaker for this group, he fostered multilateral and multicultural exchanges between scientists across faculties and disciplines.

Heinz-Wilhelm Hübers received a Dr. rer. nat. degree in physics and astronomy in 1994 from the Universität Bonn, and a habilitation degree from the Universität Stuttgart in 2009. From 1991 to 1994, he was with the MPIfR working on a THz heterodyne spectrometer for NASA’s Kuiper Airborne Observatory. During this time, he spent eight months at the National Institute of Standards and Technology in Boulder, USA, where he worked in the group of K. Evenson on far-infrared laser magnetic resonance spectroscopy of free radicals. In 1994, he joined DLR in Berlin, where he became head of the department in 2001. Since 2009, he has been professor of experimental physics at the Technische Universität Berlin and head of the department of “Experimental Planetary Physics” at DLR. A major focus of his research interest is on THz physics and spectroscopy. He is particularly interested in THz lasers and detectors, such as silicon lasers, quantum cascade lasers, and superconducting hot electron bolometers. In cooperation with a group at the electron storage ring BESSY II in Berlin, he began research on coherent THz synchrotron radiation in 1998. This led to the first demonstration of stable coherent THz synchrotron radiation in 2002. Since his time as a doctorate student, he has been involved in the development of THz heterodyne spectrometers for applications in astronomy and planetary research. In 2004, he began the development of THz systems for security applications and was a coordinator of one of the first European security research projects “TeraSec”. He is a co-investigator of the THz heterodyne receiver GREAT, which is operated on SOFIA, the Stratospheric Observatory for Infrared Astronomy, and is a spokesperson of the “Helmholtz Research School on Security Technologies.” Another research focus is on experimental planetary physics. This deals with the investigation of mineral and ice samples, which are similar to materials found, or assumed to exist, on planetary bodies. These materials are being investigated in the laboratory under simulated planetary conditions with LIBS, Raman, and FTIR spectroscopy. With this work, he is contributing, amongst others, to ESA’s ExoMars mission. Professor Hübers has received the Innovation Award on Synchrotron Radiation (2003) and the Lilienthal Award (2007).

Maurice Kimmitt graduated from Trinity College, Dublin, in 1954 and joined the Royal Radar Establishment in Malvern, UK, where his initial research was on infrared detectors. In 1957, he became part of a small group set up to study the far-infrared (FIR) emission from ZETA, Britain’s first attempt at controlled thermonuclear fusion. This research was successful and gave accurate results for both the average plasma temperature and density in ZETA, and later several other plasmas. During this research, he was the inventor of a tunable FIR detector and was part of a team that reported the first FIR laser in 1964. In 1965, he joined the Physics

Department of the then new University of Essex in a research position, where he worked on carbon dioxide lasers and a number of detectors for use throughout the infrared region. He was co-inventor of the photon drag detector, which is still widely used with lasers. In 1970, his book “Far-Infrared Techniques,” which was a brief but comprehensive description of the state of the art, was published. In 1973, Essex awarded him a PhD for a thesis on “Infrared Experimental Studies.” He continued his research on a variety of FIR lasers and detectors until taking early retirement from the University in 1982, but retaining a part-time post until 1990. Since 1983, he has participated in research over a number of years at Heriot-Watt University in Scotland, and later at ENEA, Frascati, in Italy, on free-electron lasers, and on p-Ge lasers and FIR detectors in Germany at the MPIfR in Bonn and the DLR in Berlin. From 1989, he was a frequent visitor to Dartmouth College, USA, where he became an Adjunct Professor of Physics and Astronomy in 1994. This research, which also involved a group at Oxford University, was primarily on Smith-Purcell sources for the THz region, and it continued until 2005. The Oxford and Dartmouth groups realized that observation of Smith–Purcell (S-P) emission at THz frequencies was a valuable method of studying the quality of particle beams. Professor Kimmitt has collaborated on S-P experiments at Oxford, the Technical University in Munich, Germany, at FOM in The Netherlands, at ENEA, and at SLAC in the USA, using electron beams with energies between 3 MV and 28 GV. He has travelled extensively, lecturing in countries as far apart as China and Argentina. In 2004, he was awarded the annual Kenneth J Button Prize for “Pioneering contributions to the field of far-infrared Physics” and in 2010 was the Honorary Chair of the “International Conference on Infrared, Millimeter and Terahertz Waves” held in Rome. He is a Fellow of the UK’s Institute of Physics and has been a Member of the Optical Society of America for nearly 50 years. In the January, 2012 edition of the journal “IEEE Transactions on Terahertz Science and Technology”, he was profiled as a ‘Terahertz Pioneer’.