

Gun-Sik Park · Yong Hyup Kim
Haewook Han · Joon Koo Han
Jaewook Ahn · Joo-Hiuk Son
Woong-Yang Park · Young Uk Jeong *Editors*

Convergence of Terahertz Sciences in Biomedical Systems

 Springer

Convergence of Terahertz Sciences in Biomedical Systems

Gun-Sik Park · Yong Hyup Kim
Haewook Han · Joon Koo Han
Jaewook Ahn · Joo-Hiuk Son
Woong-Yang Park · Young Uk Jeong
Editors

Convergence of Terahertz Sciences in Biomedical Systems

Editors

Prof. Gun-Sik Park
Center for THz-Bio Application Systems
and Department of Physics and Astronomy
Seoul National University
Seoul 151-742
South Korea

Prof. Yong Hyup Kim
School of Mechanical
and Aerospace Engineering
Seoul National University
Gwanak-gu 1, Seoul 151-747
South Korea

Prof. Haewook Han
Department of Electronic
and Electrical Engineering
Postech San 31 Hyojadong
Namgu Phang
790-784 South Korea

Prof. Joon Koo Han
Department of Radiology
College of Medicine
Seoul National University
103 Daehak-ro
Jongno-gu Seoul
110-799 South Korea

Prof. Jaewook Ahn
Department of Physics
KAIST, 373-1 Guseong-dong
Yuseong-gu Daejeon 305-701
South Korea

Prof. Joo-Hiuk Son
Department of Physics
College of Natural Sciences
University of Seoul
Seoulsiripdaero 163
Jeonngong-dong 90
Dongdaemun-gu Seoul
130-743 South Korea

Prof. Woong-Yang Park
Department of Biochemistry
College of Medicine
Seoul National University
103 Daehak-ro
Jongno-gu Seoul
110-799 South Korea

Prof. Young Uk Jeong
WCI Center for Quantum-Beam-based
RadiationResearch,
Korea Atomic Energy Research Institute
Daedeok-Street 1045 Yuseong, Daejeon
305-353, South Korea

ISBN 978-94-007-3964-2

ISBN 978-94-007-3965-9 (eBook)

DOI 10.1007/978-94-007-3965-9

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012931135

© Springer Science+Business Media Dordrecht 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

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

Foreword

As a Chair of the IRMMW-THz Society and as a Chair of the THz-Bio Center Advisory Committee it is with great pleasure that I support the publication and dissemination of the book “Convergence of Terahertz Sciences in Biomedical Systems” edited by Prof. Gun-Sik Park together with other five members of the THz-Bio Center recently established in Korea.

The book is the outcome of the Summer School on “Convergence of terahertz sciences in biomedical systems” which took place in *the month of June of 2011* in Korea. The book consists of 26 chapters grouped in three parts addressing the topics of THz instrumentations, THz-Bio interaction and THz imaging respectively. The final focus of the book on THz imaging is of great importance since ‘imaging’ is certainly the most promising and widespread application of THz technology covering a variety of fields from medical imaging to non-invasive diagnostics to security.

I am sure the book will provide a useful reference to researchers in the field. With its joint interdisciplinary approach to Terahertz science and technology and its tutorial character it will reach specialized readers as well as those readers, like undergraduate students, who may approach the field for the first time.

Interestingly, the publication of this book coincides with the tenth anniversary of the THz-BRIDGE International Workshop, which took place in Capri, Italy in 2002 in the frame of a research project funded by the European Union to investigate the interaction of electromagnetic radiation with biological systems at frequencies between 100 GHz and 10 THz.

The Proceedings of the THz-BRIDGE Workshop were published as a special issue of the Journal of Biological Physics (JBP-29) in 2003.

The past ten years have seen a tremendous advance in the field, which is now presented in the book ‘Convergence of Terahertz Sciences in Biomedical Systems’.

Enjoy the book!

Gian Piero Gallerano
Research Director, Radiation Sources Laboratory
ENEA-Centro Ricerche Frascati, Italy
Chair, IRMMW-THz Society

Foreword

For those researchers who have been interested in applications of Terahertz instruments in the biomedical sciences, this compendium represents the first comprehensive look at a wide variety of imaging and spectroscopy techniques as demonstrated through careful, systematic experiments, and comprehensive modeling. Significant milestones in the application of THz techniques to animal and human tissue imaging are described for the first time. Several novel and high efficacy modalities are presented, and new tools for the biomedical researcher are introduced. The compendium provides a direct linkage between medical researchers, RF and optical engineers, physicists, chemists, and biologists in cross disciplinary investigations that both draw upon, and benefit from, critical assessment, and knowledge that can only be realized through a unified approach to research of this type. Like the THz field in general, this book brings together these often very disjoint fields of study and begins to define the framework that will undoubtedly be necessary to realize a new field of THz science. It is with great expectation, and sincere regard, that I join with other members of the THz community in endorsing the publication and dissemination of this text, as a first step toward the establishment of the new field of *THz Medical Science*.

Peter H. Siegel
Senior Scientist
California Institute of Technology and
NASA Jet Propulsion Laboratory

Founding Editor-in-Chief
IEEE Transactions on Terahertz Science and Technology

Preface

Recent technological breakthrough in the field of terahertz (THz) radiation have triggered new biology and biomedicine applications. In particular, the new biological applications are based on the specific spectroscopic fingerprints of biological matter in the spectral region. Historically with the discovery of new electromagnetic wave spectra, we have always discovered new medical imaging systems. The use of terahertz wave was not fully realized due to the absence of practical terahertz sources. Now, after the successful generation of THz waves, there is a great potential regarding the use of THz waves in biomedical systems due to the resonance of these waves with biomolecules. Some examples of important results can be found from such reports as the label-free probing of the binding state of DNA and the identification of the vibrational modes of DNA components by THz time-domain spectroscopy. THz vibrations of biomolecules involve a substantial fraction of the atoms forming the molecule and motion associated with intermolecular hydrogen bond vibrations. They provide a unique fingerprint of the conformational state of the molecule and the effects of its environment. The uniqueness of THz spectroscopy in biology and biomedicine is now well proven. However, there are challenges to realizing this great potential. One immediate issue is the intensive absorption lines of water and other surrounding molecules. The bound water which influences the biological function attenuates the THz wave significantly. To exhibit the function of the cell, biomolecules, and biological tissues, the entire specimen should be in a wet condition. It is always a great feeling for scientists and engineers to foresee a great contribution to human well-being through all of our efforts to overcome the many hurdles and reach the end product. History has already shown how to realize real biomedical imaging systems, such as MRI, CT, and PET using every spectrum of electromagnetic waves. The first team effort to study THz waves in biomedicine was initiated in the EU through the THz-Bridge program in 2002. The recommendations from this program were implemented in the Center for THz-Bio Application Systems as initiated by the Ministry of Education, Science and Technology and the National Research Foundation of the Republic of Korea in 2009 (Grant No. 2009-0083512). This is a truly interdisciplinary field involving

the creation of convergence technology in which where the communication between different disciplines is the most challenging issue for success and the realizations of great work. The summer school organized by the Center for THz-Bio Application Systems at Seoul National University in 2011 aimed to accomplish such a goal. This book on the “**Convergence of Terahertz Sciences in Biomedical Systems**” stems from this summer school. The book has three parts. The first part focuses on terahertz instrumentation and describes all types of terahertz generators, including vacuum electronic sources, solid-state electronic sources, and photonic sources. Also discussed are the topics of beam generation using carbon nanotubes, device fabrication using nanotechnologies, THz antennas, THz propagation, THz focusing, THz meta-materials and ultra-wideband techniques of THz generation. The second part is entitled “THz-Bio interactions” and introduces the THz near-field microscope and several examples of interesting discoveries pertaining to neurons, genome-wide analyses, skin, and emotions. In this part nanowater and molecular modeling are introduced. The last part is called “THz clinical imaging.” It includes sections on imaging biomarkers, medical imaging, tomography imaging, quantum sensors, and intravital microscopy for THz-Bio analyses. The contents of this book is arranged to be taught to upper-classmen at the university level or to first-year graduate students. Therefore, this book will hopefully be read by any researcher who is interested in this field. Through the effort of this book, we can communicate better and reach our goal of a better understanding of the biomedical systems which use THz waves.

Gun-Sik Park
Director and Professor
Center for THz-Bio Application Systems and
Department of Physics and Astronomy
Seoul National University
Seoul, Korea
Vice-Chair, IRMMW-THz Society

Contents

Part I Terahertz Instrumentations

Principle of Terahertz Radiation Using Electron Beams	3
Young-Min Shin, Eun-Mi Choi and Gun-Sik Park	
Principle of Terahertz Free Electron Laser	51
Young Uk Jeong, Seong Hee Park and Kitae Lee	
Principle of Linear Accelerator Based fs-THz Generation and Its Application	73
Jaehun Park	
Electron Beam Sources Based on Carbon Nanotube for THz Applications.	93
Yong Hyup Kim, Tae June Kang, Wal Jun Kim, Eui Yun Jang and Jeong Seok Lee	
THz Generation and Propagation Using Femtosecond Laser.	113
Tae-In Jeon	
Recent Advances in Solid-State Electronic Terahertz Systems.	129
Jae-Sung Rieh, Dong-Hyun Kim, Kyungmin Kim and Hyunchul Kim	
Nanofabrication Techniques and Their Applications to Terahertz Science and Technology.	147
Changsoon Kim and Hwi Kim	
Sub-single Cycle Pulses of Electromagnetic Radiation.	163
Minwoo Yi and Jaewook Ahn	

Resonant Antennas on Semi-Infinite and Lens Substrates at Terahertz Frequency	181
Truong Khang Nguyen and Ikmo Park	
The Basic Principle of the Near-Field Superlens	195
Kyoungsik Kim	
Effective Material Parameter Retrieval for Terahertz Metamaterials	203
Teun–Teun Kim, Muhan Choi, Yushin Kim and Bumki Min	
Ultrabroadband Terahertz Spectroscopy	219
Inhee Maeng, Jae-Seok Kim and Hyunyong Choi	
Part II Terahertz-Bio Interaction	
Terahertz Pulse Near-Field Microscopes	233
Kiwon Moon, Meehyun Lim, Youngwoong Do and Haewook Han	
Optical Imaging and Control of Neurons	245
Yoon-Kyu Song	
Genome-Wide Analysis of THz-Bio Interaction	257
Kyu-Tae Kim and Woong-Yang Park	
Structure and Function of Skin: The Application of THz Radiation in Dermatology	281
Seong Jin Jo and Oh Sang Kwon	
Science of Water at Nanoscale	301
Sung-Jin Chang and Wonho Jhe	
Molecular Modeling of Biomolecules	313
Chaok Seok	
Emotion	337
Sukwoo Choi	
Part III Terahertz Clinical Imaging	
Imaging Biomarker	343
Hyuck Jae Choi and Joo-Hiuk Son	

Contents	xiii
Terahertz Medical Imaging	351
Joon Koo Han	
Principles of Tomography Imaging: From X-ray CT, MRI to THz Tomography	373
Chang-Beom Ahn and Hochong Park	
Introduction to Quantum Sensors in Cryogenic Particle Detection . . .	393
Yong-Hamb Kim and Sun Kee Kim	
Intravital Microscopy for THz-Bio Analysis	413
Pilhan Kim	

Part I
Terahertz Instrumentations

Principle of Terahertz Radiation Using Electron Beams

Young-Min Shin, Eun-Mi Choi and Gun-Sik Park

Abstract This part introduces high power THz coherent radiation sources that take advantage of free electron beams. Following a description of characteristics on vacuum electron devices (VEDs), fundamental radiation principle of beam-wave interaction is explained with specifying their types and applications. Conventional high power microwave VEDs such as klystrons, TWTs, gyrotrons, and FELs are described in their technical perspectives with brief overview of device characteristics. Addressing technical challenges on up-conversion-to-THz of conventional approach, this part explores the state-of-the-art micro-VEDs considered for modern THz applications such as communication, imaging, sensing, spectroscopy, and so on, which are combined with modern microfabrication technologies. Novel MEMS techniques to microminiaturize RF components such as electron gun and RF interaction circuits are also presented.

Y.-M. Shin (✉)

Department of Physics, Northern Illinois University, DeKalb, IL 60115, USA
e-mail: alcolpeter@gmail.com

Y.-M. Shin

Accelerator Physics Center (APC), Fermi National Accelerator Laboratory (FNAL),
Batavia, IL 60510, USA

E.-M. Choi

School of Electrical and Computer Engineering, Ulsan National Institute
of Science and Technology (UNIST), Ulsan, Korea

G.-S. Park

Center for THz-Bio Application Systems and Department of Physics and Astronomy,
Seoul National University, Seoul, Korea

1 Introduction

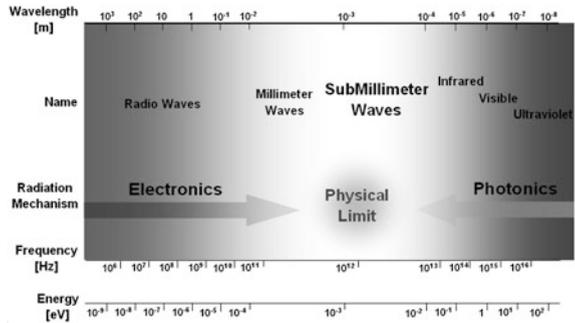
1.1 Background: Light—Backbone of Future Science and Technology

Light is one of the most intriguing objects in science because it made the most critical impact on people's lives in various ways. Electromagnetic waves can be easily converted to any type of usable energy and can carry informative data. Also, they visualize micro/macrosopic features, see through opaque substances, and analyze chemical and biological elements. They have unlimited potential for future science and technology. Creating and controlling light has been an attractive subject because it opens a new paradigm to advance science and technology to the next level. Light is electromagnetic wave and electromagnetic wave is light. EM wave is comprised of sinusoidally varying electric and magnetic fields with an oscillating frequency, which is the inverse wavelength. Long wavelength lights such as radio frequency waves are normally used for antenna communications and TV, radio broadcasting, whereas short wavelengths are normally used for bio-medical imaging applications. Note that as wavelengths become shorter, light behaves more like a particle than a wave in real life and has been accommodated to various applications with respect to those characteristics. Among EM spectrums, THz wave is very interesting and unexplored frontier, which has been not properly investigated. As for THz waves, the spectral regime ranges from 0.1 to 10 THz and the frequency of 1 THz is equal to a trillion cycles of electric and magnetic fields in one second. What makes this unique? It is the bridge spectral regime where electronics and photonics encounter, so THz waves have either characteristics, but on the other hand they are not easily generated by either electronic or photonic methodology since it exist between upper and lower limits of electronics and photonics.

1.2 Terahertz (THz) Wave

Terahertz (THz) gaps (waves), often termed “submillimeter waves” or “far-infrared (FIR)”, occupy a portion, 0.1–10 THz, of the electromagnetic (*EM*) wave spectrum between microwave and infrared. The “submillimeter” wave term, the most commonly used for the spectral region in question, restricts the discussion to frequencies above 300 GHz, corresponding to a free-space wavelength of 1 mm, while the “far-infrared” term is not entirely appropriate because its high-frequency limit is ambiguous. According to the physical unit conversion, 1 THz corresponds to a time scale of 1 ps, a wavelength of 300 μm , an electron energy of 4.1 meV, and a temperature of 47.6 K. In addition, the THz frequency range of 0.3–3 THz corresponds to 0.3–3 ps, 0.1–1 mm, 1.2–12.1 meV, and 14–140 K. Until recently, THz waves have remained relatively unexplored and basic research, new initiatives and

Fig. 1 Millimeter/submillimeter wave region in the electromagnetic wave spectrum



advanced technological developments in this spectral region have been very limited compared to the well-developed science and technology of microwave, optical and X-ray frequencies because they exit between the upper and lower physical limits of electronics and optics, as shown in Fig. 1. Typical methodologies of *EM* wave generation below the centimeter wave region ($h\nu/kT \ll 1$) and above the infrared region ($h\nu/kT \gg 1$) are respectively electronics, based on electron-transport, and optics (photonics), based on the quantum mechanical transition of electron. Both areas encounter physical limitations in the millimeter and submillimeter wave region ($h\nu/kT \sim 1$). In the absence of a compact powerful radiation source to generate controllable THz signals and efficient detectors to collect them and record information, most of the THz spectrum has not been properly developed. Therefore, developments of THz technologies related to spectroscopy, imaging and sensing are taking the world into a new scientific frontier, and, as a result, THz waves now have potential applications in physics, chemistry, medicine, biology, astronomy, meteorology, agriculture, and many other fields.

Like visible light creating a photograph, radio waves transmitting sound and X-rays allowing us to observe inside of the human body, THz waves can create images and transmit information. However, THz-rays (T-rays) have a number of advantages over other *EM* waves. The photon energies of T-rays are very low, so that they are not harmful to living organisms [1]. In case of X-rays, the photon energy is typically a few keV, a few million times higher T-ray photon energy. By mapping the transient of the electric field in amplitude and phase, absorption and dispersion spectroscopy to detect coherent THz waves in a time domain is possible. In the same way that microwave and X-ray techniques produce density images, T-rays likewise provide spectroscopic imaging within the submillimeter wave region. THz waves with rotational, vibrational, and translational response-modes to molecules, radicals, and ions can provide a range of unique information, which is difficult to obtain from optical, X-ray and NMR images. The peculiarity of THz reactivity is very effective for fingerprinting specific materials in the THz range. It provides conformational information closely related to the biological functions of molecules in tissues and cells and inaccessible by other technical means. While dielectrics are opaque to visible light and provide only low contrast

to X-rays, T-rays can easily penetrate and image inside most dielectric materials. Furthermore, T-ray imaging can provide strongly enhanced contrast because of low scattering (Rayleigh scattering) due to their longer wavelengths, whereas large amounts of scattering in conventional optical transillumination, using near-infrared pulses, tends to blur images of objects.

As previously mentioned, the T-ray system for spectroscopy, imaging, and sensing has very promising potential aspects in applications in academic, industrial, and military fields. Due to its property of appropriate penetration into dielectric materials, a THz emission microscope can be used not only for non-destructive inspection of delicate 3D-structures like semiconductor circuits (LSI, VLSI, ULSI) [2–4], but also for security inspection to detect plastic gun or bomb [5, 6]. It is also used for real-time detection of micro-leak defects in flexible plastic packing seals [7]. Likewise, a THz spectroscopic imaging system can be applied to noninvasive detection of illicit drugs by use of spectral fingerprints [5]. In pathology, it can distinctly diagnose malfunctioning cells and organs (tumors, cancers, diabetes, caries, epilepsies, etc.) [8–10]. In biology, a THz-tomography system could be utilized to investigate the structure and metabolism of cells, tissues and organisms, e.g., and water content measurement in plants and seeds [11–13]. Besides this, it can be used to verify the freezing level of foodstuffs [14]. In the scientific research area, a THz-TDS (time-domain spectroscopy) system based on a femto-second laser can also provide a stable platform to investigate various physical phenomena occurring within an infinitesimal time scale [15, 16]. Related to astronomy and meteorology, a small and light THz system can be loaded on space-satellites to study cosmic rays and atmospheric elements [17–21].

With respect to the practical aspects, THz science has encountered many challenges and the development of a convenient THz system has suffered from physical and technological limitations [22]. In modern THz technology, the most critical issues can be categorized as radiation power, signal-to-noise ratio (SNR), acquisition speed, size, target reconstruction, limited frequency bandwidth, spectroscopic data, water-absorption, scattering, and cost. In particular, power, SNR, bandwidth, and size in development of a compact powerful CW THz source have significantly affected its imaging application.

1.2.1 Radiation Power

The average power of the THz radiation source is typically of the order of the μW range (from 0.1 to 100 μW) in the case of a pulsed laser oscillator with 1 W output power. The peak power of a THz wave ranges from 100 μW to 0.1 W depending on the duty cycle of the oscillator. This power corresponds to an SNR of 104 or higher for sensing applications with single pixel detection. However, an average of mW or higher (or a peak power of kW) is preferred for a detection array system used for real-time 2D-imaging.

1.2.2 Signal-to-Noise Ratio (SNR)

THz-TDS systems can produce very high signal-to-noise ratios (SNRs) of over 100,000. Nevertheless, in practical imaging processes, a variety of factors play a synergistic role in lowering the SNR remarkably, to an acceptable limit. Some of these factors are associated with the demand to increase the imaging acquisition speed and the high absorption of biological tissue. Microwave-up-conversion of conventional electronic systems may be one of the most efficient ways of improving the SNR although these alternatives have some drawbacks.

1.2.3 Limited Frequency Bandwidth

Frequencies of conventional photoconductive dipole antenna (PDA) THz sources are below 3–4 THz [23, 24]. It has been reported that optical rectification has a wider bandwidth in excess of 30 THz. However, it is bought at the expense of THz power. An ideal THz imaging system would make it possible to measure tissue responses up into the infrared. It might broaden an observable range as well as reduce water attenuation steeply falling with an increase of the frequency over 100 GHz. Optical parametric generation of a CW THz wave provides a tunable, narrow bandwidth radiation source. The THz wavelength can be tuned from 0.7 to 2.4 THz, and the bandwidth is less than 2 MHz (instrument limitation). Other CW THz sources, using an optical down conversion, include the beating frequency from two semiconductor diode lasers on a photomixer; this provides a low cost, tunable THz source with very narrow bandwidth [25].

1.2.4 Size

Generally, typical T-ray imaging systems need areas of a few square meters excluding subsystems. However, it may be essential to cut the volume and weight of the entire system down to the scale of a tabletop or even laptop for various purposes, e.g., portable diagnosis/inspection apparatus, satellite-loading, and so on. For instance, in medical endoscopic applications a T-ray transceiver and probe would need to operate in a very tiny area as small as 1 mm².

1.3 THz Radiation Sources

In accordance with the generation of its signal waveforms, THz radiation sources are classified by pulse and CW types, as shown in Fig. 2. In the case of pulse sources, pulse with a femto-second width corresponds to the THz spectral region. Based on optical approaches, two conventional methods, photoconduction and optical rectification [26, 27], are typically utilized to create ultra-fast laser pulses.

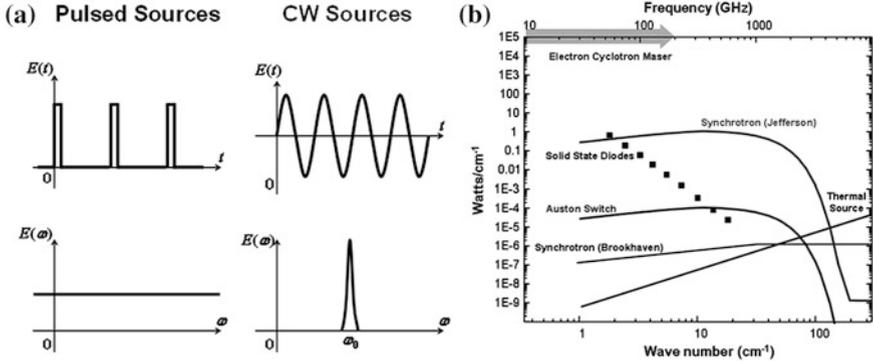


Fig. 2 a Signal wave forms of pulsed and CW THz sources. b Output power of various sources in the millimeter and submillimeter wave region

An ultra-fast Ti:sapphire laser [28] using the energy-level transition of bound electrons has a pulse duration of 100 fs and a center wavelength at 800 nm. The photoconductive and optical rectification approaches take advantage respectively of high-speed photoconductors as transient current sources for radiating antennas and electro-optic crystals as rectification media. Because of the wide bandwidth characteristic of the pulse lasers, these radiation sources are mostly addressed to spectroscopic applications in spite of their relatively low radiation power, μW – mW . The typical conversion efficiency of optical power to T-ray power ranges from 10^{-3} to 10^{-6} . The CW sources, directly generating electromagnetic waves oscillating with THz frequencies, are mostly applied to imaging/sensing systems due to their high radiation power and narrow bandwidth. They are categorized by tunable and non-tunable devices depending on whether to tune frequencies. While a tunable source can be often used for spectroscopic applications by phase-locking, the non-tunable sources are only applicable to imaging systems. Optics approaches, such as optical-down-conversion, are quantum cascade laser (QCL) [29], using inter-subband transitions of electrons with THz photon emission excited by resonant tunneling through the multiple wells of the periodic super lattice, and optically pumped terahertz laser (OPTL), using the rotational and vibrational excitations of gaseous molecules [30, 31]. Basically, the former requires a very low operating temperature and a high magnetic field, and the latter has a discrete molecular absorption band. In electronics approaches, solid-state devices, diode oscillators [32] and frequency-multipliers [33], based on conduction of free electrons in medium have fundamental limits for microwave-up-conversion to the THz region, i.e., as the frequency and power of the radiation signal are respectively restricted by the low response time of slow carriers and the high temperature gradient in the tiny medium. Recently, vacuum electronic devices (VEDs), using the convection of free electrons in a vacuum, have been newly reinvigorated as a promising breakthrough for overcoming some critical issues of other methods of THz wave generation.

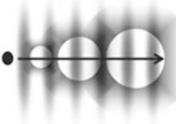
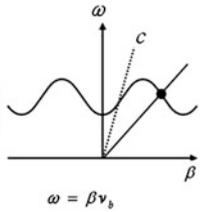
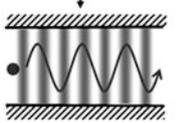
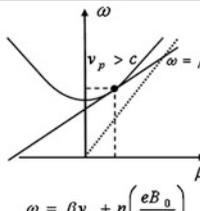
1.4 Electronic Devices Using Free Electron Beams

Vacuum electron devices were widely used as a high power radiation source for communications, plasma heatings and diagnostics, radar applications, etc., after the Second World War. However, vacuum electronics tended to fall behind solid-state electronics in the competition for the development of an efficient radiation source due to the rapid advance of semiconductor technologies after the 1970s. The merits of the solid-state devices, e.g. low voltage and high current operation capability, surpassed the bulky VEDs for the applied purposes of moderate power in the lower frequency region. Nonetheless, with demand for a powerful radiation source in the millimeter and submillimeter wave region, the classical technologies have once again been paid attention to together with advances in microfabrications. Unlike the collisional conduction transport of electrons in media, a large portion of electron energies can transit to the EM energies through collisionless convection transport in vacuum, so that VEDs can provide very high radiation efficiency. Furthermore, the lower ohmic loss and the higher breakdown threshold compared to the solid-state devices make it possible to handle extremely high energy-density in the high frequency regime at room temperature. After all, these advantages might point to the applicability of THz technology as long as some problems of miniaturizing VEDs down to millimeter and submillimeter wave scales are solved.

The VEDs are basically assorted Fundamental principles [34]. Electrons periodically circling and wiggling give rise to synchronous coherent radiation, the so-called “Bremsstrahlung radiation [35]”, when their cyclotron frequency is below the Doppler shift-frequency of EM waves interacting with the electrons. As depicted in the dispersion relation of Fig. 3, it is termed fast-wave interaction because the circuit and the electron beam modes synchronously encounter each other above the light-line. Because of their bulky size and a huge scale accelerator, electron cyclotron masers (gyrotron) [36, 37] and free electron lasers (FEL) [38–40] are not suitable for the development of a compact THz source.

Electrons passing through periodic dielectric media or waveguides retarding EM waves bring about synchronous coherent radiation, the so-called “Cherenkov radiation [35]”, which is phenomenologically similar to the shock waves from accelerating relativistic electrons. In Fig. 3, it is termed slow-wave interaction because the circuit and the electron beam modes synchronously encounter each other below the light-line. In general, Cherenkov radiation VEDs, e.g., traveling wave tube (TWT) [41, 42] and klystron [43, 44], produce relatively lower power compared to Bremsstrahlung radiation VEDs. However, these devices are suitable for high frequency applications because the linear motion of electrons is compatible with their micro-structures [45]. Microfabricated VEDs for the high-frequency applications are termed “micro-vacuum electron devices (μ VEDs) [46–48]”. For the development of a compact powerful CW source, this dissertation focuses on the μ VEDs using a Cherenkov radiation mechanism in the millimeter and the submillimeter wave region.

Fig. 3 Radiation concepts of vacuum electron devices (VEDs)

Type	Scheme	Synchronism
Cherenkov	<p>Slow Wave Medium</p> 	 <p>$\omega = \beta v_b$</p>
Bremsstrahlung	<p>Waveguide</p> 	 <p>$\omega = \beta v_z + n \left(\frac{eB_0}{\gamma m_e} \right)$</p>

2 Fundamental Principle of Vacuum Electron Devices (VEDs)

2.1 Analysis Using Inhomogeneous Maxwell Wave Equations

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= 0 & \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \vec{\nabla} \cdot \vec{B} &= 0 & \vec{\nabla} \times \vec{B} &= \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \end{aligned} \quad (1)$$

In EM waves, magnetic and electric fields are perpendicular to each other and also they are polarized perpendicular to propagating wave vectors, which can be well described by homogeneous Maxwell equations shown in Eq. (1). Including a timely varying charge in a unit volume changes these equations to an inhomogeneous form. In the ballistic analysis, there is no need to consider propagating electric and magnetic fields. However, when space charge effects are considered, these fields must be included in the analysis. These fields are described by inhomogeneous Maxwell's equations, which may be written as follows, if electron charge and current are present:

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= -\frac{\rho}{\epsilon_0} & \vec{\nabla} \times \vec{E} &= -\mu \frac{\partial \vec{H}}{\partial t} \\ \vec{\nabla} \cdot \vec{B} &= 0 & \vec{\nabla} \times \vec{B} &= -\vec{J} + \epsilon \frac{\partial \vec{E}}{\partial t} \end{aligned} \quad (2)$$

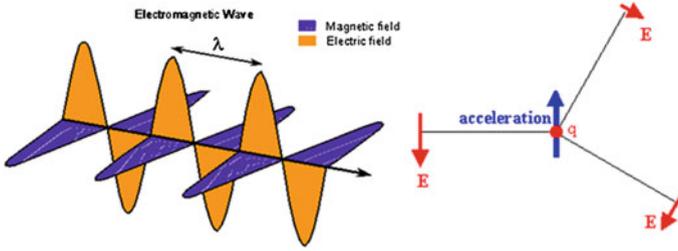


Fig. 4 Inhomogeneous EM wave model of accelerating charge

In addition,

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{H} = -\vec{\nabla} \cdot \vec{J} + \varepsilon \frac{\partial}{\partial t} \vec{\nabla} \cdot \vec{E} = 0 \quad (3)$$

So

$$\vec{\nabla} \cdot \vec{J} = -\frac{\partial \rho}{\partial t} \quad (4)$$

Now, to examine the effect of space charge on bunching, an infinitely wide electron beam will be considered first. It will be assumed that the dc space charge in the beam is neutralized by stationary ions so that potential depression problems do not exist. This simplifies the analysis a great deal because no transverse variations of quantities need to be considered. Simple mathematical operations of Maxwell's equations lead to inhomogeneous wave equation, which is

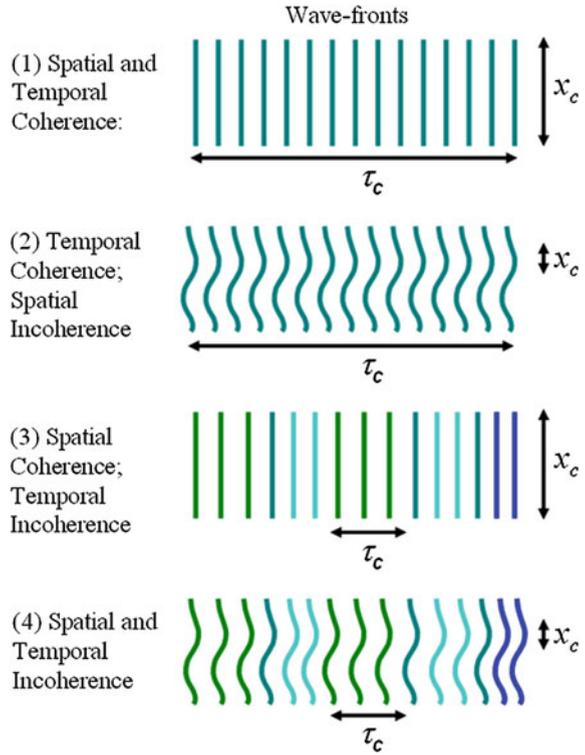
$$\nabla^2 \vec{E}_1 - \mu_0 \varepsilon_0 \frac{\partial^2}{\partial t^2} \vec{E}_1 = -\mu_0 \frac{\partial}{\partial t} \vec{J}_1 - \frac{1}{\varepsilon_0} \vec{\nabla} \rho_1 \quad (5)$$

This inhomogeneous wave equation shows how charged particles emit EM waves. In Eq. (5), spatial-temporal variation of charges in unit volume gives rise to spatial temporal variation of electric (and magnetic) fields. It should be noted that the second order differential term of a charge in time leads to the second order terms of electric fields in time and space, which indicates that an accelerated (or a decelerated) charge induces spatial propagation of timely varying electric and magnetic fields or vice versa. The equation gives conceptual understanding of radiation dynamics of charged particles. More in-depth analyses would be available elsewhere [49] (Figs. 4 and 5).

2.2 Coherent Radiation from Electron Beam

Assuming more than two charged particles participate in accelerating motion and they are charged with energies, they will emit various photons with different characteristics. Coherence is about how the emitted photons are phased in time

Fig. 5 Coherency of radiation from electron beam



and space. There are four types of coherence in terms of spatial and temporal phase since the phase-term of electromagnetic wave is function of time and space. Photons randomly propagating with all different frequencies are incoherent. If they move with different frequencies, the light is temporally incoherent. If photons have the same frequency, but move differently, it would be spatially incoherent. When the photons move all together in phase of time and space, it is called coherent radiation. Two accelerated electrons with different energies emit photons in two different time frames, these two photons have different frequencies and they independently propagate in space. There is thus no connection between the photons and individual photon intensities separately add up to total intensity. Two identical electrons with the same kinetic energy move together: they emit the exactly same photons at the same time in space. Total intensity of the coherent radiation includes crossover terms between two photon electric fields. In the electron beam, a stellar amount of electrons with the same energy simultaneously emit the same photons, so the coherent radiation has much higher intensity due to the crossover terms. Light sources using electron beam devices normally take advantage of this energy conversion from electrons to coherent photons.

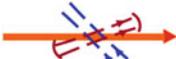
Type of radiation and synchronism	Scheme	Types of HF generators and amplifiers
<p>Cherenkov (and transition)</p> $\omega = k_{II} v$		<p>TWT, magnetron, BWO, klystron, orotron etc.</p>
<p>Bremsstrahlung</p> $\omega = k_{II} v_{II} + \Omega$		<p>Gyrotron, gyroklystron, ubitron (FEL)</p>
<p>Scattering</p> $\frac{\omega_s - \omega_i}{k_{II s} - k_{II i}} = v$		<p>Parametric devices</p>

Fig. 6 Types of VEDs

2.3 Types of Electron Beam Radiation Sources

Basically, any type of abrupt change of electron movement induces photon emission. Depending on how electrons move, the electron beam devices are categorized into three different categories. Cererenkov radiation result from linear motion, Bremsstrahlung comes from either gyrating or rotating motion, and scattering radiation from head-to-head or fender bender crossing collision. There are two big families of vacuum electron devices depending on how electrons and EM wave interact to each other. Most well known linear beam devices are klystron and traveling wave tubes and most famous cross-field devices are gyrotron and free electron lasers. Their basic operation principles and device characteristics will be briefly introduced in the next section (Fig. 6).

3 Conventional VEDs

3.1 Klystrons

Klystrons amplify RF signals by converting the kinetic energy in a DC electron beam into radio frequency power. A beam of electrons is produced by a thermionic cathode (a heated pellet of low work function material), and accelerated by high-voltage electrodes (typically in the tens of kilovolts). This beam is then passed through an input cavity. RF energy is fed into the input cavity at, or near, its natural frequency to produce a voltage which acts on the electron beam. The electric field causes the

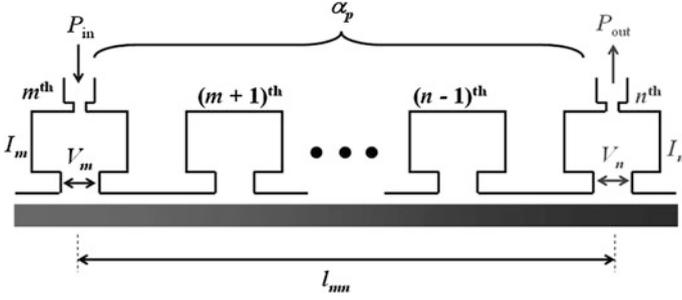


Fig. 7 Conceptual drawing of multi-cavity klystron amplifier

electrons to bunch: electrons that pass through during an opposing electric field are accelerated and later electrons are slowed, causing the previously continuous electron beam to form bunches at the input frequency. To reinforce the bunching, a klystron may contain additional “buncher” cavities. The RF current carried by the beam will produce an RF magnetic field, and this will in turn excite a voltage across the gap of subsequent resonant cavities. In the output cavity, the developed RF energy is coupled out. The spent electron beam, with reduced energy, is captured in a collector.

In Fig. 7, the small signal power gain is

$$\alpha_p = \frac{4}{(Q_e)_m(Q_e)_n} \frac{|\alpha_V|^2}{\left(\left(\frac{R}{Q}\right)_L\right)_m \left(\left(\frac{R}{Q}\right)_L\right)_n} \quad (6)$$

where $(Q_e)_m$ and $(Q_e)_n$ and $\left(\left(\frac{R}{Q}\right)_L\right)_m$ and $\left(\left(\frac{R}{Q}\right)_L\right)_n$ are respectively the external Q -values and loaded R/Q of m th and n th cavities and the voltage-gain is

$$\alpha_V = \left| \frac{\tilde{V}_n}{\tilde{V}_m} \right| = \left| \frac{\tilde{I}_n Z_n}{\tilde{V}_m} \right| = |\tilde{g}_{mm}| |Z_n| \quad (7)$$

Here, \tilde{g}_{mm} is the transconductance between the m th and n th cavities and Z_n is the cavity impedance of the n th cavity.

3.1.1 Transconductance

In Fig. 8a, the electric field restored by excessive charges at $z' = z'_1$ considered potential depression in the beam tunnel

$$E_z = -R^2 \frac{\rho_0}{\epsilon_0} (z' - z'_0) \quad (8)$$

Equation of restoring space charge force is

$$m\ddot{z}' = -eE_z = R^2 \frac{e\rho_0}{\epsilon_0} (z' - z'_0) \quad (9)$$

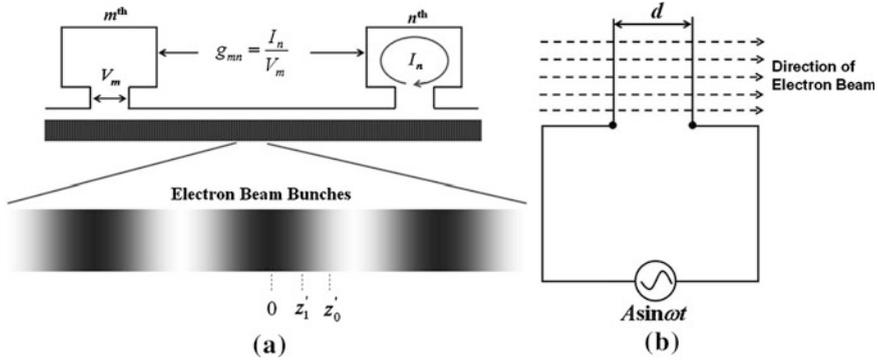


Fig. 8 **a** Electron beam bunching in the klystron amplifier **b** klystron buncher gap (d) with RF-voltage applied. Velocity modulation is produced on the electron beam

Therefore,

$$\ddot{z}' + \omega_q^2 z' = \omega_q^2 z_0' \quad (10)$$

where $\omega_q = R\omega_p$ and $\omega_p = \sqrt{e\rho_0/m\varepsilon_0}$ is the plasma frequency. R is the reduced plasma wavelength factor, which is from potential depression in the beam tunnel. The solution of Eq. (10) is

$$\dot{z}' - \dot{z}_0' = F \cos(\omega_q t + \tau) \quad (11)$$

From initial conditions, $\omega_q t_0 + \tau = 0$ and $u(d) = u_0 + (eMA/mu_0) \sin \omega t_0$, at $z = d$, as shown in Fig. 9, F is given as

$$F = \frac{eMA}{mu_0} \sin \omega t_0 \quad (12)$$

The velocity modulation on the dc electron beam with velocity u_0 is thus

$$u_{tot} = u_0 + u = u_0 - \frac{eMA}{mu_0} \sin(\beta_e z - \omega t) \cos \omega_q(t - t_0) \quad (13)$$

Substituting $z = u_0(t - t_0)$, $\omega_q(t - t_0) = \beta_q z$, and $(1/2) mu_0 = eV_0$ for Eq. (13), one has the current modulation

$$J = -\frac{1}{2} J_0 \frac{\omega}{\omega_q} \left(M \frac{A}{V_0} \right) \sin \beta_q z \cos(\beta_e z - \omega t) \quad (14)$$

$$I = -\frac{1}{2} \left(\frac{I_0}{V_0} \right) \frac{\omega}{\omega_q} MA \sin \beta_q z \cos(\beta_e z - \omega t) \quad (15)$$

After all, as shown in Fig. 9, at $z = l_{mn}$ the current modulation by m th cavity becomes

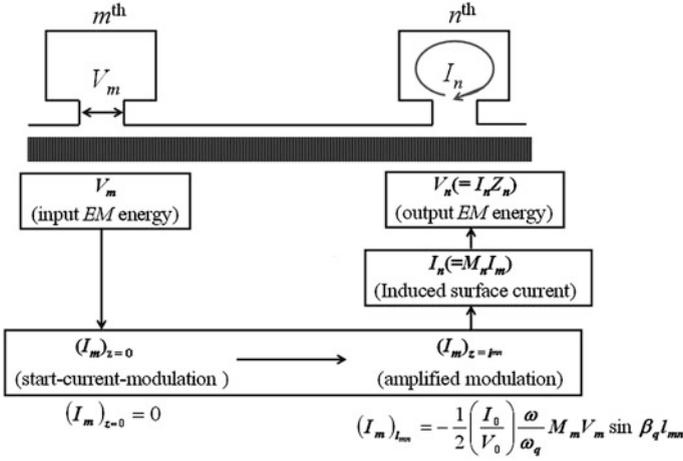


Fig. 9 Signal-amplification algorithm in the klystron amplifier

$$(I_m)_{z=l_{mn}} = -\frac{1}{2} \left(\frac{I_0}{V_0} \right) \frac{\omega}{\omega_q} M_m V_m \sin \beta_q l_{mn} \quad (16)$$

The induced current at the n^{th} cavity is thus

$$I_n = M_n I_m = \left(-\frac{1}{2} \left(\frac{I_0}{V_0} \right) \frac{\omega}{\omega_q} M_m M_n \sin \beta_q l_{mn} \right) V_m = g_{mn} V_m \quad (17)$$

where g_{mn} is the transconductance,

$$g_{mn} = \frac{1}{2} \frac{\omega}{\omega_q} \left(\frac{I_0}{V_0} \right) M_m M_n \sin \theta_{mn} \quad (18)$$

In the case of multi-cavity klystron, it is

$$\tilde{I}_n = \begin{pmatrix} I_1 \\ \vdots \\ I_p \\ \vdots \\ I_n \end{pmatrix} = \begin{pmatrix} g_{11} & \cdots & g_{p1} & \cdots & g_{m1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{1q} & \cdots & g_{pq} & \cdots & g_{mq} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{1n} & \cdots & g_{pn} & \cdots & g_{mn} \end{pmatrix} \begin{pmatrix} V_1 \\ \vdots \\ V_q \\ \vdots \\ V_m \end{pmatrix} = \tilde{g}_{mn} \tilde{V}_m \quad (19)$$

3.1.2 Cavity Impedance

The loaded n^{th} cavity can be converted to the lumped RLC circuit model. Here, the cavity impedance is defined as

$$Z_n = \left(\frac{1}{R_L} + \frac{1}{j\omega L} + j\omega C \right)^{-1} \quad (20)$$

where R_L is $1/R_L = 1/R_0 + 1/R_e + 1/R_b$ and R_0 , R_e and R_b is respectively the ohmic, the external, and the beam-loaded resistance. In the vicinity of resonance, $\omega = \omega_0 + \Delta\omega$ and $\omega_0 = \sqrt{1/LC}$,

$$\frac{1}{(\omega_0 + \Delta\omega)} \approx \frac{(1 - \Delta\omega/\omega_0)}{\omega_0} \quad (21)$$

and thus Eq. (20) becomes

$$Z_n = \frac{1}{2} \left(\frac{R}{Q} \right)_L \left(\frac{1}{1/2Q_L + j\delta} \right) \quad (22)$$

where the loaded Q -value, Q_L , is

$$Q_L = \omega R_L C = \frac{R_L}{\omega L} \quad \text{and} \quad \frac{1}{Q_L} = \frac{1}{Q_0} + \frac{1}{Q_e} + \frac{1}{Q_b} \quad (23)$$

and $\delta = \Delta\omega/\omega_0$. Here, Q_0 , Q_e , and Q_b is respectively the unloaded, the external, and the beam-loaded Q . The beam-loaded Q is

$$Q_b = \frac{1}{(R/Q)(G_b/G_0)} \quad (24)$$

Here the normalized beam conductance, G_b/G_0 , is expressed by Wessel-Berg as

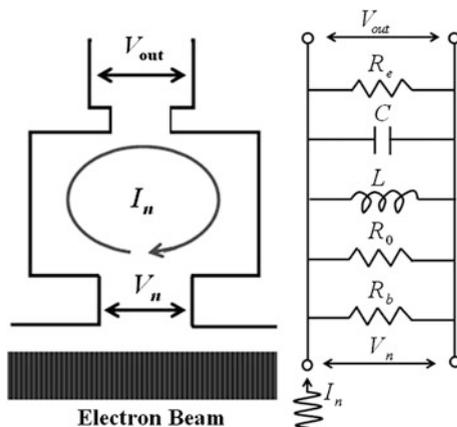
$$\frac{G_b}{G_0} = \frac{1}{8} \frac{\beta_e (M_-^2 - M_+^2)}{\beta_q} \quad (25)$$

where M_- and M_+ are the coupling coefficients for the slow, $(\beta_e + \beta_q)$, and the fast, $(\beta_e - \beta_q)$, space charge waves, respectively.

Two-Cavity Klystrons

In the two-chamber klystron, the electron beam is injected into a resonant cavity. The electron beam, accelerated by a positive potential, is constrained to travel through a cylindrical drift tube in a straight path by an axial magnetic field. While passing through the first cavity, the electron beam is velocity modulated by the weak RF signal. In the moving frame of the electron beam, the velocity modulation is equivalent to a plasma oscillation. Plasma oscillations are rapid oscillations of the electron density in conducting media such as plasmas or metals (The frequency only depends weakly on the wavelength). So in a quarter of one period of the plasma frequency, the velocity modulation is converted to density modulation, i.e. bunches of electrons. As the bunched electrons enter the second

Fig. 10 Loaded output resonator (a) conceptual model (b) equivalent RLC circuit model



chamber they induce standing waves at the same frequency as the input signal. The signal induced in the second chamber is much stronger than that in the first (Fig. 10).

Reflex Klystrons

In the reflex klystron, the electron beam passes through a single resonant cavity. The electrons are fired into one end of the tube by an electron gun. After passing through the resonant cavity they are reflected by a negatively charged reflector electrode for another pass through the cavity, where they are then collected. The electron beam is velocity modulated when it first passes through the cavity. The formation of electron bunches takes place in the drift space between the reflector and the cavity. The voltage on the reflector must be adjusted so that the bunching is at a maximum as the electron beam re-enters the resonant cavity, thus ensuring a maximum of energy is transferred from the electron beam to the RF oscillations in the cavity. The voltage should always be switched on before providing the input to the reflex klystron as the whole function of the reflex klystron would be destroyed if the supply is provided after the input. The reflector voltage may be varied slightly from the optimum value, which results in some loss of output power, but also in a variation in frequency. This effect is used to good advantage for automatic frequency control in receivers, and in frequency modulation for transmitters. The level of modulation applied for transmission is small enough that the power output essentially remains constant. At regions far from the optimum voltage, no oscillations are obtained at all. This tube is called a reflex klystron because it repels the input supply or performs the opposite function of a klystron. There are often several regions of reflector voltage where the reflex klystron will oscillate; these are referred to as modes. The electronic tuning range of the reflex klystron is usually referred to as the variation in frequency between half power points—the points in the oscillating mode where the power

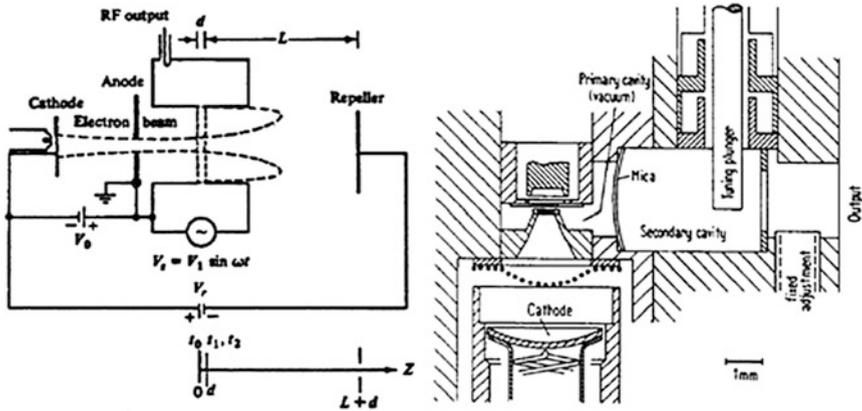


Fig. 11 Reflex klystron

output is half the maximum output in the mode. The frequency of oscillation is dependent on the reflector voltage, and varying this provides a crude method of frequency modulating the oscillation frequency, albeit with accompanying amplitude modulation as well. Modern semiconductor technology has effectively replaced the reflex klystron in most applications (Fig. 11).

Multi-Cavity Klystrons

The simple, two-cavity power klystron amplifier is not capable of high-gain or high-power output, or suitable efficiency. However, with the addition of intermediate cavities and other physical modifications, the basic two-cavity klystron may be converted into a multicavity power klystron, capable of both high-gain and high-power output. In addition to the intermediate cavities, there are several physical differences between the basic two-cavity klystron and the multicavity klystron. Klystron amplifiers have as many as seven cavities, including five intermediate cavities. The intermediate cavities improve the bunching process, resulting in increased efficiency. Adding more intermediate cavities is roughly analogous to adding more stages to an IF amplifier; that is, the overall amplifier gain is increased and the overall bandwidth is reduced if all the stages are tuned to the same frequency (Fig. 12).

Applications

Klystrons produce microwave power far in excess of that developed by solid state. In modern systems, they are used from UHF (hundreds of MHz) up through hundreds of gigahertz (as in the Extended Interaction Klystrons in the CloudSat satellite). Klystrons can be found at work in radar, satellite and wideband

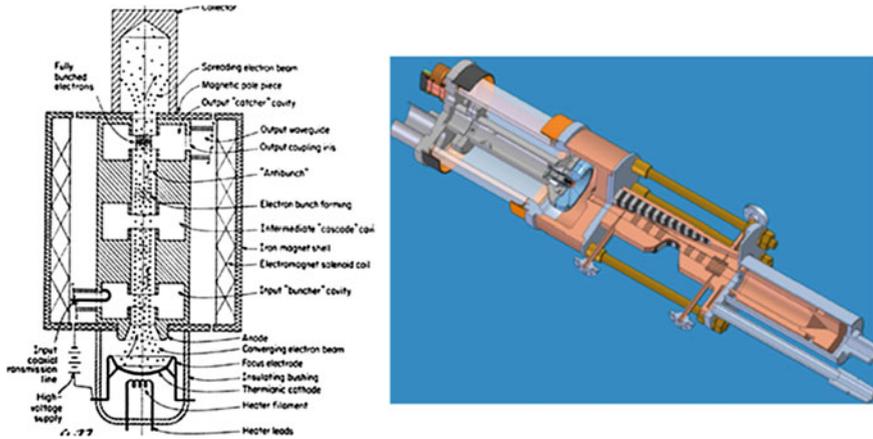


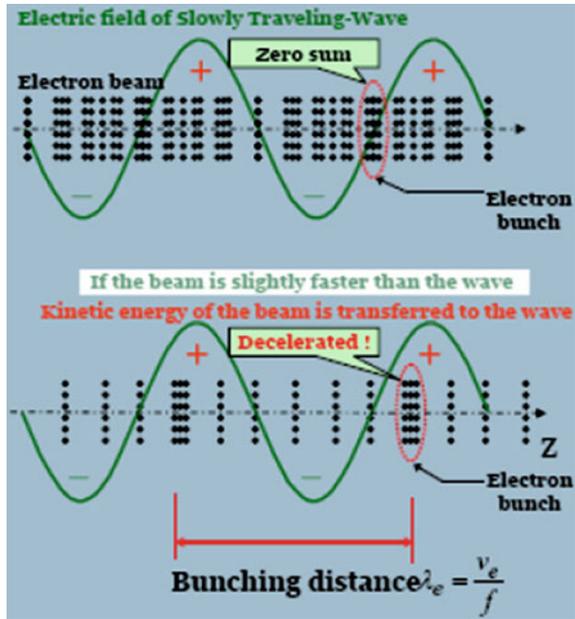
Fig. 12 Multi-cavity klystron

high-power communication (very common in television broadcasting and EHF satellite terminals), medicine (radiation oncology), and high-energy physics (particle accelerators and experimental reactors). At SLAC, for example, routinely employed klystrons have outputs in the range of 50 megawatts (pulse) and 50 kilowatts (time-averaged) at frequencies nearing 3 GHz. Popular Science's "Best of What's New 2007" [50] described a company, Global Resource Corporation, using a klystron to convert the hydrocarbons in everyday materials, automotive waste, coal, oil shale, and oil sands into natural gas and diesel fuel [51].

3.2 *Traveling Wave Tubes (TWTs)*

A traveling-wave tube (TWT) is an electronic device used to amplify radio frequency (RF) signals to high power, usually in an electronic assembly known as a traveling-wave tube amplifier (TWTA). The bandwidth of a broadband TWT can be as high as one octave, although tuned (narrowband) versions exist, and operating frequencies range from 300 MHz to 50 GHz. The voltage gain of the tube can be of the order of 70 decibels. The device is an elongated vacuum tube with an electron gun (a heated cathode that emits electrons) at one end. A magnetic containment field around the tube focuses the electrons into a beam, which then passes down the middle of an RF circuit (wire helix or coupled cavity) that stretches from the RF input to the RF output, the electron beam finally striking a collector at the other end. A directional coupler, which can be either a waveguide or an electromagnetic coil, fed with the low-powered radio signal that is to be amplified, is positioned near the emitter, and induces a current into the helix. The RF circuit acts as a delay line, in which the RF signal travels at near the same speed along the tube as the electron beam. The electromagnetic field due to the RF

Fig. 13 Operational schematics of traveling wave tubes (TWTs)



signal in the RF circuit interacts with the electron beam, causing bunching of the electrons (an effect called velocity modulation), and the electromagnetic field due to the beam current then induces more current back into the RF circuit (i.e. the current builds up and thus is amplified as it passes down). A second directional coupler, positioned near the collector, receives an amplified version of the input signal from the far end of the RF circuit. Attenuator(s) placed along the RF circuit prevents reflected wave from traveling back to the cathode. Higher powered Helix TWT's usually contain beryllium oxide ceramic as both a helix support rod and in some cases, as an electron collector for the TWT because of its special electrical, mechanical, and thermal properties [52, 53]. Helix TWTs are limited in peak RF power by the current handling (and therefore thickness) of the helix wire. As power level increases, the wire can overheat and cause the helix geometry to warp. Wire thickness can be increased to improve matters, but if the wire is too thick it becomes impossible to obtain the required helix pitch for proper operation. Typically, helix TWTs achieve less than 2.5 kW output power. The coupled-cavity TWT overcomes this limit by replacing the helix with a series of coupled cavities arranged axially along the beam. Conceptually, this structure provides a helical waveguide and hence amplification can occur via velocity modulation. Helical waveguides have very nonlinear dispersion and thus are only narrowband (but wider than klystron). A coupled-cavity TWT can achieve 60 kW output power. Operation is similar to that of a klystron, except that coupled-cavity TWTs are designed with attenuation between the slow-wave structure instead of a drift tube. The slow-wave structure gives the TWT its wide bandwidth. A free electron laser allows higher frequencies (Fig. 13).

3.3 Gyrotrons

The gyrotron is a type of free electron maser (microwave amplification by stimulated emission of radiation). It has high power at millimeter wavelengths because its dimensions can be much larger than the wavelength, unlike conventional vacuum tubes, and it is not dependent on material properties, as are conventional masers. The bunching depends on a relativistic effect called the Cyclotron Resonance Maser instability. The electron speed in a gyrotron is slightly relativistic (comparable to but not close to the speed of light). This contrasts to the free electron laser (and xaser) that work on different principles and in which electrons are highly relativistic. Electrons start to rotate when they enter perpendicularly aligned magnetic fields, so they are gyrating while propagating. Gyating electrons have a bunching motion from continuously repeated acceleration and deceleration and strongly bunched electron beams emit EM waves at the tangential direction of gyrating motion. In-depth operation principle is pretty complicated, but briefly speaking, a relativistic effect coming from gyrating motion up-shifts the beam line and it allows electrons to be synchronized with cavity resonance modes, so the electron beam can be bunched with cavity eigenfrequency. Therefore, in order to make high frequency radiation, the resonance cavity and related magnet should be small. A gyrotron consists of an electron gun, a cavity resonator, a mode converter, an output window and a collector, and requires a solenoid magnet to give a gyrating motion to the electron beam. Electrons emitted from a cathode under the action of the electric field, move in a gradually increasing magnetic field towards the cavity. In this motion, part of the energy of the electron motion along the lines of magnetic field is transformed into energy of gyration. Electrons that have a cyclotron frequency slightly below the resonant frequency of the microwave excited in the cavity are bunched and decelerated by its transverse electric field. As a result, the microwave is given the gyration energy by the bunched electrons and induces oscillation. The microwave generated within the cavity mode is converted to a wave beam by the mode converter, shaped by some mirrors, and output through the output window (Figs. 14 and 15).

3.4 Free Electron Lasers (FELs)

A free electron laser, or FEL, is a laser that shares the same optical properties as conventional lasers such as emitting a beam consisting of coherent electromagnetic radiation which can reach high power, but which uses some very different operating principles to form the beam. Unlike gas, liquid, or solid-state lasers such as diode lasers, in which electrons are excited in bound atomic or molecular states, FELs use a relativistic electron beam as the lasing medium which moves freely through a magnetic structure, hence the term free electron [54]. The free-electron laser has the widest frequency range of any laser type, and can be widely tunable [55], currently ranging in wavelength from microwaves, through terahertz radiation and infrared, to the visible spectrum, to ultraviolet, to X-rays [56].

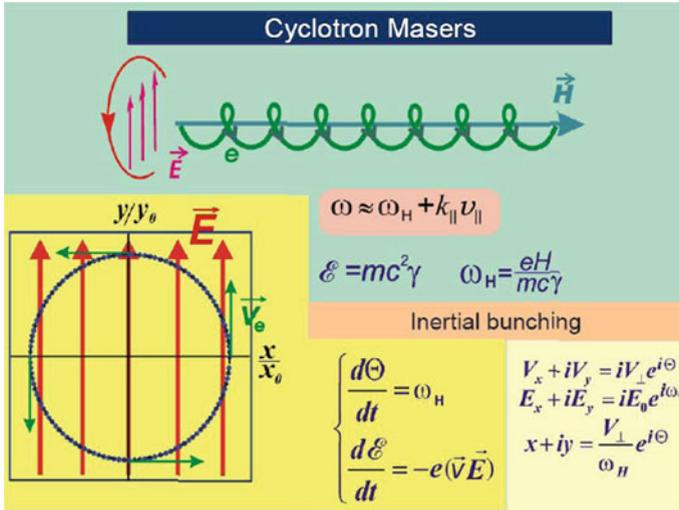
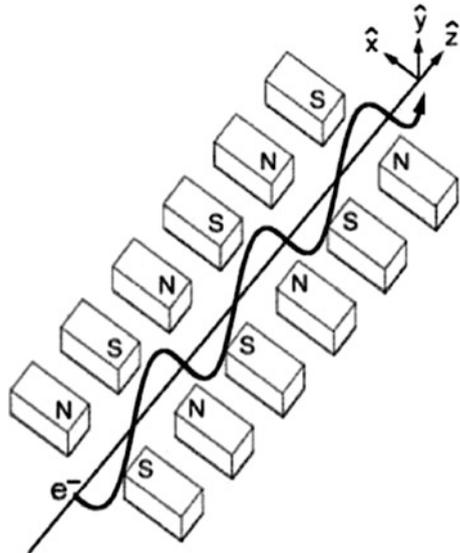


Fig. 14 Operational schematics of gyrotrons

Fig. 15 Free electron laser (FEL)



4 Current Status of THz-VEDs

The absence of an appropriate watt-level radiation source with broad spectral coverage has been an obstacle to the practical application of THz systems. Recently, micro vacuum electron devices (μ VEDs) [57–59] have received

considerable attention as a possible breakthrough for high power THz source development owing to their high energy conversion efficiency and large thermal power capacity. This approach requires linear scale-down of bulky VED elements, which necessitates the microfabrication of the associated electronic circuits resulting from the fact that as the frequency increases conventional mechanical machining becomes problematic in achieving microfeatures owing to technical limitations of tolerance, accuracy, tool-size, and surface roughness. Promising solutions will come up with a new radiation concept, more delicate MEMS fabrication, high current density sheet beam, and more accurate device assembly. The advantages of employing vacuum electron devices (VEDs) for generating coherent radiation with higher frequency are generally well accepted considering their merits in capabilities of higher power, stability, robustness, broader tunability, etc. Recently, keeping pace with the renaissance of THz science, ceaseless demand for compact submillimeter wave sources with higher powers and broader spectra, has led to the involvement of major research groups. Nevertheless, the VEDs with higher frequency which are currently available or under development [60–64] have critical shortcomings, such as bulky size, higher oscillation thresholds, lower power generation, a narrow range of tunability, stringent machining tolerances, rather complex subsystems, and so on. To cope with these problems, many research groups have been challenged to integrate VEDs with micro-electro-mechanical systems (MEMS) technologies [65–69]. Considering compact sources that employ an electron beam with relatively low voltage (i.e., in the kilovolt range), fabrication tolerances and surface roughness become more critical issues.

4.1 Issues of THz Application of μ VEDs

There are three critical factors in scale-down of electron beam devices. In the low frequency, electrons are strongly bunched and modulated. But, in the high frequency, thermal motion of the electrons disturbs the electron bunching. The electron beam thus needs very high current density for beam modulation in the THz regime. Also, the THz device should be very accurately fabricated with tight tolerance. Otherwise, it will cause huge frequency deviation. Also, their surface should be very smooth. The rough metal surface increases ohmic loss and lots of radiation power will be lost to thermal energy. The promising solutions can come up with new radiation concept, more delicate MEMS fabrication, high current density sheet beam, and more accurate device assembly (Fig. 16).

4.2 High Current Density Electron Beam Emitters

Electrons are created from cathode, and an electron gun is the device component to provide electron beams to interaction RF circuit from the cathode, as you can see

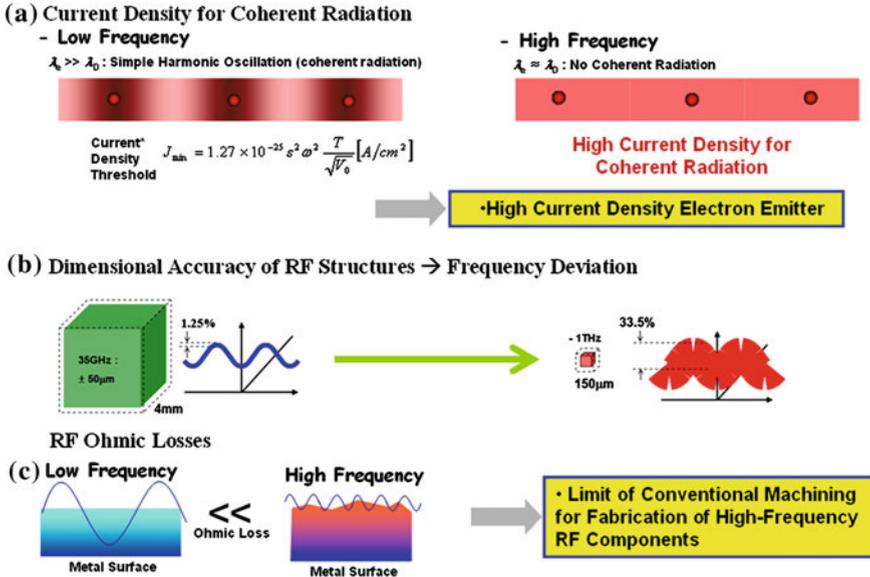


Fig. 16 Issues on VED technology of THz application

in this figure. This equation, which comes from matching Debye thermal wavelength and bunching wavelength, indicates minimum electron beam current density for coherent radiation. As you can see, this current density scales as square of frequency, so in principle ten times higher frequency needs hundred times higher current density. In practice, this criteria can be somewhat mitigated by increasing beam voltage, but still very high current density would be needed for THz coherent radiation. This table shows minimum current density and operation current density for most typical THz electron beam devices. There are two most typical ways to create a high current density electron beam.

In principle, the minimum current density J_{min} for coherent radiation of a VED at a given frequency is given by

$$J_{min} = 1.27 \times 10^{-25} s^2 \omega^2 \frac{T}{\sqrt{V_0}} \text{ (A/cm}^2\text{)} \tag{26}$$

where ω is the angular frequency (rad/s), T is the temperature (K) of the electron beam, V_0 is the voltage (V), and s is a scale factor. For conventional vacuum electron devices, the scale factor is about 10. The current threshold of Eq. (26) increases steeply with frequency ($\propto \omega^2$), so that an extremely high current density would be required for normal device operation in the terahertz and subterahertz regimes. We have thus developed a nanocomposite thermionic cathode with high emission current density, which can provide the requisite current density in the interaction circuit region via electric compression.

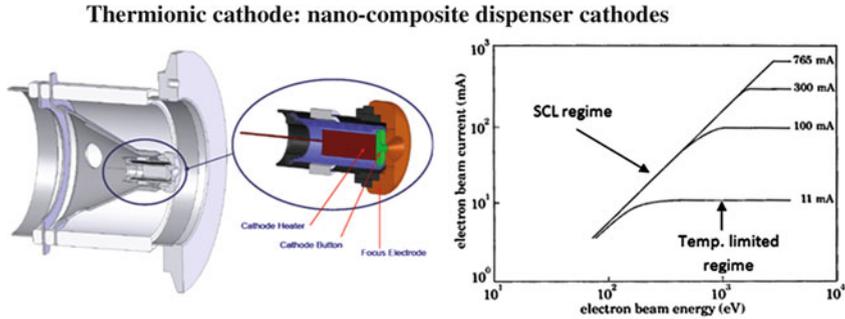


Fig. 17 Electron gun with thermionic cathode and emission characteristic graph

4.2.1 Thermionic Cathode: Nano-Composite Dispenser Cathodes

In thermionic cathodes, filament in the electron gun heats up cathodes and thermal electrons are emitted when high voltage above the energy threshold is applied to cathode surface. As you can see in this equation, current density increase has saturation in terms of temperature and voltage. Recently, people have looked into structuring the cathode surface to increase current density. Nano-composite cathodes have achieved more than 150 A/cm^2 current density most recently (Fig. 17).

4.2.2 Cold Cathode: Field Emission Arrays

Cold-cathodes are very a promising electron beam source for THz application because cold electrons have very, very small thermal perturbation. Applying high electric field to cathode tips lowers the vacuum potential barrier, which allows electrons in Fermi level to easily tunnel through the barrier and to emit to the vacuum. There are many types of cold-cathodes like moly-tips, carbon nano-tubes, and nano-wire tips, but just the most challenging issue is to improve emission uniformity and beam emittance. The nano-wires with diameters between 5 nm to 200 nm, grown via vapor-liquid-solid (VLS) method, can be used to combine photo and field emission mechanisms to achieve high brightness and modulated electron beams with integrated gates and RF transmission lines. In addition, we will grow highly oriented nanowires on a crystalline substrate and subsequently transfer them to a secondary substrate (such as a processed IC, metal surfaces etc.) while simultaneously preserving the orientation, integrity, order, shape, and fidelity of the transferred nano-wire arrays [70]. We will then coat them with a metal film such as tungsten or molybdenum for efficient field emission. This will allow us to have an array of vertically oriented ultra-sharp metal tips precisely positioned on driving circuits for controlled X-ray generation. Interplay between the two techniques (nano-composite and nano-wire) for emitters can offer risk

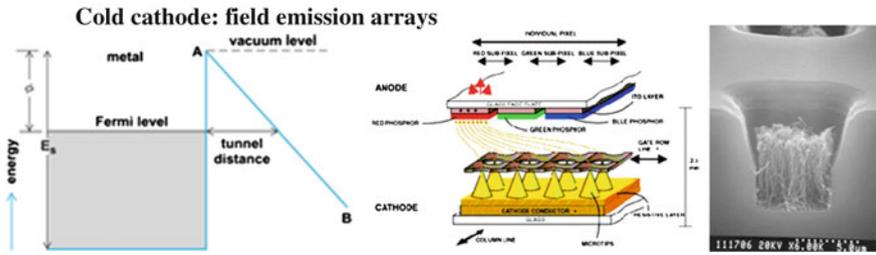


Fig. 18 Energy diagram of cold-cathode and schematic figure of FEA. SEM images of CNT arrayed emitter

mitigation and smooth transition between more mature and higher performance techniques (Fig. 18).

4.2.3 Photo-Cathode (RF gun)

A photocathode is a negatively charged electrode in a light detection device such as a photomultiplier or phototube that is coated with a photosensitive compound. When this is struck by a quantum of light (photon), the absorbed energy causes electron emission due to the photoelectric effect. Although a plain metallic cathode will exhibit photoelectric properties, the specialized coating greatly increases the effect. A photocathode usually consists of alkali metals with very low work functions. The coating releases electrons much more readily than the underlying metal, allowing it to detect the low-energy photons in infrared radiation. The lens transmits the radiation from the object being viewed to a layer of coated glass. The photons strike the metal surface and transfer electrons to its rear side. The freed electrons are then collected to produce the final image (Fig. 19).

4.3 Micro-Fabrication Technology

In principle, the radiation wavelength of a modulated electron beam is proportional to the structural size of the electronic circuit. Thus, terahertz vacuum electronic circuits require submillimeter features of a few microns to a few hundred microns, which pose significant challenges for conventional machining techniques due to tool size and fabrication tolerance requirements. In addition, conventional thermal and mechanical machining approaches have difficulty in achieving nanoscale smoothness on the circuit surface, which is critical for terahertz device operation. Recently, the MEMS techniques of lithography, etching, and deposition developed by the semiconductor chip industry have been applied to microfabrication and integration of vacuum electronic devices.

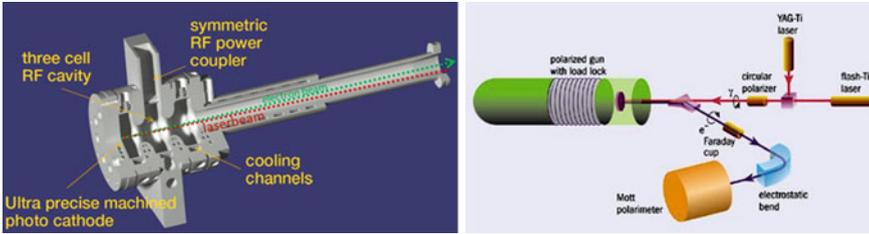


Fig. 19 Schematic figures of photocathode RF gun

4.3.1 LIGA (X-ray and UV)

The most challenging topic of THz electron beam device development is how to build micro-scale features. Conventional machining, which has been used to make meter-size tube, cannot be applied any more. So, people have been attempting to combine semi-conductor fabrication technology with tube fabrication. Most well known microfabrication technology is LIGA, a German acronym for lithography, electroplating, and molding, and DRIE, deep reactive ion etching. In LIGA, plastic film is patterned by light exposure and development and the plastic mold is washed away by chemical cleaning after metal deposition. Synchrotron light sources provide monochromatic X-ray beam for X-ray LIGA and also you can also even do LIGA process in a small clean with UV light (Fig. 20).

LIGA has almost exclusively utilized X-rays for light exposure of polymethyl-methacrylate as X-rays have a large penetration depth and are thus dimensionally compatible with millimeter wave devices. However, X-ray illumination requires a large synchrotron facility, and the beam-line is not readily accessible at all times thereby reducing its attractiveness for microwave electron beam based device fabrication. However, as the operating wavelength of vacuum electronic devices has gradually shifted down to the submillimeter wave regime, high aspect ratio lithography is no longer required for device fabrication. There have thus been widespread investigations on the application of UV light to the LIGA process. However, the insoluble crosslinked polymer (SU8) mold has always been a troublesome obstacle. As a possible solution, we have replaced SU8 with KMPR [71] in the UV LIGA process; KMPR is a chemically dissolvable, negative tone photoresist intended for MEMS applications of up to 120 μm thickness. We have succeeded in forming a 400 μm thick metal structure and in completely removing the crosslinked KMPR mold without a residue or damage to the metal deposition by a single spin coat lithography molding process. The electroformed structure is lapped and polished down to a circuit height. Figure 21 display scanning electron microscope (SEM) images of the lithography-patterned KMPR mold and the LIGA-fabricated circuit on the oxygen-free high-conductivity copper substrate. Note that the fins are gradually wider down to the substrate due to insufficient dose to the bottom layer of the mold, which requires further optimization.

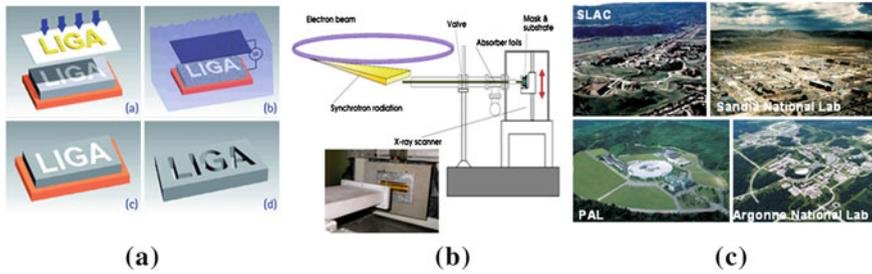


Fig. 20 X-ray LIGA (a) flow-chart (b) schematics (c) X-ray LIGA facilities

4.3.2 Deep reactive Ion Etching (DRIE)

The metalized silicon structure is formed using well known MEMS and semiconductor fabrication processes. Photoresist is spun onto the silicon substrate at a thickness. Using contact photolithography, the circuit structure is exposed and developed in the photoresist. This pattern is then etched through the full thickness of the silicon wafer. This through wafer etch is accomplished using a deep reactive ion etcher with a repeating etch (SF₆) and passivation (C₄F₈) cycle, commonly referred to as the Bosch process, shown in Fig. 22. SEM measurements show the circuit pattern on the top of the fin has a dimensional tolerance of less than 1 μm tapering down to the bottom with 5–7 μm deviation, which corresponds to a sidewall slope of ~89.6°. SEM images of the circuit sample fabricated in the DRIE process are shown in Fig. 22. This repeated etch-passivation process leaves an approximately 2 μm wide scalloping pattern on the sidewall of the etched silicon structure. In order to reduce the scallop size, the silicon structure is oxidized in a high temperature O₂ furnace until approximately 1.3 μm of silicon dioxide has formed. The silicon dioxide is then stripped completely using hydrofluoric acid. The oxidation and stripping process consumes silicon on the scallop, preferentially smoothing the sharp ridges on the sidewall, which significantly improves surface roughness to submicron scale. The silicon wafer is then furnace oxidized again to a thickness of 0.1 μm. This oxidation is to prevent direct contact of the Au conducting and bonding layer to the Si wafer during the high temperature thermocompression bond process. Gold is then sputtered at 1 μm thickness on both sides of the wafer.

4.3.3 Ultra-Precision Nano-Machining

Normally, ultra-precision machines feature air-bearing guide ways, vibration isolation technology, and laser scale position sensors. The laser scales have a resolution of 34 pico-meter, and work in concert with linear drive motors to allow position control resolution of 1 nm. Cartesian travel limits allow for work pieces of up to ~1 M mm². The machine can use either an air turbine spindle capable of

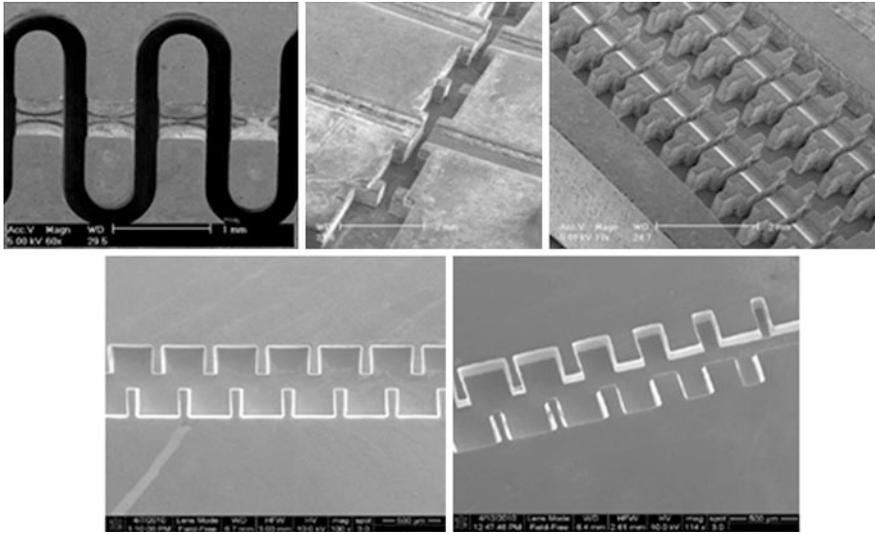
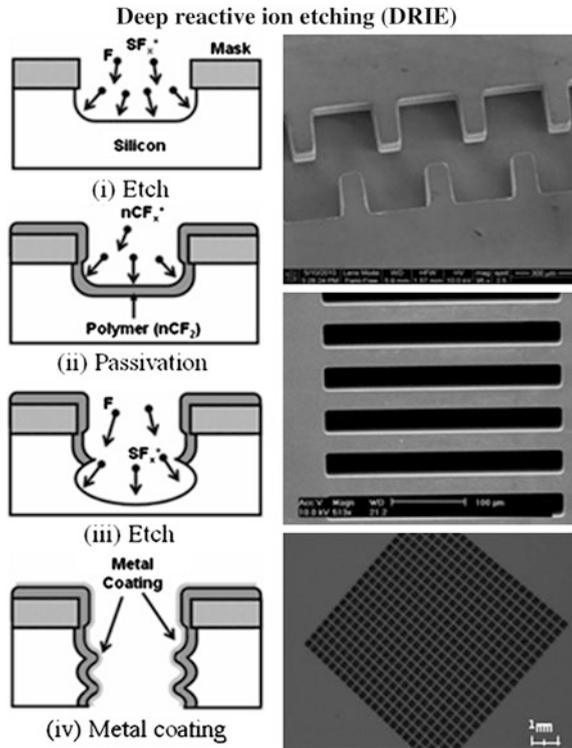


Fig. 21 SEM images of X-ray LIGA and UV LIGA fabricated THz circuits

Fig. 22 DRIE process and SEM images



a few ten thousands RPM or a fixed mandrel for scribing process [72]. The NN1000 is a smooth and steady platform designed for ultra-precision work with exceptionally fine surface finishes. The capabilities of the machine depend on the details of the work, such as materials, techniques, and tooling. Using single-crystal diamond milling and scribing tools, mirror surfaces can be machined without polishing (<10 nm R_a). Careful selection of materials, tools, and cutting strategy are critical for success. Standard tungsten-carbide tools with heavy duty coating are used with various diameters ranging from 100 to 1,300 m and cutting length to diameter ratios of 3:1 to machine THz TWT circuits. We were able to achieve tolerances finer than 1 μ m and surface finishes of 40 nm R_a . Minimum surface finish requirements are related to the EM skin depth of the material for the frequencies of interest [73]. For THz in copper, the skin depth is approximately 140 nm. Our tests have shown significant improvements in signal transmission as surface finishes improve to ~ 100 nm R_a , and diminishing returns thereafter. Problems with burr formation are a serious issue because even small burrs measuring 5 or 10 μ m are large compared to circuit dimensions. Since circuit features are small and fragile, we are very limited in what types of post processing can be done to reduce burrs and improve surface finish. We use a mild acid bath to clean and treat the parts after machining. Careful consideration must be given to acid washing, since excessive etching will reduce the surface quality (Figs. 23 and 24).

4.3.4 THz Characterization of Microfabricated Circuits

Micro-fabricated circuits are characterized by measuring THz signal transmission through the circuits. Scalar network systems can measure amplitude ratio between input and output signals, but if you are interested in measuring phase variation too, there would be not many choices, you have to use vector network analyzer. A perfectly machined circuit should have the identical RF response with computer simulation predictions. However, in practice it is very hard to make THz circuits matched with computer models. RF ohmic losses are rapidly increased if surface roughness of the fabricated circuit is larger than the RF skin depth since surface current paths on the conductor wall starts to be severely broken and damaged. It is very important to manage surface roughness below RF skin depth to minimize attenuation. MEMS fabrication provides excellent surface profile, which is much smaller than THz skin depth.

4.4 *State-of-the-Art THz VEDs*

Recently, various types of μ VEDs have received considerable attention as possible high power coherent THz sources owing to their high energy conversion efficiency and large thermal power capacity. For THz/sub-THz applications of μ VEDs, planar circuit structures using a sheet electron beam have advantages over the

Ultra-precision nano-machining

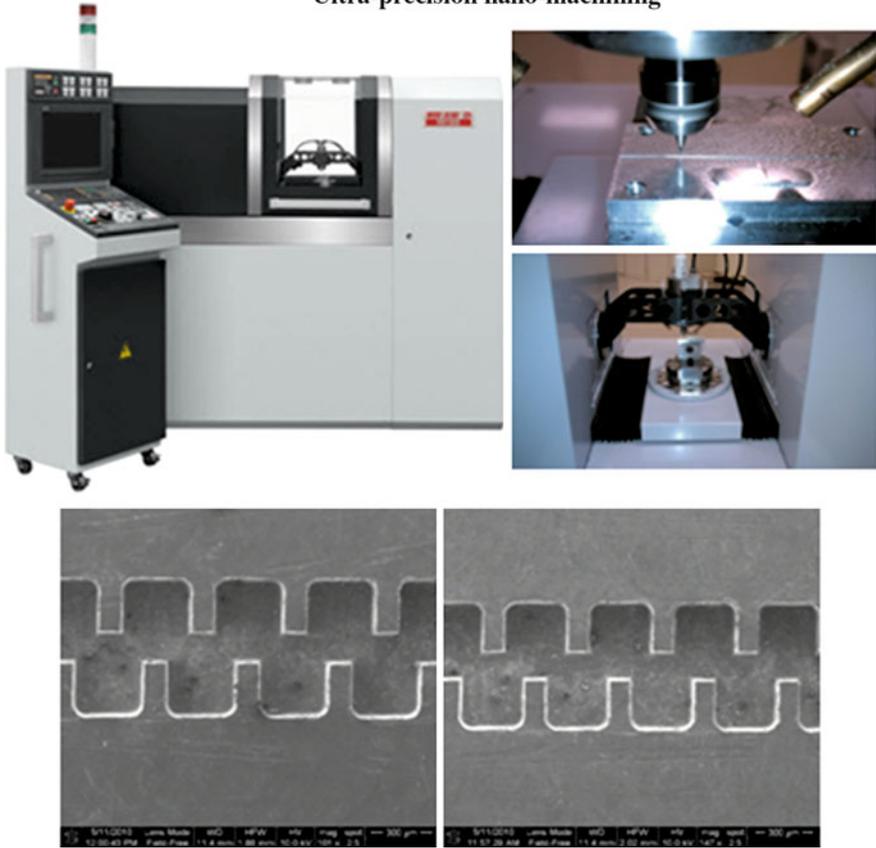
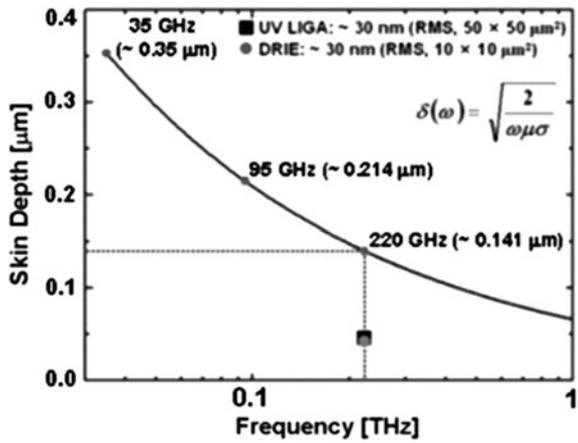


Fig. 23 DRIE process and SEM images

Fig. 24 Surface roughness vs. RF skin depth



typical round ones as their efficiency is proportional to the beam wave interaction area. Also, the beam current density for coherent beam-wave interaction is reduced with an increase of the sheet beam aspect ratio, which thereby allows compact magnetic focusing structures. Furthermore, the two-dimensional geometries of the planar sheet beam structures are more compatible with micro-fabrication technologies. This section will present the latest results and the research efforts on THz sheet beam μ VEDs for broadband and higher power applications.

4.4.1 Traveling Wave Tubes (TWTs)/Backward Wave Oscillators (BWOs)

Straight-Edge Connected Planar Helix TWTs

The most common slow-wave structure for application in TWTs is the circular helix. Nevertheless, the size of the helix scales inversely with frequency. Moreover, the circular helix is not amenable to fabrication using printed-circuit or microfabrication techniques, since for these techniques, a planar geometry is preferable [74, 75]. In this context, a planar helix SWS with straight-edge connections (PH-SEC) has been reported recently [74, 76]. The PH-SEC can be considered to be derived from a pair of parallel unidirectionally conducting (UC) screens [77], or from the rectangular helix [78]. The PH-SEC retains the broad bandwidth feature of the conventional circular helix. Besides that, the straight-edge connections can be realized easily using printed circuit fabrication or microfabrication techniques. In addition, its planar geometry is suitable for interaction with a sheet electron beam which can offer many advantages for high frequency TWTs [75]. To extend the operating frequency range of the PH-SEC above Ku-Band or so, microfabrication is the only solution to produce the required small size and three-dimensional structure. Microfabrication processes for solenoid inductors, an RF/microwave passive component that has a shape similar to that of the PH-SEC, have been reported extensively [79–81]. But the conventional multiexposure single development method adopted in these processes limits the aspect-ratio of the solenoid structure. Further, fabrication of the complete solenoid structure floating in air is also possible [75]. However such a configuration would not be preferable for TWT application due to the absence of a heat removal mechanism along the structure (Fig. 25).

Folded Waveguide and Coupled Cavity TWTs/BWOs

Among the available sources, folded waveguide traveling-wave tube (FWTWT) has advantages over other millimeter wave sources due to its moderate bandwidth ($\sim 10\%$) with power-handling capacity at higher frequencies. Also, its compatibility with planar fabrication by MEMS technologies draws an attention in sub-millimeter wave generation in folded waveguide structure to develop a miniaturized radiation sources [82–88]. In many cases, relatively high voltage

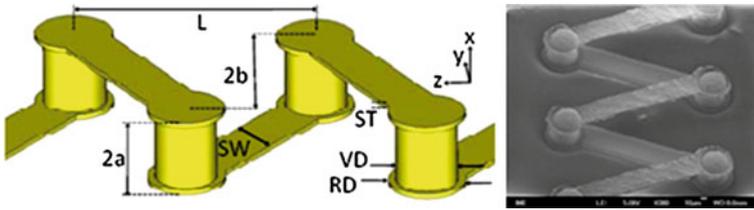


Fig. 25 THz helix TWT structure

(~ 20 kV) is used to maintain synchronism with fundamental space harmonic where the interaction impedance is the largest. However, many FWTWTs have recently been developed to operate at a relatively low voltage (< 12 kV) aiming at developing a compact millimeter-wave source with a moderate output power. In the millimeter-wave regime, the fabrication of a stacked cylindrical structure for coupled-cavity TWTs becomes difficult due to its smaller and complicated structure for brazing and welding. Therefore, it is desirable that the entire cavity structure be included in a single manufacturing process with the structure of the circuit simplified so that the number of manufacturing steps is minimized, while higher accuracy and tighter tolerance issues can be circumvented. For this purpose, there are several fabricating methods such as LIGA, micro-electro discharge machining (micro-EDM) [89], DRIE [90], wire electro-discharge grinding (WEDG) [91], etc. Here, LIGA fabrication is employed to design a ladder-core structure TWT in the millimeter-wave regime. LIGA can yield the metal features with very large aspect ratio (100 : 1) and suitable depths (3 mm). These dimensions are ideal for fabrication of RF resonant cavities with frequencies from 30 GHz to 1 THz. Details on MEMS fabrications were presented in the previous section (Fig. 26).

Staggered Double Vane Arrayed TWTs

The half-period staggering of a pair of 1-D vane arrays allows the fundamental mode to have a strong symmetric axial electric field distribution (in-phase field variation) along the electron beam propagation direction. In a conventional aligned-vane circuit, the first mode has an antisymmetric electric field along the TE direction. The resultant longitudinal field in the electron beam tunnel averages to zero, resulting in no net energy transfer from the beam to the RF wave. The input signal, thus, decays rather than undergoes the desired power amplification. In contrast, the first mode of the staggered vane structure has a symmetric field distribution, which has a sinusoidal phase variation along the axial direction. The full 3D PIC simulations predicted that an ultra-wideband TWT circuit, comprised of half-period-staggered double vane arrays (TE-mode) (Fig. 27) [92–94] produces 150 ~ 250 W output power, corresponding to ~ 3 –5.5% intrinsic electronic efficiency, overall efficiencies of $\sim 30\%$ (100% beam transmission) with depressed

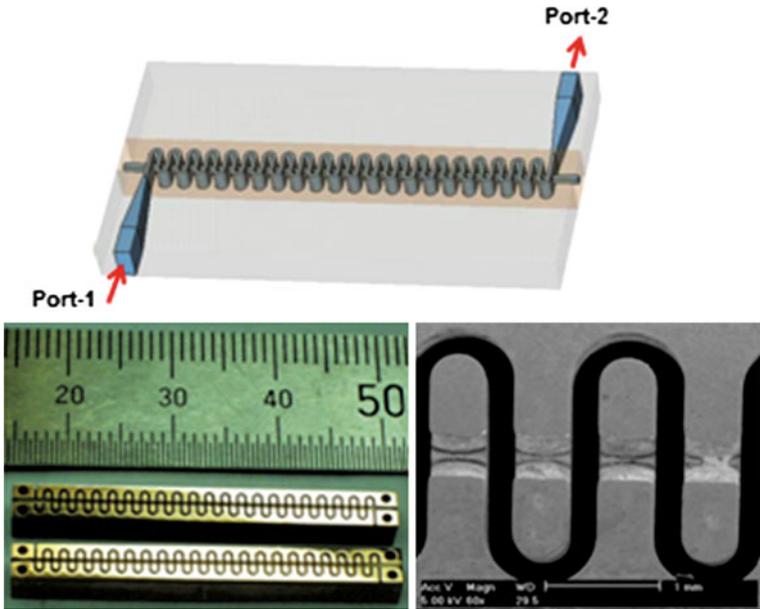


Fig. 26 THz folded waveguide TWT structure

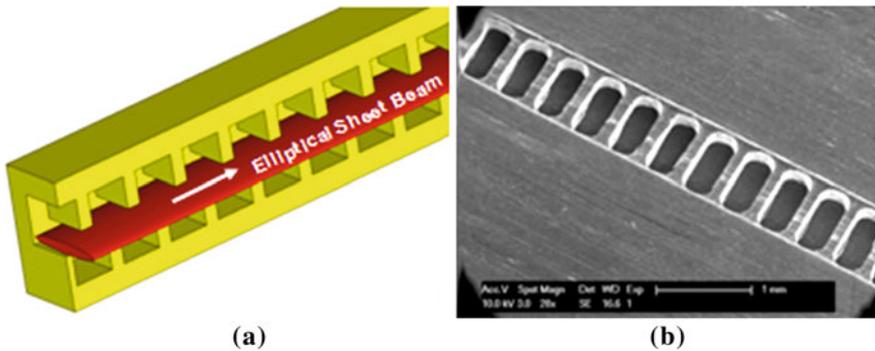


Fig. 27 Staggered double vane (Barnett-Shin) TWT circuit structure

collector (4-stage) operation, and >1000 W GHz power- bandwidth product, over 25% bandwidth at the center frequency of 0.22 THz.

Meanderline TWTs

One potential slow wave structure design for a Si-based THz or mm-wave regime TWT is the meander line. In its simplest form, the meander line consists of a

serpentine metallization on top of a planar substrate. However, in order to maximize the coupling between the electromagnetic signal and the electron beam, it is necessary to place the metallization on top of a raised meander ridge of silicon.

4.4.2 Klystrons

THz Nano-Reflex Klystron

New micromachining techniques now provide the technology to fabricate nano-scale reflex klystron oscillators with dimensions suitable for operation in the THz region of the electromagnetic spectrum. Nano-klystrons are to be used as local oscillator elements for high-resolution heterodyne spectroscopy instruments and as THz sources for bio-imagers and contraband detectors, high bandwidth communications and atmospheric chemical and biohazard probes. Initial estimates indicate power output in the range of milliwatts at 1.2 THz, which is significantly higher than that possible with other oscillators or multiplier based sources. The device requires an electron gun capable of providing up to a kA/cm^2 . Both cold and hot cathodes are being developed, with cold cathodes being the preferred choice for the eventual stand-alone device. Through a systematic study, “ropes” of multi-walled carbon nanotubes (MWNTs) have been identified as the most suitable candidate for the electron source. Using MWNTs synthesized on e-beam patterned catalyst arrays with varying parameters (bundle diameter and inter-bundle spacing), a suitable arrangement for high emission currents at low fields has been determined. We can repeatedly generate $>1.4 \text{ A}/\text{cm}^2$ at fields as low as $<3 \text{ V}/\mu\text{m}$, prior to focusing. The raw emission density can be further increased and this, in combination with beam focusing optics, is expected to achieve the required target current densities. For initial testing, a very high-resolution crossover type hot electron gun has been developed. The gun was designed to deliver $\sim 0.5 \text{ kA}/\text{cm}^2$ at the repeller of the nano-klystron. The source current density was designed for $16.2 \text{ A}/\text{cm}^2$. The gun includes three grids, one electrostatic lens and a magnetic lens and is mounted with a 0.6 THz copper constructed nano-klystron with a phosphor screen at the back end. The electron beam was successfully focused onto the phosphor screen behind the device (through the empty repeller hole) in a tunnel of $\sim 40 \mu\text{m}$ diameter. A current of 3.4 mA was recorded at the phosphor screen amounting to a current density of $\sim 270 \text{ A}/\text{cm}^2$. This gun will be used for device testing after a shaped repeller electrode has been installed (Fig. 28).

Multi-Cavity Sheet Beam Klystron

A cluster-type sheet beam klystron composed of a multicavity and multigap structure is often used in high-frequency vacuum tubes to increase cavity impedance and beam-wave coupling. A sheet beam has strong advantages over a round one in THz/sub-THz applications in klystrons and TWTs, as it can increase

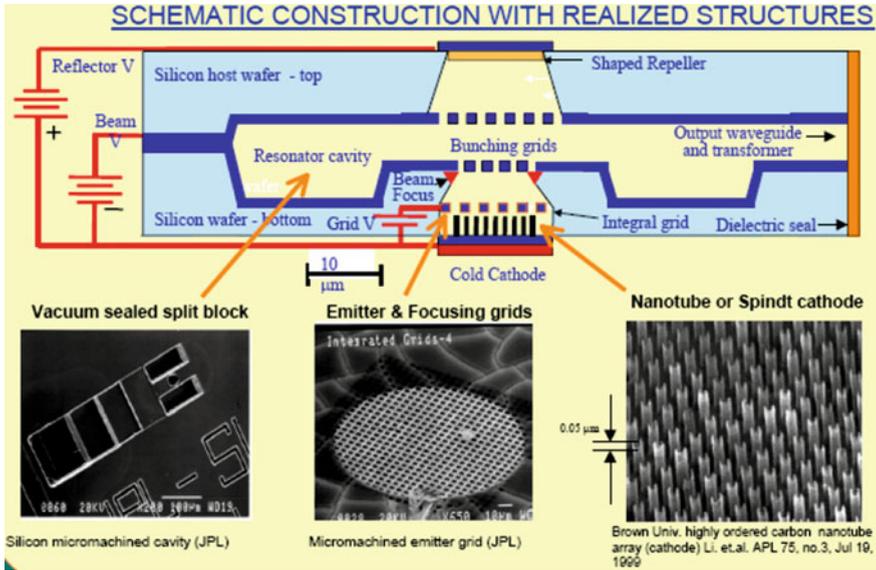


Fig. 28 THz reflex klystron using CNT FEA beam emitter (courtesy JPL)

radiation efficiency proportional to the beam wave interaction area and significantly lower current density for coherent radiation, thereby reducing the B-field requirements for beam focusing. Moreover, the resultant geometrical configuration is more compatible with microfabrication [95–102]. However, the cavity impedance (R/Q) and coupling coefficient (M) are lower in the sheet-beam cavities than in the equivalent round beam ones. A multigap structure has thus been used to compensate for insufficient R/Q and M in the intermediate and output cavities, which are designed to have uniform electric field distribution over the sheet electron beam. Such considerations led to an effort [103–107] to develop a 50 kW peak-power 2.5 kW-average-power W-band SBK (WSBK) comprised of eight cavities with a 12 : 1 aspect ratio sheet beam. The prototype WSBK was designed to have multigap cavities, and in particular, the output cavity is comprised of a seven-gap velocity-tapered structure to maximize circuit efficiency. The interaction circuit design employs eight staggetuned cavities (multigap structure) and a 12 : 1-aspect-ratio sheet electron beam (74 kV and 3.6 A) to produce 50 kW peak power (2.5 kW average) and 40 dB gain with 200 MHz instantaneous bandwidth. The output cavity is designed to have a quasi-optical (QO) external coupler utilizing optical wave superposition. The circuit design has been optimized by using a 1-D disk-model code and a 3-D particle-in-cell (PIC) solver. The iterative simulation analysis predicts that a five-gap configuration is the optimum structure for a QO-output cavity because it provides sufficient output power and stable single frequency operation without mode competition. The 3-D PIC simulation predicts that the designed circuit produces stable 50 kW output power from a 4 W input

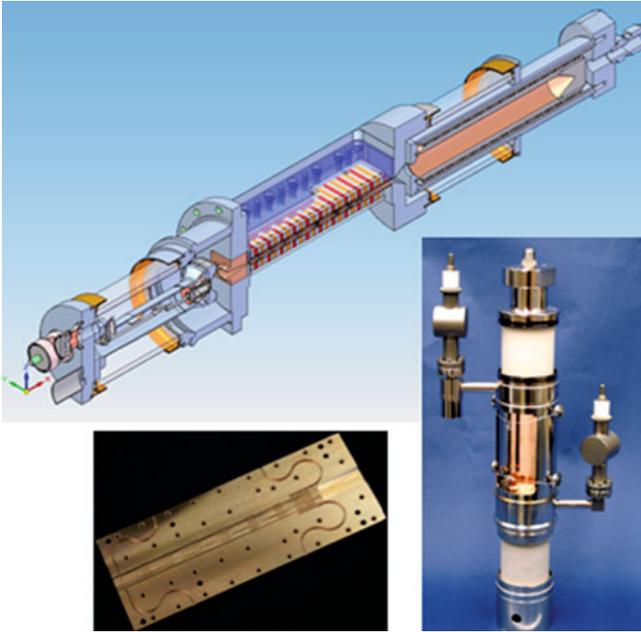


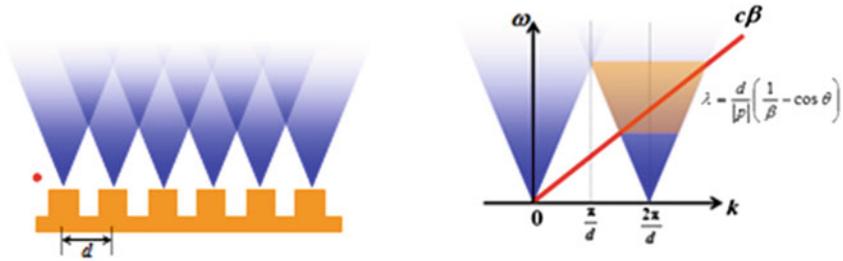
Fig. 29 Multi-cavity sheet beam klystron for W-band application

driving signal, with 40 dB gain, at 94.5 GHz. This might be the only THz device demonstrating real kilowatt level radiation power (Fig. 29).

4.4.3 Smith-Purcell Radiation (SPR) FELs

Smith-Purcell (SP) radiation is emitted when an electron passes close to the surface of a metallic grating. The radiation becomes coherent (fluence proportional to the square of the number of electrons) when the electrons are in bunches whose dimensions are smaller than the wavelength of the radiation. This has been observed in experiments in which the electrons are prebunched by an rf linac. The enhancement of the spectral intensity is accompanied by large changes in the angular and spectral distribution of the radiation when the electrons appear in periodic bunches. This is called superradiance. Recently, superradiant SP radiation has been observed from a so-called Smith-Purcell free-electron laser (SP-FEL) in which the electrons are bunched by the lasing process. As in other slow-wave structures, the electron beam in a SP-FEL interacts with an evanescent wave for which the phase velocity matches the electron velocity and amplifies it. The frequency of this wave lies below the range of SP radiation and the wave is not radiated except from the ends of the grating. However, the bunching of the electrons by the interaction with the evanescent wave enhances the ordinary

- Spontaneous SPR (Low Current Density) → Incoherent Radiation



- Stimulated SPR (High Current Density) → Coherent Radiation

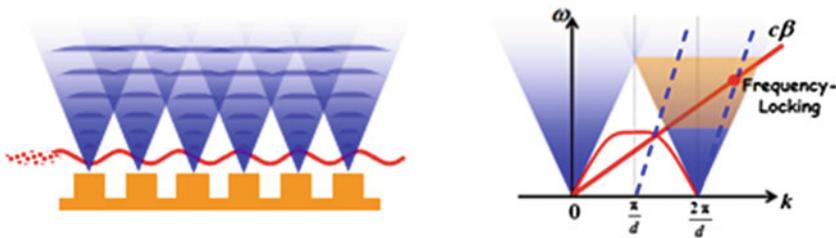


Fig. 30 Smith-Purcell radiations (SPR)—incoherent and coherent (superradiant) radiations

Smith-Purcell radiation and changes the angular and spectral distribution due to superradiant effects. The coherent superradiant SPR [108] is a few orders of magnitude larger than a spontaneous one, but its high current density transition threshold, increasing steeply with frequency, has prevented practical investigation leading to advances in THz SPR. Besides, various THz application fields constantly demand the raising of the radiation intensity. In superradiant SPR, the fundamental emission mechanism includes surface plasmon excitation [109]. There are many ways to excite superradiant SPR like using resonance cavity over the grating, replacing the metal grating with photonic crystals, using relativistic electrons, pre-bunched electron beam, multi-beam interaction, and so on. They all effectively increase radiation intensity and energy conversion efficiency of coherent SPR (Fig. 30).

4.4.4 Gyrotrons

DNP-NMR Gyrotrons

NMR spectroscopy has increased in importance as it is an element specific probe that can distinguish very small changes in the surroundings of different sites (e.g., the number of corners by which an SiO₄ unit is connected into a structure) which

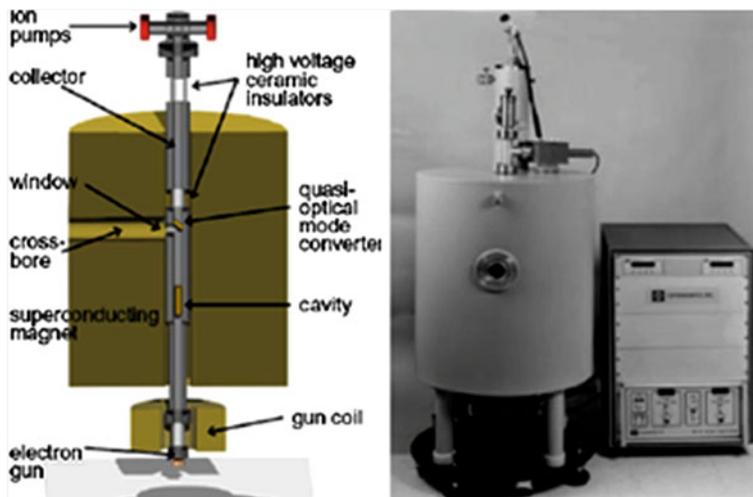


Fig. 31 DNP-NMR gyrotron

has become important throughout the sciences. Its major drawback is the intrinsically relatively weak signal due to the small thermally derived population differences between nuclear energy levels. NMR of solids was revolutionized with the implementation of cross-polarization (CP) that transferred magnetization from nuclei with high magnetic moments (e.g. ^1H) to more dilute nuclei with smaller magnetic moments (e.g. ^{13}C) that yielded a factor of ~ 4 increase in the ^{13}C NMR signal strength [110, 111]. MIT group has been continuously putting efforts on developing series of DNP-NMR gyrotrons operating at THz regimes. ~ 4 W of output power at 460 GHz had been used for DNP experiments which have been demonstrated successfully. 330 GHz gyrotron for a DNP application with higher power level has been recently developed. The CW output power of 18 W at a 10.1 keV, 190 mA electron beam. A second harmonic operation has been utilized for these THz gyrotrons which allows lower magnetic field for satisfying resonant condition to be possible [112, 113]. A gyrotron in FIR FU with a 21 T pulse magnet achieved the breakthrough of 1 THz. For convenience of the applications, CW gyrotrons (Gyrotron FU CW Series) is being developed. Gyrotron FU CW I has been developed and succeeded in the CW operation at 300 GHz under high power of 1.8 kW. The next gyrotron, Gyrotron FU CW 11 (395 GHz, 100 W) for application in NMR spectroscopy is being developed [114, 115] (Fig. 31).

THz Gyrotrons

Not only CW (a few Watt output power) THz gyrotrons, but also pulsed high power (more than kW output power) gyrotrons are being developed for specific

applications. Univ. Maryland group has recently suggested a possible application using a high power pulsed THz gyrotron to detect a concealed radioactive material remotely. Focused radiated beam from a gyrotron gives rise to a plasma breakdown at atmospheric environment if a radioactive material is present nearby although it is concealed. The difference of onset time of plasma breakdown production tells an existence of radioactive material nearby [116].

Technical development of several gyrotrons has been undertaken for achieving high power THz pulsed sources. As for a fundamental harmonic operation which directly requires to produce very strong magnetic field (~ 28 T for 1 THz), a pulsed solenoid technique has been used for achieving 1.5 kW at 1 THz recently [117]. 174 turns of a 40% Nb-40% Ti alloy wire in an outer copper shell are wired to the stainless steel pipe, which results in the solenoid inductance of 0.336 mH. The nominal voltage of ~ 3.1 kV and the current of 8 kA are applied from a pulsed solenoid power supply, and the solenoid was successfully tested up to 28 T.

Besides this pulsed solenoid technique, most efforts are to mitigate required magnetic fields for interacting circuits in gyrotrons. Typical higher order harmonic operation of gyrotrons are being considered for kW output power at THz range.

4.5 Next Generation Optoelectronic THz VEDs

4.5.1 Dielectric Cerenkov Radiation Sources

The use of a dielectric-lined waveguide structure is a means of producing narrow-band terahertz radiation in the form of electron-beam-driven coherent Cherenkov radiation wakefields. This concept builds on previously studied scenarios such as the Cherenkov maser and the Cherenkov free-electron laser. It is distinct in that it relies solely on coherent wakefield excitation instead of a microbunching instability gain process, analogous to the superradiant regime of FEL operation. The narrow bandwidth is due to the single-mode nature of the excitation, enabled by the exclusion (due to coherence) of discrete waveguide modes with wavelengths shorter than the driving electron bunch length. This allows an inherently broadband beam current profile to radiate power into a single frequency, which is selectable by appropriate choice of design parameters. Narrow-band THz coherent Cherenkov radiation driven by a sub-picosecond electron beam in a dielectric wakefield structure can be successfully adapted in the previously proven Cherenkov FEL concept to the realm of ultra-short electron beams such as are available in state-of-the-art user facilities around the world. This method has the potential to produce tunable, narrow-band, pulse-length-variable, multi-megawatt peak-power radiation at $f > 1$ THz in existing modern electron accelerators.

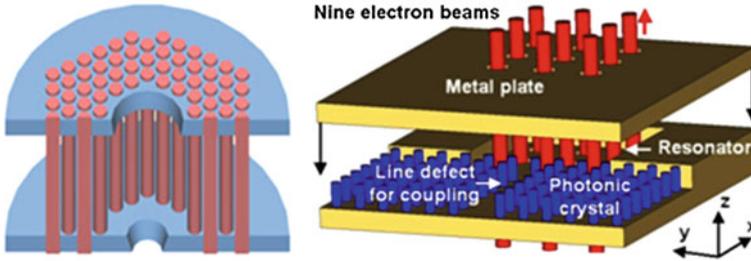


Fig. 32 Multi-beam reflex klystron using photonic crystal higher-order mode (HOM) filter

4.5.2 Multi-Beam Devices Using Photonic Crystal Resonators

A conventional reflex klystron is usually a low-power device, more so in the THz range of frequencies, where the available input electron beam power is restricted by the scaled-down size of the beam hole as well as by the power handling capability of the cavity resonator, the dimensions of which become smaller than several hundreds of microns. This makes enhancing the power capability of reflex klystrons in the terahertz range of frequencies a challenging task. Furthermore, it would be necessary to overcome two significant limiting factors that inhibit the use of higher order modes in a reflex klystron in a conventional rectangular cavity, namely, the mode distortion caused by output coupling and the mode competition of the desired modes with the lower order modes. The hybrid photonic-band-gap (PBG) cavity alleviates the limitations of a THz reflex klystron. In a hybrid PBG cavity, one of the metallic walls of a conventional metal cavity is replaced with a dielectric photonic crystal (PC) lattice by adjusting the band-gap property of which some of the lower order cavity modes could be made to lie outside the band gap of the lattice thereby excluding these modes from mode competition in a reflex klystron. A multibeam reflex klystron oscillator, employing, typically, nine electron beams, and a hybrid PBG cavity resonator, in a higher order transverse magnetic (TM) mode, typically, TM₃₃₀, for higher device powers, is developed to operate at 300 GHz (Fig. 32).

4.5.3 Metamaterial-Embedded THz VEDs

During the past few years, scientists have thus paid a great deal of attention to artificially engineered EM materials, Metamaterials [118–120], first hypothesized in 1967 by Victor Veselago, as they have explored physical properties which can be the breakthrough to jump over the insurmountable obstacles conventional electromagnetic theories confront. In particular, electron beam devices such as RF radiation sources and advanced accelerators have been considered to be well fit to

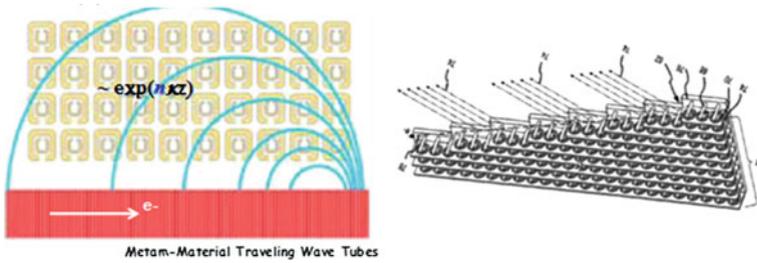


Fig. 33 Metamaterial slow wave structure (SWS) for TWT applications

the novel optoelectronic structures since negative permittivity and/or permeability distort dispersion relations of RF structures, so as to bring about extraordinary optical natures such as artificial magnetism, diffraction-free superlens, artificial plasma, localized magnetic dipole, super optical collimation, band-pass filtering, reverse Cherenkov radiation, and so on, which are widely advantageous for sheet beam devices. Considerable solution on the device stability and mode purity could be attained by architectural manipulation of optical waves. Metamaterials will provide large energy conversion efficiency, high gradients, high interaction impedance, and higher order mode (HOM) damping capacity in dynamic electron beam devices as novel beam-interactive electromagnetic structures. Extensive research on the beam-metamaterial interaction is underway to develop new physics, which will contribute to the advancement of various modern technologies. The novel optoelectronic structures appear to provide various advantages to μ VEDs in developing superluminescent THz source available for deep lithography, sub-wavelength microscopy/spectroscopy, nano-structure analysis, ultra-fast phenomena study, bio-medical imaging, radiology, oncology, and so on. Successful progress and realization of a compact THz sources will open the door to future bio-medical imaging and material science technologies (Fig. 33).

5 Summary and Conclusion

Various potential applications have been considered for compact sources of high power THz frequency electromagnetic radiation, including high resolution radar, high data rate communications, concealed weapon or threat detection, remote high resolution imaging, chemical spectroscopy, deep space research and communications, and biomedical diagnostics. VEDs have intrinsic advantages for compact high power sources at high frequencies from 100 GHz to ~ 1 THz, while there are still unsolved challenges associated with high power density and emission beam current density. To advance the coherent VED sources of Cherenkov radiation

type will require structures and strategies to handle higher power densities and high current electron beam generation and confinement. Recent research breakthroughs include novel methods to precisely fabricate miniature electromagnetic circuits, understanding rf breakdown, new high current density cathodes and improved understanding of cathode emission physics. Remaining challenges include delaying rf breakdown, innovating compact mmw-to-THz circuit configurations that are precisely dimensioned yet able to handle high average power, improving the lifetime of long-lived, low-emittance, uniformly emitting, high current density cathodes, and advances in the confinement and impact management of compact, high power electron beams. Prospects for solutions to these and other challenges are positive, owing to the emergence of powerful three-dimensional computational models for electromagnetics, beam optics, beam-wave interactions and particle kinetics, and surface physics. The radiation method using near-field interaction utilizing electron beam is considered very attractive approach to pave the way to develop high power sources for practical THz applications.

References

1. Smye, S.W., Chamber, J.M., Fitzgerald, A.J., Berry, E.: The interaction between terahertz radiation and biological tissue. *Phys. Med. Biol.* **46**, R101–R112 (2001)
2. Xu, J., Zhang, H., Yuan, T., Xu, X., Zhang, X.-C., Reightler, R., Madaras, E.: T-rays identify defects in insulating materials. In: *CLEO, CMB 2*, 2004
3. Hermann, M., Tani, M., Sakai, K., Fukasawa, R.: Terahertz imaging of silicon wafers. *J. Appl. Phys.* **91**, 1247 (2002)
4. Kiwa, T., Tonouchi, M., Yamshita, M., Kawase, K.: Laser terahertz-emission microscope for inspecting electrical faults in integrated circuits. *Opt. Lett.* **28**, 2058–2063 (2003)
5. Kawase, K., Ogawa, Y., Watanabe, Y., Inoue, H.: Non-destructive terahertz imaging of illicit drugs using spectral fingerprints. *Opt. Exp.* **11**, 2549–2554 (2003)
6. Seibert, K., Löffler, T., Quast, H., Thomson, M., Bauer, T., Leonhardt, R., Czasch, S., Roskos, H.G.: All-optoelectronic continuous wave THz imaging for biomedical applications. *Phys. Med. Biol.* **47**, 3743 (2002)
7. Zhao, G., Mars, M., Wenckebach, T., Planken, P.: Terahertz dielectric properties of polystyrene foam. *J. Opt. Soc. Am.* **19**, 1476 (2002)
8. Descour, M.R., et al.: Toward the development of miniaturized imaging systems for detection of pre-cancer. *IEEE J. Quant. Electron.* **38**, 122 (2002)
9. Arnone, D.D., Ciesla, C.M., Corchia, A., Egusa, S., Pepper, M., Chamberlain, J.M., Bezant, C., Linfield, E.H., Clothier, R., Khammo, N.: Applications of terahertz (THz) technology to medical imaging terahertz spectroscopy and applications. In: *SPIE*, vol. 3828, p. 209. Munich, Germany (1999)
10. Brucherseifer, M., Nagel, M., Haring Bolivar, P., Kurz, H., Bossenhoff, A., Büttner, B.: Label-free probing of the binding state of DNA by time-domain terahertz sensing. *Appl. Phys. Lett.* **77**, 4049 (2000)
11. Hu, B.B., Nuss, M.C.: Imaging with terahertz waves. *Opt. Lett.* **20**, 1716 (1995)

12. Ferguson, B., Wang, S., Gray, D., Abbott, D., Zhang, X.-C.: Terahertz imaging of biological tissue using a chirped probe pulse. In: Proceedings of SPIE: Electronics and Structures for MEMS II, vol. 4591, pp. 172–184. Adelaide, Australia (2001)
13. Han, P.Y., Cho, G.C., Zhang, X.-C.: Time domain transillumination of biological tissues with terahertz pulses. *Opt. Lett.* **25**, 242 (2000)
14. www.chemprocess.chalmers.se/dokument/pdf05/vinnoval/pdf
15. Löffler, T., Kreß, M., Thomson, M., Hahn, T., Hasegawa, N., Roskos, H.G.: Comparative performance of terahertz emitters in amplifier-laser-based systems. *Semicond. Sci. Technol.* **20**, S134 (2005)
16. Cook, D.J., Hochastrasser, R.M.: In: *Ultrafast Phenomena XII*. Springer Series of Chemical Physics, vol. 66, p. 197 (2000)
17. Phillips, T.G., Keene, J.: Submillimeter astronomy. *Proc. IEEE* **80**, 1662–1678 (1992)
18. Elisawitz, D. et al.: Scientific motivation and technology requirements for the SPIRIT and SPECS far-infrared/submillimeter space interferometers. In: Proceedings of SPIE, vol. 4013. Munich, Germany (2000)
19. Melnick, G., et al.: The submillimeter wave astronomy satellite: science objectives and instrument description. *Astrophys. J. Lett.* **539**(pt. 2), L77 (2000)
20. Herbst, E., Ann, E.: *Rev. Phys. Chem.* **46**, R27 (1995)
21. Waters, J.W.: Submillimeter wavelength heterodyne spectroscopy and remote sensing of the upper atmosphere. *Proc. IEEE* **80**, 1679–1701 (1992)
22. Feruson, B.: Horizons and Hurdles, Internal Report. Physics Department, Resselauer Polytechnic Institute (2001)
23. Han, P.Y., Zhang, X.-C.: Coherent, broadband mid-Infrared terahertz beam sensors. *Appl. Phys. Lett.* **73**, 3049 (1998)
24. Leitenstofer, A., Hunsche, S., Shah, J., Nuss, M.C., Knox, W.H.: Detectors and sources for ultrabroadband electro-optic sampling: experiment and theory. *Appl. Phys. Lett.* **74**, 1516 (1999)
25. Nahata, A., Yardley, J.T., Heinz, T.: Free space electro-optic detection of continuous-wave terahertz radiation. *Appl. Phys. Lett.* **75**, 2524 (1999)
26. Auston, D.H., Smith, P.R.: Generation and detection of millimeter waves by picosecond photoconductivity. *Appl. Phys. Lett.* **43**, 631–633 (1983)
27. Zhang, X.-C., Jin, Y., Hewitt, T.D., Kingsley, L.E., Weiner, M.: THz radiation by carrier transport of optical rectification ultrafast electron. *Optoelectron* **14**, 99 (1993)
28. Fattinger, C., Grischkowsky, D.: *Appl. Phys. Lett.* **54**, 490 (1989)
29. Faist, J., et al.: Quantum cascade laser. *Science* **264**, 553 (1994)
30. Vurgaftman, I., Meyer, J.R.: Optically pumped type-II interband terahertz lasers. *Appl. Phys. Lett.* **75**, 879 (1999)
31. Mueller, E.R.: Terahertz radiation: applications and sources. *The Industrial Physicist, AIP (Aug/Sep)*, 27–29 (2003)
32. Eisele, H.: Recent advances in the performance of InP Gunn devices and GaAs TUNNETT diodes for the 100–300 GHz frequency range and above. *IEEE Trans. Microw. Theory Tech.* **48**, 626 (2000)
33. Maiwald, F., et al.: Terahertz frequency multiplier chains based on planar Schottky diodes. *Proc. SPIE Soc. Opt. Eng.* **4855**, 447 (2003)
34. Landau, L.D., Lifshitz, E.M.: Landau and Lifshitz course of theoretical physics. In: *Electrodynamics of Continuous Media*, vol. 8, 2nd edn. Pargamon, New York, ch. XIV
35. Jackson, J.D.: *Classical Electrodynamics*. Wiley, New York (1998)
36. Hirshfield, J.L., Wachtell, J.M.: Electron cyclotron maser. *Phys. Rev. Lett.* **12**, 533 (1964)
37. Bajaj, V.S., et al.: Dynamic nuclear polarization at 9 Tesla using a novel 250 GHz gyrotron microwave source. *J. Magn. Reson.* **160**, 85 (2003)

38. Neil, G.R., Merminga, L.: Technical approaches for high-average-power free-electron lasers. *Rev. Modern Phys.* **74**, 685 (2002)
39. O'shea, P.G., Freud, H.P.: Free-electron lasers: status and applications. *Science* **292**, 1853 (2001)
40. Jeong, Y.U., et al.: First lasing of the KAERI compact far-infrared free-electron laser driven by a magnetron-based microtron. *Nucl. Inst. Meth. A* **475**, 4750 (2001)
41. Goebel, P.M., Liou, R.R., Menninger, W.L., Zhai, X., Adler, E.A.: Development of linear traveling wave tubes for telecommunications. *IEEE Trans. Electron Dev.* **48**, 7481 (2001a)
42. Dibb, D.R., Le Borgue, R.H.: 45 GHz TWT—recent advances. In: *International Vacuum Electronics Conference*, Monterey, CA (2000)
43. Symons, R.S., Arfin, B., Boesenberg, R.E., Ferguson, P.E., Kirshner, M., Vaughn, J.R.M.: An experimental clustered cavity klystron. In: *International Electron Devices Meeting Digest*, p. 153. Washington DC (1987)
44. Gelvich, E.A., Borisov, L.M., Zhary, Y.V., Zakudayev, A.D., Pobedonostev, A.S., Poognin, V.I.: The new generation of high power multi-beam klystrons. *IEEE Trans. Microw. Theory Tech.* **41**, 15 (1993)
45. Goebel, D.M., Liou, R.R., Menninger, W.L., Zhai, X., Adler, E.A.: Development of linear traveling wave tubes for telecommunications applications. *IEEE Trans. Electron Dev.* **48**, 74 (2001b)
46. Granastein, V.L., Parker, R.K., Armstrong, C.M.: Vacuum electronics at the dawn of the twenty-first century. *Proc. IEEE* **87**, 702–716 (1999)
47. Abrams, R.H., Levush, B., Mondelli, A.A., Parker, R.K.: Vacuum electronics at the dawn of the twenty-first century. *Proc. IEEE* **87**, 702–716 (1999)
48. Calame, J.P., Abe, D.K.: Applications of advanced materials technologies to vacuum electronics. *Proc. IEEE* **87**, 840–864 (1999)
49. Gitinz J.F.: *Power Traveling Wave Tubes*. Elsevier, New York (1965)
50. PopSci's Best of What's New 2007. http://www.popsci.com/popsci/flat/bown/2007/green/item_67.html. Retrieved 28 Feb 2010
51. US Patent 7629497—Microwave-based recovery of hydrocarbons and fossil fuels Issued 8 Dec 2009
52. 1997 Industrial Assessment of the Microwave Power Tube Industry—US Department of Defense
53. Beryllium oxide properties. http://en.wikipedia.org/wiki/Beryllium_oxide
54. Duke university free-electron laser laboratory. <http://www.fel.duke.edu/>
55. Duarte, F.J. (ed.): *Tunable Lasers Handbook*. Academic, New York (1995) Chapter 9
56. SLAC National Accelerator Laboratory. <http://home.slac.stanford.edu/pressreleases/2009/20090421.htm>
57. Nusinovich, G.S., Cooke, S.J., Botton, M., Levush, B.: Wave coupling in sheet- and multiple-beam traveling-wave tubes. *Phys. Plasmas* **16**, 063102 (2009)
58. Ives, R.L.: Microfabrication of high-frequency vacuum electron devices. *IEEE Trans. Plasma Sci.* **32**, 1277 (2004)
59. Zhang, H., Wang, J., Tong, C., Li, X., Wang, G.: Numerical studies of powerful terahertz pulse generation from a super-radiant surface wave oscillator. *Phys. Plasmas* **16**(12), 123104 (2009)
60. Ives, R.L., Kory, C., Read, M., Neilson, J., Caplan, M., Chubun, N., Schwartzkopf, S., Witherspoon, R.: Development of backward-wave oscillators for terahertz applications. In: *Terahertz for Military Applications*. Orlando, FL (2003)
61. Chen, L., Guo, H., Chen, H.-Y., Chu, K.-R.: *IEEE Trans. Plasma Sci.* **28**(3), 626 (2000)
62. Naumenko, V.D., Schumann, K., Variv, D.M.: Miniature 1 kW, 95 GHz magnetrons. *Elec. Lett.* **35**(22), 1960 (1999)
63. Hargreaves, T.A., Scheitrum, G.P., Bemis, T., Higgins, L.: *IEDM 93*, 349 (1993)
64. O'Shea, P.G., Freund, H.P.: *Science* **292**(8), 1853 (2001)
65. Ehrfeld, W., Schmidt, A.: *J. Vac. Sci. Tech. B* **16**(6), 3526 (1998)

66. Ruhmann, R., Pfeiffer, K., Falenski, M., Reuther, F., Engelke, R., Grutzner, G.: SU-8 - a high performance material for MEMS applications. *Polym. MEMS* **45** (2002)
67. Johnson, D.W., Jeffries, A., Minsek, D.W., Hurditch, R.J.: Improving the process capability of SU-8, part II. *J. Photopolym. Sci. Tech.* **14**(5), 689 (2001)
68. O'Brien, G.J., Monk, D.J., Najafi, K.: Device and process technologies for MEMS and microelectronics II. In: Chiao, J.-C. (ed.) *Proc. SPIE*, 315, vol. 4592 (2001)
69. Bhardwaj, J., Ashraf, J., MacQuarrie, A.: In: *Symposium on Microstructures and Microfabricated Systems*, Electrochemical Society (1997)
70. Liu K.X., Chiang C.-J., Heritage J.P.: Photoresponse of gated *p*-silicon field emitter array and correlation with theoretical models. *J. Appl. Phys.* **99**, 034502 (2006)
71. Technical data sheet (Ver 4.1) at the Kayaku-Microchem Web site (<http://www.kayakumicrochem.jp>)
72. Digital Technology Laboratory. 3805 Faraday Ave., Davis, CA, 95618. (530) 746-7400. <http://www.dtlab.com>
73. Shin, Y.-M., Barnett, L.R., Gamzina, D., Luhmann, N.C. Jr, Field M., Borwick, R.: Terahertz vacuum electronic circuits fabricated by UV lithographic molding and deep reactive ion etching. *Appl. Phys. Lett.* **95**, 181505 (2009)
74. Carlsten, B.E., Russell, S.J., Earley, L.M., Krawczyk, F.L., Potter, J.M., Ferguson, P., Humphries, S.: Technology development for a mmwave sheet-beam TWT. *IEEE Trans. Plasma Sci.* **33**(1), 85–93 (2005)
75. Srivastava V.: THz vacuum microelectronic devices. In: *International Symposium on Vacuum Science and Technology (IVS2007)*
76. Kory, C.L., Ives, L., Read, M., Phillips, P.: Novel TWT interaction circuits for high frequency applications. In: *Proceedings of the IEEE International Vacuum Electronics Conference*, pp. 51–52 (2004)
77. Arora, R.K., Bhat, B., Aditya, S.: Guided waves on a flattened sheath-helix. *IEEE Trans. Microwave Theory Tech.* **MTT-25** 71–72 (1977)
78. Fu, C.F., Wei, Y.Y., Wang, W.X., Gong, Y.B.: Dispersion characteristics of a rectangular helix slow-wave structure. *IEEE Trans. Electron Devices* **55**(12) (2008)
79. Yoon, J.-B., Han, C.-H., Yoon, E., Kim, C.-K.: Monolithic fabrication of electroplated solenoid inductors using three-dimensional photolithography of a thick photoresist. *Jpn. J. Appl. Phys.*, pt. 1, **37**(12B) 7081–7085 (1998)
80. Yoon, J.-B., Kim, B.-I., Choi, Y.-S., Yoon, E.: 3-D construction of monolithic passive components for RF and microwave ICs using thick-metal surface micromachining technology. *IEEE Trans. Microwave Theory Tech.* **51**(1), 279–288 (2003)
81. Kim, Y.-J., Allen, M.G.: Surface micromachined solenoid inductors for high frequency applications. *IEEE Trans. Comp. Packag. Manufact. Technol.* **21**(1) 26–33 (1998)
82. Mann, C.M.: Fabrication technologies for terahertz waveguide. In: *Proceedings of 6th International Conference Terahertz Electronics* 46–49 (1998)
83. Moon, S.W., Mann, C.M., Maddison, B.J., Turcu, I.C.E., Allot, R., Huq, S.E., Lisi, N.: Terahertz waveguide components fabricated using 3D X-ray microfabrication technique. *Electron. Lett.* **32**, 1794–1795 (1996)
84. Collins, C.E., Miles, R.E., Pllard, R.D., Steenson, D.P., Digby, J.W., Parkhurst, G.M., Chamberlain, J.M., Cronin, N.J., Davies, S.R., Bowen, J.W.: Technique for micro-machining millimeter-wave rectangular waveguide. *Electron. Lett.* **34**, 996–997 (1998)
85. Collins, C.E., Miles, R.E., Digby, J.W., Parkhurst, G.M., Pollard, R.D., Chamberlain, J.M., Steenson, D.P., Cronin, N.J., Davies, S.R., Bowen, J.W.: A new micro-machined millimeter-wave and terahertz snaptogther rectangular waveguide technology. *IEEE Microwave Guided Wave Lett.* **9**, 63–65 (1999)
86. Song, J.J., Bajikar, S., Kang, Y.W., Kustom, R.L., Mancini, D.C., Nassiri, A., Lai, B., Feinerman, A.D., White, V.: LIGA fabrication of mm-wave accelerating cavity structures at the Advanced Photon Source (APS). In: *Proceedings of Particle Accelerator Conference*, pp. 461–463. (1997)

87. Dohler G., Gagne, D., Gallagher, D., Moats, R.: Serpentine waveguide TWT,” in Proceedings of technical digest international electron devices meeting. pp. 485–488 (1987)
88. Choi, J.J., Armstrong, C.M., Calise, F., Ganguly, A.K., Kyser, R.H., Park, G.S., Parker, R.K., Wood, F.: Experimental observation of coherent millimeterwave radiation in a folded waveguide employed with a gyrating electron beam. *Phys. Rev. Lett.* **76**, 4273–4276 (1996)
89. Reynaerts, D.: A versatile technology for silicon by electro-discharge machining. *Int. J. Jpn. Soc. Precision Eng.* **33**(2), 114–128 (1999)
90. Madou, M.: Fundamentals of microfabrication. CRC, Boca Raton (1997)
91. Masuzawa, T.: Wire electro-discharge grinding for micro-machining. *Ann. CIRP (International Institute for Production Engineering Research)* **34**, 431–434 (1985)
92. Shin, Y.-M., Barnett, L.R.: Intense Wideband Terahertz Amplification Using Phase-Shifted Periodic Electron-Plasmon Coupling. *Appl. Phys. Lett.* **92**, 091501 (2008)
93. Shin, Y.-M., Barnett, L.R., Luhmann Jr, N.C.: Strongly confined plasmonic wave propagation through an ultra-wideband staggered double grating waveguide. *Appl. Phys. Lett.* **93**, 221504 (2008)
94. Shin, Y.-M., Barnett, L.R., Luhmann, N.C. Jr.: Phase-shifted traveling wave tube circuit for ultra- wideband high power submillimeter wave generation. *IEEE Trans. Elec. Dev.* **56**, 706 (2009)
95. Booske, J.H., McVey, B.D., Antonsen Jr, T.M.: Stability and confinement of nonrelativistic sheet electron beams with periodic cusped magnetic focusing. *J. Appl. Phys.* **73**(9), 4140–4155 (1993)
96. Carlsten, B.E., Earley, L.M., Krawczyk, F.L., Russell, S.J., Potter, J.M., Ferguson, P., Humphries, S., Jr.: Stability of an emittance-dominated sheet-electron beam in planar wiggler and periodic permanent magnet structures with natural focusing. *Phys. Rev. ST—Accel. Beams* **8** (6) 062001-1–062001-14 (2005)
97. Carlsten, B.E.: Modal analysis and gain calculations for a sheet electron beam in a ridged waveguide slow-wave structure. *Phys. Plasmas* **9**(12), 5088–5096 (2002)
98. Carlsten, B.E., Russell, S.J., Earley, L.M., Haynes, W.B., Krawczyk, F., Smirnova, E., Wang, Z.-F., Potter, J.M., Ferguson, P., Humphries, S.: MM-wave source development at Los Alamos. In: Proceedings of 7th Workshop High Energy Density High Power RF, AIP Conference, vol. 807, pp. 326–334 (2006)
99. Nguyen, K.T., Wright, E.L., Pershing, D.E., Levush, B.: Design of a G-band sheet-beam extended-interaction klystron. In: Proceedings of IEEE International Vacuum Electronics Conference, pp. 298–299 (2009)
100. Larsen, P.B., Abe, D.K., Cooke, S.J., Levush, B., Pchelnikov, Y.N.: Experimental characterization of a ka band sheet-beam coupled-cavity slow-wave structure. In: Proceedings of IEEE International Vacuum Electronics Conference, pp. 224–225 (2009)
101. Shin, Y.-M., Barnett, L.R., Luhmann Jr, N.C.: Phase-shifted traveling-wave-tube circuit for ultrawideband high-power submillimeterwave generation. *IEEE Trans. Electron Device* **56**(5), 706–712 (2009)
102. Shin, Y.M., So, J.K., Jang, K.H., Won, J.H., Srivastava, A., Park, G.S.: Superradiant terahertz Smith–Purcell radiation from surface Plasmon excited by counterstreaming electron beams. *Appl. Phys. Lett.* **90**(3), pp. 031502-1–031502-3 (2007)
103. Scheitrum, G., Caryotakis, G., Jensen, A., Burke, A., Haase, A., Jongewaard, E., Neubauer, M., Steele, B.: Fabrication and testing of a W-band sheet beam klystron. In: Proceedings of IEEE IVEC, 15–17 May 2007, pp. 1–2
104. Burke, A., Besong, V., Granlund, K., Jensen, A., Jongewaard, E., Phillips, R., Rauenbuehler, K., Scheitrum, G., Steele, R.: W-band sheet beam klystron PCM focusing design. In: Proceedings of IEEE International Vacuum Electron Conference, IEEE International Vacuum Electron Sources, 2006, pp. 485–486
105. Scheitrum, G., Caryotakis, G., Burke, A., Jensen, A., Jongewaard, E., Neubauer, M., Phillips, R., Steele, R.: W-band sheet beam klystron research at SLAC. In: Proceedings of International Vacuum Electron Conference, IEEE International Vacuum Electron Sources, pp. 481–482 (2006)

106. Scheitrum, G.: Design and construction of a W-band sheet beam klystron. In: Proceedings of 7th Workshop High Energy Density High Power RF, Kalamata, Greece, pp. 120–125 (2005)
107. Scheitrum, G., Caryotakis, G., Burke, A., Jensen, A., Jongewaard, E., Krasnykh, A., Neubauer, M., Phillips, R., Rauenbuehler, K.: W-band sheet beam klystron design. In: Proceedings of International Conference Infrared Millimeter Waves, 12th International Conference Terahertz Electronics, September 27–October 1, 2004, pp. 525–526
108. Urata, J., Goldstein, M., Kimmitt, M.F., Naumov, A., Platt, C., Walsh, J.E.: Superradiant Smith-Purcell emission. *Phys. Rev. Lett.* **80**, 516 (1998)
109. Ritchie, R.H.: Plasma losses by fast electrons in thin films. *Phys. Rev.* **106**, 874 (1957)
110. Becerra, L.R., Gerfen, G.J., Temkin, R.J., Singel, D.J., Griffin, R.G.: Dynamic nuclear polarization with a cyclotron resonance maser at 5 T. *Phys. Rev. Lett.* **71**(21) 3561–3564 (1993)
111. Bajaj, V.S., Hornstein, M.K., Kreischer, K.E., Sirigiri, J.R., Woskov, P.P., Mak-Jurkauskas, M.L., Herzfeld, J., Temkin, R.J., Griffin, R.G.: 250 GHz CW gyrotron oscillator for dynamic nuclear polarization in biological solid state NMR. *J. Magn. Reson* **189**(2), 251–279 (2007)
112. Hornstein, M.K., Bajaj, V.S., Griffin, R.G., Kreischer, K.E., Mastovsky, I., Shapiro, M.A., Sirigiri, J.R., Temkin, R.J.: Second harmonic operation at 460 GHz and broadband continuous frequency tuning of a gyrotron oscillator. *IEEE Trans. Electron Devices* **52**(5), 798–807 (2005)
113. Hornstein, M.K., Bajaj, V.S., Griffin, R.G., Temkin, R.J.: Continuous-wave operation of a 460 GHz second harmonic gyrotron oscillator. *IEEE Trans. Plasma Sci.* **34**(3), 524–533 (2006)
114. Idehara, T., Kosuga, K., Agusu, L., Ikeda, R., Ogawa, I., Saito, T., Matsuki, Y., Ueda, K., Fujiwara, T.: Continuously frequency tunable high power sub-THz radiation source – Gyrotron FU CW VI 600 MHz DNP-NMR spectroscopy. *J. Infrared Milli. Terahz Waves* (2010). doi:[10.1007/s10762-010-9643-y](https://doi.org/10.1007/s10762-010-9643-y)
115. Idehara, T., Kosuga, K., Agusu, L., Ogawa, I., Takahashi, H., Smith, M.E., Dupree, R.: Gyrotron FU CW VII for 300 MHz and 600 MHz DNP-NMR spectroscopy. *J. Infrared Milli. Terahz Waves* (2010). doi:[10.1007/s10762-010-9637-9](https://doi.org/10.1007/s10762-010-9637-9)
116. Granatstein, V.L., Nusinovich, G.S.: Detecting excess ionizing radiation by electromagnetic breakdown in air. *J. Appl. Phys.* **108**, 063304 (2010)
117. Glyavin, M.Y., Luchinin, A.G., Golubiatnikov, G.Y.: Generation of 1.5 kW, 1 THz Coherent radiation from a gyrotron with a pulsed magnetic field. *Phys. Rev. Lett.* **100**, 015101 (2008)
118. Nader, E., Ziolkowski, R.W.: *Metamaterials: physics and engineering explorations*. IEEE Press, Wiley (2006)
119. Zouhdi, S., Sihvola, A., Vinogradov, A.P.: *Metamaterials and plasmonics: fundamentals, modelling, applications*. Springer, New York (2008–2012)
120. Smith D.R.: What are electromagnetic metamaterials? *Novel electromagnetic materials* (2006-06-10)

Principle of Terahertz Free Electron Laser

Young Uk Jeong, Seong Hee Park and Kitae Lee

Abstract This part introduces free-electron lasers (FELs) as high-power terahertz (THz) sources. The main features of the FELs are wavelength tunability and high-power capability. Brief principle of the FEL is described by explaining the electron dynamics in a undulator and the stimulated amplification of the radiation due to the interaction between wiggling electron beam and undulator radiation. The status of the operating THz FELs and recent developments of table-top THz FELs are summarized.

1 Introduction

Laser, invented in early 1960s [1], has played a marvelous role in the advance of science and technology for human being during such a short period. Laser was already permeated not only into areas of the advanced science and technology, but also into our daily life unawares. Such a surprise can be explained by the simple causal relation: the laser, new light source, enables us to recognize the reality, just as we see a great part of things through the light. How can we put too cheap a valuation on or have our apathy to the worth of new tools, in spite that we are confronted with the fierce battle of science technology near future?

A laser in the range of terahertz is a useful tool to study and research properties of living life or molecular compounds, existing with us [2, 3]. It can be used as a precise tool in the area of competitive semiconductor industries to explore the fine

Y. U. Jeong (✉) · S. H. Park · K. Lee
WCI Center for Quantum-Beam-based Radiation Research,
Korea Atomic Energy Research Institute, Daejeon, Korea
e-mail: yujung@kaeri.re.kr

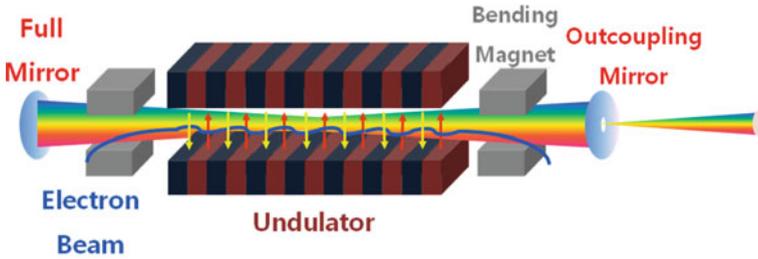


Fig. 1 Schematic view of an oscillator-type free electron laser (FEL)

structure of semiconductors. It is, ultimately, expected that it links to develop new structure or materials. Besides, there are immense applications in many areas, such as, medical science, environmental science, isotope separation, nuclear physics and engineering, and other basic material science. The range of terahertz is wide compared to that of visible wavelengths, of which lasers has been developed since 1960s. Possibilities of development of advanced technology and new knowledge due to the advent of this new light source may be higher than we expect.

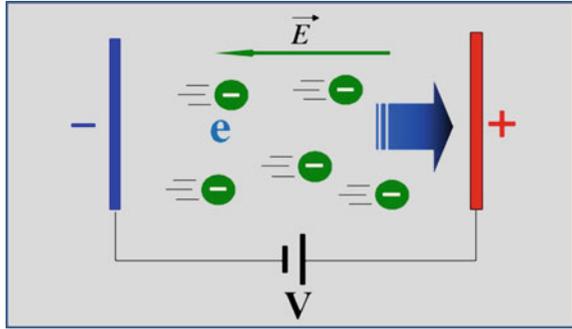
A laser light is monochromatic and not divergent. It has also the power stronger than other light sources. However, in general, conventional lasers cannot produce all colors or a certain color that we need. Each and every conventional laser has its own property, such as, wavelength and bandwidth. Up to now, a variety of lasers, which can be hardly memorized one by one, has developed and being developed.

A free electron laser (FEL) has new characteristic unlike the conventional quantum lasers [4, 5]. One is that it can generate any color as you wish. FELs, especially, fit for light sources in the ranges, terahertz or X-ray, in which it is quite difficult to produce with the conventional quantum lasers so far.

The FEL is a classical device which amplifies electromagnetic wave by stimulated emission, using relativistic electrons. Actually electrons in the FEL are not really free. They are just moving in the free space but governed by electric or magnetic forces. But the term of free electrons has been used to discriminate the bound electrons in atoms and molecules. To generate the FEL radiation, we need an accelerator and a periodic magnetic structure which is called ‘undulator’ or ‘wiggler’. The relativistic electrons from the accelerator passing through the undulator are subject to transverse acceleration, which results in emission of undulator radiation. In certain condition the undulator radiation can be amplified to be the FEL radiation. Usually FELs in the terahertz range operate in the oscillator configuration [6], but those in X-ray range operate in the amplifier configuration due to the absence of high-reflectivity materials for mirrors [7]. A schematic view of the FEL in oscillator configuration is shown in Fig. 1.

There are many kinds of electron accelerators. They can be classified into electrostatic, radio frequency, induction, and plasma accelerators according to the acceleration mechanism. More simply we can divide them as linear and circular accelerators depending on their shapes. In any case, almost all

Fig. 2 Electrons accelerating in the electrostatic field



accelerators use electric field to propel electrons to high speeds. And the high-speed electrons are usually controlled by magnetic field in the accelerators. Therefore, if you can recall the most famous two scientists, Charles-Augustin de Coulomb (14 June 1736–23 Aug 1806) and Hendrik Antoon Lorentz (18 July 1853–4 Feb 1928), and their outstanding contributions to the electromagnetism, Coulomb’s law and Lorentz force, respectively, the basic physics of the electron accelerators can be understood easily. An electron in the electric field, \vec{E} , feels a force proportional to the magnitude of the electric field and its charge, e (Fig. 2). We can simply write the equation of motion for the electron acceleration in the electric field as follows.

$$\begin{aligned} \vec{F} &= e\vec{E} = \frac{d\vec{P}}{dt} = m_e \frac{d\vec{v}}{dt} \quad (\text{for non-relativistic case, } v \ll c) \\ &= m_e \frac{d(\gamma\vec{v})}{dt} \quad (\text{for relativistic case, } v \approx c) \end{aligned}$$

where \vec{P} is the momentum of the electron, m_e is the electron mass, \vec{v} is the velocity of the electron, c is the light velocity in vacuum, and γ is the Lorentz factor which is expressed as $(1 - \beta^2)^{-1/2}$ with $\beta = v/c$.

When an electron is accelerated by the electric potential of V , then its kinetic energy is expressed by 1 eV (electron volt), which is an energy of $\sim 1.6 \times 10^{-19}$ J. Usually we use the unit of eV for the energy of electrons in the accelerator. Then the electron energy and Lorentz factor has the simple relation of

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} = 1 + \frac{KE}{m_e c^2} \cong 1 + \frac{KE[\text{MeV}]}{0.511[\text{MeV}]},$$

where KE is the kinetic energy of the electron and $m_e c^2$ is its rest-mass energy.

Recently radio-frequency (RF) linear accelerators are the most commonly used type of the electron accelerators (Fig. 3). Usual RF linear accelerators consist of electron guns, accelerating RF cavities and magnets for bending and focusing/defocusing. In the RF accelerators, the electrons are accelerated by high-power

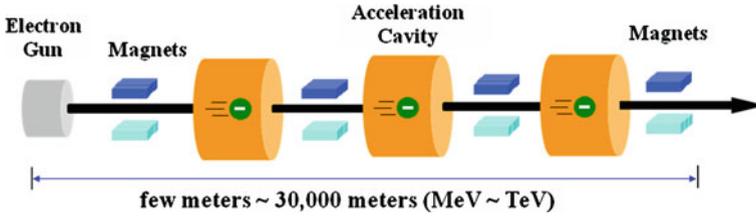


Fig. 3 Schematics of a radio-frequency linear accelerator

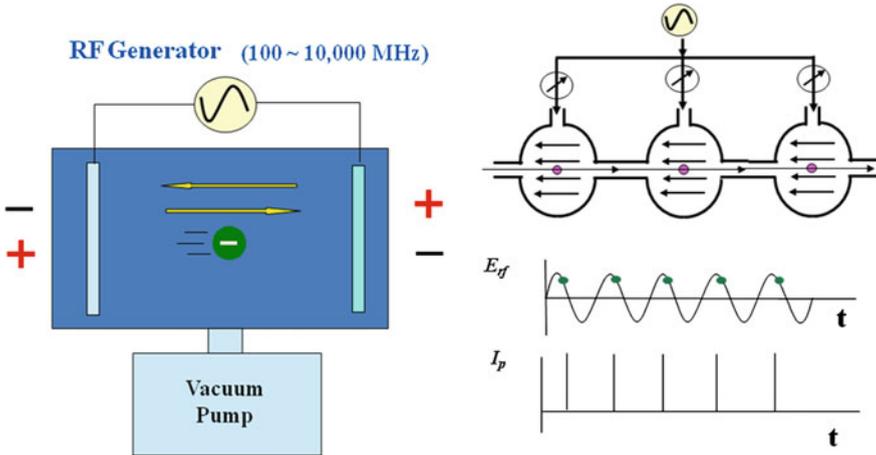


Fig. 4 Acceleration of bunched electron beams by using a radio-frequency linear accelerator

radio-frequency waves in high quality-factor cavities. The RF accelerators generate short-length and periodic-bunched electron beams as shown in Fig. 4.

When the speed of electrons is close to the light velocity, magnetic fields are more effective to change the motion of the high-energy electrons than electric fields. Therefore almost all undulators and devices for electron optics in the accelerators use magnetic fields. One can choose permanent magnets or electromagnets for generating undulator fields. There are planar or helical undulators depending on the configurations of the magnetic periodic structure. The planar undulators generate simple alternating magnetic fields through the electron pass as shown in Fig. 5. The radiation generated from the planar undulator is linearly polarized.

In the case of the helical undulator, as shown in Fig. 6, magnetic structure has helical configuration through the axis of electron pass. For both cases, important parameters to determine the radiation wavelength are undulator period, λ_w , and magnetic induction on the center axis of the undulator, B_w . The number of undulator periods, N_w , is a crucial parameter for the power of the undulator radiation and FEL gain. The key components of FELs, accelerators and undulators, are simply introduced for rough understanding. I will summarize this section by

Fig. 5 Magnetic configuration of a planar undulator

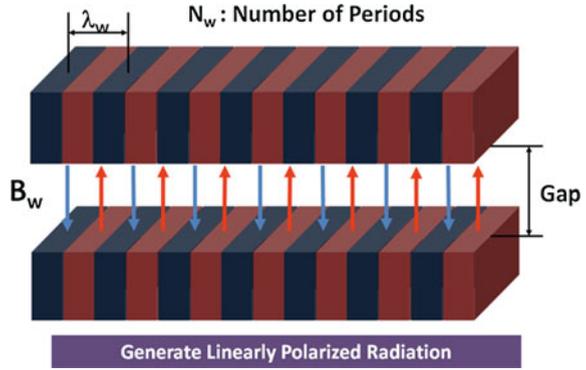
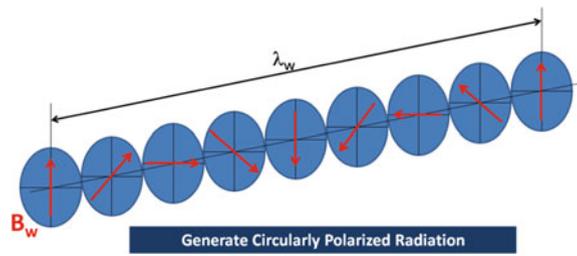


Fig. 6 Magnetic configuration of single period of a helical undulator



noting main characteristics of the FELs, for example wavelength tunability and potential to high average and peak power.

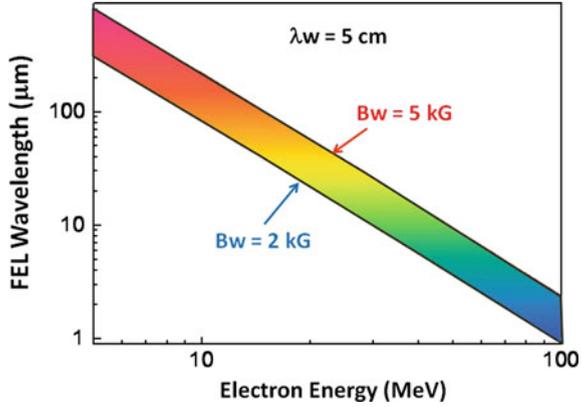
The most important feature of the FEL is wavelength tunability. FELs can generate any wavelength of electromagnetic waves in principle. There are FELs operating in the wavelengths from millimeters to angstroms. The wavelength of the undulator radiation, λ , is expressed with the electron energy and undulator parameters.

$$\lambda = \frac{\lambda_w}{2\gamma^2} \left(1 + \frac{K^2}{2} \right) \tag{1}$$

where $K \cong 0.934B_w\lambda_w$ with the dimensions of tesla and centimeters. We can generate any wavelength of radiation just by choosing proper electron energy and undulator parameters. As you can see from Fig. 7, FELs can cover wide range of the wavelengths by changing the electron energy. Precise tuning of the wavelength can be done by changing undulator magnetic field strengths.

Unlike solid-state, liquid, or gas lasers, the active medium of the FEL is just relativistic electron beam in a vacuum. There is no damage on the FEL medium, which means no power limit in the FEL output. During the cold war between USA and USSR, plans for destroying adversary's ICBMs were considered with mega-watt class FELs. Still there is an on-going project of US Navy on the laser weapon with a high-power FEL to protect ships. The most powerful THz source is a FEL in the Budker Institute of Nuclear Physics, Russia, which has the average

Fig. 7 Calculated FEL wavelength depending on electron energy and undulator magnetic field strength



and peak power of 400 W and 0.6 MW, respectively. If we consider the other THz sources having average powers much less than 1 W, the FEL has superior capability in power.

2 Fundamental Principle of FEL

2.1 Dynamics of Relativistic Charged Particles in Undulator

To understand the dynamics of electrons in undulator field, we need to recall the famous Lorentz force and the equation of motion for relativistic particles.

$$\vec{F} = e \left(\vec{E} + \frac{1}{c} \vec{v} \times \vec{B} \right) = \frac{d\vec{P}}{dt} = \frac{d(\gamma m_e \vec{v})}{dt} \quad (2)$$

The simplest expression of the undulator magnetic field is as follows.

$$\vec{B} = B_w \sin(k_w z) \vec{y}, \quad \text{for an ideal planar undulator,} \quad (3)$$

and

$$\vec{B} = B_w [\sin(k_w z) \vec{x} + \cos(k_w z) \vec{y}], \quad \text{for an ideal helical undulator,} \quad (4)$$

where \vec{x} , \vec{y} and \vec{z} are the unit vectors of Cartesian coordinate, as shown in Fig. 8, and $k_w \equiv \frac{2\pi}{\lambda_w}$ is the wavenumber of the undulator. The magnetic field of the planar undulator has only y-axis component and electrons are moving to the z-axis in both cases.

Let's calculate the electron motion in an ideal planar undulator, which is the simplest example. The equation of motion of the electron in the ideal planar undulator field expressed by Eq. (3) can be written as follows, when we neglect the radiation loss during the motion, which means that γ is constant.

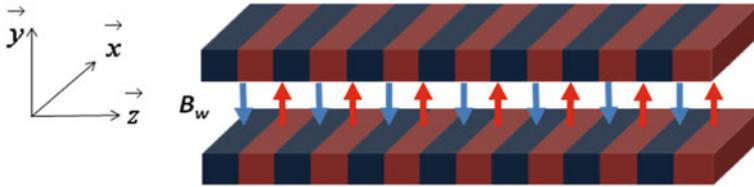


Fig. 8 Cartesian coordinates for describing magnetic field of ideal planar undulator

$$\frac{d\vec{P}}{dt} = \frac{d}{dt}(\gamma m_e \vec{v}) = \frac{e}{c}(\vec{v} \times \vec{B})$$

$$\gamma m_e \frac{d}{dt}(\vec{v}) = \frac{e}{c}(\vec{v} \times \vec{B})$$

In this case the, only x and z components of the acceleration can exist.

$$\dot{v}_x = -\frac{v_z \Omega}{\gamma} \sin(k_w z)$$

where $\Omega \equiv \frac{eB_w}{m_e c}$.

$$\dot{v}_z = \frac{v_x \Omega}{\gamma} \sin(k_w z)$$

Then we can calculate x and z components of the electron velocities as follows.

$$v_x = \frac{v_0 \Omega}{\gamma \omega_0} \cos(k_w z) = \frac{v_0 \Omega}{\gamma \omega_0} \cos(\omega_0 t) \quad (5)$$

$$v_z = v_0 - \frac{v_0 \Omega^2}{4\gamma^2 \omega_0^2} \cos(2\omega_0 t) \quad (6)$$

where we used the assumptions and the relations of $v_x \ll v_z$, $z \approx v_z t \equiv v_0 t$ and $\omega_0 \equiv k_w v_z$. The electron orbit in z-x plane can be expressed as follows.

$$x = \frac{v_0 \Omega}{\gamma \omega_0^2} \sin(\omega_0 t)$$

$$z = v_0 t - \frac{v_0 \Omega^2}{4\gamma^2 \omega_0^3} \cos(2\omega_0 t)$$

From the above relations, we can see the electron shows sinusoidal motion in the perpendicular plane to the undulator field with a phase difference of $\pi/2$. From the relation between the transverse component of the velocity and the initial electron velocity, we can understand that the unduator K-value, which is shown in Eq. (1), shows deflecting ability of the undulator.

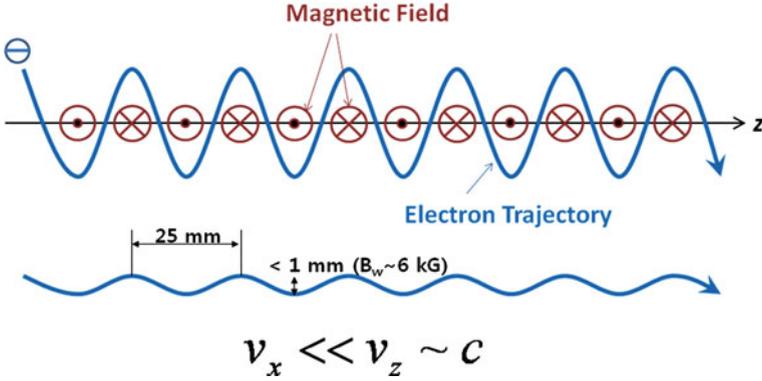


Fig. 9 Trajectory of an electron passing through an ideal planar undulator (*upper part*). The real deviation of the electron motion in the transverse direction is estimated to be less than 1 mm for a case of $B_w = 6$ kG, $\lambda_w = 25$ mm, and $E = 7$ MeV (*lower part*)

$$\frac{v_x}{v_0} = \frac{\Omega}{\gamma\omega_0} = \left(\frac{eB_w}{m_e c^2 k_w} \right) \frac{1}{\gamma} \equiv \frac{K}{\gamma}$$

where $K \equiv \frac{eB_w}{m_e c^2 k_w} \cong 0.934\lambda_w[T]B_w[\text{cm}]$. The trajectory of an electron passing through the ideal planar undulator shows sinusoidal motion along the z - x plane, as illustrated in Fig. 9. The real transverse amplitude of the electron trajectory is much smaller than the period of the undulator field, when the electron energy is close to the light velocity. Also the x -component of the electron velocity, v_x , is much smaller than the z -component of the electron velocity, v_z .

2.2 Coherent and Stimulated Radiation in FEL

When charged particles are accelerated or decelerated, there should be the emitting of radiation. Especially, for the case of the radial acceleration of the charged particles, $\vec{a} \perp \vec{v}$, it is called 'synchrotron radiation'. We can generate the synchrotron radiations by using bending magnets, undulators, and/or wigglers. The synchrotron radiation is generated to the forward tangential direction of the orbit of the relativistic particle, as shown in Fig. 10. The divergence angle of the radiation is inversely proportional to the Lorentz factor of the particle. Therefore we can generate more directional radiation with a higher energy of charged particles. Usually synchrotron radiations generated by bending magnets show wide spectrum over X-rays to millimeters. A typical relation between photon energy and brightness of the synchrotron radiation generated by a bending magnet is shown in Fig. 11.

When we use undulators instead of bending magnets for generating radiation, the spectral width of the radiation decreases as the number of undulator periods increases as shown in Fig. 12. The spectral power of the undulator radiation increases as the number of undulator periods increases. These are easily

Fig. 10 Synchrotron radiation generated from a relativistic electron moving in a bending magnet

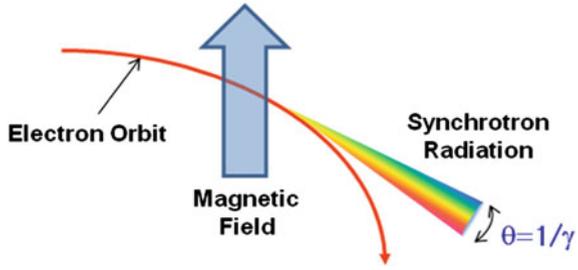
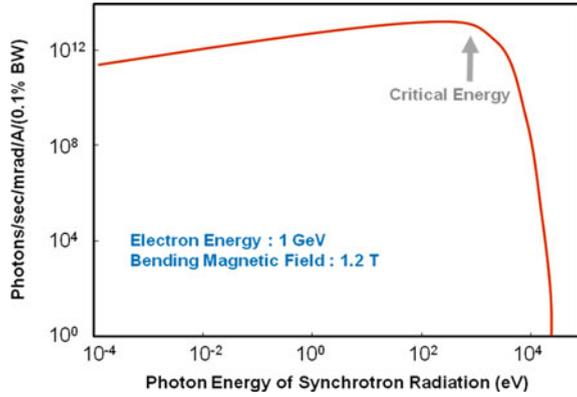


Fig. 11 Typical synchrotron radiation spectrum for the case of 1-GeV electron energy and 1.2-T bending magnet



understood by comparison with the case of multi-slit interference. Each periodic orbits of the electron motion are the sources of the radiation, and those multi sources have analogy to the multi-slits. Let's derive the wavelength of the undulator radiation. The energy exchange between the electron and the electromagnetic wave can be caused by the transverse component of the electron velocity. The rate of the electron energy change is

$$\frac{d\varepsilon}{dt} = m_e c^2 \frac{d\gamma}{dt} = -e\vec{v} \cdot \vec{E},$$

where \vec{E} is the electric field of the wave. And if we recall the motion of the undulating electrons in Eq. (5), the electromagnetic wave has electric field of x component.

$$\vec{E} = E \cos[\omega(z/c - t)]\vec{x}$$

Note that $dz = v_z dt$. Then we can rewrite the above equation as follows.

$$\begin{aligned} \frac{d\varepsilon}{dz} &= -\frac{e}{v_z} \vec{v} \cdot \vec{E} \\ &\cong -\frac{eK}{2\gamma} E \cos[k_w z + \omega(z/c - t)] \end{aligned} \tag{7}$$

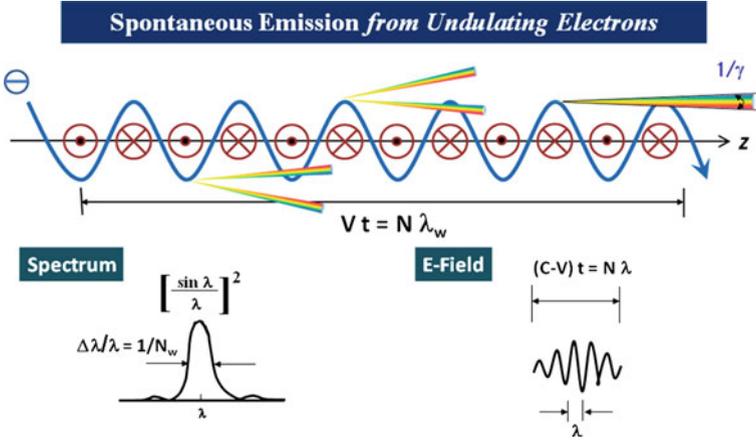


Fig. 12 Spontaneous emission from undulating electrons

The condition for maximum energy transfer from the electron to the electromagnetic wave through the whole undulator is simply obtained at the phase term of Eq. (7) is equal to constant. That is

$$d[k_w z + \omega(z/c - t)] = 0.$$

This is so-called synchronism condition in the FEL. We can write the relation as follows with the relation of $dz = v_z dt$.

$$k_w + \frac{\omega}{c} - \frac{\omega}{v_z} = 0$$

If we rewrite the above relation, we can simply understand the synchronism condition takes place when the wave advances the electron by one wavelength at one undulator period as shown in Fig. 13.

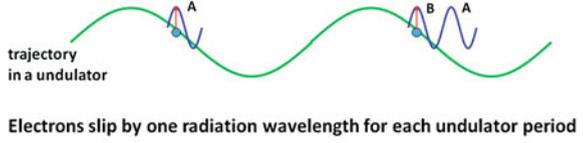
$$\frac{c - v_z}{v_z} = \frac{\lambda}{\lambda_w}, \quad (8)$$

where $\lambda = 2\pi c/\omega$, which is the wavelength of the undulator radiation in Eq. (1). From the Eq. (8), we can get the Eq. (1).

The above relation is simply obtained by the synchronism condition of the undulator radiation by one electron to satisfy the maximum energy transfer. When N electrons generate such waves through the undulator, the total power of the radiation, P , can be expressed by the following summation of each wave like Fig. 14.

$$P = |E_1 e^{-i\phi_1} + E_2 e^{-i\phi_2} + \dots + E_N e^{-i\phi_N}|^2 = P_0 \left| \sum_{j=1}^N e^{-i\phi_j} \right|^2,$$

Fig. 13 Illustrating synchronism condition of the undulator radiation and its wavelength



$$\lambda(\theta) = \frac{\lambda_w}{2\gamma^2} \left(1 + \frac{K^2}{2} + \gamma^2 \theta^2 \right)$$

Fig. 14 Total radiation power by N electrons

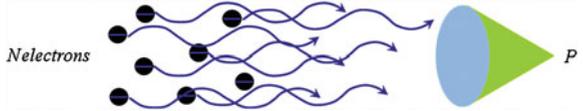
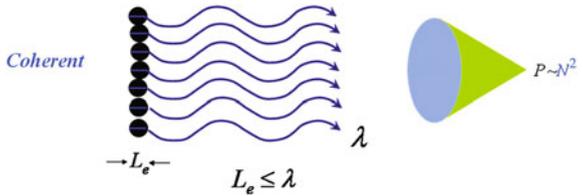


Fig. 15 Coherent radiation by a short-bunch electron beam, when the bunch length is smaller than the radiation wavelength



where E_i and ϕ_i is the electric field amplitude and phase of the wave by i th electron. And radiation power by single electron, $P_0 = |E_1|^2 = |E_2|^2 = |E_i|^2$. Then we can get the following relation.

$$P = NP_0 + P_0 \sum_{k,q} \cos(\phi_k - \phi_q)$$

$$= NP_0 + N(N + 1)P_0F,$$

where F is called form factor of the electron beam, $F \cong [\int \cos(kz)S(z)dz]^2$, and $S(z)$ is the electron-beam's spatial distribution.

Usually the number of electrons, N , in the electron beam is much bigger than 1. So, we can simply write as follows.

$$P \approx NP_0 + N^2P_0F \tag{9}$$

The first term of the above relation, NP_0 , denotes the incoherent contribution of the waves. Usual case as described in Fig. 14, the form factor, F , is zero due to the random-phase relations between the all electrons. In the special case of short-bunched electron beams having the bunch length, L_e , is less than the radiation wavelength, λ , as shown in Fig. 15, there can be the contribution of the second-term of the Eq. (9). The contribution of the coherent term to the total power is proportion to N^2 , and the power enhancement is significant compared with incoherent power. Usually the number of the electrons in a bunch is $10^6 - 10^9$ depending

Fig. 16 FEL lasing process by bunching of the randomly distributed electron beam

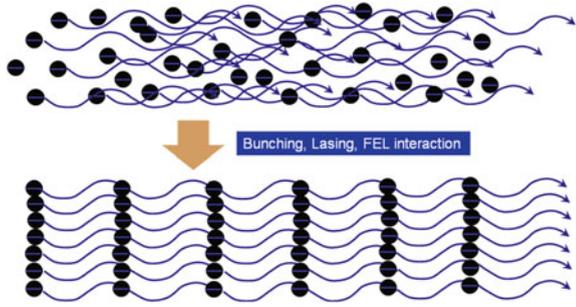
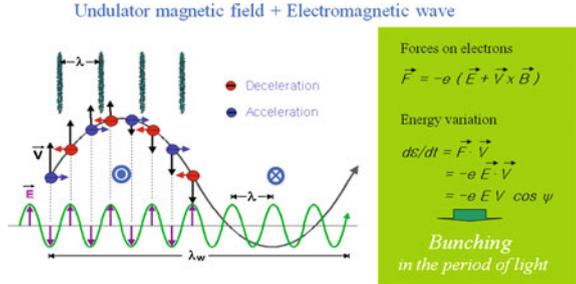


Fig. 17 Illustrating FEL bunching process with the interaction between wave's electric field and undulating electrons



on its bunch charges. In the THz range, the electron beam having a bunch length less than ~ 0.3 mm can generate much more intense radiation than that by a long-bunched electron beam.

The principle of the FEL lasing is based on the coherent process by the interaction between electrons and waves in undulator. In the FEL, the electron beam is usually longer than the generating wavelength. In a proper condition of electron beam and undulator, the electrons can be re-distributed as periodic micro-bunches, as shown in Fig. 16. The periodically distributed micro-bunched electron beam generates coherent, directional, and powerful laser output, which is the FEL lasing.

The bunching process of the FEL is illustrated in Fig. 17. Due to the interaction between electric field of the wave and undulating electrons, there should be energy modulation of the electron beam with the period of the radiation wavelength, which results in the density modulation of the electron beam. The density modulated electron beam means micro-bunched structure of the electron beam to generate the FEL output.

The single-pass incremental gain is the power ratio between the output and input waves by single-passing electron beam through the FEL. If we consider the case of the early-stage of the lasing process with small-amplitude signals, the gain, G , can be expressed as following relation.

$$G = \frac{32\sqrt{2}\pi^2\lambda^{3/2}\lambda_w^{1/2}}{S_{\text{mode}}} \frac{K^2}{(1 + K^2/2)^{3/2}} \frac{I_p}{I_A} N_w^3 f(s), \quad (10)$$

where S_{mode} is the cross-sectional size of the radiation mode in the FEL, I_p is the peak current of the electron beam, I_A is the Alfvén current having a value of 17000 A, and $f(s)$ is the lineshape function of the radiation

$$f(s) = \frac{d}{ds'} \left(\frac{\sin s'}{s'} \right)^2 \text{ and } s' = \frac{1}{2}s.$$

From the gain equation in Eq. (10), we can see that the gain of the FEL depends on undulator parameters, electron beam parameters, radiation mode size, and line broadening of the radiation. In general, the requirements for accelerators and undulators are less severe for long-wavelength FELs than short-wavelength FELs. The FELs in hard X-ray range should have extreme conditions of electron beam and undulators. And they need more than hundred meters in length. In the case of THz FELs, the size and the cost are much smaller and cheaper than those of short-wavelength FELs. Korea Atomic Energy Research Institute (KAERI) is under developing a table-top THz FEL for the practical applications of security inspection, which will be discussed on Sect. 4 of this chapter.

3 Status of THz FELs

The FEL is considered to be an ideal radiation source due to its high power, monochromatic spectral characteristics, and tunability. Among various radiation sources, only the FEL is considered to reach 1 MV/cm electric field in the THz range with a compact size, which is needed for contemplated nonlinear and non-equilibrium studies of the radiation-matter interaction such as electronic transition in solid, collective modes in large molecules, cell damage, and optical response of meta-materials.

Irrespective of the superior properties of the FEL over other radiation sources in the THz range, constructions have begun in the late 1980s due to less-interest in the THz range over difficulties in technology and budget. It is very recently that many groups begin to construct the THz-FEL. There are several FELs operating in the THz range. Among them, some noticeable facilities are summarized in Table 1.

The first FEL in THz range has been constructed at UCBS (Center for Terahertz Science and Technology) as a user facility [9], which provides tunable THz radiation in the range of 120 GHz to 4.8 THz with peak output power of 500 W to 5 kW and operates in a quasi-CW mode. The FEL is based on Van de Graaff accelerator with a hot-cathode electron gun and the e-beam energy is changed by varying the voltage on the terminal of the accelerator. Unlike other FELs, the output consists of a pulse with duration of a few microseconds without micropulse. It has been utilized in semiconductor fabrication and processing, which led to great synergy among radiation researchers and material scientists (Fig. 18).

The FELIX [10] at Rijnhuizen in Netherlands covers the spectral range from 1.2 to 100 THz. The peak power exceeds 100 MW at micropulse repetition rate of 25, 50, or 1000 MHz. For experiments susceptible to sample heating, a single

Table 1 Specifications of THz FELs in operation [8]

Specifications	UCSB	FELIX	Stanford	ENEA	NovoFEL	KAERI
λ (μm)/ f (THz)	63–2500/ 0.12–4.8	3–250/ 1.2–100	15–80/ 3.75–20	2000–3500/ 0.09–0.15	40–240/ 8–1.1	100–300/ 1–3
Width	–	2–10 ps	6–100 cycles	50 ps	10–60 ps	25–40 ps
Rep. rate	–	11.8 MHz	25, 50, 1000 MHz	3 GHz	22 MHz	2.8 GHz
Energy	–	1 μJ	1–50 μJ	0.5 μJ	20 μJ	0.1 μJ
Width	1–20 μs	0.5–5 μs	5 μs	4 μs	CW	4 μs
Rep. Rate	0–7.5 Hz	0–20 Hz	10 Hz	1–10 Hz	–	20 Hz
Average power	5–100 mW	<1 W	<1 W	4–40 mW	<500 W	<20 mW
Peak power	<15 kW	<100 MW	<500 kW	10 kW	<1 MW	<3 kW

Fig. 18 FEL at UCSB**Fig. 19** European FEL center at FELIX

micropulse can be sliced out of the pulse train using a fast optical switch. The maximum micropulse duration is about 100 cycles, which results in a minimum bandwidth of 0.4% in FWHM. The primary application of the facility is in the studies of the vibrational dynamics of molecules, liquids, and solids, but applications requiring an intense source such as photo-ablation of tissue, near-field microscopy or linear spectroscopy of small-sized samples are equally applicable (Fig. 19).

The Stanford picosecond FEL [11] Center covers the spectral range from 3.75–20 THz. The output consists of bursts of micropulses. The duration of macropulse is 5 μ s with micropulses in duration from 700 fs to 10 ps. The repetition rate of the macropulse is 20 Hz. The FEL is driven by a superconducting radio-frequency accelerator. FELs with superconducting linacs, with lower RF losses, are capable of higher macropulse duration and average power (Fig. 20).

Recently, Russia has developed a high power FEL in the spectral range of 1–3 THz, named NovoFEL [12]. Its average power can reach up to 400 W, which makes it possible to investigate highly nonlinear interaction of THz radiation with

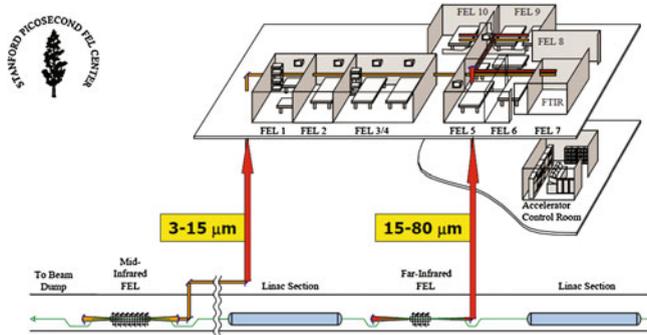


Fig. 20 Layout of the Stanford picosecond FEL center

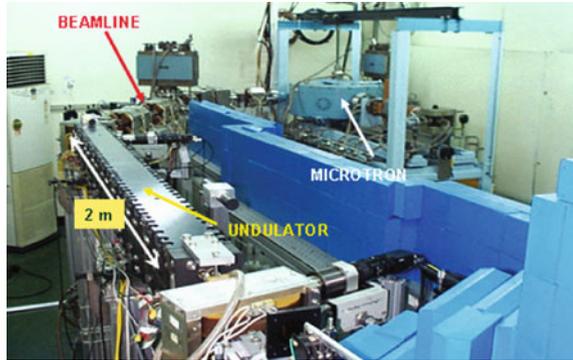
Fig. 21 NovoFEL at the Budker Institute of Nuclear Physics



matters. The system is based on Energy Recovery Linac (ERL) with normal conducting RF conductor. It consists of two phases, one is for high power radiation at THz range, the other for Infrared radiation by boosting up the electron beam energy. The peak power reaching up to 1 MW enables ablation with burning, which makes it possible to perform nondestructive transfer of complex molecular systems into aerosol phase, such as crystal minerals (marble), fullerenes, DNA. This provides new opportunities investigating many fundamental phenomena in plasma physics, aerodynamics, chemistry, material processing and modification, and biology (Fig. 21).

KAERI has developed compact FEL operating in the spectral range of 1–3 THz [6], which is the first compact FEL in THz range. The system is fitted in $3 \times 5 \text{ m}^2$ area. For an electron accelerator, a microtron with a thermionic electron gun has been used, which enables FEL to be compact. The macropulse with duration of $4 \mu\text{s}$ consists of micropulse trains which have 25–40 ps in duration at 2.8 GHz repetition rate. The peak power reaches up to 3 kW. The high average power makes it possible to take an image on a large, wet biological specimen. It is also great interest to investigate the effect of the high power THz radiation on biological cell and nonlinear response of meta-materials (Fig. 22).

Fig. 22 FEL at KAERI



4 Prospect of Table-Top THz FELs

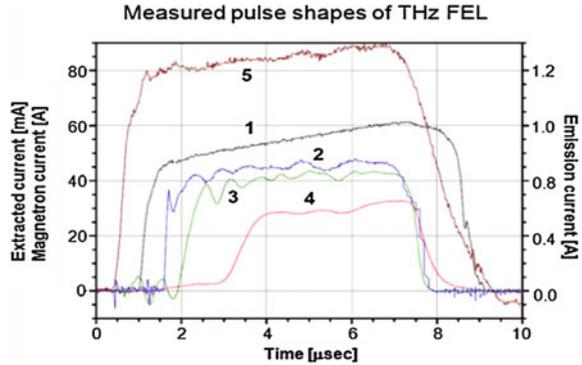
A Free electron laser is coherent light with high-peak power and/or high-average power and is easy to tune its wavelength in broadband range. There is no such a tool which can span wide spectral ranges with high power. Main obstacle to utilize it to real life or to spread out it to many sites for applications is the size and cost. Several groups [13, 14] put many efforts on reducing the size to a table top and on looking up ideas to specify for particular applications.

For a table-top FEL, all components, such as, accelerator, undulator, resonator, RF system, cooling and vacuum system, and so on, should be fitted on an optical table of $2 \times 4 \text{ m}^2$ or less while keeping the desired output. In THz range, an electron beam of several MeV in energy is suitable for an undulator of a few cm period. The longer the undulator period, the higher the electron beam energy, vice versa. In general, higher energy electrons correspond to longer in linacs or bigger in circular accelerator. At lower energy, the beam quality is mostly poor compared to that at higher energy, which is closely related to the gain of FEL. High peak current with low emittance can be generated by using the photocathode gun, but the average power may be limited by their reputation rate. The average power may be increased by producing high average current from the gun using other scheme. The specification of the system should be optimized as the requirements for the end-users.

4.1 Acceleration System

A microtron is small in size and selectable in energy. In addition, it provides good quality of electron beams with low emittance and high peak current, but, low average current. The electron emission from the cathode and the capture efficiency of electrons for acceleration is closely related with the synchronism condition in a microtron as well as the temperature of the cathode. The emission current is increased due to the temperature growth of cathode and the secondary emission

Fig. 23 Measured pulse shapes of THz FEL and Electron beam currents, Emission current, Magnetron pulse current. 1 Emission current [A], 2 Accelerated current [mA] at the 12th orbit, 3 Beam current [mA] at the entrance of the undulator, 4 FEL lasing power signal (macropulse) (in rel. units), 5 Magnetron pulse current [A]



due to the back bombardment of electrons. Non-captured electrons hit the RF cavity wall followed by heating up the cathode. To accelerate all captured electrons to the desired energy, the applied RF power should be increased as the current is increased.

As a RF source, a magnetron is the cheapest and the smallest RF source. But its frequency jittering and drift degrades the FEL performance. In KAERI THz FEL driven by magnetron [15, 16], the frequency jittering was suppressed by developing the RF frequency feedback system combined with the temperature control of cathode. Another issue is high voltage/current modulator, which has to provide $\sim 10\%$ increase of the RF power during each macropulse, as shown in Fig. 23. The vacuum device is somewhat easier to adjust the pulse shape, but it is bulky! Recently the pulse shaping was demonstrated using a solid state modulator of the size of $0.64 \times 0.53 \times 1.09 \text{ m}^3$ in ScandiNova (Sweden, www.sc-nova.com, private communication). The microtron is a dispersive system and the dispersion becomes large when electron beams are extracted from 180° of circular orbit from the RF cavity. Therefore, a beam transfer line should be necessary to inject the dispersion-free beam into the undulator. Its length can be shorter with better electron beam with lower emittance and narrower energy spread. The operating microtron-based THz FEL at KAERI has the size of $2.7 \times 5 \text{ m}^2$ and many efforts to reduce the size further are underway. New design may fit on a table of $1.6 \times 2.6 \text{ m}^2$ as shown in Fig. 24 and the major components for the size reduction are the beam transfer line and the undulator, compared to the existing one [17]. The designed beam optics consists of two 45° bending magnets and six quadrupoles [18].

A RF injector itself can now produce good quality of electron beams. The accelerating structure may be combined with a RF gun or a DC gun. Main concern is how to keep the emittance low while the current remains high. The photocathode gun can generate electron beams with low emittance and high peak current, in expense of low average current and low duty factor. It can generate electron beams of 200 pC with a normalized emittance below 0.4 mm mrad [19].

Spring-8 and KEK in Japan has developed a compact C-band RF linac with high quality diode-type DC gun using crystal cathodes. C-band accelerator

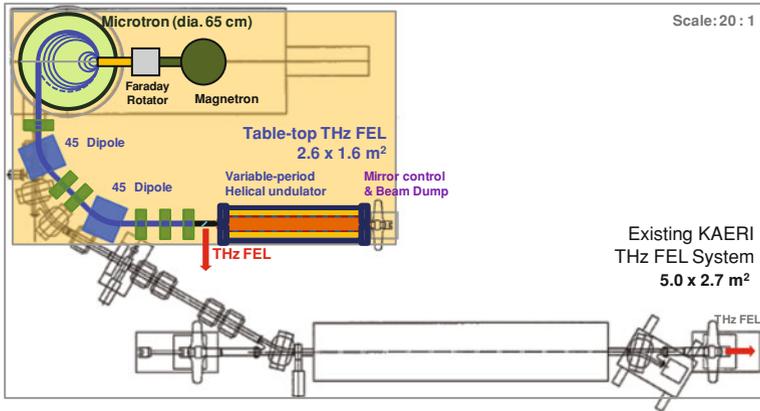
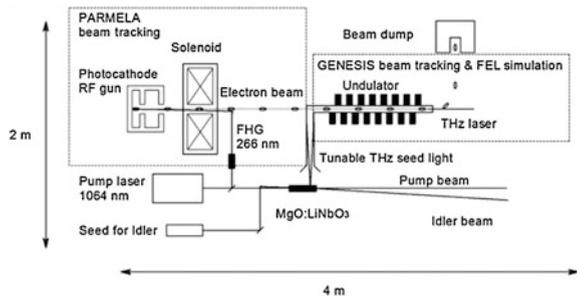


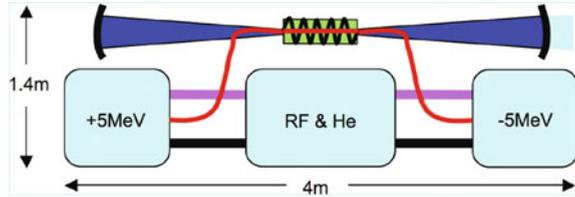
Fig. 24 Comparison between the existing KAERI THz FEL and Newly designed Table-top THz FEL

Fig. 25 Schematic of the THz FEL amplifier at Kyoto University [21]



technology immediately adapted to medical applications and the length of a 6 MeV C-band RF linac for medical therapy [20] is about 30 cm. However, we should consider the size of the RF klystron and modulator, which is 2.5 m length, 2 m in height, and ~60 cm in width. To use as an accelerator for THz FELs, some modification is needed to improve the emittance and the energy spread, including lens system. There are several groups to operate THz FELs driven by RF linacs, but most of them are about or greater than 10 m in length. KU-FEL team in Kyoto University has developed a THz FEL of $2 \times 4 \text{ m}^2$ using a RF injector with thermionic RF gun in S-band [21]. They proposed THz FEL amplifier seeded by tunable THz light. After successful developments of superconducting RF linacs and energy-recovery linacs, many efforts to reduce the size of accelerators as well as the shielding size are made world-widely. Colson et al. [22] suggested $1.4 \times 4 \text{ m}^2$ sized energy-recovery THz FEL using 5 MeV Superconducting cavity injector. However, it becomes bigger when considering the size of liquid helium refrigerator even though it doesn't need the shielding with such a high power THz light (Figs. 25, 26).

Fig. 26 Schematic of a energy-recovery THz FEL [22]



4.2 Insertion Device

According to the resonant condition for FEL lasing, THz light can be generated when an electron beam of several MeV in energy is traveling through the undulator of a few-cm period. Since the FEL gain increases as the number of periods in undulator, the shorter period can reduce the size. Permanent magnet can realize cm-scale period, but it requires complicate gap adjustment system. Instead, KAERI had developed hybrid planar undulator of 2 m long, which can adjust the magnetic strength by the current by the coils winding around each modules consisting of several periods with closed loop [23]. Recently, they have developed a variable-period undulator proposed by Vinokurov et al. [24]. By using the property of magnet, i.e., repulsive forces at the mid-plane of the iron poles equally bisected, the undulator period can be varied by changing the position of one end horizontally. The designed helical undulator can adjust its period from 23 to 26 mm and total length is less than 1 m with the number of periods of 30 [18]. In case of FEL amplifier, the minimum length of undulator is determined by seed laser power.

4.3 Optical Resonator

The size of optical resonator is limited by the diffraction limit of THz light if it propagates and evolves in free space. In case of permanent-magnet or hybrid-type undulators, the gap is too small for THz light to propagate through a few meter-long. It is natural to use the waveguide to increase the transmission. The size of waveguide should be properly designed to reduce the waveguide loss depending on the type of undulators: a parallel plate for planar undulators, a cylindrical for helical undulators. Typically, the waveguide loss is increased as the frequency increases, for a given size of waveguide. Reduction of waveguide loss is critical for the compactness of overall system while keeping the desired output power. The dielectric coated metallic waveguide may be reduced significantly depending on the ratio of the radius to wavelength. However, it is hard to use the waveguide for high average power FEL due to heat problem. It may need to design special cooling system.

The second issue is the resonator mirrors. Most of FEL oscillators put their resonator mirrors at the tangential line of the bending magnets located in both

sides. Actually, the trajectory of electron beam is detoured to give the space for mirrors, which means more length of the beamline. The electron beam has to pass through the side hole in the waveguide extended to the resonator mirrors. Newly designed THz FEL at KAERI will use a mesh mirror as an outcoupling mirror and a full mirror in downstream combined with the beam dump. The interval and thickness of mesh determines the efficiency of outcoupling depending on the wavelength. For 8% outcoupling efficiency, 10 μm meshes with interval of 100 μm is enough. For most of metallic meshes, the estimated saturation temperature due to electron beams passing through is far less than half of melting point. A 45° thin foil mirror with hole for electron beam will be located in upstream of the mesh mirror to extract THz FEL.

The FEL amplifier, proposed at Kyoto University, does not need the resonator mirror, but tunable seed laser system. For high average power FEL using energy-recovery SC RF linacs, the optical cavity length may be much longer than undulator length with large resonator mirrors.

5 Summary and Conclusion

Since a 1 THz photon has the energy of 4 meV corresponding to temperature of 48 K, (cf. For an optical laser at $\lambda = 800$ nm, 18,000 K), a THz pulse is non-ionizing and can be thought of as transient magnetic and electric fields, which can be used for a non-destructive pump source. A table-top THz FEL will be a tool to discover new phenomena in protein or DNA, and magnetic materials.

References

1. Maiman, T.H.: Stimulated optical radiation in ruby. *Nature* **187**(4736), 493–494 (1960)
2. Xie, A., van der Meer, A.F.G., Austin, R.H.: Excited-state lifetimes of far-infrared collective modes in proteins. *Phys. Rev. Lett.* **88**, 018102 (2002)
3. Austin, R.H., Roberson, M.W., Mansky, P.: Far-infrared perturbation of reaction rates in myoglobin at low temperatures. *Phys. Rev. Lett.* **62**, 1912–1915 (1989)
4. Elias, L.R., Fairbank, W.M., Madey, J.M., Schwettman, H.A., Smith, T.I.: Observation of stimulated emission of radiation by relativistic electrons in a spatially periodic transverse magnetic field. *Phys. Rev. Lett.* **36**(13), 717 (1976)
5. Deacon, D.A.G., Elias, L.R., Madey, J.M.J., Ramian, G.J., Schwettman, H.A., Smith, T.I.: First operation of a free-electron laser. *Phys. Rev. Lett.* **38**(16), 892–894 (1977)
6. Jeong, Y.U., Lee, B.C., Kim, S.K., Cho, S.O., Cha, B.H., Lee, J., Kazakevitch, G.M., Vobly, P.D., Gavrilov, N.G., Kubarev, V.V., Kulipanov, G.N.: First lasing of the KAERI compact far-infrared free-electron laser driven by a magnetron-based microtron. *Nucl. Instrum. Method A* **475**, 47–50 (2001)
7. Emma, P., Akre, R., Arthur, J., Bionta, R., Bostedt, C., Bozek, J., Brachmann, A., Bucksbaum, P., Coffee, R., Decker, F.-J., Ding, Y., Dowell, D., Edstrom, S., Fisher, A., Frisch, J., Gilevich, S., Hastings, J., Hays, G., Hering, Ph, Huang, Z., Iverson, R., Loos, H., Messerschmidt, M., Miahnahri, A., Moeller, S., Nuhn, H.-D., Pile, G., Ratner, D., Rzepiela,

- J., Schultz, D., Smith, T., Stefan, P., Tompkins, H., Turner, J., Welch, J., White, W., Wu, J., Yocky, G., Galayda, J.: First lasing and operation of an ångstrom-wavelength free-electron laser. *Nat. Photonics* **4**, 641–659 (2010)
8. Sherwin, M.S., Schmuttenmaer, C.A., Bucksbaum, P.H. (eds.): Opportunities in THz sciences, report of a DOE-NSF-NIH workshop, Arlington, VA, 12–14 Feb 2004
 9. Ramian, G.: The new UCSB free-electron lasers. *Nucl. Instrum. Method A* **318**, 225 (1992)
 10. van der Meer, A.F.G.: FELs, nice toys or efficient tools? *Nucl. Instrum. Method A* **528**, 8 (2004)
 11. Berryman, K.W., Crosson, E.R., Ricci, K.N., Smith, T.I.: Coherent spontaneous radiation from highly bunched electron beams. *Nucl. Instrum. Method A* **375**, 526 (1996)
 12. Bolotin, V.P., Vinokurov N.A., Kayran, D.A., Knyazev, B.A., Kolobanov, E.I., Kotenkov, V.V., Kubarev, V.V., Kulipanov, G.N., Matveenko, A.N., Medvedev, L.E., Miginsky, S.V., Mironenko, L.A., Oreshkov, A.D., Ovchar, V.K., Popik, V.M., Salikova, T.V., Serebnyakov, S.S., Skrinisky, A.N., Shevchenko, O.A., Scheglov, M.A.: Status of the Novosibirsk THz FEL. In: Proceedings of the 2004 FEL Conference, Trieste, Italy, 29 Aug–3 Sep 2004
 13. Doria, A., Asgekar, V.B., Esposito, D., Gallerano, G.P., Giovenale, E., Messina, G., Ronsivalle, C.: Long wavelength compact-FEL with controlled energy–phase correlation. *Nucl. Instrum. Method A* **475**, 296 (2001)
 14. Gallerano, G.P., Biedron, S.: Overview of Terahertz Radiation Sources, In: Proceedings of the 2004 FEL Conference, 216–221 (2004)
 15. Kazakevitch, G.M., Jeong, Y.U., Pavlov, V.M., Lee, B.C.: Stabilization of the microtron-injector for a wide-band compact FIR FEL. *Nucl. Instrum. Method A* **528**, 115–119 (2004)
 16. Jeong, Y.U., Kazakevitch G.M., Lee, B.C., Kim, S.K., Cho, S.O., Gavrilov, N.G., Lee, J.: Status and prospects of a compact FIR FEL driven by a magnetron-based microtron. *Nucl. Instrum. Method A* **483**, 195–199 (2002)
 17. Jeong, Y.U., Kazakevich, G.M., Park, S.H., Lee, K., Mun, J., Jang, K.H., Sunwoo, J., Kim, K.N., Kim, H.N., Jang, J.S., Lee, B.C., Cha, Y.-H., Cha, H., Kim, D.H., Cha, B.H.: Design status of a KAERI table-top THz free-electron laser. Presented in 33rd International FEL conference, Shanghai, China, 22–26 Aug 2011
 18. Park, S.H., Jang, K.H., Lee, K., Mun, J., Kazakevich, G., Miginsky, S., Jeong, Y.U.: Microtron-based beam transfer system for a compact, high power THz FEL. In: 33rd international FEL conference, Shanghai, China, 22–26 Aug 2011
 19. Ding, Y., Brachmann, A., Decker, F.-J., Dowell, D., Emma, P., Frisch, J., Gilevich, S., Hays, G., Hering, Ph., Huang, Z., Iverson, R., Loos, H., Miahnahri, A., Nuhn, H.-D., Ratner, D., Turner, J., Welch, J., White, W., Wu, J.: Measurements and simulations of ultralow emittance and ultrashort electron beams in the linac coherent light source. *Phys. Rev. Lett.* **102**, 254801 (2009)
 20. Kamino, Y., et al.: development of a new concept automatic frequency controller for an ultra-small C-band linear accelerator guide. *Med. Phys.* **34**, 243–249 (2007)
 21. Ohgaki, H., Kii, T., Masuda, K., Bakr, M.A., Higashimura, K., Kinjo, R., Yoshida, K., Ueda, S., Sonobe, T.: Research activities on a MIR-FEL and table-top THz generation in Kyoto University. *J. Korean Phys. Soc.* **57**(2), 344–348 (2010)
 22. Colson, W.B., Grimm, T.L., Smith, T.I.: Compact THz free electron lasers for laboratory laser-matter interactions. In: Research trends in laser matter interaction, La Jolla, California, 7–8 May 2010
 23. Jeong, Y.U., Lee, B.C., Kim, S.K., Cho, S.O., Cha, B.H., Lee, J., Vobly, P.D., Kolokolnikov, Y.M., Mihaylov, S.F., Kulipanov, N.K.: Short-period equipotential-bus electromagnetic undulator for a far infrared free-electron laser. *Nucl. Instrum. Method A* **407**, 396–400 (1998)
 24. Vinokurov, N.A., Shevchenko, O.A., Tcheskidov, V.G.: Variable-period permanent magnet undulators. *Phys. Rev. Special Top-Accel. Beams* **14**, 040701–040708 (2011)

Principle of Linear Accelerator Based fs-THz Generation and Its Application

Jaehun Park

The generation, detection, and application of intense THz radiation have drawn a great attention recently. The dramatically enhanced energy and peak electric field of the coherent THz radiation can be generated by coherent superposition of radiated fields emitted by ultrafast electron bunches or by optical rectification of intense near-IR laser pulses in a LiNbO₃ crystal with the pulse front tilted by a diffraction grating. THz field strengths on the order of ~ 1 MeV/cm has great potential for applications in nonlinear optical phenomena and in the study of high E-field effects on the materials in the condensed phase with ultrafast temporal resolution.

1 Introduction and Background

Terahertz radiation is a part of the electromagnetic spectrum located between the microwaves and the infrared (Fig. 1). The word “tera” indicates the 10^{12} hertz for the frequency. The Terahertz (THz) spectral range typically covers frequencies between 0.1 and 10 THz. THz electromagnetic (EM) radiation has potential applications in many fields including biology, medicine, physics, chemistry and material sciences. Researchers in different disciplines use different units to describe the THz ranges. As seen in Fig. 1, THz corresponds to 33 cm^{-1} , $300\ \mu\text{m}$, 4.1 meV , or 1 ps . However, actual applications have been limited by the lack of suitable THz light sources and its detection methods. Recent developments of THz technology based on table-top femtosecond lasers may have solved these problems [1–8]. Coherent THz radiation with dramatically enhanced energy, peak power,

J. Park (✉)

Pohang Accelerator Laboratory, POSTECH, Pohang 790784, Korea
e-mail: jaehunpa@postech.ac.kr

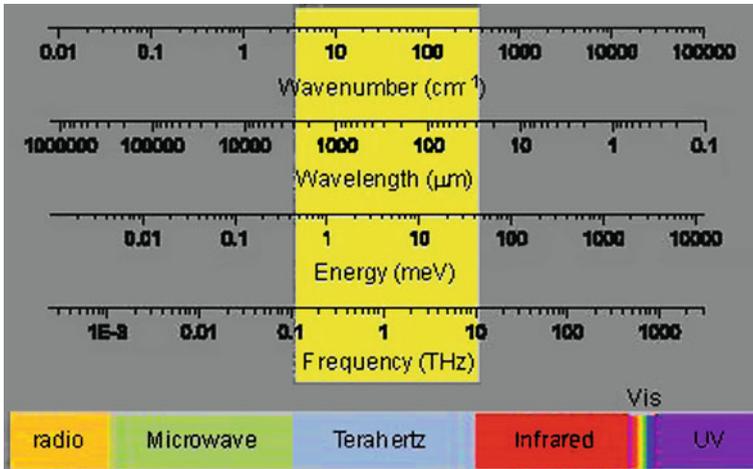


Fig. 1 The terahertz region of the electromagnetic spectrum

and peak electric field is generated by coherent superposition of radiated fields emitted by ultrafast electron bunches [9–11]. This intense THz light source offers the opportunity to investigate nonlinear optical phenomena in the THz EM region [1, 4, 11, 12].

This part presents THz generation, Transport, detection methods based on the LINAC, the design of the fs-THz beamline, and some results from electro-optic sampling [13, 14] for THz detection.

2 THz Generation

In the fs-THz beamline (Fig. 2) at Pohang Accelerator Laboratory (PAL), a laser-driven cathode RF-gun is used to generate electron bunches. An RF-gun cavity consists of a half-cell and a full-cell. The coupling between the waveguide and the full cell is accomplished using a coupling slot on the full cell cavity. The full cell also includes tuners and a pumping port. RF power from the full cell couples to the half cell through an iris by electric coupling. The coupling coefficient is 1.26 and the quality factor of the gun is 9,000. The gun is operated in π -mode at 2,856 MHz RF frequency; multipole suppression is achieved by means of cavity symmetrization. The electron beam from the photocathode RF-gun passes through a solenoid containing eight disc-shaped coils for transverse emittance compensation. Two S-band accelerating structures are used to accelerate the electrons up to 75 MeV, and two chicane-type bunch compressors reduce the temporal width of the electron bunch. Titanium foil or a copper mirror is used to achieve coherent transition radiation at THz frequencies.

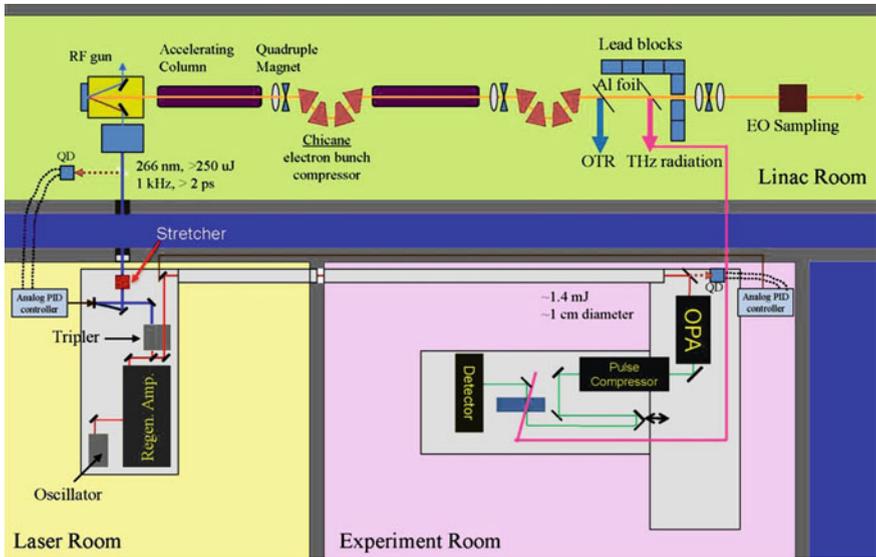


Fig. 2 The layout of the fs-THz beamline at PAL. *AC* accelerating column, *QM* quadruple magnet, *QD* quadrant detector, *C* chicane; *OTR* optical transition radiation, *CTR* coherent transition radiation, *AF* aluminum foil, *EOS* electro optic sampling, *OPA* optical parametric amplifier, *PC* pulse compressor, *APC* analog PID controller [15]

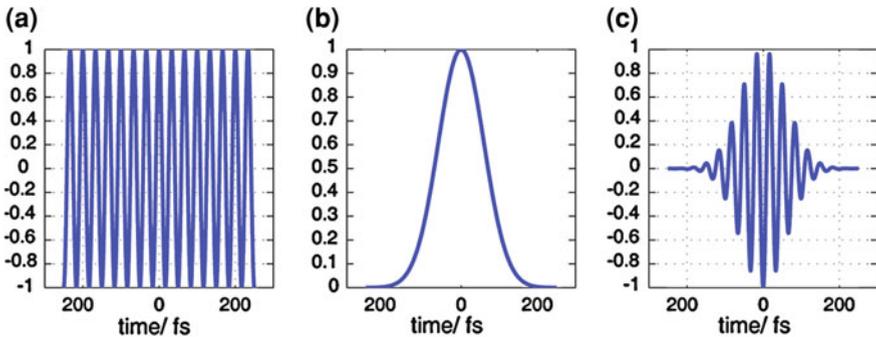
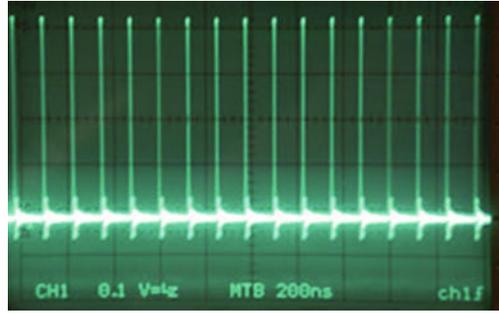


Fig. 3 **a** Time evolution of the electric field. **b** Gaussian envelope function. **c** Time evolution of the electric field in a Gaussian envelope function

2.1 Laser Basics

A continuous wave (CW), as seen in He:Ne laser, is an electromagnetic wave of constant amplitude and frequency of infinite duration (Fig. 3a). Electric field can be defined as

Fig. 4 Pulse train of the ultrashort pulses of the Ti:sapphire oscillator



$$E = \text{Re}(E_0 \exp(i\omega_0 t)).$$

The time evolution of the continuous electric field is an unlimited cosine function (Fig. 3a), however, when it is multiplied by an envelope function like a Gaussian envelope function ((Fig. 3b), the electric field exists only within the envelope (Fig. 3c), which is called light pulse. Then a Gaussian light pulse can be defined as

$$E = \text{Re} (E_0 \exp(i\omega_0 t) \exp(-\Gamma t^2))$$

and its time evolution is shown in (Fig. 3c). Γ is the shape function of the Gaussian envelope function. The full width half max (FWHM) of the envelope is called pulse duration or pulse width. Picosecond (ps) or femtosecond (fs) pulse means the FWHM of the pulse envelope is picosecond or femtosecond, not the time between the two pulses. Repetition rate is defined as the number of pulses within 1 s. The characteristic frequency of 80 MHz represents the repetition rate of the Ti:sapphire oscillator laser. Periodic ultrashort light pulses are called the pulse train (Fig. 4).

Figure 5 shows the comparison of the electric fields in different frequency ranges. Although the duration of all the pulse envelopes are same (500 fs), one can see the big difference of the electric fields. Figure 5a shows a 500 nm ($20,000 \text{ cm}^{-1}$) visible pulse and Fig. 5b shows a 5 μm ($2,000 \text{ cm}^{-1}$) mid-infrared (IR) pulse. Since the periodic time of the electric field is much shorter than the time window, it is hard to see the field clearly. In the far-IR range (50 μm , 200 cm^{-1}), due to the comparatively large period of the electric field (more than 100 femtosecond), one can see around ten cycles of the electric field (Fig. 5c). However, Fig. 5d shows only less than a single cycle of the electric field in the THz range due to the long wavelength of 500 μm (20 cm^{-1}), which corresponds to $\sim 1.7 \text{ ps}$.

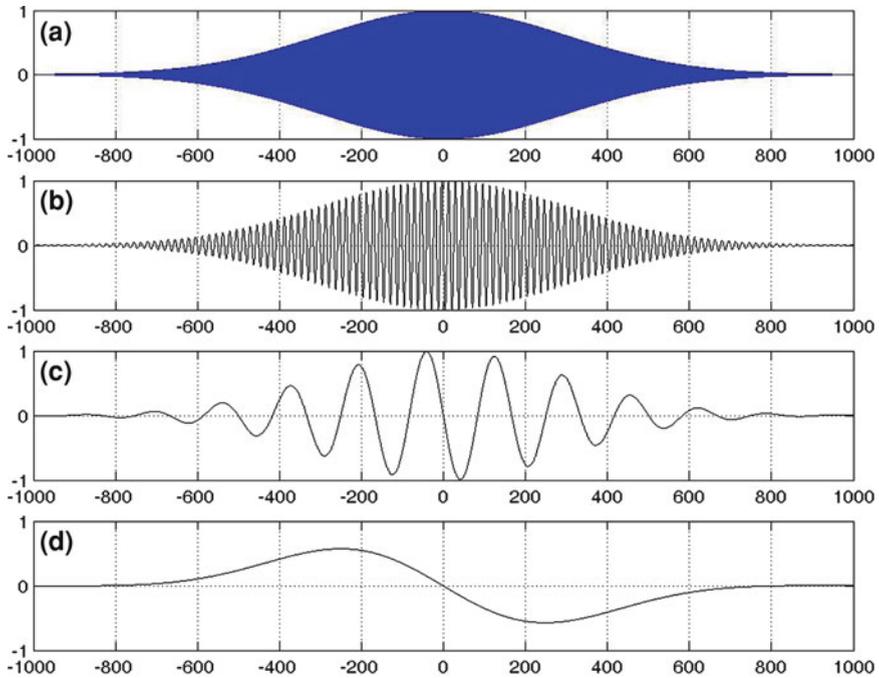


Fig. 5 The comparison of the electric fields in different frequency ranges. FWHM of each envelope is 500 fs. **a** Visible pulse, $20,000\text{ cm}^{-1}$, **b** mid-IR, $2,000\text{ cm}^{-1}$, **c** far-IR, 200 cm^{-1} , **d** THz, 20 cm^{-1}

2.2 Laser Systems

A Ti:Sapphire laser amplifier system is used for a photocathode RF-gun and the pump/probe experiments. The laser system consists of an oscillator, a regenerative amplifier, a third-harmonic generator, and a UV pulse stretcher. The Ti:Sapphire oscillator typically provides a 80 MHz train of pulses with its spectrum centered around 800 nm and a bandwidth of several tens of nm. $\sim\text{nJ}$ s of pulse energy out of the oscillator is amplified up to several mJ at the repetition rate of 1 kHz in the regenerative amplifier (Fig. 6). Before going to the amplifier, these pulses are first stretched to $\sim 200\text{ ps}$ in the pulse stretcher to increase the pulse duration $\sim 4,000$ times, which reduces the peak intensity, so that they may be amplified without damaging the amplifier optics. Then, the stretched pulses are sent to a Ti:sapphire regenerative amplifier, which is pumped by a Nd:YLF laser firing at 1 kHz. The amplified pulses are sent to a pulse compressor to restore the pulses to the pulse duration of $\sim 100\text{ fs}$ with the final energy of several mJ.

As shown in Fig. 2, the intensity of the 800 nm pulses is to be divided by a beam splitter; the more intense (60%) pulse is used to generate UV pulses at 266 nm by third harmonic generation and the less intense (40%) pulse is directed

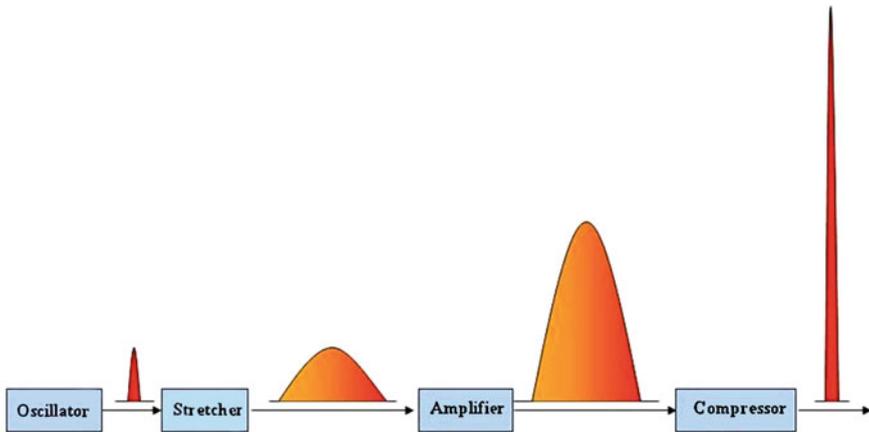
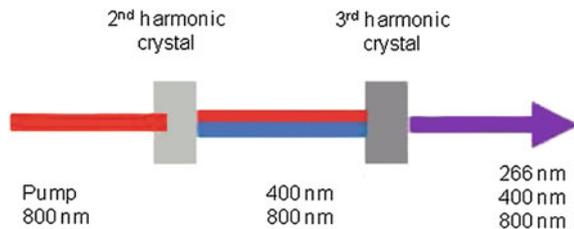


Fig. 6 Schematic of the Ti:sapphire chirped pulse amplifier system

Fig. 7 A typical configuration for frequency tripling



to an experiment room for pump-probe experiments and THz detection by electro-optic sampling. Third harmonic generation or frequency tripling is a process of nonlinear frequency conversion where the resulting optical frequency is three times that of the input laser beam. Frequency tripling is usually realized with frequency doubling ($\chi^{(2)}$) of the input beam in the nonlinear crystal such as a BBO crystal and subsequent sum frequency generation of the fundamental and frequency doubled waves in the second nonlinear crystal. (Fig. 7)

2.3 RF-gun and Accelerator

UV pulses are required to overcome the work function (the electron binding energy) of a metal (eg. Cu or Mg) cathode in a Radio-Frequency (RF) gun (Fig. 8) and to generate electron bunches. The work function is the minimum energy needed to liberate an electron from the surface of a matter. Photoelectron is emitted from the material if the photon energy greater than the work function is absorbed by the material. (photoelectric effect) The kinetic energy of the emitted



Fig. 8 Photocathode RF-gun and the accelerating column

photoelectrons depends only on the energy or frequency of the photons absorbed. The intensity of the photons should be increased to generate more number of electrons emitted.

If emitted electrons from the photocathode are accumulated within a short bunch length, each electron will experience the repulsive forces from the other electrons since they have the same electrical charge. (space charge effect) These repulsive forces may affect the longitudinal and transverse dynamics in the accelerator. Two pairs of fused silica prisms arranged in a pulse compressor geometry may be used to reduce the space charge effect in the electron bunches; these prisms provide negative group velocity dispersion to stretch the pulses up to several ps. Pulse duration of the UV pulses can be varied by changing the distance between the two prism pairs to find an optimum value for a low-emittance electron beam. Since the laser beams travel rather long distances to the RF-gun, an active beam pointing stabilization system that consists of a quadrant detector, a proportional-integral-derivative (PID) controller and a piezo-electric mirror may be employed to minimize the spatial movement at the photocathode.

To synchronize the laser with the RF systems, two frequency dividers are used [15]. First, a master RF at 2856-MHz is divided by 36 to obtain 79.33 MHz, where the Ti:sapphire oscillator is locked by using a cavity length stabilization system (Synchrolock-AP). A second frequency divider is used to get ~ 1 kHz, which triggers the pump laser of the regenerative amplifier. The photodiode signal of the oscillator is used as a reference clock to control the delay and timing of the high voltages at the Pockel's cell in the regenerative amplifier cavity.

~ 5 ps (FWHM) electron bunches are produced from the photocathode RF-gun with the charges up to 0.5 nC and 2.5 Hz repetition rate at the Source Development Lab of Brookhaven National Laboratory (BNL). [11] The electrons are accelerated to 120 MeV and the electron bunches are compressed to below 1 ps in

Table 1 Relationship between electron bunch charge and bunch length

Bunch charge (nC)	Bunch length after RF-Gun (ps)	Bunch length after chicane (fs)
0.2	0.5	75
0.5	2	150

combination with the chicane. THz pulse energies as high as 100 μJ have been measured when the charge approaches ~ 1 nC. However, with the high charges of the electron bunches, bunch length cannot be reduced down to hundreds of fs. fs-THz program at PAL pursuits generating intense THz pulses with femtosecond pulse duration to study ultrafast dynamics of materials in THz ranges. The LINAC of the fs-THz beamline at PAL runs at 50 MeV energy and produces electron bunches with total charges of 0.2–0.5 nC. The relation between the bunch charge and bunch length is summarized in Table 1 [15]. The frequency bandwidth and the pulse duration of the THz output will be optimized when the electron bunch length is minimized. Therefore, the bunch charge should be reduced to obtain short bunch length considering the space-charge effect.

2.4 Transition Radiation

Transition radiation occurs when an electron crosses a boundary between two different media [16–18]. The transition radiation is coherent at wavelengths that exceed the longitudinal length of the electron bunch. Accordingly, fs-THz pulses with an intensity that scales with N^2 , where N is the number of electrons in the bunch, can be generated by coherent transition radiation (CTR) with THz pulse duration comparable to the electron bunch length.

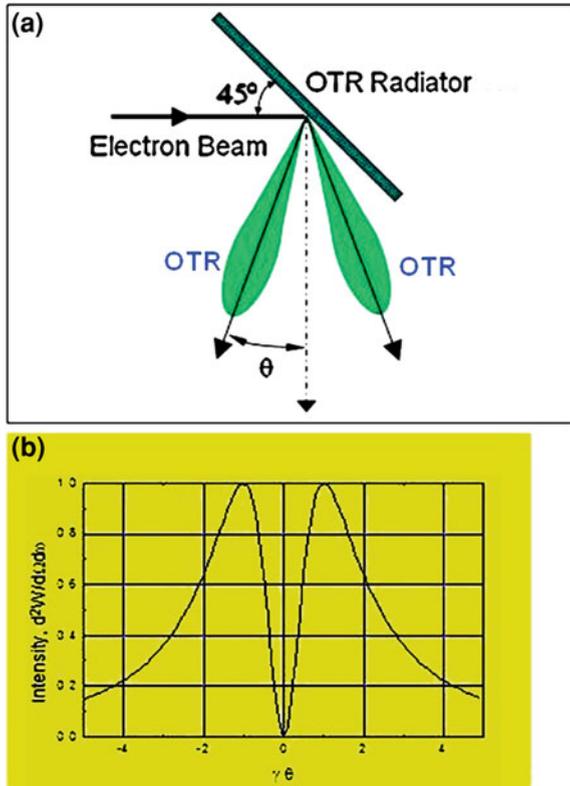
For relativistic electrons ($\beta \equiv v/c \approx 1$) passing through the boundary between vacuum and a perfect conductor, the angular distribution of the spectral energy per unit frequency interval is [16, 17]

$$E(\omega, \theta) = \frac{e^2}{\pi^2 c} \frac{\beta^2 \sin^2 \theta}{(1 - \beta^2 \cos^2 \theta)^2}$$

where θ is the angle between the electron trajectory and the emitted radiation, ω the angular frequency of the radiation, v the velocity of the electron, c the speed of light. Since the emitted radiation has the radial polarization, a characteristic feature of the transition radiation is that the intensity at $\theta = 0$ is 0 due to the destructive interference. The angular distribution shows a maximum at the angle $\theta_{\max} \sim 1/\gamma$, where γ is the Lorentz factor (Fig. 9).

The CTR target is a metal foil such as Ti with a thickness of 1 μm or Cu mirror, and the diameter of the electron beam is ~ 500 μm at the target. The radiated THz pulses are extracted from the LINAC vacuum pipe through a wedged CVD diamond window or crystalline quartz window.

Fig. 9 Transition radiation and its angular intensity distribution



3 THz Transport

Propagation of the THz CTR from the metal target to the sample in the experiment room should be investigated carefully when the source and the end points are separated by long distance. A successful Transport of the THz radiation with minimal intensity loss requires careful examination while accounting for the large diffraction effect originating from the long wavelength of THz radiation. *THz-Transport* code developed by Bernhard Schmidt at Deutsches Elektronen-Synchrotron (DESY) [19] is a useful tool to optimize the parameters of each optical component so that maximal transmittance is achieved at the sample position over the THz spectral range. A layout of the beam line optics and characteristics of the THz radiation propagating along the THz transfer line at PAL is shown in Fig. 10 [15]. As a radiation source, CTR generated from an electron beam with the Lorentz factor γ of 100, longitudinal length 100 fs, and transverse radius of 1 mm impinged onto the 1 in. diameter target is assumed. Intensity distributions of the 0.5 THz at five focusing mirrors are shown in the top row, where the size of each square panel corresponds to the actual dimension of the optical element, and the

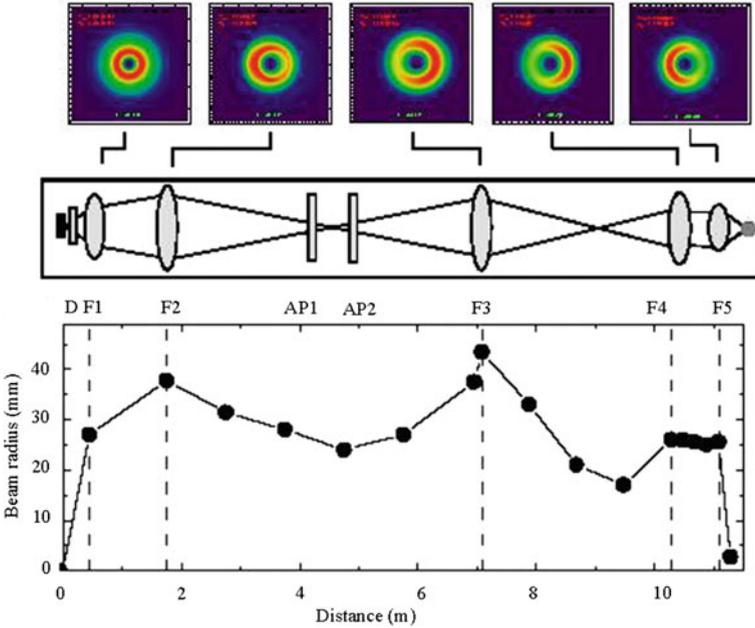


Fig. 10 A layout of the beam line optics and characteristics of the THz radiation propagating along the THz transfer line. *Top* Intensity distributions (0.5 THz) at the positions of focusing mirrors. *Middle* Optics arranged from the source point through diamond window (D) and five focusing mirrors (F1–F5), represented by lenses, to the experimental spot. *Bottom* Effective beam radii estimated at several representative positions along the beam propagation

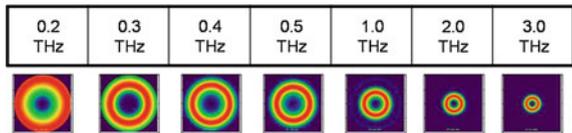
layout of the optics is shown in the middle panel. The details of the optics are summarized in Table 2. The wedged diamond window next to the metal target is approximated as a clear aperture of the same size. This window has a transmittance of about 70% in the THz region; Five focusing mirrors (F1–F5), which are 90° off-axis elliptic or parabolic mirrors, are positioned to collimate or focus the beam properly, while F5 was chosen arbitrarily to demonstrate the tight focus at the experiment position located 200 mm from F4.

Size of the optics, focal lengths, and surface types are optimized with the positions of the optics fixed. Great care should be taken to ensure high transmittance at low THz frequencies, where the diffraction effect is significant (Fig. 11). Besides the source and the objective focal points, the beam is loosely focused at two points, one between F2 and F3 and the other between F3 and F4. In the bottom row of Fig. 10, the effective beam radii estimated at several representative positions along the beam propagation are shown. The THz radiation strongly diverges at the source point and its divergence is reduced by F1, and then focused by F2. With the parabolic mirror (F2) of focal length 2,000 mm, the beam radius is only reduced to ~ 25 mm at the focal point. The radiation becomes almost collimated

Table 2 Parameters of optical elements installed in the THz transfer line

	Position (mm)	Focal length (mm)	Radius (mm)
Diamond window	85		10
F1 (Parabolic)	450	550	50
F2 (Parabolic)	1,750	2,000	75
AP1	2,950		50
AP2	3,950		50
F3 (Elliptical)	7,080	1,700	75
F4 (Parabolic)	10,280	2,000	50
F5 (Parabolic)	11,080	200	50

Fig. 11 The two-dimensional intensity distributions at the diamond window at different frequencies



with a radius of 25 mm after F4. THz radiation after the F4 mirror can be tightly focused when using a parabolic mirror with appropriate focal length.

The transmittance of the THz Transport system changes with frequency. Although the transmittance at >1 THz is greater than 80%, but the finite sizes of the optical elements limit the transmittance at lower frequencies (Fig. 12a). Nevertheless, transmittance to the experimental point can be maintained at an acceptable value even at 0.2 THz. Figure 12b shows the intensity spectra which peak at ~0.5 THz, and decrease sharply at lower frequencies; this is a characteristic of the electron beam used for the THz generation and the diffraction by the finite sized metal screen used. In spite of the great loss of the intensity between F2 and F3, the spectral shape is nearly the same throughout the Transport system.

4 THz Detection

4.1 Detectors

Pyroelectric detector, goley cell, and cryogenic bolometer are the typical detectors to measure the power of the THz radiation. Pyroelectric detectors are thermal detectors working with a thermally isolated chip which is covered by a black absorption coating. This coating converts the THz radiation falling on the chip to heat. The change in temperature modifies the polarization of the material and this polarization change gives rise to a voltage across the material. The Goly cell is an opto-acoustic detector and consists of a gas-filled enclosure with an IR and THz absorbing material and a flexible membrane. When THz radiation is absorbed, the heated gas expands and results in the deformation of the membrane. A light

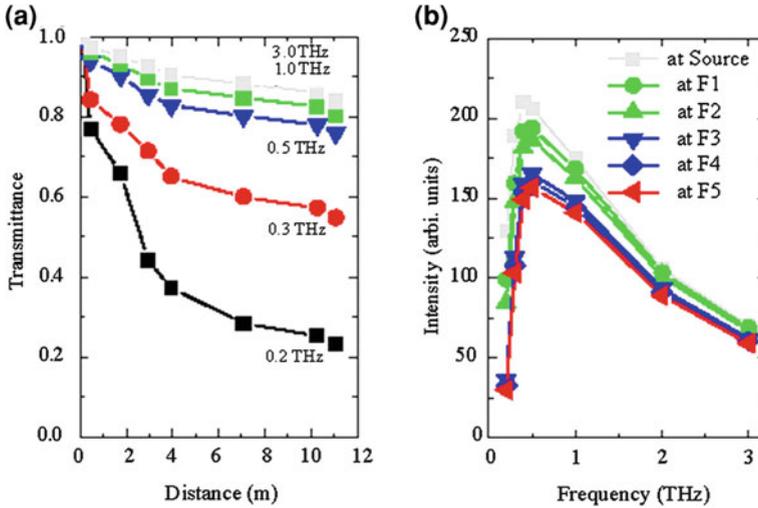


Fig. 12 Spectral characteristics of the coherent transition radiations during the beam propagation. **a** Beam line transmittance for representative frequencies as a function of the distance from the source point. **b** Intensity spectra estimated at the positions of the focusing mirrors [15]

emitting diode (LED) emits through re-focussing optics and onto the mirrored back surface of the chamber containing the absorbing membrane. This radiation is reflected back through the lower half of the optics via a grating and re-focussed onto a photodiode, and motion of the membrane produces a change in the signal on the photodiode [20]. The Golay cell has high sensitivity and a flat response over a very broad range of frequencies (0.2–20 THz). A bolometer consists of an absorptive element, such as a thin layer of metal or semiconductor, connected to a body of constant temperature (reservoir) through a thermal link. THz radiation falling on the absorptive element raises its temperature above that of the reservoir. The temperature change can be measured directly with an attached resistive thermometer. Most bolometers use semiconductor or superconductor as absorptive elements and these devices can be operated at liquid He temperatures enabling significant sensitivity improvement.

Figure 13 shows the transverse profile of the electron bunch at the YAG screen (Yttrium Aluminum Garnet, $Y_3Al_5O_{12}$), which is located 30 cm upstream of the CTR target, and the beam profile of the generated THz pulses at the CTR target. The THz beam profile is taken with a $LiTaO_3$ based pyroelectric array detector (Pyrocam III, Spiricon, inc.). The measured image matches well with a simulated data discussed above. Since the intense THz output easily saturates the bolometer, the THz pulse energy generated from the CTR target can be measured with a Golay cell or pyroelectric detector and a lock-in amplifier. With the beam charge of ~ 1 nC, 100 μ J single-cycle THz radiation with electric field strength of approaching 1 MeV/cm has been measured at BNL [11].

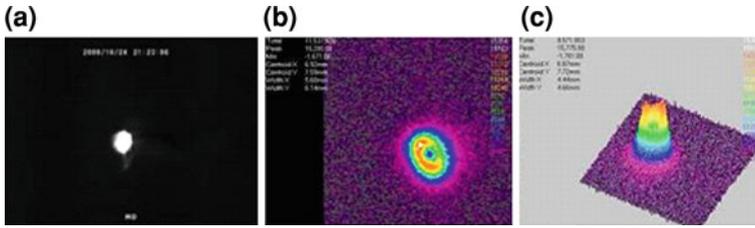


Fig. 13 Image of the electron bunch at the YAG screen (a) and Image of the THz output out of the CTR target (b), (c) measured at PAL

4.2 Electro-Optic Sampling

The THz radiation can be detected by electro-optic (EO) sampling method [3, 13, 14] in 1 (or 0.5) mm thick $\langle 110 \rangle$ ZnTe nonlinear crystal (Fig. 14). EO sampling is based on the Pockels effect in which an applied voltage causes the crystal to become birefringent [3]. If the 800 nm gate pulse travels through the nonlinear crystal at the same time as one point in the THz pulse, the THz field induces birefringence in the crystal to rotate the polarization of the gate pulse in proportion to the THz field amplitude, and the direction of rotation is proportional to the sign of the field. By scanning a delay line, the entire pulse amplitude can be mapped out. The signal is collected with a quarter-wave plate, a Wollaston polarizer, a pair of balanced photodiodes, and a lock-in amplifier phase locked to an optical chopper. The THz beam path needs to be enclosed and purged with dry air or nitrogen to minimize THz absorption by water vapor. Figure 14b shows the time domain THz transients and their power spectra.

4.3 Single Shot EO Sampling

Single shot EO sampling is adopted to solve the problem of intrinsic time jittering of THz pulse and of slow data acquisition rate due to the low repetition rate of the THz beam generated from LINAC. Time jittering of THz pulse makes the conventional EO sampling with good time resolution impossible. The THz temporal E field of these intense pulses can be measured by single shot EO sampling with chirped laser pulses [11, 21, 22]. A small fraction of the linearly chirped Ti:sapphire laser output is used as a gate or probe. The THz pulse and the linearly polarized probe pulse are copropagated through a $[110]$ ZnTe crystal, and the probe pulse goes through a quarter wave plate and a polarizer analyzer (Fig. 15). The time evolution of the THz E field is transformed into an intensity modulation on the probe spectrum. By measuring the spectrum difference of the probe beam with and without the THz pulse, the time evolution of the THz E field is obtained without moving the time delay (Fig. 16).

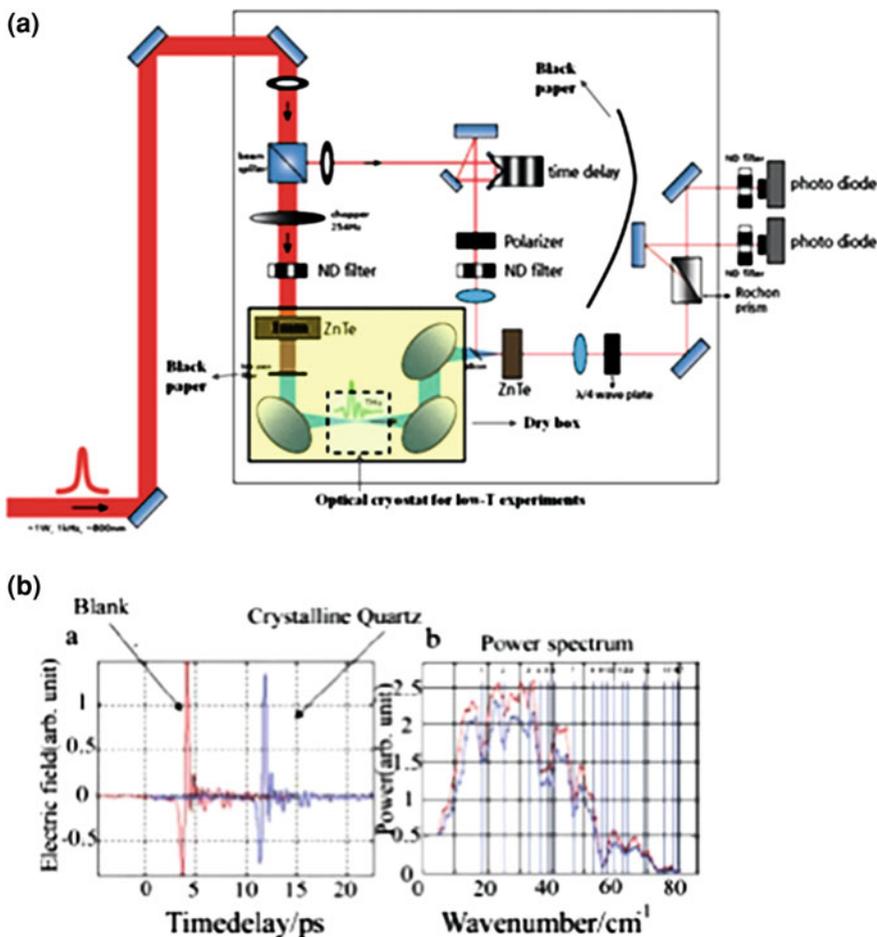


Fig. 14 a EO sampling detection scheme (upper). b Comparison of time domain THz transients of blank and crystalline quartz window and power spectra of THz pulses (red blank, blue crystalline quartz window) (lower)

5 Application

5.1 Time Domain Spectroscopy (TDS)

Figure 14a shows a schematic diagram of THz time-domain spectroscopy (THz-TDS) [3, 8, 23, 24]. Although the measurements are made in the time domain, this technique is not time resolved measurements. By Fourier transformation into the frequency domain, Fig. 14b, the THz frequency dependent absorption coefficient $\alpha(\omega)$, refractive index $n(\omega)$, permittivity $\epsilon(\omega)$, and susceptibility $\chi(t)$ of materials

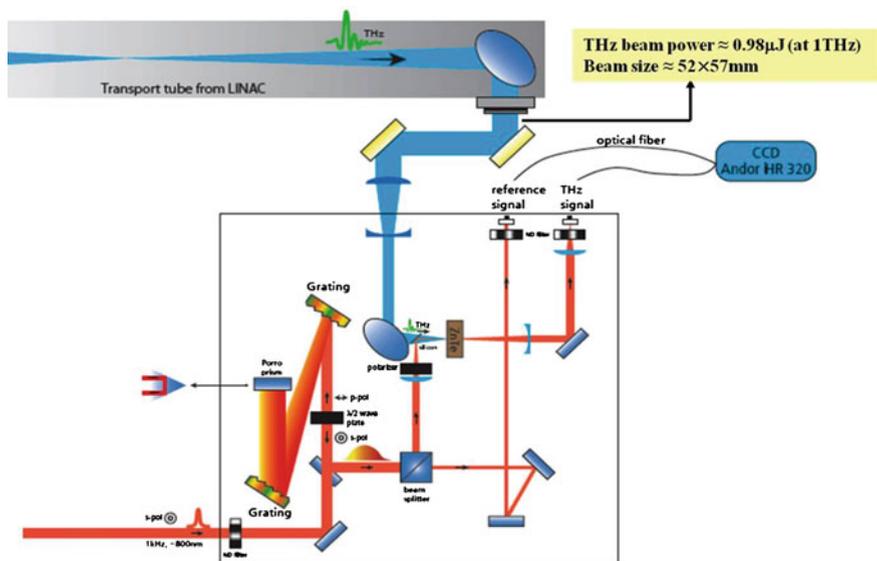


Fig. 15 Schematic of experimental setup of single shot EO measurement with a chirped probe beam

can be measured directly without resorting to the Kramers–Kronig relation since phase and amplitude of the THz field can be measured simultaneously using THz time domain spectroscopy [3] (Fig. 17).

5.2 Time Resolved Terahertz Spectroscopy (TRTS)

Time-resolved terahertz spectroscopy [2, 5, 7, 26] is a specific case of optical pump-probe methods (Fig 18). For a typical pump-probe measurement, ultrafast optical pulses excite a sample, and subsequent photo induced changes of sample properties are examined by short probe pulses as a function of delay time. Figure 19 shows the experimental setup of time resolved THz spectroscopy. After ultrafast optical pulses excite the sample, the transient change of the sample is probed by THz pulses by EO sampling or by single shot EO sampling method [3]. Since the energy of the probing photons is low, this method is mainly sensitive to intraband absorptions of solid or to the collective low-frequency solvent response in a liquid.

Recently, the development of high-energy tabletop THz pulse sources and the accelerator based THz source has enabled nonlinear transmission studies on semiconductors as well as measurements with THz pump and near-IR or mid-IR probe pulses [27]. Under an intense THz pulse irradiation, the interband excitation that results in the generation of excitons occurs, although the THz photon energy

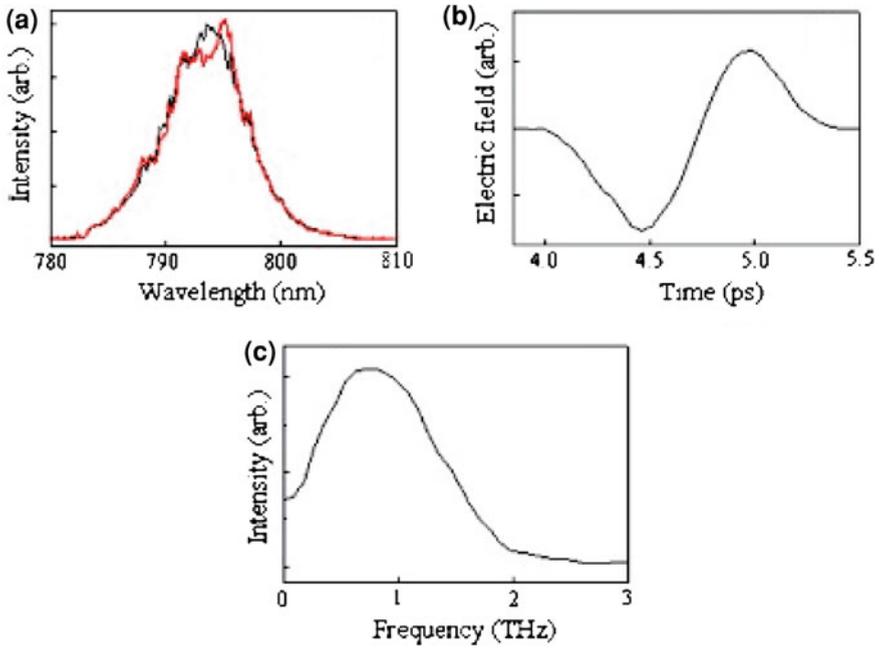
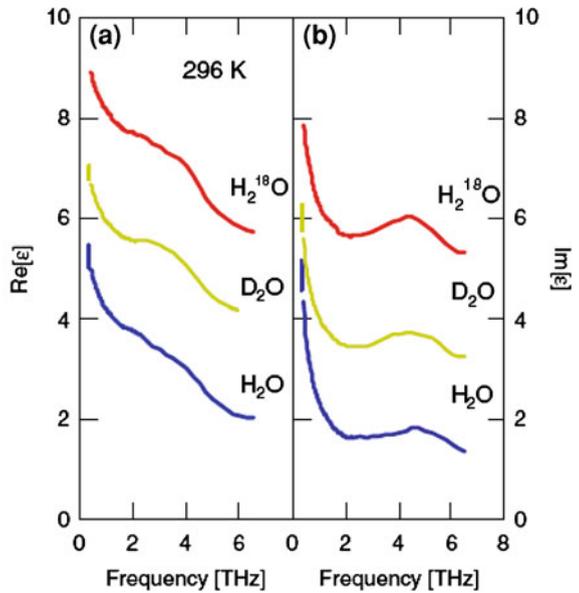


Fig. 16 **a** Raw data, single-shot probe pulse spectra with (*red*) and without (*black*) a copropagating THz pulse. **b** Measured single-shot THz waveform. **c** Fourier transform of the waveform of **a** [11]

Fig. 17 **a** Real and **b** imaginary parts of the complex dielectric constants of liquid H₂O, D₂O, and H₂¹⁸O at 296 K. The absolute values of the dielectric constants of liquid D₂O and H₂¹⁸O are shifted by 2 and 4, respectively. [25]



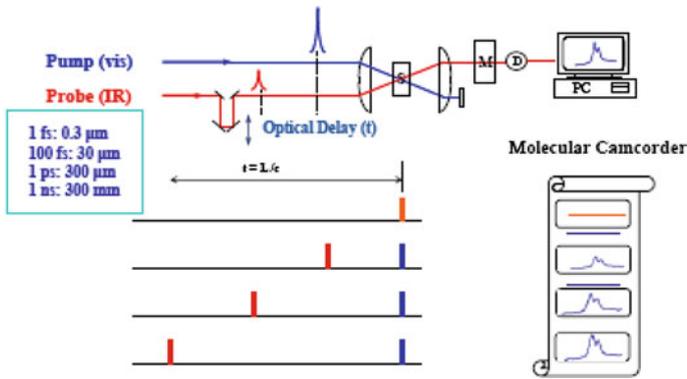


Fig. 18 Schematic of experimental setup of time resolved measurement

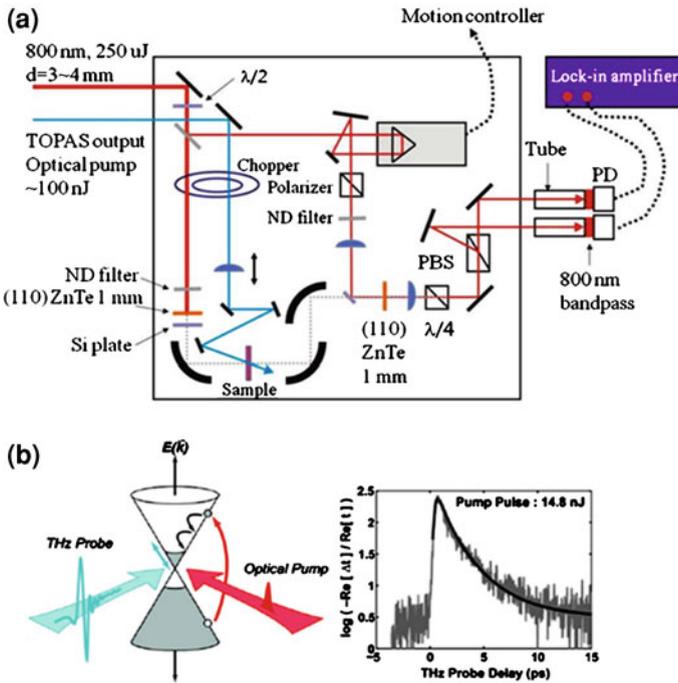


Fig. 19 a Experimental setup of time resolved THz spectroscopy. b The ultrafast relaxation and recombination dynamics of photogenerated electrons and holes in epitaxial graphene. The measured change in the real part of the complex amplitude transmission (gray) and the theoretical fit (black) for pump pulse energy of 14.8 nJ [5]

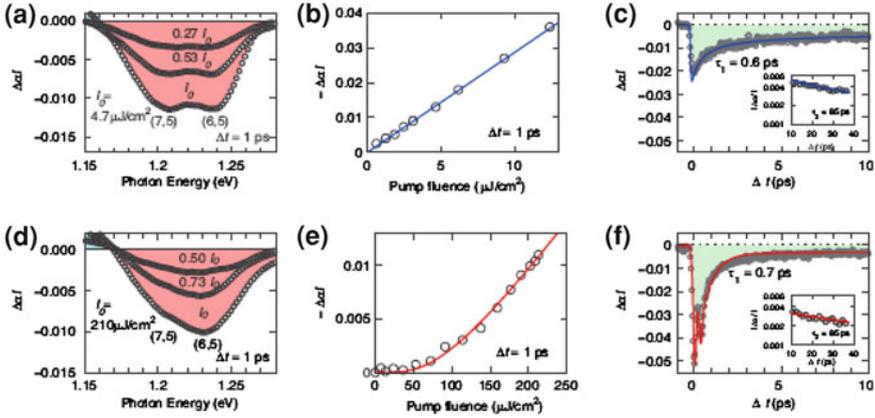


Fig. 20 Intense terahertz pulse induced exciton generation in carbon nanotubes. Results for the near-IR-pump ($\hbar\omega_{\text{pump}} = 1.55$ eV) (a)–(c) and THz-pump experiments ($\hbar\omega_{\text{pump}} \sim 4$ meV) (d)–(f) [28]

(~ 4 meV) is much smaller than the gap energy of SWNTs (~ 1 eV) (Fig. 20). THz field strengths on the order of up to 1 MeV/cm can open up the field of nonlinear optics with single-cycle THz pulses and will be a powerful tool to study high E-field effects in condensed phase with ultrafast temporal resolution.

6 Summary

The design of the fs-THz beamline including a fs-regenerative laser amplifier system, Generation, Transport, detection of the THz, and the experimental methods such as TDS and TRTS are described. Recent development of high-energy tabletop THz pulse sources and the accelerator based coherent THz source has enabled nonlinear transmission studies as well as THz pump and near-IR or mid-IR probe pulses. This intense THz source can open up the new field of nonlinear science with single-cycle THz pulses and provide a wonderful opportunity to understand nonperturbative light-matter interactions in condensed phase in a time-resolved manner.

Acknowledgments This work was supported by National Research Foundation of Korea (NRF) Grant funded by the Korean Government (grant code: 2011-0001291). The work at the PLS were supported in part by the Ministry of Education, Science and Technology (MEST) and POSTECH.

References

1. Bartel, T., Gaal, P., Reimann, K., Woerner, M., Elsaesser, T.: *Opt. Lett.* **30**, 2805 (2005)
2. Baxter, J.B., Schmuttenmaer, C.A.J.: *Phys. Chem. B* **110**, 25229 (2006)
3. Beard, M.C., Turner, G.M., Schmuttenmaer, C.A.J.: *Phys. Chem. B* **106**, 7146 (2002)
4. Danielson, J.R., Lee, Y.-S., Prineas, J.P., Steiner, J.T., Kira, M., Koch, S.W.: *Phys. Rev. Lett.* **99**, 237401/1 (2007)
5. George, P.A., Strait, J., Dawlaty, J., Shivaraman, S., Chandrashekar, M.V.S., Rana, F., Spencer, M.G.: *Nano Lett.* **8**, 4248 (2008)
6. Hendry, E., Schins, J.M., Candeias, L.P., Siebbeles, L.D.A., Bonn, M.: *Phys. Rev. Lett.* **92**, 196601/1 (2004)
7. Suzuki, T., Shimano, R.: *Phys. Rev. Lett.* **103**, 057401/1 (2009)
8. Tielrooij, K.J., Garcia-Araez, N., Bonn, M., Bakker, H.J.: *Science* **1006**, 328 (2010)
9. Carr, G.L., Martin, M.C., McKinney, W.R., Jordan, K., Neil, G.R., Williams, G.P.: *Nature* **420**, 153 (2002)
10. Doria, A., Gallerano, G.P., Giovenale, E., Messina, G., Spassovsky, I.: *Phys. Rev. Lett.* **93**, 264801/1 (2004)
11. Shen, Y., Watanabe, T., Arena, D.A., Kao, C.C., Murphy, J.B., Tsang, T.Y., Wang, X.J., Carr, G.L.: *Phys. Rev. Lett.* **99**, 043901/1 (2007)
12. Yeh, K.L., Hoffmann, M.C., Hebling, J., Nelson, K.A.: *Appl. Phys. Lett.* **90**, 171121/1 (2007)
13. Nahata, A., Weling, A.S., Heinz, T.F.: *Appl. Phys. Lett.* **69**, 2321 (1996)
14. Zhang, X.C., Jin, Y., Ma, X.F.: *Appl. Phys. Lett.* **61**, 2764 (1992)
15. Park, J., Kim, C., Lee, J., Yim, C., Kim, C.H., Lee, J., Jung, S., Ryu, J., Kang, H.-S., Joo, T.: *Rev. Sci. Instrum.* **82**, 013305/1 (2011)
16. Frank, I.M., Ginzburg, V.L.J.: *Phys. USSR* **9**, 353 (1945)
17. Happek, U., Sievers, A.J., Blum, E.B.: *Phys. Rev. Lett.* **67**, 2962 (1991)
18. Leemans, W.P., Geddes, C.G.R., Faure, J., Toth, C., Van Tilborg, J., Schroeder, C.B., Esarey, E., Fubiani, G., Auerbach, D., Marcellis, B., Carnahan, M.A., Kaindl, R.A., Byrd, J., Martin, M.C.: *Phys. Rev. Lett.* **91**, 074802/1 (2003)
19. Casalbuoni, S., Schmidt, B., Schmueser, P., Arsov, V., Wesch, S.: *Phys. Rev. Spec. Top.-Accel. Beam*, **12**, 1–3 (2009)
20. Golay, M.J.E.: *Rev. Sci. Instrum.* **18**, 347 (1947)
21. Jiang, Z., Zhang, X.C.: *Appl. Phys. Lett.* **72**, 1945 (1998)
22. Kim, K.Y., Yellampalle, B., Rodriguez, G., Averitt, R.D., Taylor, A.J., Glowina, J.H.: *Appl. Phys. Lett.*, **88**, 041123/1, (2006)
23. Fattinger, C., Grischkowsky, D.: *Appl. Phys. Lett.* **54**, 490 (1989)
24. Flanders, B.N., Cheville, R.A., Grischkowsky, D., Scherer, N.F.J.: *Phys. Chem. B* **100**, 11824 (1996)
25. Yada, H., Nagai, M., Tanaka, K.: *Chem. Phys. Lett.* **473**, 279 (2009)
26. Haran, G., Sun, W.-D., Wynne, K., Hochstrasser, R.M.: *Chem. Phys. Lett.* **274**, 365 (1997)
27. Hebling, J., Hoffmann, M.C., Hwang, H.Y., Yeh, K.-L., Nelson, K.A.: *Physical Review B: Condensed Matter and Materials Physics* **81**, 035201/1 (2010)
28. Watanabe, S., Minami, N., Shimano, R.: *Opt. Express* **19**, 1528 (2011)

Electron Beam Sources Based on Carbon Nanotube for THz Applications

Yong Hyup Kim, Tae June Kang, Wal Jun Kim, Eui Yun Jang
and Jeong Seok Lee

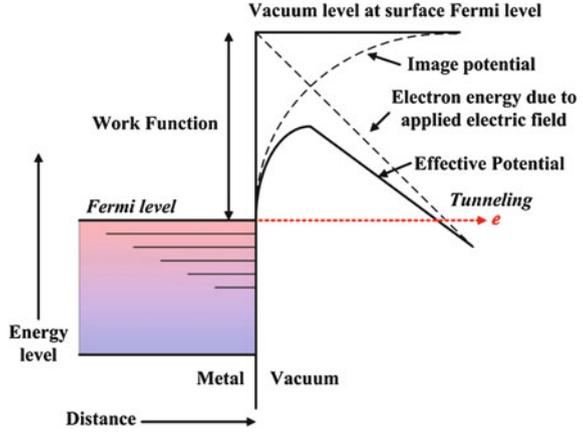
1 Introduction

Performance of electron emitter plays an essential role in the detection and generation of electromagnetic wave signals. It becomes technologically challenging from spanning terahertz (THz) applications, due to the lack of sufficient power sources. The vacuum THz amplifier, such as a travelling wave tube (TWT) or a klystron is practically used to increase the output power. The characteristic of these amplifiers is mainly represented by the performance of electron beam source which has to deliver a sufficient current in order to allow an amplification of the THz signal.

The electron sources mostly have used thermionic emission in which electrons are emitted from heated filament (hot cathode). However, the thermionic emission requires high thermal energy for heating the emitter to an operating temperature, resulting in additional power dissipation (most of which is lost as conduction heat or radiation) and distinct time delays in starting operation. Field emission (FE) is an alternative method to generate an electron beam. FE mechanism is basically a quantum effect with a sufficiently high external electric field, electrons near the Fermi level can tunnel through the energy barrier and escape to the vacuum level. Compared to thermionic emission, the FE could be preferable for the THz electron source because no heating is required and high emission current is readily generated by applying an external electric field. Thus the cathode used in field emission is often called “cold cathode” and has considered as a strong candidate of high current electron source.

Y. H. Kim (✉) · T. J. Kang · W. J. Kim · E. Y. Jang · J. S. Lee
Department of Physics and Astronomy, School of Mechanical and Aerospace Engineering,
Seoul National University, Gwanak-gu 1, Seoul 151-747, Republic of South Korea
e-mail: yongkim@snu.ac.kr

Fig. 1 Schematic draw of field emission mechanism based on quantum tunneling



Carbon nanotube (CNT) has emerged as a promising class of electron field emitters in terms of a low threshold electric field for emission and a high emission current density. In this part, we will briefly introduce the fundamentals of field emission theory, CNTs and their physical properties. Discussion on CNT cold cathode in THz devices will also be described. The emphasis is on the fabrication methods for various type of planar and point emitters, the cutting-edge emission performances of the CNT cathodes, and their relations with the underlying materials properties.

2 Fundamentals of Field Emission

Field emission is defined as emission of electrons from surface of a condensed phase into another (usually vacuum) by an external electric field. Wood first observed this phenomenon in 1897. The field emission basically consists of tunneling of electrons through deformed potential barrier at the surface as shown in Fig. 1. In this sense, it is different from thermionic emission or photoemission in which electrons are given sufficient energy to overcome the potential barrier. Potential barrier in field emission deforms strongly so that unexcited electrons can be emitted through the deformed thin potential barrier.

Tunneling is a purely quantum mechanical phenomenon, thus field emission is not explained by classical physics. Fowler and Nordheim firstly calculated tunneling probability on metal surface using quantum mechanics in 1928. The field emission current density (J) from the Fowler–Nordheim equation is given by equation [1].

$$J = 1.4 \times 10^{-6} \frac{E_{loc}^2}{\phi} \exp\left(\frac{9.87}{\phi^{0.5}}\right) \exp\left[-6.53 \times 10^7 \frac{\phi^{1.5}}{E_{loc}}\right] \quad (1)$$

As shown in the equation, the field emission current density depends on local electric field E_{loc} and work function (ϕ). Local electric field can be expressed as $E_{loc} = \beta E = \beta V/d_0$ where V is an applied voltage, d_0 is an inter-electrode distance, and E represents applied electric field. The field emission enhancement factor (β) is determined solely by geometrical shape, such as aspect ratio, curvature of tip. Therefore, there are two possible approaches to enhance the current density; (1) increasing field emission enhancement factor and (2) reducing work function of emitter material. Many researches using CNT materials have focused on the first approach to build up high performance field emitter due to geometrical advantage of CNTs (i.e., high aspect ratio), based on extraordinary field emission properties.

3 Carbon Nanotubes

Carbon nanotubes (CNTs) are allotropes of carbon with a cylindrical nanostructure, discovered in 1991 by Iijima [2]. The name is derived from their long, hollow structure with the walls formed by one-atom-thick sheets of sp^2 -bonded carbon, called graphene. These carbon sheets are rolled at specific and discrete chiral angles, and the combination of the rolling angle and radius decides the nanotube electrical properties, such as metallic or semiconducting. Nanotubes are categorized as single-walled nanotube (SWNT) and multi-walled nanotube (MWNT) depending on the number of walls composing the tube. MWNT is constructed with a set of concentric SWNTs with different diameters while having an inter-tube distance of 3.4 Å. The MWNT is usually considered as a zero-gap metal.

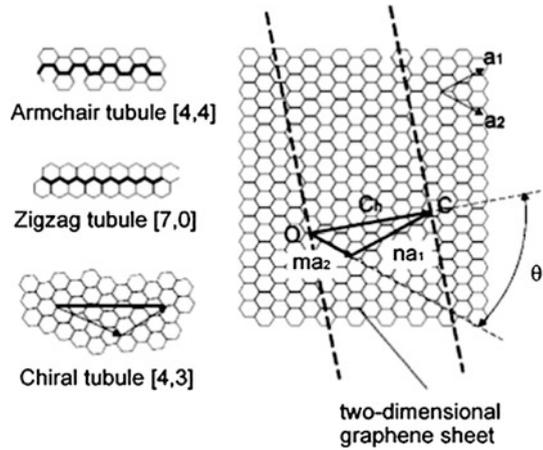
3.1 Material Properties

Large amount of effort has been devoted to evaluate the physical properties of CNTs, but a lot of work should be done because CNTs exhibit diverse physical properties depending on the diameter, length, and chirality. The following subsections briefly summarize the electric, thermal, and mechanical properties of nanotubes known up to now.

(1) Electrical property

Electrical property of CNT depends on its diameter and chiral vector. Band gap of CNT is inversely proportional to the diameter. The way how the graphene sheet is wrapped is represented by a pair of indices (n, m) as shown in Fig. 2. The integers n and m denote the number of unit vectors along two directions in the honeycomb crystal lattice of graphene. When $m = 0$, the nanotubes are called zigzag nanotubes, and when $n = m$, the nanotubes are called armchair nanotubes. Otherwise, they are called chiral. For chiral vector, if $n = m$ or $n - m = 3q$ where

Fig. 2 The chiral vector defined on the hexagonal lattice of carbon atoms [5]



q is a nonzero integer, the CNTs are metallic with a small band gap, otherwise the CNTs are semiconductors. Metallic CNT is theoretically able to carry an electric current density of $4 \times 10^9 \text{ A cm}^{-2}$ [3], which is 1,000 times greater than those of metals such as copper. In 1998, Frank et al. [4] observed that the CNTs have significantly high ballistic conductance, which enabling them to carry massive current without essential heating. It was experimentally confirmed that the current density could exceed 10^7 A cm^{-2} .

(2) Thermal property

Outstanding thermal conductivity of CNTs has been widely used to demonstrate novel thermal dissipation and interfacial material. Hone et al. [6] reported that the thermal conductivity of a SWNT rope at room temperature is in the range of $1,800\text{--}6,000 \text{ W m}^{-1} \text{ K}^{-1}$. Che et al. [7] numerically showed that the thermal conductivity of a $(10, 10)$ nanotube approaches $2,980 \text{ W m}^{-1} \text{ K}^{-1}$. Berber et al. [8] reported a relationship between thermal conductivity and temperature, and theoretically confirmed that the mean free path of the phonon is dominant factor on significantly high thermal conductivity value of $6,600 \text{ W m}^{-1} \text{ K}^{-1}$.

(3) Mechanical property

CNTs have exceptional mechanical properties, such as tensile strength and elastic modulus, which results from the covalent sp^2 -bonds between the individual carbon atoms. MWNT was tested to have a tensile strength of 63 GPa [9]. Peng et al. [10] revealed that individual CNT shells have strengths of up to $\sim 100 \text{ GPa}$, which is in good agreement with quantum/atomistic models. Since carbon nanotubes have a low density of $1.3\text{--}1.4 \text{ g cm}^{-3}$, its specific strength of $48,000 \text{ kN m kg}^{-1}$ is the highest among known materials, compared to high-carbon steel of 154 kN m kg^{-1} [11].

3.2 *Synthesis*

Carbon nanotubes are generally produced by three main techniques, arc discharge, laser ablation and chemical vapor deposition (CVD). In this section, different techniques for nanotube synthesis and their current status are briefly explained.

(1) **Arc-discharge**

When arc-discharging between two graphite rods of electrodes in an enclosure that is usually filled with inert gas (helium, argon), the carbon is vaporized from the anode rod and then condensed on the relatively cold cathode rod. Such condensed carbons are composed of multi-walled carbon nanotubes and carbon particles. SWNTs can be synthesized by mixing different metals into the anode. The yield for this method is up to 30% by weight and it produces both single- and multi-walled nanotubes with lengths of up to 50 μm with few structural defects.

(2) **Laser ablation**

Smalley's group in Rice University, USA, used the laser ablation method to synthesize CNTs. A pulse laser vaporizes a graphite target in a reactive chamber that is maintained at 1,200°C. The vaporized carbons are condensed on a cold collector. The resulting material is composed of MWNTs and carbon particles. Also, by mixing various catalytic metals such as cobalt and nickel on to the graphite target, SWNTs can be synthesized with a controllable diameter determined by the reaction temperature. However, it is a more expensive method than either arc discharge or chemical vapor deposition.

(3) **Chemical vapor deposition**

CVD synthesis refers that reactive carbon atoms diffuse to catalytic nanoparticles and then crystallizing into tubes. The synthesis of CNTs through CVD method is essentially a two-step process consisting of a catalyst preparation step followed by the actual synthesis of the nanotube. The catalyst is generally prepared by depositing a transition metal such as iron, cobalt, nickel onto a substrate and then using thermal annealing to induce catalyst particle nucleation. Thermal annealing results in cluster formation on the substrate, from which CNTs will generally grow at a specific temperature in the range of 650–900°C. The diameters of CNTs are related to the size of the metal nanoparticles. This can be controlled by annealing condition. Moreover, the CVD method is capable of selective growth of either SWNTs or MWNTs by tuning processing parameters such as temperature, size of catalytic nanoparticles, and species of carbon sources. The CVD method is an economically efficient method on the industrial scale to synthesize SWNTs and MWNTs.

3.3 Dispersion

CNTs have poor solubility in many kinds of solvents which limits their processability for diverse applications. There are two strategies for dispersing CNTs: physical modification (non-covalent treatment) or chemical modification (covalent treatment). The well-dispersed CNTs in solution allow us to manipulate and extend CNTs into a desired structure with various wet-based assembly methods. In this section, physical and chemical modification for the dispersion of CNTs will be summarized.

(1) Physical modification

The non-covalent surface treatment by polymer wrapping or surfactant coating has been widely used in the preparation of both aqueous and organic solvents to obtain highly concentrated solution with individually dispersed CNTs [12]. It is basically van der Waals, hydrophobic or $\pi - \pi$ interactions, and this modification is particularly interesting because it does not perturb the electronic structure of CNTs. In practice, CNTs are dispersed by adding them into an aqueous solvent with a surfactant such as sodium dodecyl sulphate (SDS), lithium dodecyl sulfate (LDS). By encasing CNTs with functional polymers, such as cross-linked copolymer polystyrene-blockpolyacrylic acid (PS-b-PAA) is also used for non-covalent treatment [13]. This encapsulation may enhance CNT dispersions in various organic solvents. After the non-covalent surface treatment, sonication may help to debundle nanotubes by steric or electrostatic repulsions.

(2) Chemical modification

Covalent functionalization on CNT surface with chemical functional groups effectively improves their dispersion properties in various solvents by enhancing wetting and adhesion properties. CNTs are functionalized by irreversible attachment of appendage on the sidewalls and/or on the tips. Strong acid treatments with sulfuric acid (H_2SO_4) or nitric acid (HNO_3), or mixture of acids, are usually used to improve the wettability so that CNTs can be well-dispersed even in water [14]. Though useful chemical properties can be additionally embedded by means of functional groups, aggressive chemical functionalization might introduce structural defects so it degrades the intrinsic CNT properties. For instance, electrical properties of SWNT are perturbed by covalent functionalization because double bonds between carbon atoms are irreversibly lost.

3.4 Applications

The special nature of carbon combines with the molecular perfection of SWNTs to endow them with exceptional electrical, mechanical and chemical properties. These superior characteristics of CNTs have been stood in the spotlight of various applications.

CNTs have high potential for improving the mechanical, thermal, and electrical properties of composite materials, which has inspired researchers to use them as reinforcements for metal and ceramic matrices, as well as polymer matrices [15, 16]. Excellent electrical properties, such as ballistic transport behavior can be useful for developing single electron transistor [17]. High conductivity and capability of high current density are excellent for an electrical wire, anti-static electricity, magnetic shield, field emission source, etc. [15, 18]. Thermal conductivity of CNTs has inspired many researchers to demonstrate outstanding thermal dissipation and interfacial materials [19]. Furthermore the interesting combination of electrical, mechanical and chemical properties of CNTs has led to wide-range investigations of their potential in future applications, such as highly sensitive detectors, flexible electronics, and energy storage/harvesting devices [15, 20, 21].

Of these applications, CNTs and their assembled structures has attracted considerable attentions for the use of electron emitter in field emission. Since de Heer et al. [22] had attempted to apply CNTs for field emitter, the diverse efforts for developing field emission display (FED), X-ray tube and electron amplifier have been carried out [23]. Recently these superior field emission properties of CNTs are spanning to THz applications. The following section covers on the fabrication methods of various types of CNT emitter, and their field emission performances.

4 CNT Array and Its Assembled Structures for Field Emission

CNTs have been investigated as a promising cold-cathode due to their excellent field emission properties. However, since individual CNTs are able to only emit up to 1 μA [24–26], CNT array and its assembled structures have been required for high-power field emitters. There exist various CNT structures in macroscopic scale, including fiber, yarn, sheet, film, vertical-aligned (VA) array, and sponge-like form. Among them, fiber, yarn, and VA-array can be useful for field emission due to their geometric aspect ratio, as discussed in Sect. 2. These structures are categorized into planar (VA-array) and point (fiber and yarn) CNT emitter. This section covers an overview of field emission of planar and point CNT emitters, the state-of-art emission performances, and their relations with the underlying material properties.

4.1 CNT Planar Emitter

Large-scale synthesis of VA-CNT structure, for the first time, was achieved by using a method based on CVD catalyzed by iron nanoparticles embedded in mesoporous silica [27]. Afterward many researchers have fabricated VA-CNT structures by different methods [22, 28–31]. Figure 3 provides typical images of

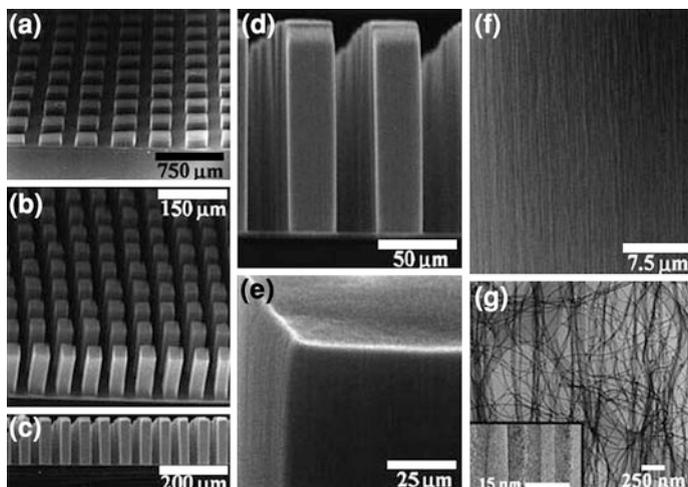
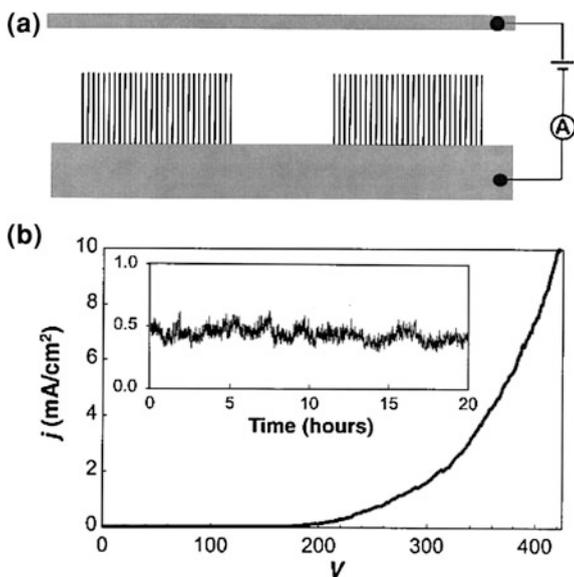


Fig. 3 Secondary electron microscopy (SEM) images of VA-CNT structures synthesized on porous silicon substrates [31]. **a–f** SEM images according to different pattern size, magnification, and angle view. **g** Transmission electron microscopy (TEM) image of MWNTs in a VA-CNT structure. The *inset* is a high-resolution TEM image

Fig. 4 VA-CNT structures as electron field emitter arrays [31]. **a** Experimental setup: the cathode consists of a porous silicon substrate with VA-CNT structures. An aluminum-coated silicon substrate serves as the anode and is kept 200 μm away from the VA-CNT structure. **b** Current density (j) versus voltage (V) characteristics of the sample. The *inset* shows the emission current stability [32]



VA-CNT structures. These VA-CNT structures have been directly used for a cathode of field emission. In 1996, a high-intensity electron gun from VA-CNT structures was demonstrated by de Heer and co-workers [22]. The VA-CNT structures were fabricated by template-assisted filtration method. The current densities greater than 100 mA/cm^2 have been realized.

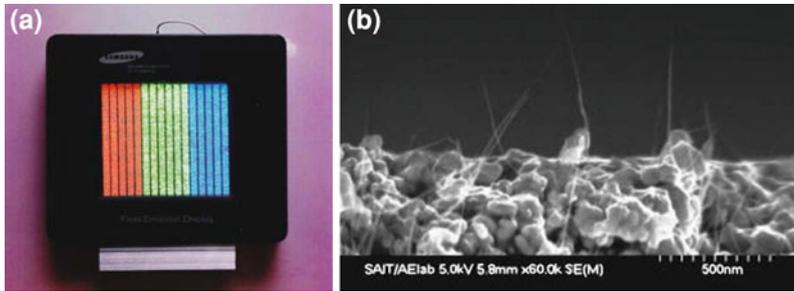


Fig. 5 **a** Emitting image of fully sealed SWNT-FED at color mode with *red, green, and blue* phosphor columns. **b** Cross-sectional SEM image of CNT cathode [23]

Fan et al. [31] used patterned VA-CNT structures for an electron field emission array. The current densities from the array reached 10 mA/cm^2 at $\sim 5 \text{ V}/\mu\text{m}$ as shown in Fig. 4. The emission current is found to be linearly proportional to the number of nanotube blocks on the samples. In addition, the CNT arrays emit electrons stably over a test period of 20 h at a current density of $\sim 0.5 \text{ mA/cm}^2$.

Choi et al. [23] demonstrated high-brightness field emission display using SWNTs by screen-printing method (see Fig. 5a). The SWNTs are vertically aligned using paste squeeze and surface rubbing techniques as shown in Fig. 5b. The turn-on fields of $1 \text{ V}/\mu\text{m}$ and field emission current of 1.5 mA at $3 \text{ V}/\mu\text{m}$ ($90 \mu\text{A/cm}^2$) were achieved. Li et al. [33] also fabricated planar field emitters by screen-printing MWNTs paste on glass substrates with indium-tin-oxide (ITO) film. The turn-on field is less than $3 \text{ V}/\mu\text{m}$ and the emission current density can approach 1 mA/cm^2 at $6.5 \text{ V}/\mu\text{m}$. The screen-printing method is easy to fabricate and scale-up planar CNT emitters, but there is a fatal problem regarding to degassing of organic vapor.

Meanwhile, Zhou group reported a novel method for fabrication of VA-CNT structure as field emitter using electrophoresis and activation process (see Fig. 6a) [34]. The cathodes have significantly enhanced field emission current density ($1,500 \text{ mA/cm}^2$) and long-term stability under high operating voltages. The application of this CNT electron source is demonstrated for high-resolution X-ray imaging as shown in Fig. 6b.

4.2 CNT Point Emitter

CNT point emitter using CNT assembled structure is a highly effective field emission source because the screening effect is negligible in the configuration of point geometry [35]. The fabrication method of CNT point emitter can be divided into two types: solid-state and liquid-state spinning. Solid-state spinning is of assembling as-synthesized CNTs, without wet process, into a yarn or fiber through twisting or drawing process, whereas liquid-state spinning is fabricating CNT fiber

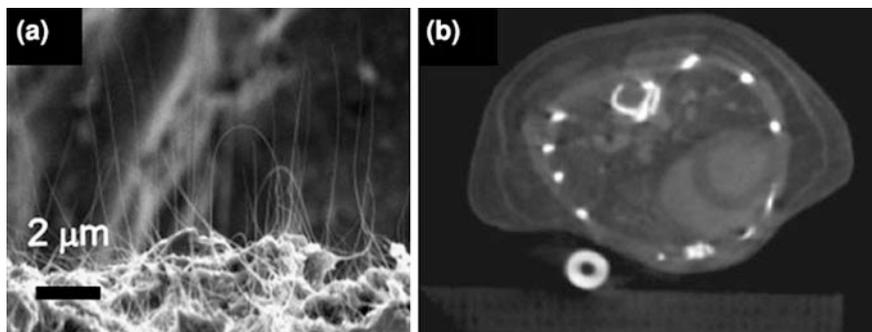


Fig. 6 **a** Cross-sectional SEM image of the CNT cathode after the activation process. **b** Micro-computed tomography (micro-CT) image of anesthetized mouse using the CNT based micro-focused X-ray source [34]

from dispersive solution based on various phenomena, such as coagulation, capillary condensation, gelation, and so forth.

(1) Solid-state spinning

Jiang et al. [36] reported, for the first time, spinning of the continuous CNT yarns through directly drawing CNTs from VA-CNT structure, which is similar process of silk production. Figure 7b shows a 100- μm -high, freestanding CNT array held by adhesive tape. The indentation at the top of the array marks the region that is being turned into a yarn 30 cm long and 200 μm wide. Continuous yarns can be drawn out from super-aligned CNT arrays in which the CNTs are aligned parallel to one another and are held together by van der Waals interactions to form bundles (Fig. 7c). The yarns usually appeared as thin ribbons composed of parallel threads that have diameters in the range of several hundreds of nanometers (Fig. 7d).

Zhang et al. [37] improved this technique by fabricating MWNT yarns by twisting nanotube sheet, which is also of highly aligned and interconnected CNT networks. SEM image of Fig. 8 shows the MWNT yarn assembly during spinning process, in which MWNTs ~ 10 nm in diameter were simultaneously drawn from the MWNT forest and twisted. The direction of drawing was orthogonal to the original nanotube direction and parallel to the substrate. The twisting process improves mechanical strength due to mechanical coupling between CNTs.

Using these CNT assembled structures, highly effective field emitters have been demonstrated. Wei et al. reported an efficient method to fabricate field emitter by cutting a continuous MWNT yarn into segments [38–40]. The cross section of each segment was composed of open-ended MWNTs which serve as field emitters. The emission current can reach several milli-amperes, and the emitters can work stably for a long time. Chen et al. fabricated point emitters using a MWNT yarn which was treated by ethylene glycol. The point emitter showed a very high emission current of 3.01 mA and good emission stability of over 20 h. Large field

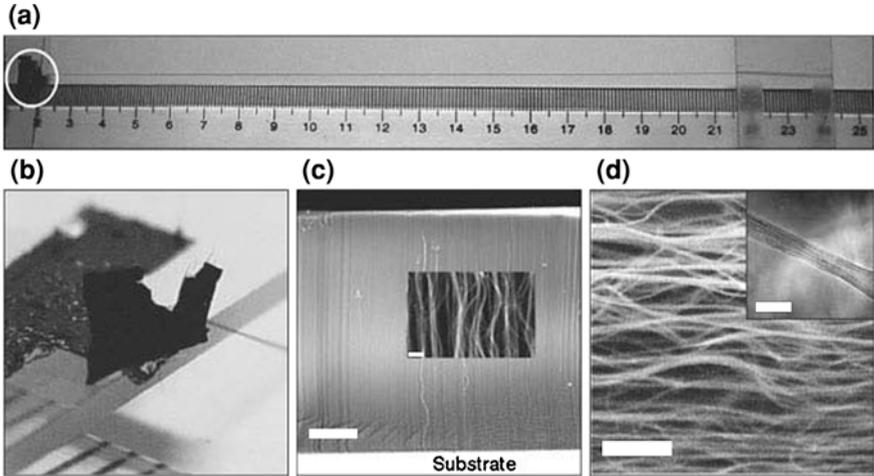
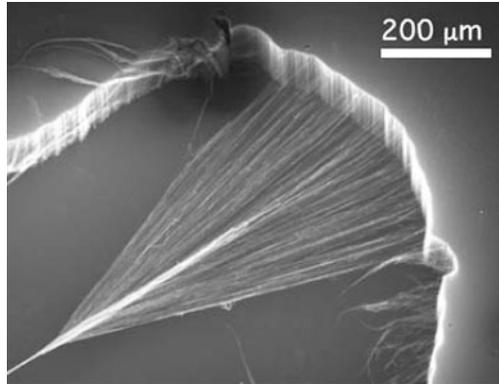


Fig. 7 Carbon nanotube yarns. **a** A carbon nanotube yarn being continuously pulled out from a free-standing carbon nanotube array, which is shown enlarged in **(b)** (roughly $\times 8$ magnification). **c** SEM images of a carbon nanotube array grown on a silicon substrate, showing the super-alignment of carbon nanotubes (scale bars: $100\ \mu\text{m}$; *inset*, $200\ \text{nm}$). **d** SEM image of the yarn and the *inset* show TEM image of a single thread of the yarn (scale bars: $500\ \text{nm}$; *inset*, $100\ \text{nm}$) [36]

Fig. 8 SEM images of a carbon nanotube yarn in the process of being simultaneously drawn and twisted during spinning from a nanotube forest outside the SEM [37]



enhancement factor caused by the large aspect ratio of sharp tip, and tight bonding of neighboring MWCNTs via the ethylene glycol treatment are attributed to these excellent field emission properties [38].

Despite of these achievements, solid-state spinning CNT fibers have limited applicability due to its cumbersome processes to construct a field emission device. The spun fibers should be manually attached to the substrate that is usually coated with adhesive material by trial and error to fabricate point emitter. Moreover it is quite difficult to attach a CNT yarn perpendicular to a substrate. These processes might also lead to CNT point emitter that is poor in electrical contact and weak in the mechanical strength between substrate and CNT yarn.

(2) Liquid-state spinning

Liquid-state spinning has an advantage to produce CNT fibers with high purity because it allows selection and purification process of CNTs. Liquid-state spun CNT point emitter have high alignment as well as high packing density due to capillary flow and condensation effects. In addition, liquid-state spinning can employ electroplating technique to tightly bond between CNTs and substrate.

Vigolo et al. [41] reported, for the first time, macroscopic SWNT fibers through a coagulation spinning process. A surfactant-stabilized SWNT solution is injected into a rotating polyvinyl alcohol (PVA)/water, as a coagulant, bath. The PVA displaces the surfactant and induces flocculation of the SWNTs into a long ribbon structure where SWNTs are aligned along the flow direction of the coagulant. The prepared ribbon structures are rinsed with water in order to remove the PVA polymer, and then dried. The ribbon is transformed into fibrous structure due to capillary condensation during dry process. Kozlov et al. [42] fabricated polymer-free SWNT fiber through either acid- or base-coagulation spinning. When surfactant-stabilized aqueous SWNT dispersion is injected through a syringe tube into a rotating acid or base-filled bath, the SWNTs flocculate spontaneously into a fibrous structure. Although the resulting fibers have electrically conductive rather than PVA-coagulated fibers, their mechanical properties such as tensile strength, modulus and toughness) are weak. Fuming sulfuric acid enables to form liquid-crystalline phase of SWNTs at a high concentration [43]. The anions of the sulfuric acid surround individual SWNTs, which promotes their ordering into a well-aligned phase. The prepared SWNT dispersion is extruded by conventional liquid-state spinning technique into continuous SWNT fibers. The ordered SWNT dispersion renders high alignment of SWNT fiber during the extrusion process.

Jang et al. [44] fabricated pure macroscopic SWNT fibers by using dip-coating method without any additive or additional electrical equipment or complex apparatuses. A sharp tip is immersed into the SWNT colloidal solution (upper inset in Fig. 9), and then withdrawn out of the solution at a constant velocity. After the tip of the wire is raised above the surface of the solution, the SWNT fibers with the diameter of about 10–20 μm (lower inset in Fig. 9) are continuously formed. This method only utilizes micro-fluidics including capillary condensation, capillary flow and surface tension, which result in self-assembly and self-alignment of SWNT colloids. Since the prepared fiber is made of pure SWNT only, it is flexible to modify the properties of the fiber by using a post-treatment for various applications.

Liquid-state spinning process can make CNT ceramic composite fiber. Kim et al. [45] reported macroscopic one-dimensional WO_3/CNT composite fibers made by liquid-state spinning and electroplating technique. A sharpened tungsten tip is immersed into the electrolyte solution (Fig. 10a), which is prepared by dissolving CNTs and sodium tungstate dihydrate (STD) in dimethylformamide (DMF) solvent. Upon applying bias to the tungsten anode with the container bottom grounded, negatively charged CNTs that are acid-treated gather around the anode along with the WO_4^{2-} ions. After electroplating process, the tip is pulled up

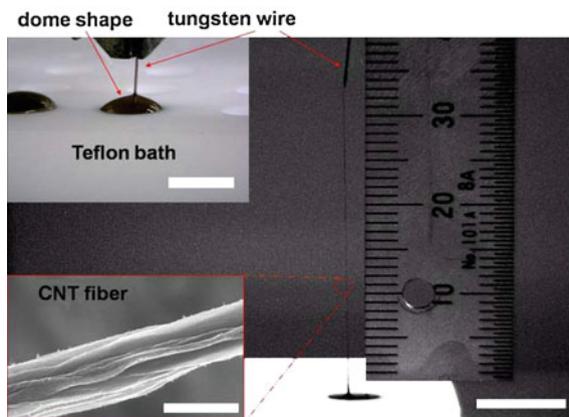


Fig. 9 Optical image showing the tungsten wire immersed into the SWNT colloidal solution forming a dome shape (*upper inset*; scale bar = 5 mm) and the SWNT fiber being formed continuously from the solution during the withdrawal process (*main panel*; scale bar = 10 mm). Scanning electron micrograph of *lower inset* showing the magnified SWNT fiber in the *dotted circle* (scale bar = 10 μm) [44]

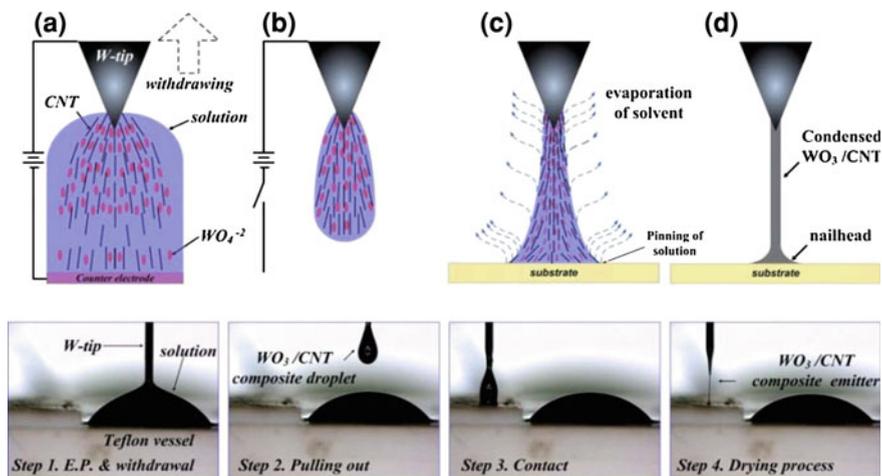
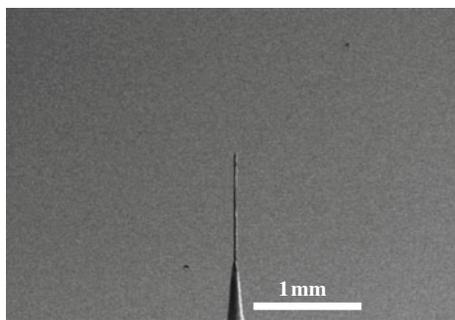


Fig. 10 Schematic of the procedure for fabricating one-dimensional WO_3/CNT composite structure by crystal-like growth. **a** Immersing tungsten W-tip into stable CNT suspension and electrolyte in DMF and beginning electroplating with bias applied between W-tip and counter electrode. **b** Pulling up WO_3/CNT composite from the solution at a faster rate to finish the growth. **c** Contacting WO_3/CNT composite on Teflon substrate to avoid curling effect. **d** Drying of solvent for condensed WO_3/CNT composite, resulting in a vertically aligned composite emitter. The *bottom row* frames show the digital snapshots of each step of the procedure [45]

Fig. 11 SEM image of vertically aligned WO_3/CNT composite emitter on substrate (tungsten tip) [45]



at a constant speed with a column of the solution of CNTs and WO_4^{2-} clinging to the tungsten tip. After the desired length of the macroscopic one-dimensional WO_3/CNT composite is reached, the electrode is quickly pulled up to complete the electroplating, as shown in Fig. 10b. The tip of the withdrawn WO_3/CNT column is carefully placed on a Teflon plate and is made to touch its surface (Fig. 10c). As shown in Fig. 10d, the solvent in the exposed column of solution evaporates, leading eventually to the formation of a condensed macroscopic one-dimensional CNT structure having a diameter of about 20 μm . In this process, the contact that mechanically constrains free end of the column is a key to obtain vertically aligned composite emitter along the axis of the tungsten tip.

However, undesirable nail head shape at the apex of the composite emitter is formed after drying process. The nail head is not desirable for high current field emission. The nail head was removed by an electrical discharge machining (EDM) as shown in Fig. 11. Moreover, the EDM treatment could dramatically reduce the series resistance between W-tip and the emitter via high level of Joule heating caused by the arc discharge.

The field emission current of the composite emitter reached the maximum level of 10 mA (Fig. 12a) and showed high stability of field emission at high current as shown in Fig. 12b. The emission current density of 5 mA corresponds to approximately 1,590 A/cm^2 since the emitter diameter is about 20 μm . The incredibly high current level of better than 10 mA of the emitter could be attributed to uniform geometric length of individual CNTs in the composite, large aspect ratio, high conductivity of the composite and low contact resistance between substrate and CNTs emitter.

5 Summary and Outlooks

In spite of planar CNT emitters fabricated by screen printing and chemical vapor deposition have attracted considerable attentions for field emission display in earlier research, the point emitters are generally preferred for applications in THz devices. Because focusing spot size for electron emission and field emission

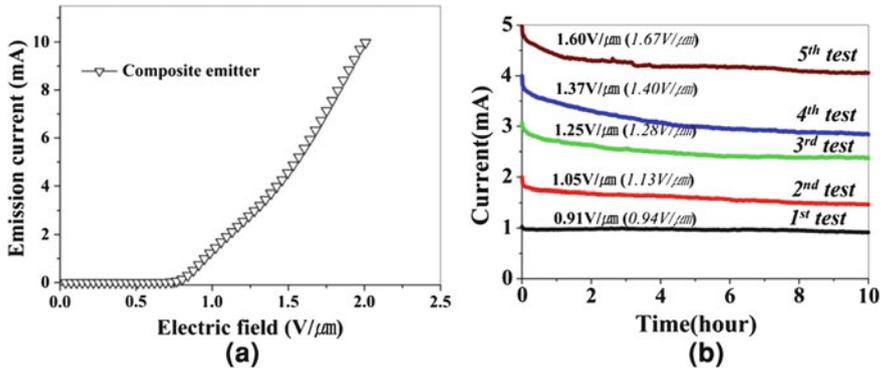


Fig. 12 a I–V plots and (b) field emission stability of the composite emitter. Boldface numbers and the numbers in the *round bracket* indicate applied electric field during stability test and the electric field which is required to obtain initial current level after the stability test [45]

properties are better than those of planar emitters, based on geometrically high aspect ratio of the point emitter. Moreover, the point emitter could be a highly effective field emission source because the screen effect does not significantly appear in the point emitter configuration. The macroscopic CNT point emitter was demonstrated by solid-state spinning technique using MWNTs; however this method requires an adhesive and manual manipulation. These cumbersome processes may lead to CNT point emitter that is poor in electrical contact and weak in the mechanical strength between substrate and CNT fiber so that result in a low emitter performance and emitter stability with time. The liquid-state spun point emitter has various advantages, such as securing high purity, packing density and length directional alignment of CNTs at the emitter. The CNT composite point emitter fabricated by this liquid-state approach shows the best field emission properties in terms of current density and total current up to now.

Most of the published CNT field emission research so far has focused on field emission display. The performance of the CNT cathodes under high emission current and high voltage has not been adequately exploited except a few published studies. Even though the recent outstanding researches provide possible routes to develop a high power electron source, they are still less satisfactory for the use of cathode utilized in THz applications. The cutting-edge field emission performance based on CNT materials are summarized in Table 1. Future researches on field emitter will focus on developing novel field emission cathode that are capable of high current density as well as high total current. Both field emission characteristics are prerequisites for THz devices. Visible goal of these researches will be a current density of 100 A/cm² with a total current of several tens of milli-ampere. To accomplish this work, it might be necessary to overcome non-uniform field emission on a single point emitter tip, and bunching of these emitters could be considered to achieve such a high total current.

Table 1 The cutting-edge field emission emitter based on CNT materials

References	Current density and total current	Operating voltage	Focal spot diameter	Stability	Remarks
[46]	1.5 μA	10 kV	–	1.5 μA for 1.3 h	DC mode Triode
[47]	140 mA/cm ² (28 mA peak emission, cathode area of 0.2 cm ²)	~60 kV	3.2 mm	6 mA (30 mA/cm ²) for 24 h	Pulse mode Triode
[48]	50 μA	21 kV	10 μm	–	DC mode Diode
[49]	760 mA/cm ² (6 mA, cathode area of 0.007 cm ²)	60 kV	150 \times 30 μm (ellipse)	–	Pulse mode Triode
[50]	1 A/cm ² (10 mA, cathode area of 0.007 cm ²)	40 kV	50 μm	200 μA (12 h)	DC mode Triode
[51]	100 mA/cm ²	40 kV	35 μm	300 μA (16 h)	DC mode Triode
[52]	165 mA/cm ² (1.3 mA, cathode area of 0.007 cm ²)	7 kV	30 μm	Degradation From 700 to 400 μA (24 h)	DC mode Triode
[53]	3.18 mA/cm ²	25–40 kV	80 μm	100 μA (24 h)	DC mode Triode
[54]	10 mA/cm ²	40 kV	5 μm	–	DC mode Triode
[45]	3,000 A/cm ² (10 mA, cathode area of 3.14 \times 10 ⁻⁶ cm ²)	3–4 kV	–	From 5 mA (1,590 A/cm ²) to 4 mA (1,273 A/cm ²) for 10 h, respectively	DC mode

References

1. Fowler, R.H., Nordheim, L.: Electron emission in intense electric fields. *Proc. R. Soc. Lond. A* **119**(781), 173–181 (1928)
2. Iijima, S.: Helical microtubules of graphitic carbon. *Nature* **354**(6348), 56–58 (1991)
3. Hong, S., Myung, S.: Nanotube electronics: a flexible approach to mobility. *Nat. Nano.* **2**(4), 207–208 (2007)
4. Frank, S., et al.: Carbon nanotube quantum resistors. *Science* **280**(5370), 1744–1746 (1998)
5. Lau, A.K.-T., Hui, D.: The revolutionary creation of new advanced materials—carbon nanotube composites. *Comp. Pt. B Eng.* **33**(4), 263–277 (2002)
6. Hone, J., et al.: Thermal conductivity of single-walled carbon nanotubes. *Phys. Rev. B* **59**(4), R2514 (1999)
7. Che, J., et al.: Thermal conductivity of carbon nanotubes. *Nanotechnology* **11**(2), 65 (2000)
8. Berber, S., Kwon, Y.-K., Tománek, D.: Unusually high thermal conductivity of carbon nanotubes. *Phys. Rev. Lett.* **84**(20), 4613 (2000)
9. Yu, M.-F., et al.: Strength and breaking mechanism of multiwalled carbon nanotubes under tensile load. *Science* **287**(5453), 637–640 (2000)
10. Peng, B., et al.: Measurements of near-ultimate strength for multiwalled carbon nanotubes and irradiation-induced crosslinking improvements. *Nat. Nano.* **3**(10), 626–631 (2008)
11. Minus, M., Kumar, S.: The processing, properties, and structure of carbon fibers. *JOM J. Min. Metals Mat. Soc.* **57**(2), 52–58 (2005)
12. Vaisman, L., Wagner, H.D., Marom, G.: The role of surfactants in dispersion of carbon nanotubes. *Adv. Colloid Interface Sci.* **128–130**, 37–46 (2006)
13. Kang, Y., Taton, T.A.: Micelle-encapsulated carbon nanotubes: a route to nanotube composites. *J. Am. Chem. Soc.* **125**(19), 5650–5651 (2003)
14. Ros, T.G., et al.: Surface oxidation of carbon nanofibres. *Chem. Eur. J.* **8**(5), 1151–1162 (2002)
15. Baughman, R.H., Zakhidov, A.A., de Heer, W.A.: Carbon nanotubes—the route toward applications. *Science* **297**(5582), 787–792 (2002)
16. Andrews, R., Weisenberger, M.C.: Carbon nanotube polymer composites. *Curr. Opin. Solid State Mater. Sci.* **8**(1), 31–37 (2004)
17. Tans, S.J., Verschueren, A.R.M., Dekker, C.: Room-temperature transistor based on a single carbon nanotube. *Nature* **393**(6680), 49–52 (1998)
18. Li, Z., et al.: Microstructure of carbon nanotubes/PET conductive composites fibers and their properties. *Compos. Sci. Technol.* **66**(7–8), 1022–1029 (2006)
19. Xu, J., Fisher, T.S.: Enhancement of thermal interface materials with carbon nanotube arrays. *Int. J. Heat Mass Transf.* **49**(9–10), 1658–1666 (2006)
20. Artukovic, E., et al.: Transparent and flexible carbon nanotube transistors. *Nano Lett.* **5**(4), 757–760 (2005)
21. Hu, R., et al.: Harvesting waste thermal energy using a carbon-nanotube-based thermo-electrochemical cell. *Nano Lett.* **10**(3), 838–846 (2010)
22. de Heer, W.A., Châtelain, A., Ugarte, D.: A carbon nanotube field-emission electron source. *Science* **270**(5239), 1179–1180 (1995)
23. Choi, W., et al.: Fully sealed, high-brightness carbon-nanotube field-emission display. *Appl. Phys. Lett.* **75**(20), 3129 (1999)
24. Zhang, J., et al.: Efficient fabrication of carbon nanotube point electron sources by dielectrophoresis. *Adv. Mater.* **16**(14), 1219 (2004)
25. Wang, Z.L., et al.: In situ imaging of field emission from individual carbon nanotubes and their structural damage. *Appl. Phys. Lett.* **80**(5), 856 (2002)
26. Wei, Y., et al.: Stability of carbon nanotubes under electric field studied by scanning electron microscopy. *Appl. Phys. Lett.* **79**(27), 4527 (2001)
27. Li, W.Z., et al.: Large-scale synthesis of aligned carbon nanotubes. *Science* **274**(5293), 1701–1703 (1996)

28. Wei, B.Q., et al.: Microfabrication technology: organized assembly of carbon nanotubes. *Nature* **416**(6880), 495–496 (2002)
29. Hata, K., et al.: Water-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes. *Science* **306**(5700), 1362–1364 (2004)
30. Thong, J.: High-current field emission from a vertically aligned carbon nanotube field emitter array. *Appl. Phys. Lett.* **79**(17), 2811 (2001)
31. Fan, S., et al.: Self-oriented regular arrays of carbon nanotubes and their field emission properties. *Science* **283**(5401), 512–514 (1999)
32. Ando, Y., Iijima, S.: Preparation of carbon nanotubes by arc-discharge evaporation. *Jpn. J. Appl. Phys. Pt. 2 Lett.* **32**(1A–B), L107–L109 (1993)
33. Li, J., et al.: Field emission characteristic of screen-printed carbon nanotube cathode. *Appl. Surf. Sci.* **220**(1–4), 96–104 (2003)
34. Xiomara, C.-C., et al.: A carbon nanotube field emission cathode with high current density and long-term stability. *Nanotechnology* **20**(32), 325707 (2009)
35. Jo, S.: Effect of length and spacing of vertically aligned carbon nanotubes on field emission properties. *Appl. Phys. Lett.* **82**(20), 3520 (2003)
36. Jiang, K., Li, Q., Fan, S.: Spinning continuous carbon nanotube yarns. *Nature* **419**(6909), 801 (2002)
37. Zhang, M., Atkinson, K.R., Baughman, R.H.: Multifunctional carbon nanotube yarns by downsizing an ancient technology. *Science* **306**(5700), 1358–1361 (2004)
38. Lee, C.J., et al.: Field emission characteristics of point emitters fabricated by a multiwalled carbon nanotube yarn. *Nanotechnology* **20**(31), 315201 (2009)
39. Wei, Y.: Efficient fabrication of field electron emitters from the multiwalled carbon nanotube yarns. *Appl. Phys. Lett.* **89**(6), 063101 (2006)
40. Zakhidov, A.A., et al.: Field emission of electrons by carbon nanotube twist-yarns. *Appl. Phys. A Mater. Sci. Proc.* **88**(4), 593–600 (2007)
41. Vigolo, B., et al.: Macroscopic fibers and ribbons of oriented carbon nanotubes. *Science* **290**(5495), 1331–1334 (2000)
42. Kozlov, M.E., et al.: Spinning solid and hollow polymer-free carbon nanotube fibers. *Adv. Mater.* **17**(5), 614–617 (2005)
43. Ericson, L.M., et al.: Macroscopic, neat, single-walled carbon nanotube fibers. *Science* **305**(5689), 1447–1450 (2004)
44. Jang, E.Y., et al.: Macroscopic single-walled-carbon-nanotube fiber self-assembled by dip-coating method. *Adv. Mater.* **21**(43), 4357–4361 (2009)
45. Kim, W.J., et al.: Better than 10 mA field emission from an isolated structure emitter of a metal oxide/CNT composite. *ACS Nano* **5**(1), 429–435 (2010)
46. Sugie, H., et al.: Carbon nanotubes as electron source in an x-ray tube. *Appl. Phys. Lett.* **78**(17), 2578–2580 (2001)
47. Yue, G.Z., et al.: Generation of continuous and pulsed diagnostic imaging x-ray radiation using a carbon-nanotube-based field-emission cathode. *Appl. Phys. Lett.* **81**(2), 355–357 (2002)
48. Matsumoto, T., Mimura, H.: Point x-ray source using graphite nanofibers and its application to x-ray radiography. *Appl. Phys. Lett.* **82**(10), 1637–1639 (2003)
49. Cheng, Y., et al.: Dynamic radiography using a carbon-nanotube-based field-emission x-ray source. *Rev. Sci. Instrum.* **75**(10), 3264–3267 (2004)
50. Zhang, J., et al.: A nanotube-based field emission x-ray source for microcomputed tomography. *Rev. Sci. Instrum.* **76**(9), 094301–094304 (2005)
51. Liu, Z., et al.: Carbon nanotube based microfocus field emission x-ray source for microcomputed tomography. *Appl. Phys. Lett.* **89**(10), 103111-3 (2006)
52. Kawakita, K., et al.: Development of microfocused x-ray source by using carbon nanotube field emitter. *Rev. Sci. Instrum.* **77**(5), 054302-6 (2006)

53. Tolt, Z.L., et al.: Carbon nanotube cold cathodes for application in low current x-ray tubes. *J. Vac. Sci. Technol. B* **26**(2), 706–710 (2008)
54. Heo, S.H., Ihsan, A., Cho, S.O.: Transmission-type microfocus x-ray tube using carbon nanotube field emitters. *Appl. Phys. Lett.* **90**(18), 183109-3 (2007)

THz Generation and Propagation Using Femtosecond Laser

Tae-In Jeon

There are several methods to generate and detect THz electromagnetic radiation whose frequency lies between the microwave and infrared regions of the spectrum. For example photoconductive switching, optical rectification, photomixing, quantum cascade lasers, and free electron laser are widely used methods that can generate THz beams. Meanwhile, photoconductive switching (antenna), electro-optics sampling, and bolometers are widely used detection meters. Recently air-breakdown-coherent THz generation and detection have been developed [1, 2]. In this section, I will explain commonly used generation methods, such as photoconductive switching and optical rectification. Also, the photoconductive switching (antenna) and electro-optics sampling methods will be explained because they are commonly used to detect THz beams.

1 THz Generation

1.1 Photoconductive Antenna

When a femtosecond laser pulse illuminates semiconductors, such as Si or GaAs, and its energy ($E = hf$ or $= hc/\lambda$) is bigger than that of the band-gap of the semiconductors, photo induced carriers, such as like electrons and holes, are generated. For example, if the Ti:sapphire laser beam, whose wavelength is 800 nm with 100 femtoseconds pulse width, the energy of a photon of the laser beam can be calculated by $E = (6.626 \times 10^{-34} \text{ Js}) \times (3 \times 10^8 \text{ m/s}) / (800 \times 10^{-9} \text{ m}) = 2.48 \times 10^{-19} \text{ J}$. The unit of [J] can be converted to the unit of [eV] by

T.-I. Jeon (✉)
Division of Electrical and Electronics Engineering,
Korea Maritime University, Busan, South Korea
e-mail: jeon@hhu.ac.kr

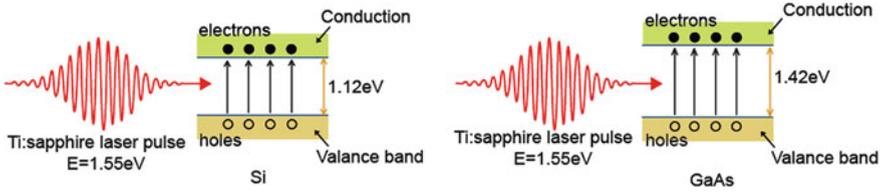


Fig. 1 The absorption of a laser energy results in the generation of an electron-hole pair

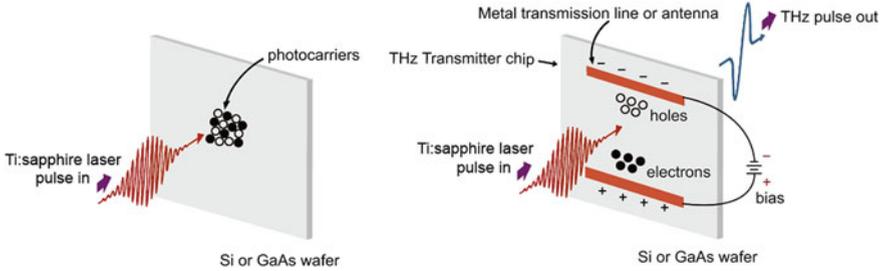


Fig. 2 a Generated photocarriers by femtosecond laser pulse. **b** Radiated THz pulse by the photocurrent. The electrons and holes are moved to positive and negative charged transmission lines respectively

dividing the electronic charge by 1.6×10^{-19} [C]. The energy of a photon of the laser beam is 1.55 [eV], whereas the band gaps of Si and GaAs are 1.12 and 1.42 [eV], respectively. Therefore, the semiconductor materials absorb the photo energy and create carriers, as shown in Fig. 1.

The generated carriers are randomly distributed on the semiconductor surface. If an electric field, which comes from the DC bias connected to two metal transmission lines (or metal dipole antenna), is applied to the carriers, the electrons move to a positive transmission line and the holes move to a negative transmission line. When the carriers move in one direction, the photocurrent, $J(t)$, is generated. The photocurrent can be expressed as

$$J(t) = N(t)e\mu E, \tag{1}$$

where N denotes the density of the photocarriers, which depends on the laser pulse shape and the carrier lifetime, e is the fundamental charge, μ is the mobility of electron, and E is the bias electric field. Figure 2 shows the schematic view of generated photocarriers by femtosecond laser pulse and the radiated THz pulse by photocurrent.

The THz radiation amplitude is proportional to the time derivative of the transient photocurrent as $E_{THz}(t) \propto \frac{\partial J(t)}{\partial t}$. Since only the N parameter depends on time for Eq. (1), the THz radiation amplitude can be expressed as $E_{THz}(t) \propto \frac{\partial N(t)}{\partial t}$.

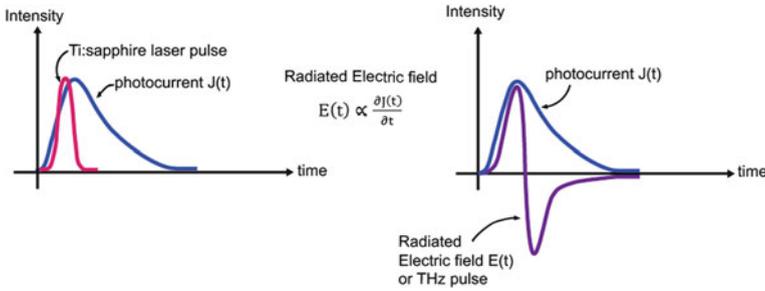


Fig. 3 Temporal intensity of the laser pulse, photocurrent, and radiated THz pulse

Table 1 Characteristics of commonly used semiconductors for the THz transmitter or receiver chips

Photoconductive materials	Carrier lifetime [ps]	Electron mobility [cm ² /(V · S)]	Band gap [eV]
LT-GaAs	0.3	150–200	1.42
GaAs	100	8,500	1.42
Si	0.8–20.0	1	1.12
Ion-implanted SOS	0.6	30	1.12

Figure 3 shows the amplitude of the laser pulse, photocurrent, and radiated THz pulse [3].

Although the width of the Ti:sapphire laser pulse is very short (approximately less than 100 fs), the induced photocurrent can be extended by a few picoseconds. In order to generate a very short THz pulse, an ultra-short laser pulse and very short carrier lifetime are required. Also, in order to generate a strong THz pulse (big THz amplitude), a high laser beam power, high mobility, and high bias are needed. Because the two metal transmission lines are very closely located (usually 5 microns for the dipole antenna and few tens micron for the transmission line), they can be easily shortened for the high laser beam power and high bias. Therefore, a high mobility of semiconductors is an important parameter to make the THz transmitter chip avoid short circuits. Table 1 shows the characteristics of commonly used semiconductors for the THz transmitter or receiver chips [4].

1.2 Optical Rectification

When ultra-short laser pulses are focused on second-order nonlinear materials, such as ZnTe, GaAs, and InP, a frequency dependent polarization $P(\omega)$ is created by using the interaction of two optical photons at different frequencies (ω_1 and ω_2) of the ultra-short laser pulse. Therefore, the polarization $P(t)$ can be written as a power series of the electric field (ultra-short laser beam).

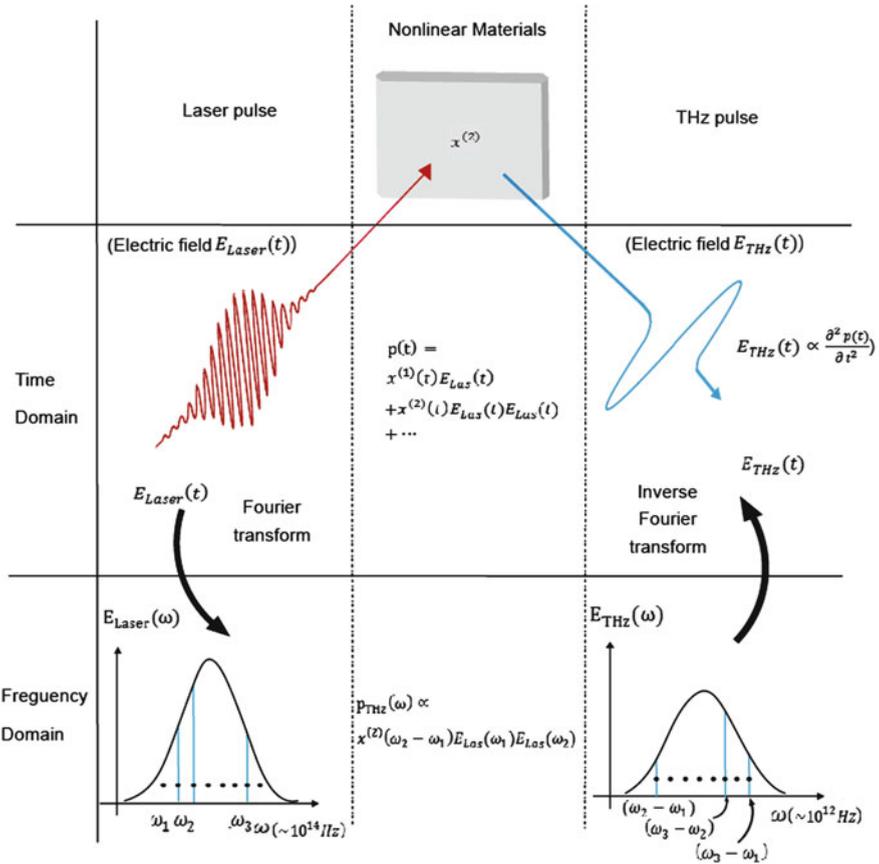


Fig. 4 The relationship between laser pulse and THz pulse by optical rectification (modified Ref. [5])

$$P(t) = \chi^{(1)}(t)E(t) + \chi^{(2)}(t)E(t)E(t) + \chi^{(3)}(t)E(t)E(t)E(t) + \dots \quad (2)$$

The frequency dependent polarization $P(\omega)$ is converted from $P(t)$ by Fourier transformation. Finally, the THz radiation amplitude is proportional to the second order time derivative of the $P(t)$ as

$$E_{THz}(t) = \frac{\partial^2 P(t)}{\partial t^2}. \quad (3)$$

The spectrum of the short pulse has a very broad bandwidth, whereas the spectrum of the broad pulse has a very narrow bandwidth. Therefore, a femto-second laser pulse has many frequency components (very broad spectrum bandwidth). Any different frequency components make a new THz frequency component of the THz pulse. For example, the frequencies of the laser pulse, such as $\omega_1, \omega_2,$ and $\omega_3,$ make the frequencies of the THz pulse become $\omega_2 - \omega_1, \omega_3 - \omega_2,$

and $\omega_1 - \omega_3$ as shown in Fig. 4. Because of the very broad spectrum bandwidth of the ultrafast laser pulse, the THz spectrum bandwidth generated is mathematically enough to convert the THz pulse in the time domain to an inverse Fourier transformation.

Meanwhile, similar to optical rectification, the photo-Dember effect is another way to generate the THz pulse without bias on the semiconductor surface. The generated photocarriers due to the absorption of laser beam can be diffused toward the semiconductor. The diffusion of the electrons and holes depends on the mobility of the carriers. Since electrons have higher mobility than the holes, the electrons go farther than the holes. The brief separation between the electrons and holes in semiconductors induces a transient photo-Dember field that generates electromagnetic radiation that has a THz frequency component.

2 THz Detection

2.1 Photoconductive Switching (Antenna)

As explained by the energy relationships between the femtosecond laser pulse and the band gap of the semiconductor material, if the energy of the Ti:sapphire laser pulses are bigger than that of Si or GaAs, the photocarriers are generated on the semiconductor surface. Whenever the laser pulses are exposed at the gap of the dipole antenna, which exists on the semiconductor surface, the generated photocarriers connect the dipole antenna (the photo conductive switch transitions to on state). The ammeter can detect the induced current when the incoming THz electric field from the back side of the receiver chip is applied. However, the carrier lifetime is very short because of the recombination between the electrons and holes or the trapping of the carriers in the semiconductor surface. If the carrier lifetime is much shorter than the THz pulse, the photo conductive dipole antenna detects only a point of the THz electric field. Therefore, the short carrier lifetime is a very important parameter to have a good THz detector. Table 1 also shows the carrier lifetime for several semiconductor materials.

Ion-implantation is a very useful technique to reduce the carrier lifetime by the carrier trapping in the defected surface. After the carrier disappears by the recombination or the trapping, the photo conductive switch transitions to an off state. Whenever the laser pulses arrive at the dipole antenna gap, the photoconductive switch is repeated on and off. Figure 5 shows the working principle of the dipole antenna both with and without laser pulse when the THz pulse enters the antenna. The right side of Fig. 5 shows the equivalence circuit. The generated current by the dipole antenna denotes a current source.

Because one laser pulse can detect the DC current at a moment of the THz electric field (THz pulse), the laser pulses (or THz pulses) have to shift in order to detect another DC current at another moment of the THz electric field. In order to make the time shift, the laser beam path is controlled by a motor controlled delay line.

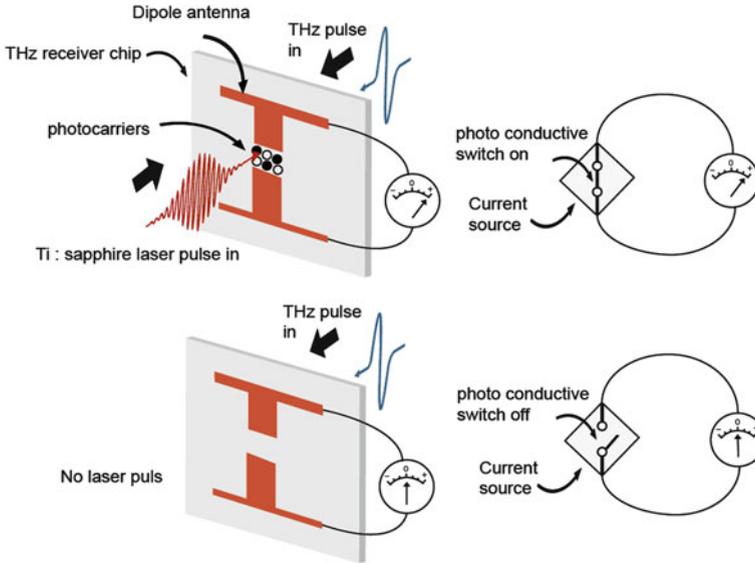


Fig. 5 Working principle of dipole antenna both with and without laser pulse. The right side shows the equivalence circuit. The generated current by the dipole antenna denotes a current source

Figures 6 and 7 show the mechanism of the THz pulse detection. The first bunch of laser pulses detects a DC current at the “A” point of the first bunch of THz pulses. After the time shift, the second bunch of laser pulses detects another DC current at the “B” point of the second bunch of THz pulses. Using this sampling method, various points, such as “C” and “D”, can be detected. This detection mechanism is called the sampling method.

For example, if a laser beam bath shifts (extends) 10 micron (5 micron round trip), the time shift (delay) is 33.33 fs ($10 \times 10^{-6}(\text{m})/3 \times 10^8(\text{m/s})$). When 1000 data are taken, the measured time duration is 33.33 ps which is enough time to measure the THz pulse (1 THz corresponds 1 ps).

2.2 Electro-Optics Sampling

An electro-optical (EO) crystal such as ZnTe, is good birefringent material for THz frequency. When the THz pulse (electric field) enters the EO crystal, the refractive indexes of the EO crystal change according to the applied THz electric field. Therefore, if the THz pulse and the laser pulse enter at the same time, the incident THz beam make a birefringence of the EO crystal, which leads to a change in the optical polarization of the laser pulse. Figure 8 shows a schematic diagram of a typical setup of EO sampling and the THz pulse detection by sampling method. The THz pulse and laser pulse are propagated in the same direction along the same axis. Usually, an undoped and polished Si plate is used to co-

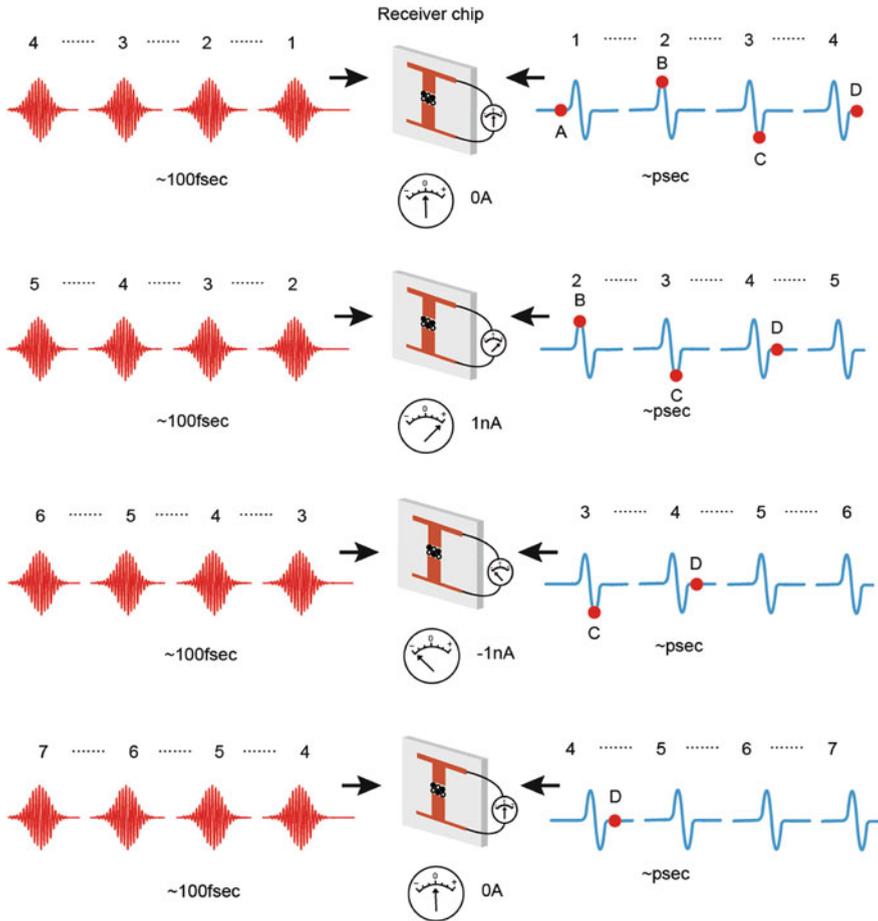


Fig. 6 The principle of THz pulse detection with photo conductive antenna by sampling method

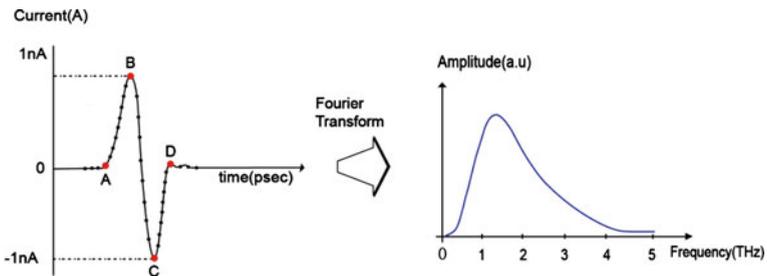


Fig. 7 The measured THz pulse by sampling method and its spectrum

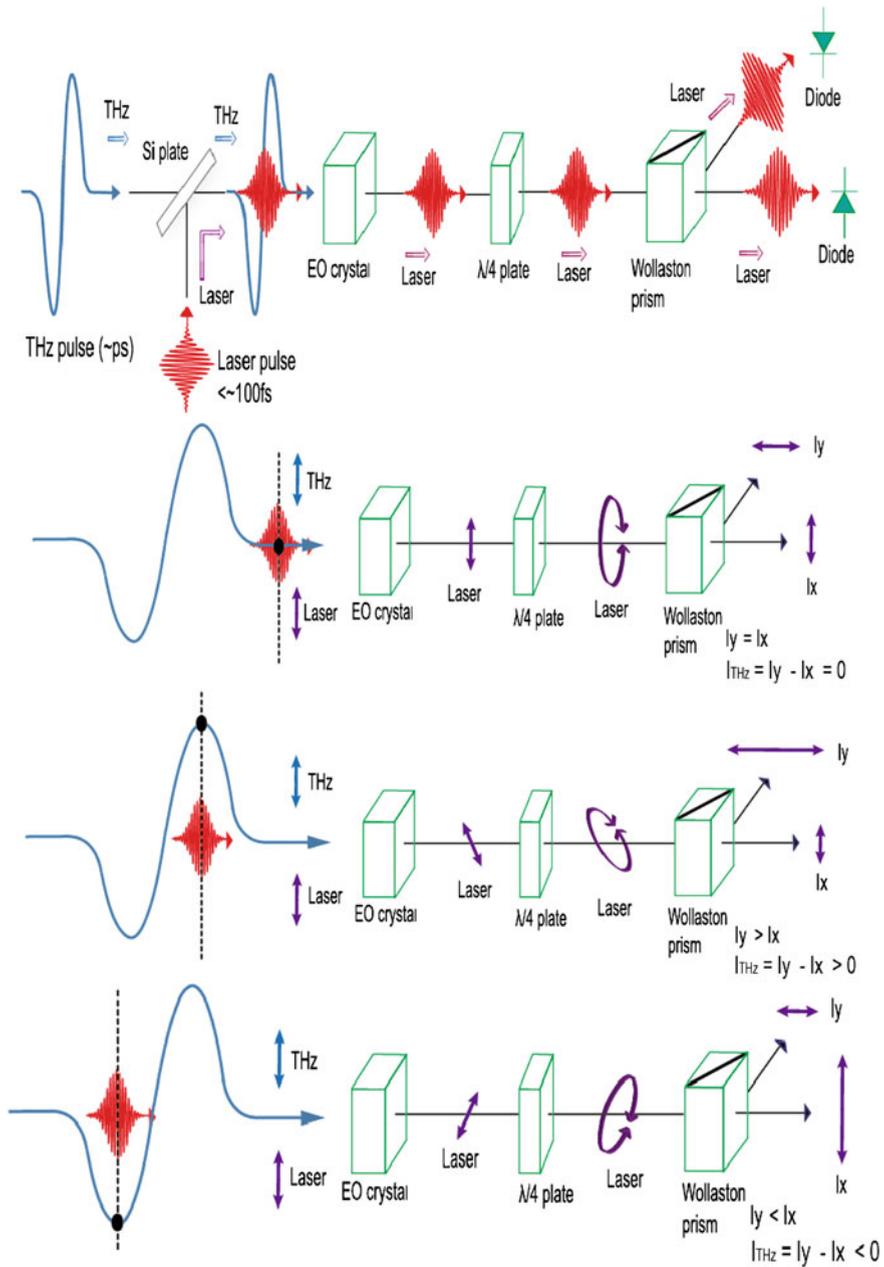


Fig. 8 The principle of THz pulse detection with electro-optics sampling rectification (modified Ref. [5])

propagate the two THz and laser pulses (beams) because the THz beam can be propagated without any propagation loss and the laser beam can be reflected with a very small reflection loss. After the EO crystal, the laser beam changes its polarization, and the polarization of the laser beam depends on the polarity of the THz pulse. The $\lambda/4$ plate makes the linearly polarized laser beam go through circular or ellipsoid polarization. A Wollaston prism separates the laser beam into two x and y components, which are sent to a balanced photodiode. The intensities of I_x and I_y depend on the polarization of the laser beam. Finally, each of the photodiodes measure the voltage signal proportional to the laser beam intensities of the I_x and I_y polarized components. Therefore, the intensity of the THz pulse (I_{THz}) depends on the difference between I_x and I_y . As explained in the sampling method, the entire data of the THz pulse can be measured by the time shifted THz pulse or laser pulse, as shown in Fig. 8.

3 THz System

There are many ways to build a transmitter and receiver for the THz system. A widely used THz system for the THz time-domain spectroscopy (THz-TDS) is the photoconductive antenna type transmitter and receiver because they have more power and greater sensitivity than any other transmitter and receiver combinations. Figure 9 shows the THz-TDS system setup with the photoconductive antenna type transmitter and receiver. Silicon lens are used to collimate the outgoing THz beam at the transmitter and recollimate the beam incident upon the receiver. Because the optoelectronic receiver is gated synchronously with the excitation of the transmitter, the two beam paths between the beamsplitter and the receiver chip should be the same length. The THz pulse is obtained by scanning the relative time delay between the laser excitation pulse and the laser detection pulse. The measurement delay line is controlled through a programmable motor controller and computer.

The laser generated burst of THz radiation is emitted from the transmitter chip in a cone normal to the interface between the laser exposed semiconductor surface and air. This burst is then collected and collimated by a crystalline silicon lens attached to the back side of the chip. The low-frequency components are more spread out than the high-frequency components, as indicated in Fig. 10. After collimation by the silicon lens, the THz beam propagates and diffracts to a parabolic mirror where the THz beam is recollimated into a highly directional beam. The polarization of the THz beam depends on the direction of the photocarrier, as shown in Fig. 5. Therefore, the dipole antenna of the receiver chip should be a line to the polarization of the THz beam, as shown in Fig. 10.

Many rotational and vibrational frequencies of water vapor lie in the THz spectra regions. With free space propagation of the THz beam, interaction with ambient water vapors in the air becomes a significant factor on the signal quality, as shown in Fig. 11. In order to eliminate the effects of vapor on the THz beam path, the THz system is located in an airtight dry box.

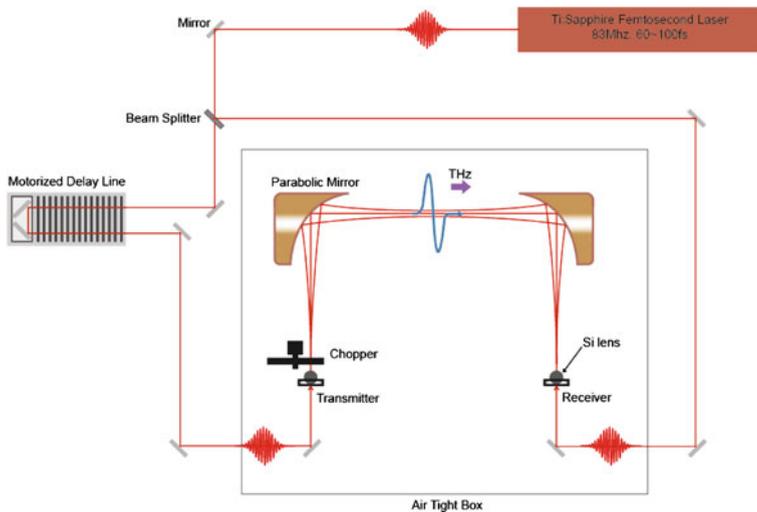


Fig. 9 THz time-domain spectroscopy system setup with the photo conductive antenna type transmitter and receiver chips

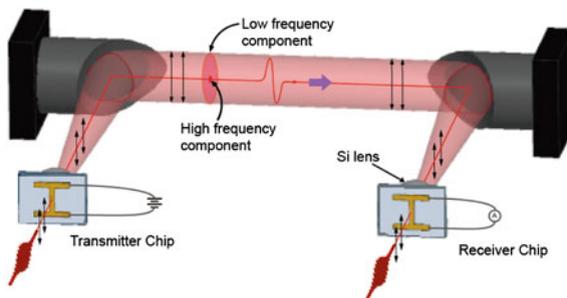


Fig. 10 The polarization (*vertical arrows*) relationship between laser beam and THz beam

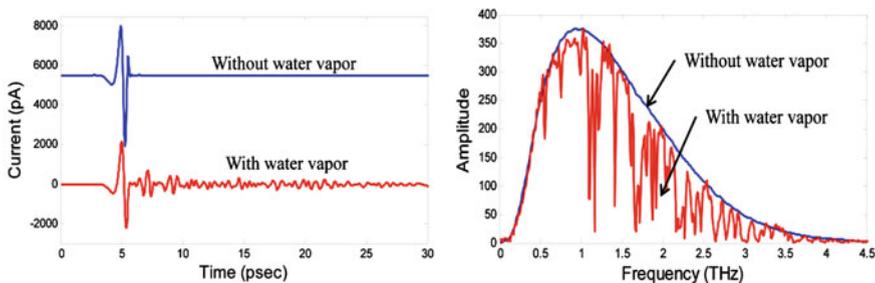


Fig. 11 Comparison of THz pulse and spectrum with and without water vapor in the THz beam path

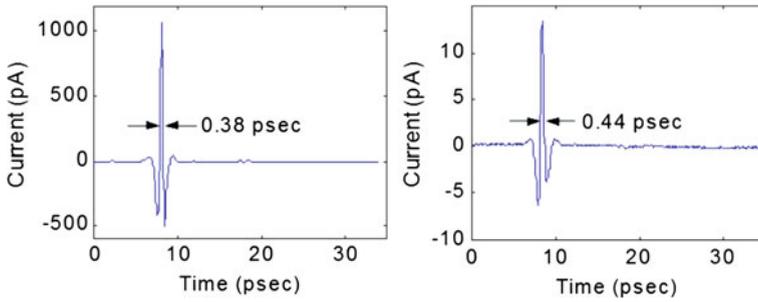


Fig. 12 The measured THz pulses without sample (*left*) and with highly doped 10- μ m thick conducting polymer sample (*right*)

Table 2 THz waveguides done by Grischkowsky group

Stainless-Steel Circular Tube [7]		Brass Rectangular Tube [8]	
High-Density Plastic Ribbon [9]		Parallel Plate Metal Waveguide [10]	
Copper Coaxial Cable [11]		Metal Wire [12]	
Metal Sheet [13]		Coated Metal Sheet [14]	

Since one of the advantages of THz-TDS is a very high signal-to-noise ratio (S/N), the THz-TDS characterizations are performed by measuring THz pulses transmitted (output) through a high conductive sample. These transmitted THz pulses are then compared to the measured THz pulses (reference) with no sample in place, as shown in Fig. 12, where the sample is a highly doped 10- μ m thick conducting polymer [6]. When the measured reference THz pulse is done by averaging seven individual measurements, the S/N of the reference THz pulse is increased up to 40,000:1. With the sample placed in the THz beam path, the measured amplitude of the transmitted pulse, which is the average of 27 individual measurements, has dropped by approximately 100. However, the S/N is still 100. The S/N is good enough to analyze the characteristics of the sample.

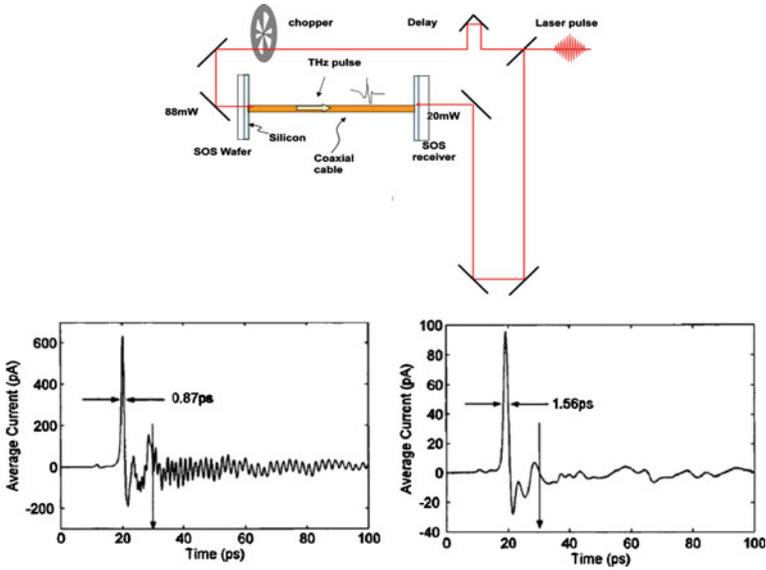


Fig. 13 The experimental setup to propagate the THz beam through coaxial cable (*upper*) and measured THz pulses of 15-mm-length (*lower left*) and 105-mm-length (*lower right*) [11]. Copyright © 2004, American Institute of Physics

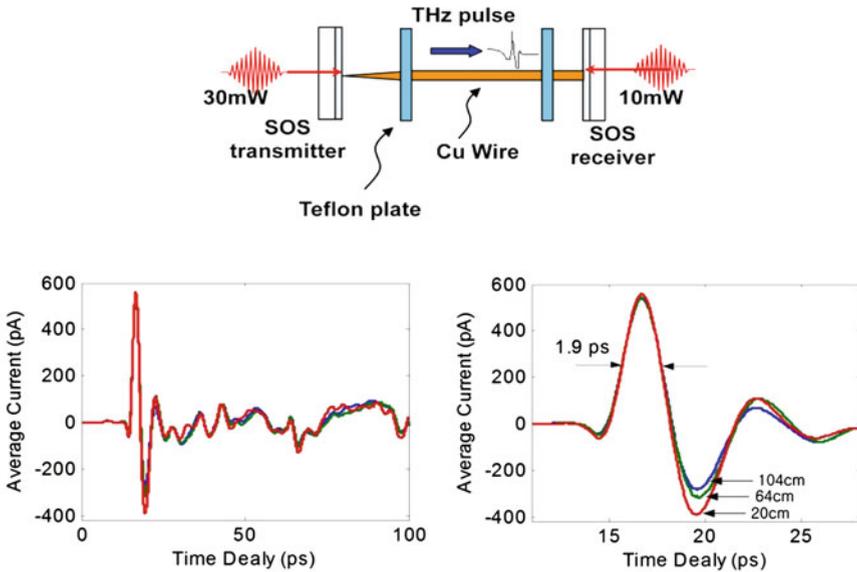


Fig. 14 The experimental setup to propagate THz beam on metal wire (*upper*) and measured THz pulses with different lengths (*lower left*) and expanded view (*lower right*) [12]. Copyright © 2005, American Institute of Physics

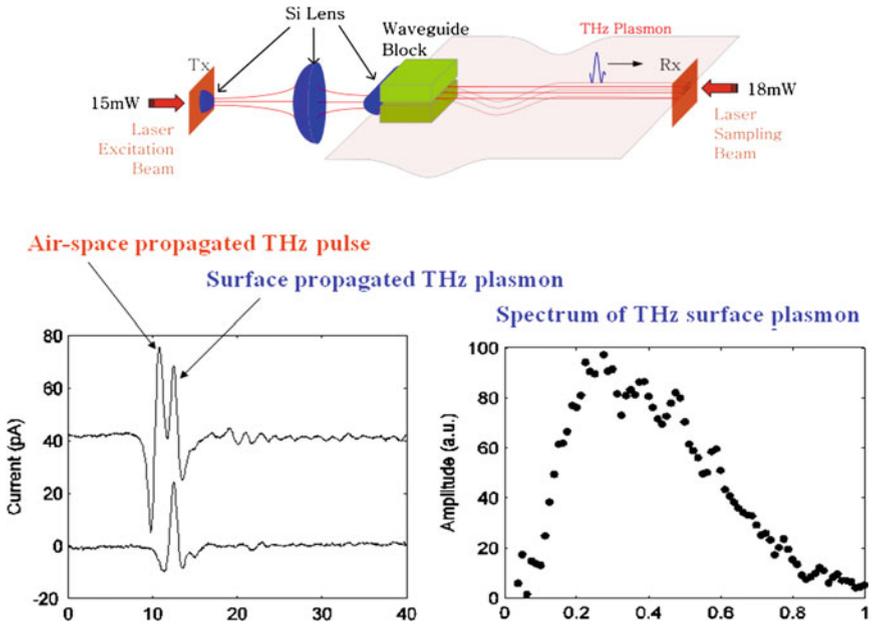


Fig. 15 The experimental setup to propagate THz beam on metal sheet (*upper*) and measured THz pulses (*lower left*) and spectrum of TSP (*lower right*) [13]. Copyright ©2006, American Institute of Physics

4 THz Propagation by Waveguides

The THz beam can be propagated by waveguides, as shown in Table 2. The first THz waveguide propagation was done by Grischkowsky’s group in 1999 through a stainless-steel circular tube. The freely propagated THz beam through air was focused on the entrance of the circular tube by silicon lens. The rectangular tube, plastic ribbon, and parallel plate metal waveguides (PPWGs) were also carried out using the silicon lens to couple the THz beam in and out of the waveguides. However, coaxial cables and metal wires were used to couple the THz beam in and out of the waveguides, which could strongly detect THz beams.

Among the waveguides, PPWG has a low propagation loss and good guiding properties, but the width of the parallel plate should be wider than an air gap in order to keep the TEM mode. Reference [10] shows the measured THz pulses without PPWG (reference) and with a 250-mm-long “Ω” shape for a 90 μm air gap between the two plates. The propagated pulses clearly exhibit no dispersive pulse broadening. Since the THz beam can be focused in a submicron air gap between the two parallel plates, the PPWG is used to couple the THz beam to the metal plate.

The copper coaxial cable has very good guiding properties because the outer copper tube covers the inner metal wire. The THz beam propagates through Teflon, which is located between the inner and outer conductors. The insulating

material (Teflon) absorbs the THz beam for a long range. Figure 13 shows the measured THz pulses for 15-mm- and 105-mm-long coaxial cables. The rapid oscillation observed in the 15-mm-long coaxial cable indicates the onset of the TE_{11} mode [11], which has a high propagation loss. Therefore, there is no TE_{11} mode for 105-mm-long coaxial cable measurement. Only the TEM mode THz pulse can be detected.

The THz beam propagation on the single metal wires has a remarkably low loss and low group velocity dispersion, while all the other modes vanish almost immediately upon excitation due to their high attenuation. Only the TEM mode can be propagated on the wire surface [15]. Figure 14 shows the experimental setup and the transmitted THz pulses for 20-cm-, 64-cm-, and 104-cm-long copper wires, which are of 0.52 mm diameter [12]. The main peaks and S/N are almost identical, demonstrating the very low attenuation and low group velocity dispersion of the THz surface wave. However, the THz waves propagate as weakly guided surface waves due to the finite conductivity of the metal. Therefore, when the wire is bent, most of the THz waves radiate in the air. This weakly guiding property will limit practical applications.

When the electromagnetic waves are guided along the metal-dielectric interface, the surface waves are called surface Plasmon polaritons (SPPs). Usually infrared or visible frequency electromagnetic waves are used to generate the SPP. Recently, an experimental study of THz pulses propagating on the boundary between air and metal have been conducted [13]. The measured THz surface plasmons (TSP) has a much higher attenuation and a much reduced spatial extent of the TSP evanescent field than predicted by theory because the metal surface is not extremely flat and there are optically smooth surfaces. Figure 15 shows the experimental setup and the measured separated air-space propagated THz pulses and TSP (upper pulses). Only the TSP pulse (lower pulse) was measured for 14.5-cm-long aluminum sheet surface propagation.

References

1. Xie, X., Dai, J., Zhang, X.-C.: Coherent control of THz wave generation in ambient air. *PRL* **96**, 075005 (2006)
2. Dai, J., Xie, X., Zhang, X.-C.: Detection of broadband Terahertz waves with a laser-induced plasma in gases. *PRL* **97**, 103903 (2006)
3. Duvillaret, L., Garet, F., Roux, J.-F., Couta, J.-L.: Analytical modeling and optimization of Terahertz time-domain spectroscopy experiments using photoswitches as antennas. *IEEE J. Sel. Top. Quantum Electron.* **7**, 615 (2001)
4. Sakai, K.: *Terahertz Optoelectronics*. Springer, Berlin (2005)
5. Mittleman, D., Cheville, A.: THz Generation and Applications. Ultrafast optics textbook on <http://frog.gatech.edu/ultratext.html>
6. Jeon, T.-I., Grischkowsky, D.: Electrical characterization of conducting polypyrrole by THz time-domain spectroscopy. *Appl. Phys. Lett.* **77**(16), 2452 (2000)
7. McGowan, R.W., Gallot, G., Grischkowsky, D.: Propagation of ultra-wideband, short pulses of THz radiation through sub-mm diameter circular waveguides. *Opt. Lett.* **24**, 1431 (1999)

8. Gallot, G., Jamison, S.P., McGowan, R.W., Grischkowsky, D.: THz waveguides. *J. Opt. Soc. B.* **17**, 851 (2000)
9. Mendis, R., Grischkowsky, D.: Plastic ribbon THz waveguides. *J. Appl. Phys.* **88**, 4449 (2000)
10. Mendis, R., Grischkowsky, D.: THz interconnect with low loss and low group velocity dispersion. *IEEE Micro. Wire. Com. Lett.* **11**, 444 (2001)
11. Jeon, T.-I., Grischkowsky, D.: Direct optoelectronic generation and detection of subps electrical pulses on sub-mm coaxial transmission lines. *Appl. Phys. Lett.* **85**, 6092 (2004)
12. Jeon, T.-I., Zhang, J., Grischkowsky, D.: THz Sommerfeld wave propagation on a single metal wire. *Appl. Phys. Lett.* **86**, 161904 (2005)
13. Jeon, T.-I., Grischkowsky, D.: THz Zenneck surface wave (THz surface plasmon) propagation on a metal sheet. *Appl. Phys. Lett.* **88**, 061113 (2006)
14. Gong, M., Jeon, T.-I., Grischkowsky, D.: THz surface wave collapse on coated metal surfaces. *Opt. Express* **17**, 17088 (2009)
15. Wang, K., Mittleman, D.M.: Metal wires for terahertz wave guiding. *Nature* **432**, 376 (2004)

Recent Advances in Solid-State Electronic Terahertz Systems

Jae-Sung Rieh, Dong-Hyun Kim, Kyungmin Kim and Hyunchul Kim

1 Introduction

The terahertz (THz) band generally indicates the spectrum range that falls on roughly between the traditional electrical and optical frequency bands. Its definition varies over different sources, but one widely accepted definition is the frequency range of 0.1–10 THz. In terms of the wavelength in free space, it corresponds to the range of 3–0.03 mm. In terms of the more readily accepted band definition in the scientific community, the THz band partly include the mm-wave band (1 mm–1 cm, or 30–300 GHz) for the lower side and the infrared band (0.3 mm–750 nm, or 1–400 THz) for the upper boundary, while it covers the entire range of sub-millimeter band (0.1–1 mm, or 300 GHz–3 THz). It is notable that the range includes some scientifically significant points, one example being the frequency that corresponds to a photon energy equal to kT at room temperature (~ 6.2 THz).

Compared to the neighboring electrical and optical bands, which have been extensively exploited for all different range of applications, the THz spectrum has long remained as a territory only scarcely explored. For this reason, the spectrum has been widely called the ‘terahertz gap’. A couple of reasons can be considered as major barriers against the full utilization of this frequency band.

Firstly, it is quite challenging to develop devices that can reliably generate, detect, or properly process the THz signals. In particular, the generation of THz signal with sufficient output power is very difficult. It would be fair to say that the method of generating THz *pulses* is relatively less challenging, owing to the well-established femto-second laser technology. On the other hand, the generation of continuous wave (CW) THz signal, which is more useful for the majority of

J.-S. Rieh (✉) · D.-H. Kim · K. Kim · H. Kim
School of Electrical Engineering, Korea University, Seoul, Korea
e-mail: jsrieh@korea.ac.kr

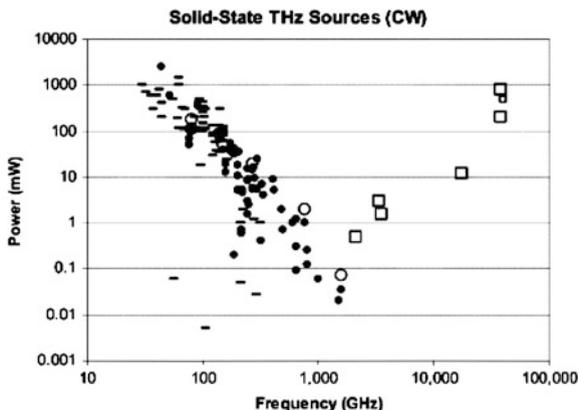


Fig. 1 Output power versus frequency for various solid-state CW THz sources [1]

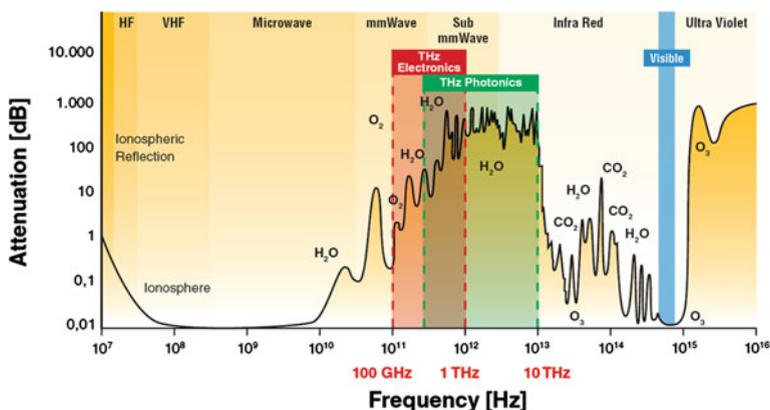


Fig. 2 Attenuation of the electromagnetic wave in the atmosphere (dB/km) [2]

applications, still remains as a major challenge. This fact is well reflected in Fig. 1, in which the output power of CW THz signals generated by various solid-state sources is plotted as a function of the frequency [1]. The output power rapidly drops as the frequency approaches the THz band from either side of the spectrum, clearly visualizing the terahertz gap.

Secondly, the attenuation level of electromagnetic waves in the earth atmosphere is relatively higher for the THz band. Typically the attenuation is a result of the resonance of molecules in the atmosphere with the electromagnetic wave at a given frequency range. This implies that the atmosphere molecules have various resonance modes, typically due to rotational or vibrational transitions, that correspond to the THz frequency range. Figure 2 illustrates the attenuation as a function of the frequency, which reaches well beyond the 100 dB/km level near 1 THz [2]. This imposes a challenge in exploring THz band for applications that need propagation over extensive distances.

The fact that the THz band has long remained less exploited does not necessarily imply the band is less attractive. On the contrary, it indicates open opportunities for new scientific and engineering developments when properly approached. In fact, the THz band has lots of potentially attractive unique properties for practical applications. Various materials such as papers, clothes, plastics, wood panels, and leather are transparent to the THz wave unlike to visible light. On the other hand, the THz wave is easily absorbed by water-containing materials such as human body, which results in high contrast against dry background materials. Also, the THz frequency range corresponds to the resonance frequencies of various molecules in the atmosphere and the space, and thus will show unique spectroscopic patterns with those molecules. Besides, the THz wave is safe as it will not cause harmful photoionization in biological tissues.

With these distinguishing properties given, the THz band retains a great potential for applications that range from security and medical imaging, spectroscopy, plasma diagnosis, chemical and biological sensors, product quality inspection, to astronomy and atmospheric studies. Another advantage of the THz band that attracts increasing interest today is its immense bandwidth available for broadband wireless communication [3]. The channel capacity (C) is proportional to the available bandwidth (B) as indicated by the Shannon's theorem ($C = B \log_2(1 + SNR)$), revealing an obvious advantage over the traditional microwave band for wireless applications.

It is noted that there are two approaches for the THz band development. One is the 'downward' approach from the traditional optics, for which the reduction of operation frequency is needed. The other one is the 'upward' approach from the traditional electronics, which requires the increase of operation frequency. Historically, THz studies were initiated by the optical community and thus the related researches of today are still dominated by the downward approach. However, such method requires rather bulky optical systems that involve lasers and optical lenses and/or mirrors, and its application is limited from many practical aspects. On the other hand, the electrical community has recently found the THz band practical and attractive than ever, strongly promoting the upward approach. Especially, the interest based on solid-state electronic devices is growing fast, owing to the rapidly improving operation speed of semiconductor device technologies. This would enable the realization of compact and affordable THz systems with integrated devices and circuits.

In this article, an overview of the recent development of the solid-state electronic approach for THz study and application is provided, with focus on the semiconductor technologies and circuits currently available for the possible implementation of the THz communication systems. In Sect. 2, various options to implement THz systems will be reviewed, while Sect. 3 provides a brief introduction to diode-based passive solid-state electronics approach. In Sect. 4, the current status and the comparison of semiconductor transistor technologies will be presented, followed by Sect. 5 that describes the recent advances in THz solid-state circuits.

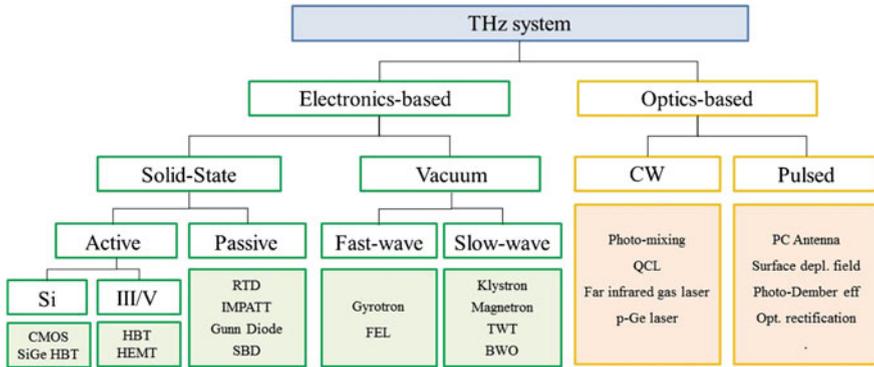


Fig. 3 Various approaches for THz system implementation

2 Various Technology Options for THz Systems

There have been efforts to implement THz systems based on various technical approaches, which are summarized in Fig. 3 [4]. As mentioned earlier, it can be largely divided into two major approaches: optical and electronic.

The optical approach can be again grouped into two methods: pulse-based and CW-based ones. The pulse-based approach is mostly based on the well-developed femto-second laser technique and currently the most widely employed for various THz applications including TDS and imaging. The very short optical pulses from a femto-second laser incident on a photoconductive (PC) antenna can generate THz pulses. The optical pulse also can interact with the surface of semiconductor and generate THz pulses due to the surface depletion field or as a result of the photo-Dember effect. For detector side, the optical pulses from a femto-second laser can be utilized to detect THz pulses by sampling with PC antenna or optical rectification with electro-optic crystals.

CW THz signal, which is more useful for most THz applications, can be effectively generated by the photo-mixing method. Two closely located optical waves, typically generated by CW lasers, can be combined and then generate CW signal as a result of the nonlinearity of the combiner, which is basically a mixer. The frequency of the generated CW signal is the difference of the two frequencies of the incident waves and can be controlled to fall on THz band by adjusting the source frequencies. Although it is rather compactly implementable, the output power is still limited far below mW range, and the frequency stability still remains to be improved. Quantum cascade laser (QCL) can generate higher power CW signals above mW range, owing to the power combining as a result of the cascaded inter sub-band electron transitions. However, it still needs to operate at cryogenic temperature to generate THz range of output signals. Some of the more conventional lasers can extend the output frequency down to THz band, including gas lasers based on CH_3OH , CH_3F , NH_3 , or CH_2F_2 , and solid-state lasers based on

p-type germanium, although they have not been readily adopted for THz applications yet.

The electronic approach mostly involves CW signals and can be based on either vacuum or solid-state devices. Vacuum electronic devices are typically based on electron beams travelling within a vacuum tube. THz signals are obtained by controlling the kinetics of electrons in the tube, from which CW output power over 10 W up to nearly 1 MW can be achieved beyond 100 GHz depending on the kinds of the devices. Klystrons, traveling wave tubes (TWTs), and backward oscillators (BWOs) are based on operation principle that involves electron beams bunched by the modulation of the electron speed along the tube, which result in the generation of CW THz signals. While these devices exploit the longitudinal motion of electrons along the tube, magnetrons involve the electrons moving in the radial direction inside the tube, which interact with the cavities formed along the wall and generate CW signals.

Gyrotrons and free electron lasers (FELs) are categorized as fast-wave devices as they involve electrons travelling at relativistic speed, as opposed to the other vacuum devices mentioned earlier which are considered as slow-wave devices. Gyrotrons utilize electrons in gyration motion along the tube, the radiation from the electrons being captured as a CW THz source. FEL can generate a wide frequency range of CW signals including the THz band, which is based on the coherent radiation from accelerated electrons in a wiggling motion along the tube. The electrons in FEL are free in a sense that they are not bound to any solid-state material or gas molecules, and also no stimulated emission is involved in the operation, which are major differences from other types of lasers.

The approaches described so far, based on optical method and vacuum electronics, have been widely accepted so far as they can provide high power THz sources and, in some cases, efficient detection. However, a major problem for these methods is their bulky size required. This large form factor seriously deters their application to commercial products, along with the potentially high cost involved. In that sense, another approach based on solid-state electronic devices seems to be a strong candidate for the implementation of compact and affordable THz systems. Although they lag behind in terms of performance, both output power and frequency, there have been significant improvements for these types of devices recently, which will be the major topic of the remaining part of this article.

3 Diode Technologies for THz Systems

The solid-state electronic devices for THz applications can be divided into two groups: passive and active. The passive devices are mostly diodes while transistors represent the active devices. These two groups of devices are discussed separately over the following sections.

Diodes can operate at higher frequencies and have already been widely adopted for various THz systems in both signal generation and detection. Diodes for signal

generation typically make use of the negative resistance established across a two terminal device and there are various kinds depending on the specific method of obtaining the negative resistance [5]. IMPACT ionization Transit Time (IMPATT) diodes exploit the avalanche delay and transit-time delay for negative resistance. When a strong reverse bias is applied across a p–n diode, avalanche current will be induced with a finite delay, which occurs as a result of the delay in the impact ionization itself and the time required for the generated carriers to reach the terminals of the device. With a proper design of the diode dimension, the total delay time for the current can be tuned to exactly match the half cycle of the AC voltage signal externally supplied, effectively leading to a negative resistance. If the diode with negative resistance is externally terminated with a proper combination of inductance and capacitance, oscillation with controlled frequency can be achieved and can be used as a signal source. The oscillation frequency with IMPATT diodes can exceed a few hundred GHz, based on both Si and III–V semiconductors (such as GaAs and InP). The output power near 1 W can be obtained around 100 GHz, while it tends to drop as frequency increases, reaching a few hundreds of GHz.

Gunn diodes make use of Gunn Effect to obtain the negative resistance. For many of useful III–V semiconductors, including GaAs and InP, a satellite valley is formed near the main valley in the E-k relation for the electrons. Hence, when a large bias is applied across a diode based on such semiconductors, some of the electrons originally residing in the main valley are excited and gain sufficient energy required for transition to the satellite valley. Once such transition occurs, the average mobility of the electrons in the device is suddenly degraded as the electrons in the satellite valley tend to show much lower mobility than those in the main valley, leading to a current reduction with increased voltage. This will cause a negative resistance and can be used for oscillation and signal generation. Output power from Gunn diodes reaches around a few hundreds of mW around 100 GHz, while its frequency can reach a few hundred GHz.

Resonant Tunneling Diodes (RTDs) make use of (mis)alignment of the energy levels in adjacent quantum wells, which are created by an alternating stack of two types of semiconductors with difference energy bandgaps. Depending on the alignment of the energy levels, the amount of current flowing across the stack will vary. If the alignment degrades as the applied voltage increases, which actually can happen with proper design of the device, negative resistance can be obtained. While the output power of RTDs is smaller than IMPATT or Gunn diodes, limited below 1 mW around 100 GHz, operation frequency near 1 THz can be achieved.

Some types of diodes can be used for the detection of THz signals. Schottky Barrier Diodes (SBDs) are one of the most popular choices for the diode detection of THz signals. Being a majority carrier device, SBDs exhibit much faster switching time than p–n diodes owing to smaller reverse recovery time. A cutoff frequency f_T exceeding 4 THz has been recently reported for Si-based SBDs [6] and f_T of 11.1 THz has been reported for GaAs-based SBDs [7]. It is noted, though, that the cutoff frequency of diodes cannot be directly compared with that of transistors. For transistors, f_T indicates the frequency beyond which the gain

becomes smaller than unity, while f_T of diodes is simply a reverse of RC time constant with a factor of 2π multiplied ($f_T = 1/(2\pi RC)$). If we apply the definition of transistor f_T to diodes, it will be always zero because the gain of diodes will be smaller than unity at any frequency point. Another advantage of SBDs is the fact that they can be implemented with conventional semiconductor technologies, enabling the integration with remaining part of a given system.

Other passive devices for THz detection include superconductor-insulator-superconductor (SIS) tunnel junction mixer and hot electron bolometer (HEB). The operation of SIS mixers is based on the photon-assisted tunneling across the junction. With the presence of THz radiation, the device shows nonlinear I-V characteristics, which can function as a mixer with LO and serve as a THz detector. SIS mixers can cover the frequency range up to around 1 THz [8]. HEB is basically a microbridge structure made of superconductor material, contacted by metal at the terminals. When THz signal is incident on the device, the heating will cause a transition between superconducting and normal states, leading to a change in the resistance along the device. The resistance change can be taken as a detection of THz signal. The operation speed of HEBs can reach up to a higher frequency of several THz [9]. Although SIS mixers and HEBs exhibit superior noise performance than SBDs, they typically need to operate at cryogenic temperature while SBDs can operate at room temperature, which is a major drawback of these rather exotic devices for commercial applications.

4 Transistor Technologies for THz Systems

We have seen from the description above that diodes can function deep into the THz band for both generation and detection of THz signals. However, diodes have an inherent limit of being a passive device, which critically limits the range of their applications. On the other hand, the transistors, an active device, provide versatile electronic functions mainly owing to their ability to amplify signals, which leads to critical performance enhancement in electronic systems including the THz system.

4.1 Transistor Versus Diode

The advantage of the transistor-based systems over the diode-based ones can be easily seen with the heterodyne receiver shown in Fig. 4 as an example [10]. The availability of transistors enables the inclusion of a low noise amplifier (LNA) at the very first stage of the system (Fig. 4a). This will not only enhance the overall gain of the system, but, more importantly, can maintain noise figure (NF) of the system to a sufficiently low level since the total noise level of a system is dictated by NF of the first stage. In addition, following the LNA stage, an *active* mixer can be employed, which will effectively suppress the noise from the following IF

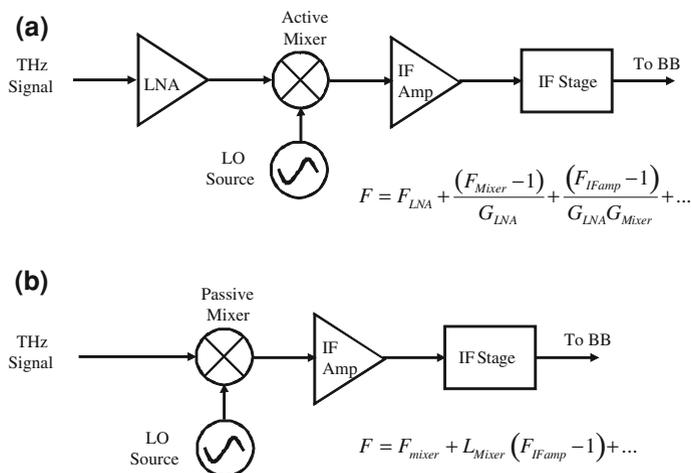
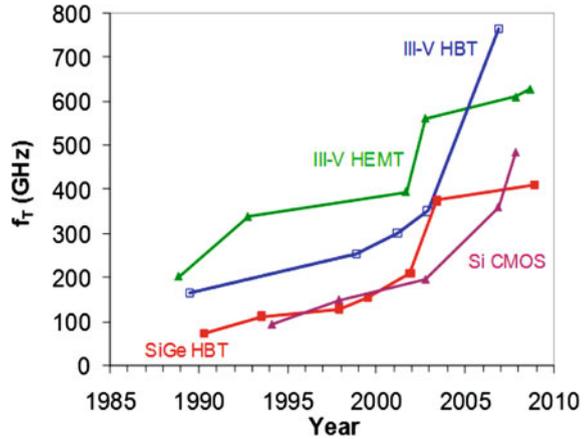


Fig. 4 Receiver block diagram for: **a** transistor-based system; **b** diode-based system

blocks with its own gain larger than unity, as suggested by the Friis's noise formula (included in the figure). The overall gain of the receiver will be also further increased with the gain from the active mixer. On the other hand, for the diode-based system (Fig. 4b), NF of the entire system will be determined by that of the mixer, which is significantly larger than that of LNA in general. Further, the noise from the IF block will not be suppressed but rather magnified by a factor equivalent to the loss of the passive diode-based mixer. Finally, the overall gain will be degraded with the inclusion of the passive mixer, which exhibits a conversion loss instead of gain. It is noted that there are some cases in which active mixers shows a conversion loss instead of gain. However, even in those cases, the loss will be much smaller than that of passive mixers, and thus its effect on the overall noise and gain of the receiver will be more favorable than the case of passive mixers.

The availability of transistors greatly enhances the performance of the transmitters, too, by enabling the inclusion of a power amplifier at the final stage, which further boosts the output power level from the preceding stage. It is also important to note that the transistor-based circuits have an additional advantage of easier integration with other blocks than the diodes. As mentioned before, the diodes introduced earlier are not quite compatible with the general-purpose semiconductor technologies except for SBDs. Hence, *transistor* technology is highly favored for the upward electronic approach to the THz systems. One major drawback of transistors for THz applications has been the relatively lower operation speed of the transistors. Recently, however, there has been a significant enhancement in the transistor speed owing to the continued scaling and material innovation, which has rendered the transistor-based THz system a very realistic option. The details are described in the following discussion.

Fig. 5 Trend of the increase of operation frequency for various transistor technologies



Modern semiconductor transistor technologies can be largely categorized into two groups: III–V compound semiconductor technologies and Si technologies. The representative high-speed devices for the former are Heterojunction Bipolar Transistor (HBT) and High Electron Mobility Transistor (HEMT), while those for the latter are SiGe HBT and Si Metal–Oxide–Semiconductor Field Effect Transistor (MOSFET). The speed performance of these devices has been continuously improved over the past decades and now they all have entered the regime of several hundreds of GHz. Figure 5 depicts the trend of the device speed evolution of in terms of the cutoff frequency f_T [10]. The trend shows that the III–V devices tend to exhibit superior operation speed than Si devices, although the gap is closing up. This places the III–V HBTs and HEMTs on the front line for the THz applications. It is noteworthy to briefly mention that another measure of the device operation speed, the maximum oscillation frequency f_{max} , is a more relevant parameter than f_T for most analog and RF circuit designs. However, f_{max} is more sensitive to the details of device layout and its extraction is less reliable than f_T , making f_T a more popular measure for comparisons between technologies.

4.2 III–V Transistor Technologies

While both HBT and HEMT are based on III–V semiconductor systems typically represented by GaAs or InP, each device has its own pros and cons in terms of device characteristics. HBT is basically a bipolar transistor, where the bandgap of the emitter is intentionally made larger than that of the base. In this way, the current gain can be larger, or more importantly, it is possible to make the base layer thinner and more heavily doped. This will result in higher speed and lower noise level. HBTs tend to show higher current driving capability and larger transconductance g_m , which is typical characteristics of bipolar transistors, than HEMTs. The epitaxial

structure and fabrication steps of HBTs, however, are more complicated as they are basically a vertical device. As of today, the highest reported f_T at room temperature for HBT is 765 GHz [11]. The device, reported by University of Illinois at Urbana Champaign, is based on InP technology and its speed can be further pushed up to 845 GHz when the device is cooled down to $T = -55$ C.

HEMT is basically a Field Effect Transistor (FET) that employs an undoped channel region, which would significantly improve the mobility of the carriers travelling in the channel with reduced scattering. This will lead to improvement in both speed and noise performance of the device, especially at low temperature. The carriers will instead be supplied by an adjacent heavily doped layer which has a wider bandgap than the channel for better carrier confinement. As HEMTs are intrinsically lateral devices, the performance tends to be highly dependent on the lateral patterning of the gate electrode, for which E-beam lithography is typically employed. Its epitaxial structure and fabrication steps, on the other hand, are relatively simple. HEMTs used to be considered as the fastest device until mid-2000's, when the record HBT performance rapidly increased. The record performance of HEMT as of today is f_T of 628 GHz, achieved with an InP-based device as reported by MIT [12]. In terms of f_{max} , the operation speed of HEMTs has already exceeded 1 THz [13].

The excellent speed performance of the III–V devices mainly arises from the superior electron transport characteristics in the III–V semiconductor materials. The low-field electron mobility and the pronounced electron ballistic transport property exceed those of Si, leading to the higher operation frequency of n-type III–V devices. It is noted though that the transport properties of holes for III–V semiconductors is typically worse than those of Si, explaining for the rare adoption of p-type III–V devices for practical applications. III–V devices also tend to show higher breakdown voltage and thus superior power performance than Si devices, as a result of the larger bandgap of typically employed III–V materials. Finally, it is noteworthy that the substrate employed for III–V devices, typically GaAs and InP, shows a much higher resistivity (10^7 – 10^9 Ω cm) than that of Si (typically 10 Ω cm for microwave and mm-wave applications). This will significantly suppress the loss of passive devices formed on the substrate, which is a major advantage especially for high frequency applications including THz ones.

4.3 Si-Based Transistor Technologies

In spite of the excellent high frequency performance, the III–V devices suffer from relatively poor reliability compared to the Si counterparts. The typically encountered non-planar structure and high defect density in the active region of the III–V devices tend to cause leakage current due to traps, raising the reliability concerns. Further, the fabrication cost of the III–V technology is on the unfavorable side, mainly due to the expensive wafers with small diameter, costly epitaxial growth steps, and the need for low throughput E-beam lithography process.

Si-based transistors, which are represented by SiGe HBTs and Si MOSFETs, have shown rapid operation speed improvement in recent years, now exhibiting comparable performance as the III–V devices. Moreover, their development is mostly driven by the industry, implying a shorter time to the production line than the III–V devices case. SiGe HBT is basically a variant of Si bipolar transistors, which contains a small amount of Ge in the base region for bandgap reduction. The advantages of the smaller base bandgap in HBTs are as mentioned earlier and apply to SiGe HBTs as well. Compared to Si MOSFETs, it provides higher current drivability, larger g_m , superior $1/f$ noise property, and superior device matching. Besides, SiGe HBTs show higher operation frequency than Si MOSFETs for a given lithography node, since it is a vertical device whose operation speed is dictated by the vertical rather than lateral dimension. Typically, SiGe HBTs show a similar device speed performance with Si MOSFETs based on 2–3 more advanced node. The record performance of SiGe HBTs at this moment is f_T of 410 GHz [14], while f_{max} of SiGe HBTs has recently reached 500 GHz [15].

Si MOSFETs are by far the most dominant device for modern electronics. Until quite recently, however, Si MOSFETs have not been considered as a serious contender for the RF arena. This was because carrier transport characteristic of Si is not as good as that of III–V semiconductors and the gate length was not extremely scaled down either. Nonetheless, the continued scaling of Si MOSFETs, driven by the strong need for high-speed digital applications, has brought about sub-100 nm era and consequently advanced the RF properties of the device. That has eventually led to the impressive f_T of 485 GHz for an n-type MOSFET [16]. Along with the additional merit that they can be easily integrated with other circuit blocks with low manufacturing cost, Si MOSFETs have emerged as a competent candidate for the THz applications. It is fair to state that the widely appreciated cost advantage of Si MOSFETs is valid only when the production volume is large, since the cost for the phase shift masks required for the advanced lithography is significant.

5 Recent Advances in Transistor-based THz Circuits

The recent advances in transistor technologies described above have enabled them eligible for the implementation of solid-state electronic THz systems with active components. The most popular THz system based on the transistors is the heterodyne system, which provides functions of up-conversion and down-conversion of frequency bands and can be adopted for various applications. Three basic circuit blocks that compose the heterodyne system are the amplifier, oscillator, and mixer. There have been a good amount of reports on these circuits operating at the THz band based on various transistor technologies. Moreover, THz transmitter and receiver circuits that integrate these circuit blocks on a single chip have also been reported. In this section, some of the key results are highlighted.

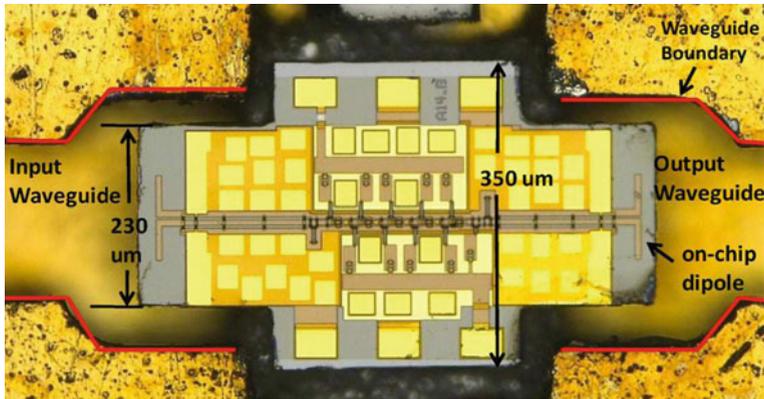


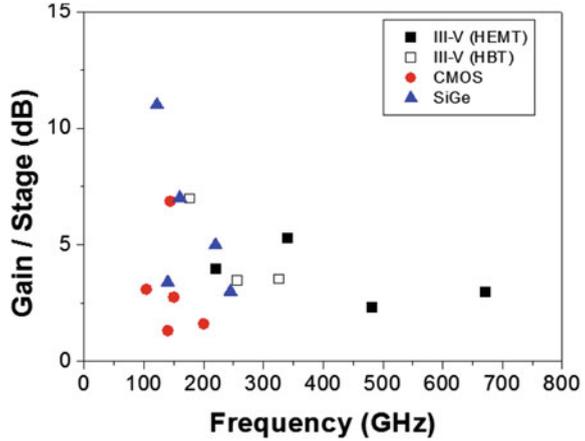
Fig. 6 Chip photo of NGC 670 GHz LNA in a packaging structure [17]

5.1 Amplifiers

Amplifiers are probably the most important circuit block in heterodyne systems. For receivers, a low noise amplifier (LNA) is typically employed at the very first stage, while a power amplifier (PA) is incorporated at the last stage of transmitter for the final boost of output power transmitted. Additional amplifiers can be inserted wherever additional gain is required. For high frequency applications, amplifiers demand the most stringent requirement for the transistor operation frequency among the three circuit blocks mentioned above. To achieve a gain of several dB per stage, which is typically the case, f_{max} needs to be at least twice of the amplifier operation frequency assuming the typical gain roll-off of -6 dB/octave. If losses and possible mismatches are taken into account, the required device operation frequency will be even higher.

Such strong dependence of the amplifier operation frequency on the device speed naturally renders the amplifiers based on III–V technologies highly competitive for THz applications. Figure 6 shows the chip photo of the recently reported HEMT LNA by Northrop–Grumman Company (NGC) [17]. It exhibit a power gain higher than 7 dB around 670 GHz, which is the highest operation frequency as of today for any amplifier based on semiconductor. It is noted that the chip area is only $655 \mu\text{m} \times 375 \mu\text{m}$, which clearly shows the advantage of solid-state electronic approach in terms of the compactness for the implementation of THz systems. The thickness of the semiconductor chips are typically around $100 \mu\text{m}$. The operation frequency of Si-based amplifiers continues to grow as well. An LNA based on SiGe HBT operating up to 245 GHz with 12 dB gain has recently been reported by IHP [18], which is based on SiGe HBT technology that exhibits device f_T and f_{max} of 300 and 500 GHz, respectively. Figure 7 shows recently reported amplifiers operating beyond 100 GHz based on various transistor technologies. The gain per stage is largely limited below 5 dB except for some cases, indicating the challenge to extract high gain from the current technology.

Fig. 7 Gain per stage versus frequency for amplifiers operating beyond 100 GHz



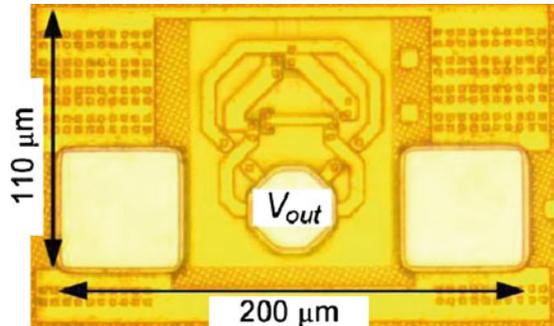
While the highest operation frequencies are being obtained from HEMT technologies owing to high f_{max} of HEMT devices, the plot indicates the improving operation frequency of Si-based amplifiers as well.

5.2 Oscillators

Oscillators provide local oscillation (LO) to mixers for both receivers and transmitters. They also can serve as a THz source for various applications, which proves themselves as an inevitable circuit block for THz systems. An oscillator is typically implemented by terminating a network with negative resistance by a proper combination of inductance and capacitance. Negative resistance can be achieved on the circuit level in many different ways by properly connecting conventional transistors, unlike the case of passive devices where special diodes should be designed for negative resistance on the device level. It is noted that voltage controlled oscillators (VCOs) are useful for many applications for which the oscillation frequency can be externally tuned by adjusting the bias voltage.

Required device operation frequency for oscillators is not as stringent as for the case of amplifiers. A gain barely enough to compensate for the loss along the feedback loop in oscillators suffices for triggering oscillation. Besides, not only the fundamental but also harmonic oscillation frequencies are available at the output. This makes it possible to obtain oscillation frequencies even higher than the device cutoff frequency when properly designed. In terms of the fundamental oscillation, 346 GHz is currently the highest frequency reported so far for transistor-based oscillator [19]. The push-push technique is widely adopted for high frequency application as it takes the second harmonic at the output, while the triple-push technique is also attracting recent interest, which takes the third-harmonic at the output. These harmonic techniques provide high oscillation frequencies with a given technology, although they commonly suffer from low output

Fig. 8 Chip photo of 480 GHz oscillator based Si CMOS technology [21]



power level. For push–push oscillators, an oscillation frequency of 410 GHz was achieved with Si CMOS technology by University of Florida [20]. More recently, 482 GHz has been achieved with a Si CMOS triple-push oscillator by Cornell University [21], which is the highest oscillation frequency in the open literature as of today from any semiconductor transistor technology. Figure 8 shows the chip photo of the 482 GHz oscillator, which occupies only $110\ \mu\text{m} \times 200\ \mu\text{m}$ including probing pads. Figure 9 shows the output power versus oscillation frequency for oscillators operating beyond 100 GHz based on various transistor technologies and circuit techniques. Overall trend observed in the plot is the decreasing output power with increasing oscillation frequency. It is also clear that harmonic techniques, such as push–push, triple-push, and even quadruple-push, boost the oscillation frequency, but with rather suppressed output power.

5.3 Mixers

Mixers serve to down-convert RF signal down to IF band or up-convert IF signal up to RF band. Active mixers are based on transistors and they provide not only the basic function of frequency conversion, but also add gain to the system, an extra contribution. In some cases, active mixers exhibit negative gain, or loss, especially at high frequencies, but it will be small as mentioned earlier. Passive mixers are typically based on diodes and their operation frequencies tend to be larger than that of active mixers, although they inherently show a loss. Sometimes mixers can be implemented with transistor without bias voltage applied, which are also regarded as passive mixers as they do not provide gain as the transistors are under the passive operation mode.

Mixers have been a popular circuit block for the very early stage in the traditional THz heterodyne receivers based on diodes to provide the downward frequency conversion to ease the signal detection. However, they were basically passive mixers and the performance was limited as described earlier. Active mixers operating at the multiple hundred GHz frequency range are still rare and only limited number of results have been reported. Figure 10 shows the trend of mixers operating at

Fig. 9 Output power versus frequency for oscillators operating beyond 100 GHz

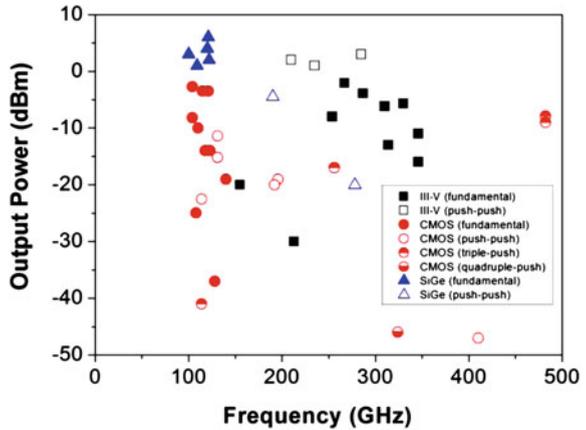
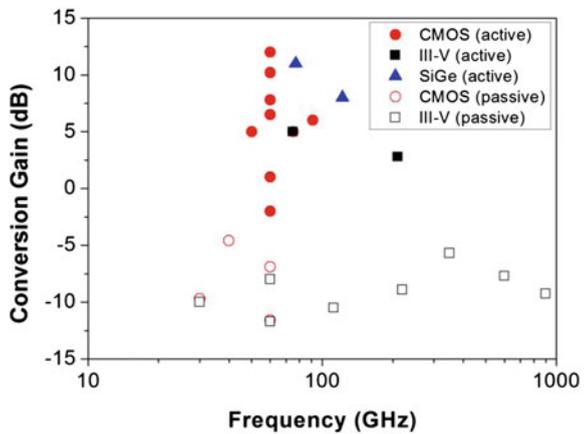


Fig. 10 Conversion gain versus frequency for mixers operating at mm-wave range and beyond



mm-wave band and beyond in terms of conversion gain and frequency. Active mixers show higher conversion gain but the operation frequency is rather limited. A mixer reported by Fraunhofer Institute based on a HEMT technology showed a positive conversion gain of 2.8 dB at 215 GHz [22], which is currently the highest frequency for active mixers. Also presented in the plot for comparison of conversion gain (loss) are some of the passive mixers, including those based on semiconductor diodes. Diode-based mixers can easily operate beyond 1 THz.

5.4 Integrated Circuits

As the techniques for THz unit circuit blocks grow increasingly matured, integrated circuits, typically receivers and transmitters, operating at THz band are also being developed based on various semiconductor technologies. The additional challenges for implementing THz integrated circuits include the correct matching

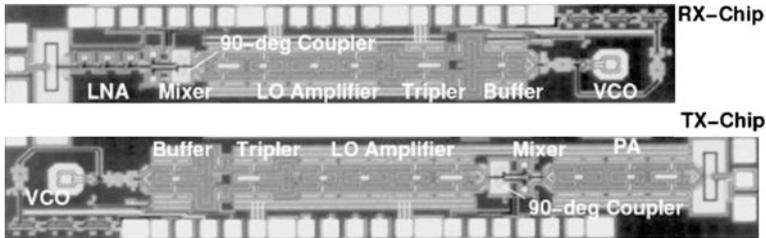


Fig. 11 Chip photo of integrated receiver and transmitter based on SiGe BiCMOS technology [23]

between the unit circuit blocks incorporated, since the chances for mismatch become higher with increasing operation frequency due to device model mismatch issues. Besides, even very small dimensional deviations in passive components result in a significant change in the characteristics, leading to aggravated mismatch and degraded performance. Despite such challenges, some remarkable results have been reported, some of which are introduced below as examples.

Based on a SiGe BiCMOS technology with f_T/f_{max} of 260/380 GHz, 160 GHz receiver and transmitter chip sets were developed by Wuppertal University [23]. The circuits integrate LNA, mixer, VCO for receiver chip and PA, mixer, and VCO for transmitter chip, while frequency multipliers and dividers are also included for both chips. They are designed to handle the quadrature signal to enable quadrature modulation for communication applications with higher spectral density. The receiver exhibits a conversion gain up to 29 dB without IF amplifier, and the transmitter shows saturated output power reaching 5 dBm. The chip photo is shown in Fig. 11.

CMOS-based integrated transceivers of today are typically based on 65 nm CMOS technology with f_{max} around 250 GHz. 140 GHz transmitter and receiver chip sets were recently reported by UCLA [24]. The transmitter is composed of a 140 GHz fundamental VCO and a PA, the modulation being achieved by controlling the gain of the PA with the base-band signal. The receiver includes an LNA, an envelope detector, and an IF amplifier, which performs direct demodulation of the modulated signal without down-conversion. 2.5 Gbps data link was demonstrated with the circuits. Overall, CMOS-based transceivers are still lagging the SiGe BiCMOS-based counterpart in terms of complexity and performance levels, while useful wireless data transfer have been demonstrated notably with lower supply voltage and DC power dissipation.

As for III–V HEMT and HBT technologies, it is interesting to note that the focus has been still with high performance unit circuit blocks so far, only a few works on integrated transceivers based on those technologies having been reported. One example is integrated receiver and transmitter chipsets based on HEMT technology operating at 220 GHz [25]. The heterodyne receiver incorporating an on-chip antenna is composed of a mixer and an LNA, with a 55 GHz LO being externally supplied and then frequency-doubled. The measured RX conversion gain is 3.5 dB with NF of 7.4 dB. The transmitter exhibits output power up to

−7.1 dBm and data link at 12.5 Gbps was also demonstrated. It is noted that the achieved NF is remarkably low compared to Si-based receivers, which typically show NF well beyond 10 dB even at lower frequencies. The reported conversion gain and output power, though, are rather moderate, but it should be noted that the employed HEMT technology is not the best kind of III–V technologies available as of today. State-of-the-art InP technologies are now exhibiting f_{max} larger than 1 THz [13] and III–V transceivers with superior performance are expected to emerge in the near future.

6 Conclusion

In this article, possible approaches for THz systems have been reviewed, followed by a more detailed discussion on the solid-state electronic approaches based on both diodes and transistors. It is clear from the discussions that the solid-state electronic approaches, especially the one based on transistor technology, have several prime advantages including the compactness, cost, and compatibility with other electronic part. Although they are still yet to be further improved in terms of operation frequency and output power, these advantages are critical for commercial application of THz systems. Since the transistor-based solid-state electronic THz systems are in their infant stage, more results with much better performance are expected to arrive soon.

References

1. Crowe, T.W., Bishop, W.L., Porterfield, D.W., Hesler, J.L., Weikle II, R.M.: Opening the terahertz window with integrated diode circuits. *IEEE J. Solid-State Circuits* **40**, 2104 (2005)
2. Pfeiffer, U.R., Ojefors, E., Lissauskas, A., Roskos, H.G.: Opportunities for silicon at mmwave and terahertz frequencies, In: *IEEE Bipolar/BiCMOS Circuits Technol. Meet.* 149–156 (2008)
3. Rieh, J.-S., Jagannathan, B., Greenberg, D.R., Meghelli, M., Rylyakov, A., Guarin, F., Yang, Z., Ahlgren, D.C., Freeman, G., Cottrell, P., Haramé, D.: SiGe heterojunction bipolar transistors and circuits toward terahertz communication applications. *IEEE Trans. Microw. Theory Tech.* **52**, 2390–2408 (2004)
4. Rieh, J.-S., Jeon, S., Kim, M.: An overview of integrated THz electronics for communication applications. In: *2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)* 1–4 (2011)
5. Miles, R., Zhang, X.-C., Eisele, H., Krotkus, A.: *Terahertz Frequency Detection and Identification of Materials and Objects*. Springer, Dordrecht (2007)
6. Chen, S.-M., Fang, Y.-K., Juang, F.R., Yeh, W.-K., Chao, C.-P., Tseng, H.-C.: Terahertz schottky barrier diodes with various isolation designs for advanced radio frequency applications. *Thin Solid Films* **519**, 471–474 (2010)
7. Thomas, B., Maestrini, A., Beaudin, G.: A low-noise fixed-tuned 300–360-GHz sub-harmonic mixer using planar schottky diodes. *IEEE Microw. Wirel. Compon. Lett.* **15**, 865–867 (2005)
8. Karpov, A., Miller, D., Rice, F., Stern, J.A., Bumble, B., LeDuc, H.G., Zmuidzinas, J.: Low noise 1 THz–1.4 THz mixers using Nb/Al–AlN/NbTiN SIS junctions. *IEEE Trans. Appl. Supercond.* **17**, 343–346 (2007)

9. Khosropanah, P., Gao, J.R., Laauwen, W.M., Hajenius, M., Klapwijk, T.M.: Low noise NbN hot electron bolometer mixer at 4.3 THz. *Appl. Phys. Lett.* **91**, 221111–221111-3 (2007)
10. J.-S. Rieh and D.-H. Kim.: An overview of semiconductor technologies and circuits for terahertz communication applications, In: 2009 IEEE GLOBECOM Workshops. 1–6 (2009)
11. Snodgrass W., Hafez, W., Harff, N., Feng, M.: Pseudomorphic InP/InGaAs heterojunction bipolar transistors (PHBTs) experimentally demonstrating $f_T = 765$ GHz at 25C increasing to $f_T = 845$ GHz at -55 C. In: Technical Digest of International Electron Devices Meeting. Late news 2006
12. Kim, D.-H., del Alamo, J.A.: 30-nm InAs pseudomorphic HEMTs on an InP substrate with a current-gain cutoff frequency of 628 GHz. *IEEE Electron Device Lett.* **29**, 830–833 (2008)
13. Lai R., Mei, X.B., Deal, W.R., Yoshida, W., Kim, Y.M., Liu, P.H., Lee, J., Uyeda, J., Radisic, V., Lange, M., Gaier, T., Samoska, L., Fung, A.: Sub 50 nm InP HEMT device with F_{max} greater than 1 THz. In: IEEE International Electron Devices Meeting. 609–611 (2007)
14. Geynet, B., Chevalier, P., Vandelle, B., Brossard, F., Zerounian, N., Buczko, M., Gloria, D., Aniel, F., Dambrine G., Danneville, F., Dutartre, D., Chantre, A.: SiGe HBTs featuring $f_T > 400$ GHz at room temperature. In: IEEE Bipolar/BiCMOS Circuits and Technology Meeting. 121–124 (2008)
15. Heinemann, B., Barth, R., Bolze, D., Drews, J., Fischer, G.G., Fox, A., Fursenko, O., Grabolla T., Haak, U., Knoll, D., Kurps, R., Lisker, M., Marschmeyer, Ru, x, cker, H., Schmidt D., Schmidt J., Schubert, M.A., Tillack B., Wipf, C., D. Wolansky, D.W., Yamamoto, Y.: SiGe HBT technology with f_T/f_{max} of 300 GHz/500 GHz and 2.0 ps CML gate delay. In IEEE International Electron Devices Meeting. 30.5.1–30.5.4 (2010)
16. Lee, S., Jagannathan, B., Narasimha, S., Chou, A., Zamdmer, N., Johnson, J., Williams, R., Wagner, L., Jonghae Kim, Plouchart, J.O., Pekarik, J., Springer, S., Freeman, G.: Record RF performance of 45-nm SOI CMOS technology. In: IEEE International Electron Devices Meeting. 255–258 (2007)
17. Deal, W.R., Leong, K., Radisic, V., Sarkozy, S., Gorospe, B., Lee, J., Liu, P.H., Yoshida, W., Zhou, J., Lange, M., Lai, R., Mei, X.B.: Low noise amplification at 0.67 THz using 30 nm InP HEMTs. *IEEE Microw. Wirel. Compon. Lett.* **21**, 368–370 (2011)
18. Mao, Y., Schmalz, K., Borngräber, J., Scheytt, J.C.: A 245 GHz CB LNA in SiGe. In: *Eur. Microw. Integr. Circuits (EuMIC)* 224–227 (2011)
19. Radisic, V., Mei, X.B., Deal, W.R., Yoshida, W., Liu, P.H., Uyeda, J., Barsky, M., Samoska, L., Fung, A., Gaier, T., Lai, R.: Demonstration of sub-millimeter wave fundamental oscillators using 35-nm InP HEMT technology. *IEEE Microw. Wirel. Compon. Lett.* **17**, 223–225 (2007)
20. Seok, E., Cao, C., Shim, D., Arenas, D.J., Tanner, D.B., Hung C.-M., O, K.K.: A 410 GHz CMOS push–push oscillator with an on-chip patch antenna. In: IEEE International Solid-State Circuits Conference. 472–473 (2008)
21. Momeni, O., Afshari, E.: High power terahertz and millimeter-wave oscillator design: a systematic approach. *IEEE J. Solid-State Circuits* **46**, 583–597 (2011)
22. Kallfass, I., Massler, H., Leuther, A., Tessmann, A., Schlechtweg, M.: A 210 GHz dual-gate FET mixer MMIC with >2 dB conversion gain, high LO-to-RF isolation, and low lo-drive requirements. *IEEE Microw. Wirel. Compon. Lett.* **18**, 557–559 (2008)
23. Pfeiffer, U.R., Ojefors, E., Yan, Z.: A SiGe quadrature transmitter and receiver chipset for emerging high-frequency applications at 160 GHz. In: International Solid-State Circuits Conference. 416–417 (2010)
24. Xu, Z., Gu, Q.J., Wu, Y.-C., Tang, A., Lin, Y.-L., Chen, H.-H., Jou, C., Chang, M.C.F.: D-band CMOS transmitter and receiver for multi-giga-bit/sec wireless data link. In: IEEE Custom Integrated Circuits Conference. 1–4 (2010)
25. Abbasi, M., Gunnarsson, S.E., Wadefalk, N., Kozhuharov, R., Svedin, J., Cherednichenko, S., Angelov, I., Kallfass, I., Leuther, A., Zirath, H.: Single-chip 220-GHz active heterodyne receiver and transmitter MMICs with on-chip integrated antenna. *IEEE Trans. Microw. Theory Tech.* **59**, 466–478 (2011)

Nanofabrication Techniques and Their Applications to Terahertz Science and Technology

Changsoon Kim and Hwi Kim

1 Introduction

The capability to create nanostructures—structures whose features sizes are smaller than 100 nm—has been essential in the development of modern science and technology. It has played crucial roles in dramatically improving the performance of various devices and systems. The most notable example is perhaps electronic integrated circuits, where the performance enhancement, such as higher operational speed, lower power consumption, and higher throughput, is attributed to the continual decrease in transistor size. The nanofabrication capability has also made possible the demonstration of novel devices, such as photonic band-gap lasers [1], photonics crystal fibers [2], microfluidic devices [3], and nanoelectromechanical systems (NEMS) [4]. Furthermore, the discovery of new scientific phenomena occurring at the nanoscale would not have been possible without the development of nanofabrication techniques [5, 6, 7].

Despite its potential for applications ranging from biology, medical sciences, and material characterization to tomographic imaging, security, astrophysics, and computing, the field of terahertz (THz) science and technology has not been as developed

C. Kim (✉)

Department of Nano Science and Technology,
Graduate School of Convergence Science and Technology,
Seoul National University, Seoul, Korea
e-mail: changsoon@snu.ac.kr

C. Kim

Advanced Institutes of Convergence Technology, Suwon, Korea

H. Kim

Department of Electronics and Information Engineering,
College of Science and Technology, Korea University, Sejong Campus,
Jochiwon-eup, Yoengi-gun, Chungnam 339-700, Korea
e-mail: hwikim@korea.ac.kr

as that exploiting other portions of electromagnetic spectrum [8, 9]. This is mainly due to the lack of low-cost, high-performance, miniaturized THz sources, detectors, and modulators. However, extensive research efforts have recently made significant advances in the development of THz devices, where nanofabrication played crucial roles [8, 9, 10, 11, 12]. For example, in the development of a quantum cascade laser operating at 4.4 THz, the thicknesses of GaAs (10–20 nm) and AlGaAs (0.6–4.3 nm) layers composing repeating units of quantum wells were controlled to precisely engineer the inter-miniband transitions in the quantum wells [13].

In this chapter, we focus on two recent publications on THz photonics—one proposing a novel THz surface plasmon waveguide based on graphene, the other experimentally investigating the interaction between THz waves and a metallic nano-slit. First, we discuss the operational principles of the proposed waveguide (Sect. 2), and provide a brief introduction to the nano-slit experiment (Sect. 3). Then, by presenting step-by-step descriptions of fabrication processes for creating a nanostructure required for each example, we explain various micro- and nano-fabrication techniques (Sect. 4), followed by conclusion (Sect. 5).

2 Terahertz Graphene Surface Plasmon Waveguides

The Nobel Prize in Physics for 2010 was awarded to Andre Geim and Konstantin Novoselov for pioneering experimental research on graphene, a single layer of carbon atoms arranged in a honeycomb lattice. Owing to its exceptional and unusual electronic properties such as very high charge carrier mobilities ($> 15,000 \text{ cm}^2/\text{V} \cdot \text{s}$) [6, 14], anomalous quantum Hall effect [6, 14], and its charge carriers being massless Dirac Fermions [15, 16], graphene has recently been one of the most popular subject in condensed-matter physics. Also, graphene-based devices that have been demonstrated include field-effect transistors [17], solar cells [18], light-emitting devices [18], and THz sources [18]. In 2011, a proposal of using graphene as a one-atom-thick platform for THz photonic integrated circuitry was published [19], whose operational principles are discussed in this section.

Due to the complicated many body interactions of electrons in graphene, which can be described by the Kubo formula [20, 21], the complex conductivity of graphene (σ) is highly dependent on the angular frequency ω , charge carrier scattering rate Γ , temperature T , and chemical potential μ_c . In particular, it is known that the imaginary part of the graphene conductivity, $\text{Im}(\sigma)$, can be negative or positive in the THz frequencies, depending on μ_c which can be tuned by varying the electric field at the graphene layer. When $\text{Im}(\sigma) > 0$ ¹, the real part of the electric permittivity is negative, meaning that the graphene monolayer behaves as a one-atom-thick “metal” layer [19]. Therefore, a graphene monolayer sandwiched by dielectrics can

¹ We use the convention of expressing a harmonic field by a complex exponential as: $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, \omega)e^{-i\omega t}$.

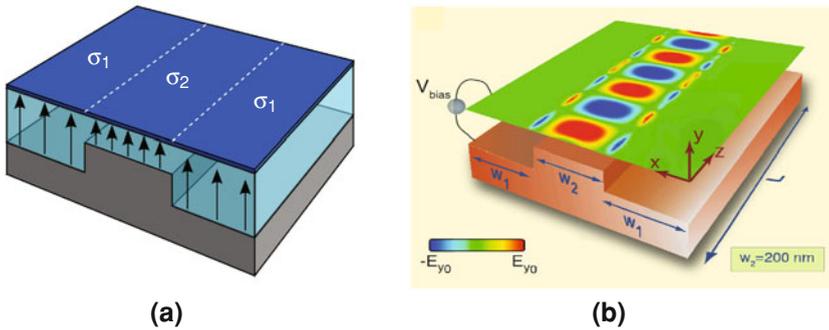


Fig. 1 **a** Terahertz (THz) graphene surface (SP) plasmon waveguide. The blue layer represents a graphene monolayer, and σ_i is the complex conductivity of each region. When a static voltage applied between the substrate (gray) and the graphene is such that $\text{Im}(\sigma_1) < 0$, $\text{Im}(\sigma_2) > 0$ and $\text{Im}(\sigma_3) > 0$, THz wave is confined in the graphene monolayer in the middle region. **b** Numerical simulation result showing the y-component of the electric field of THz wave with the frequency of 30 THz (reproduced from [19] with permission)

support transverse magnetic (TM) surface plasmon polariton (SPP) modes confined in the graphene layer, as in a conventional metal slab waveguide [22].

In addition to the vertical (i.e., perpendicular to the graphene surface) confinement of a THz SPP wave, a lateral confinement can also be obtained, if we can locally vary μ_c across the graphene. Figure 1 shows a proposed THz waveguide consisting of a graphene monolayer placed above a conducting substrate with an uneven surface profile, and a dielectric spacer between the graphene and the substrate [19]. When a static (DC) voltage is applied between the graphene and the substrate, the electric field is varied across the graphene, creating a pattern of μ_c that resembles the substrate profile. If the bias voltage and the device geometry are such that the imaginary part of conductivity in the middle region of the graphene is positive ($\text{Im}(\sigma_2) > 0$) and that in the other regions is negative ($\text{Im}(\sigma_1) < 0$), the graphene layer behaves as a dielectric-metal-dielectric structure. As a result, this device functions as a waveguide that tightly confines a THz SPP wave, which is confirmed by numerical simulations (Fig. 1b). The waveguide can be divided into two (Fig. 2a) or bent (Fig. 2b) and still maintain the guiding property for a THz SPP wave. Furthermore, the ability to locally control σ and hence the electric permittivity in a graphene monolayer opens up the possibility for realizing metamaterials and transformation optical devices [19].

3 Terahertz Field Enhancement by a Metallic Nano-Slit

In 1998, Ebbesen et al. found that when light is incident on a thin metal film with an array of subwavelength holes, the spectrum of the zeroth-order transmission efficiency (defined as the power of the transmitted light that is collinear with the

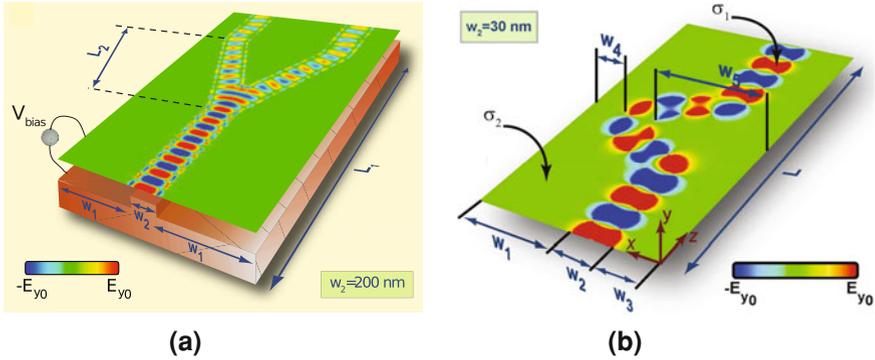


Fig. 2 THz SP MIM Y-branch (a) and bended waveguide (b). Reproduced from [19] with permission

incident light normalized to that of the incident light) exhibits sharp peaks, whose maximum, to their surprise, exceeds the area of the holes divided by the total illuminated area [5]. This phenomenon, referred to as extraordinary optical transmission (EOT), has stimulated extensive experimental and theoretical studies on the enhanced transmission characteristics of a metal film with a subwavelength aperture or its array. The enhanced transmission has been explained by resonant excitation of surface plasmons [23], Fabry-Perot-type resonance [24], and dynamical diffraction theory [25].

In most studies conducted on the enhanced transmission, metal films have feature sizes such as film thickness, hole diameter, and slit width that are similar to or larger than the skin depth.² The same is true for other photonics applications of metallic micro or nanostructures, such as transformation optics [27], metamaterials [28], superlens [29], solar cells [30], and optical antenna [31].

Recently, Seo et al. experimentally investigated the interaction of an electromagnetic wave and a metallic nanostructure—the near-field enhancement and associated enhancement in transmission, in particular—whose feature sizes are much smaller than the skin depth [32]. Specifically, they measured the transmission of a THz wave through a free-standing multilayer consisting of $1.2 \mu\text{m SiO}_2/0.5 \mu\text{m SiN}/60 \text{ nm Au}$ with the width of a slit in the Au varied from $14 \mu\text{m}$ down to 70 nm (Fig. 3). The measurement was performed by THz time-domain spectroscopy discussed in Chaps. 3 and 12. The results show that the near-field enhancement spectra obtained for $f = 0.1 - 1.1 \text{ THz}$ exhibit $1/f$ -dependence, whose maximum value of ~ 800 was measured for the 70-nm -wide slit.

The $1/f$ -dependence can be explained by the following simple argument. When the metal film thickness is smaller than the skin depth, an incident electromagnetic wave induces the electric current flowing *inside* the metal film, leading to the

² The skin depth is the distance that an electromagnetic wave travels in a metal before its amplitude is reduced by a factor of $1/e$ [26].

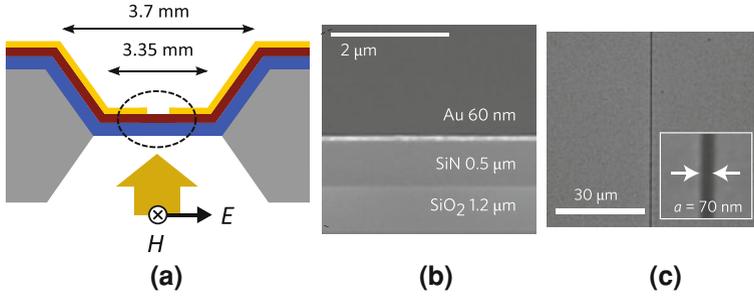


Fig. 3 **a** Cross-section schematic of the free-standing multilayer of $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{Au}$, where a nano-slit is defined in the Au layer. *Large arrow* represents the propagating direction of incident THz waves. Scanning electron microscope (SEM) images showing the cross-section **b** and top **c** views (reproduced from [32] with permission)

charge accumulation at the slit edges. The accumulated charge creates the electric field in the slit, whose magnitude increases as the slit width decreases. In fact, the Au film with a slit can be viewed as a capacitor that is charged by an incident THz wave. When the electric current induced by the incident wave is $I = I_0 e^{-i\omega t}$, the amount of the accumulated charge is

$$Q = \frac{I_0}{i\omega} e^{-i\omega t}, \tag{1}$$

which is inversely proportional to f like the field enhancement factor.

This work showed that a THz wave with a wavelength λ can be highly concentrated, by a factor of ~ 1000 , in a nano-gap whose width and thickness are $\sim \lambda/30,000$. It also raised an interesting question as to what the ultimate enhancement factor would be as the gap width keeps decreasing further. The free-standing Au film can be used to study THz nonlinear processes and realize ultra-high-sensitivity detection of nano-sized particles. The nano-gap field enhancement provided a physical basis for recent demonstration of a metamaterial with a ultra-high refractive index at THz frequencies [33].

4 Fabrication of Nanostructures for Terahertz Applications

To realize the devices discussed in the previous sections, successful employment of fabrication techniques is of critical importance. Some of the requirements that need to be satisfied for the graphene waveguide are: (i) the feature of the conducting substrate—the width of the raised strip in particular—must be accurately patterned; (ii) the top surface of the spacer layer must be flat; (iii) a defect-free graphene monolayer must be formed on top of the space layer. For the THz field enhancement experiment, the free-standing multilayer of $\text{SiO}_2/\text{SiN}/\text{Au}$ must be

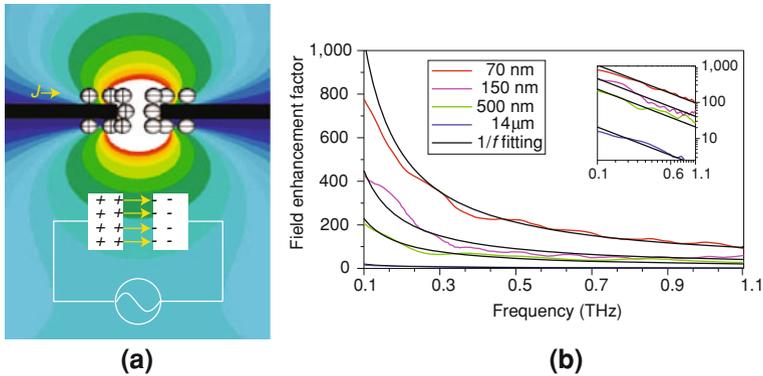


Fig. 4 **a** Electric current (J) induced by an incident THz wave leads to charge accumulation at the slit edges, which in turn enhances the electric field in the slit. Shown below is an equivalent circuit model. **b** Field enhancement factor versus frequency of THz waves for different slit geometries: (width, height) = (70, 60 nm), (150, 150 nm), (500, 60 nm), and (14, 17 μm), for *red*, *pink*, *green*, and *blue* traces, respectively. The *black lines* represents $1/f$ -fits to the traces. Inset: log–log plot of the field enhancement factor. Reproduced from [32] with permission

formed, and a slit whose width is smaller than 100 nm must be to be precisely defined in the Au (Fig. 4).

The most common method of fabricating structures with feature sizes on the micrometer or nanometer scales is by successive applications of thin-film formation and its selective removal processes. This method is often called the top-down approach. You may think that the range of structures that can be made using this approach is quite limited. However, the combinations of thin-film formation and its selective (i.e., patterned) removal can create fairly complex structures, and are the basis for manufacturing of commercial applications, such as central processing units (CPUs) in modern computers and microelectromechanical systems (MEMS) devices [34, 35]. In this section, we discuss various techniques for the formation of thin films and their patterning by presenting step-by-step descriptions of fabrication processes for the devices described in the previous sections.³

4.1 Terahertz Graphene Surface Plasmon Waveguides

In this section, we will go through processes to fabricate a structure shown in Fig. 1. The overall process is shown in Fig. 5. We start with a piece of highly doped silicon wafer, which is the substrate for the waveguide structure. A silicon

³ The sequence of fabrication processes in Sect. 4.1 is a possible method to experimentally realize the device proposed in Ref. [19], and that in Sect. 4.2 may differ in detail from the actual processes used in Ref. [32].

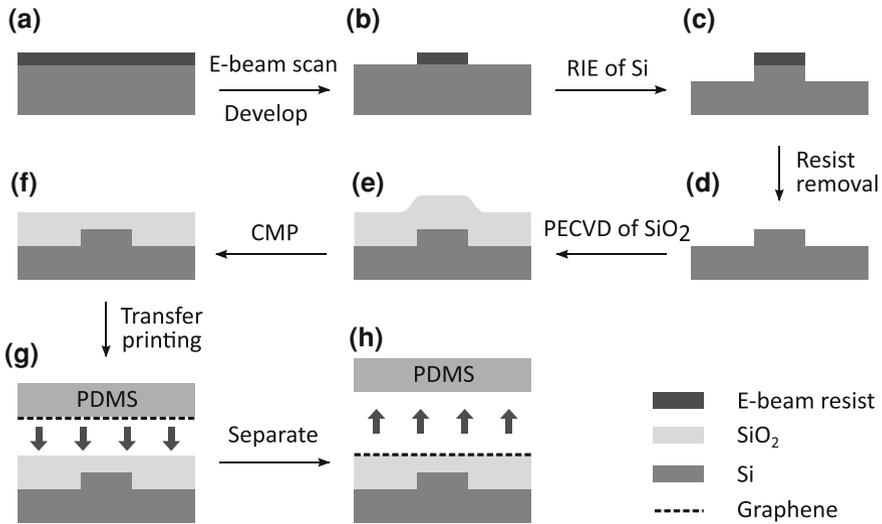


Fig. 5 Sequence of processes to fabricate a THz grapheme waveguide proposed in Ref. [19]. *Step 1* Formation of a thin-film of electron-beam resist on the substrate by spin coating, (a). *Step 2* Electron-beam lithography to pattern the resist layer (b) → (c). *Step 3* Reactive ion etch to define the raised strip in Si, (c) → (d). *Step 4* Plasma-enhanced chemical vapor deposition of SiO₂ on the patterned Si, (d) → (e), followed by chemical-mechanical polishing for planarization, (e) → (f). *Step 5* CVD growth of a graphene monolayer on a separate substrate and its transfer by a stamping process onto the waveguide substrate, (f) → (h)

wafer is typically about 500- μ m thick and atomically flat. First, we need to pattern the Si wafer into a raised strip as shown in Fig. 1. This type of patterning is commonly achieved by a process called **photolithography** [36, 37]. The first step of photolithography is to form (or deposit) a thin film of a light-sensitive polymer, called photoresist, on a substrate. In most cases, the photoresist is deposited by **spin coating**, where a polymer solution is dispersed onto the substrate, which is subsequently spun at a high angular speed (> 1000 rpm) [37]. As the spin speed is maintained, the polymer solution is distributed across the surface and solvent evaporates, resulting in a uniform thin film.

→ *Step 1* Formation of a thin-film of photoresist on the substrate by spin coating.

The next step is to create a pattern in the uniformly coated photoresist. The photoresist is exposed to a pattern of ultraviolet (UV) light, by irradiating the sample through a photomask that is in contact with or in close proximity to the photoresist, or by projecting an image of the photomask onto the photoresist by an optical system (Fig. 6). The UV exposure induces a chemical change in the polymer, and therefore creates across the polymer film a contrast in its solubility in a specific solvent (called developer). The pattern in the photoresist is then obtained by immersing the sample in a developer, where in a positive photoresist, the polymer in the exposed area is dissolved away, while in a negative photoresist, it

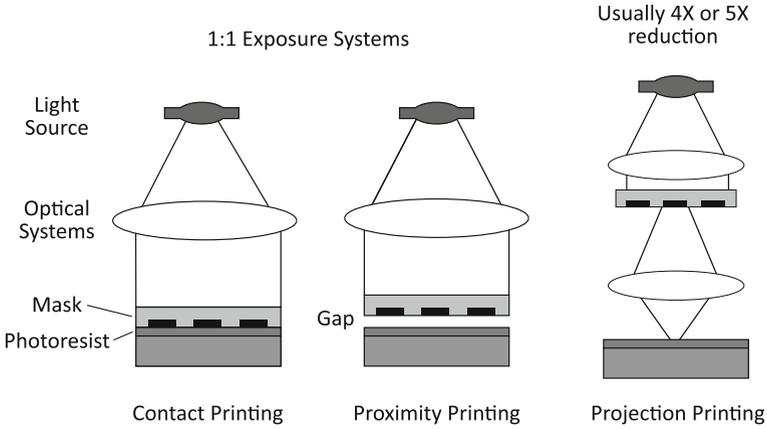


Fig. 6 UV exposure in photolithography (reproduced from [37] with permission)

remains on the substrate. The minimum feature size that can be patterned by photolithography is determined by optical diffraction and is approximately the wavelength of UV light that is used. Therefore, for the waveguide fabrication, we have to use a photolithography system equipped with a light source whose emission wavelength (λ) is shorter than ~ 200 nm—for example, an ArF excimer laser with $\lambda = 193$ nm.

Alternatively, we can use **electron-beam lithography**. Electron-beam (e-beam) lithography is similar to photolithography, but uses electrons, instead of photons, to expose a resist (in this case, called e-beam resist) layer without a mask [38, 39]. Electrons, which are emitted by thermionic or field emission, are accelerated by a DC voltage, and a focused electron beam is scanned across the resist layer. Since the de Broglie wavelength of an electron is given by

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m_0qV}}, \quad (7.2)$$

where h is the Planck constant, p is the electron momentum, m_0 is the electron rest mass, q is the elementary charge, and V is the applied acceleration voltage, the resolution is not limited by diffraction; $V = 10$ keV corresponds to $\lambda \simeq 0.01$ nm. Rather, it is mainly affected by the so-called proximity effect, arising from scattering processes between energetic, incident electrons and the resist materials [38, 39]. Feature sizes as small as 10 nm are routinely fabricated using today's e-beam lithography system. Unlike photolithography, an e-beam writing process is serial, presenting the major drawback of low throughput. However, since e-beam lithography can generate arbitrary patterns with high resolution in a maskless fashion, it has become the most frequently used nanolithographic tool for many science and engineering applications. For this reason, we continue to discuss our fabrication processes with e-beam lithography:

- Step 1 (revised) Formation of a thin-film of *electron-beam resist* on the substrate by spin coating, (a) in Fig. 5.
- Step 2 Electron-beam lithography to pattern the electron-beam resist layer, (a) → (b) in Fig. 5.

Now, we can transfer the photoresist pattern into the underlying material (in this case, silicon) by **etching** it with the photoresist pattern serving as an etch mask, as shown in Fig. 5c [37]. Etching can be classified as wet or dry. In wet etching, the sample is immersed into a solution containing reactive species (called etchant), which converts by chemical processes the material to be etched into a soluble species. Dry etching uses highly reactive gaseous species or plasma, and the etching mechanism is chemical, physical (sputtering off the material to be etched by bombarding it with energetic species, usually ions), or combination of both. Most wet etching processes are isotropic, meaning that the etch rate in different directions is the same (we will see an exception in Sect. 4.2). Therefore, to define the raised strip with vertical sidewalls as shown in Fig. 1, we employ a reactive ion etching (RIE), a dry etching process where very reactive ionic species are produced in a plasma and anisotropic etching is achieved by DC electric field that vertically delivers them onto the sample surface [37, 38].

- Step 3 Reactive ion etch to define the raised strip in Si, (b) → (c) in Fig. 5.

A 200-nm-thick layer of silicon dioxide is then formed on the sample. There are several techniques to form a thin layer, and which technique to choose mainly depends on the material to be deposited, the layer thickness, and the desired degree of crystallinity. In **evaporation** process, the source materials are heated above their boiling or sublimation point in a vacuum chamber (pressure $< 10^{-5}$ Torr), and the evaporated atoms or molecules are condensed on the sample which is placed above the source materials [37, 39]. One method to heat a source material is resistive heating, where a source container is heated by passing electrical currents through it or a filament surrounding the container. This process is called **thermal evaporation**, which can be used to deposit a layer of metals (e.g. Ag, Au, Al, Mg), and organic molecules (e.g. pentacene [40], ZnPc [40], Alq₃ [41, 42]). Another heating mechanism that is commonly used is electron bombardment. In **electron-beam evaporation**, high-energy electrons are directed onto the source material by magnetic fields, and heat up the source material by dissipating their kinetic energy [37, 39]. Since this process can deliver higher thermal power to the source material than resistive heating, electron-beam evaporation has a wider range of applicable materials including Pt, Ni, Ge, MgF₂, and SiO₂.

Thin-layer formation by evaporation proceeds in two steps, evaporation and condensation, which are both physical processes in a sense that materials being deposited undergo phase changes only, from solid to vapor, and then to solid. Hence, evaporation is regarded as a physical process. In **chemical vapor deposition** (CVD), gaseous species are introduced to a chamber to participate in chemical reactions, whose final product is the material to be deposited [37, 39]. Chemical reactions take place on the sample surface, which is typically heated to

provide the gaseous reactants with sufficient thermal energy required for the chemical reactions to occur. To facilitate the chemical reactions, sometimes a plasma is created, where the reactant gases are converted into highly reactive species including ionized molecules or atoms, and free radicals. This technique is called **plasma-enhanced chemical vapor deposition** (PECVD), and it is widely used in semiconductor manufacturing [37]. We use PECVD to deposit the 200-nm-thick SiO₂ layer on the sample (Fig. 5e). During the process, deposition conditions, such as the gas flow rates, substrate temperature, and chamber pressure, need to be controlled so that a dense, defect-free, conformal film can be obtained. Since the thickness of the SiO₂ layer is comparable to the height of the ridge, the top SiO₂ is not going to be flat. Considering that we need to have a monolayer of graphene on top of the SiO₂ layer, our sample needs to be planarized. This planarization issue is very often encountered in multilayer fabrication of silicon integrated circuits and MEMS devices. A process called chemical-mechanical polishing (CMP) [37] is commonly employed for this purpose (Fig. 5f).

→ Step 4 Plasma-enhanced chemical vapor deposition of SiO₂ on the patterned Si, followed by chemical-mechanical polishing for planarization, (d) → (e) → (f) in Fig. 5.

Now, we are done except the single layer of graphene on top. Rather than depositing the graphene monolayer directly on the Si/SiO₂ structure, we grow it on a separate substrate and transfer it onto the SiO₂ layer (Fig. 7). The reason for this will be clear as we discuss the process. We first obtain a monolayer of graphene using CVD, where a single or a few layers of graphene is synthesized from CH₄ and H₂ on a Ni-coated substrate (e.g. Si/SiO₂/Ni) placed in a hot wall furnace (~1000°C). Other techniques that may be used include mechanical exfoliation of graphite crystals [6], epitaxial growth on silicon carbide [43], and chemical [44, 45] or electrochemical [46] reduction of graphene oxide, but the CVD method is well-suited for monolayer growth over a large area (~2 cm × 2 cm). The graphene monolayer can then be transferred onto the Si/SiO₂ structure that we obtained in Step 4 (target substrate) by the following method: (i) Polydimethylsiloxane (PDMS) prepolymer is poured over the Ni-coated substrate and is cured by UV irradiation (Fig. 7c); (ii) the PDMS layer, along with the graphene monolayer attached to it, is released from the substrate by wet etching the Ni layer (Fig. 7d); (iii) the graphene layer is transferred onto the target substrate by inducing an ‘intimate contact’ between the substrate and graphene (Fig. 5g), followed by separation of the PDMS from the substrate (Fig. 5h). It is van der Waals interactions between graphene and SiO₂ that enable the transfer of the graphene monolayer. Hence, the ‘intimate contact’ means that the graphene needs to be brought to an atomically close proximity to the SiO₂ surface, which is facilitated by low Young’s modulus (~3 MPa) of PDMS. The transfer of the graphene monolayer by the stamping process completes the fabrication of the graphene waveguide structure proposed in Ref. [19].

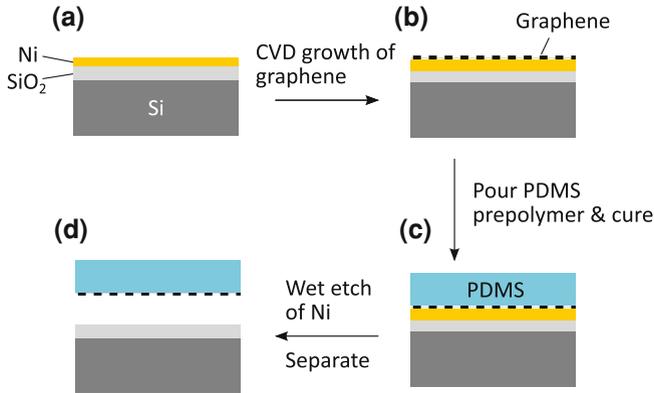


Fig. 7 Schematic of the processes to obtain a PDMS stamp with a graphene monolayer. A graphene monolayer CVD-grown on a Ni-coated substrate is lifted off with PDMS upon wet etching of the Ni layer

→ Step 5 CVD growth of a graphene monolayer on a separate substrate and its transfer by a stamping process onto the waveguide substrate obtained in Step 4, (f) → (g) → (h) in Fig. 5.

You may think that the method of transferring the graphene monolayer described above is somewhat low-tech; it does not involve sophisticated instruments and resembles an ordinary stamping process. However, owing to this simplicity, along with the fact that transfer relies only on van der Waals (or in other cases covalent or metallic) bond formation occurring at room temperature, many conceptually similar patterning techniques have been demonstrated and applied to a wide range of areas such as inorganic [47, 48, 49] or organic electronics [50, 51, 52], and cell biology [53]. The earliest demonstration is perhaps **microcontact printing**, where molecules such as alkanethiols are transferred from a PDMS stamp with a relief pattern onto a substrate to form a self-assembled monolayer (SAM) in areas contacted by the stamp [53]. The patterned SAM can be used as a resist in wet etching or a template for a selective deposition of materials including organic semiconductors [50], conducting polymers, metals, and proteins. In **cold-welding techniques** [51, 52], a thin film of metal is transferred from a stamp to a substrate [51, 54, 55] or from a substrate to a stamp [51] via metallic bond formation between metal atoms on the stamp and substrate surfaces. Since cold-welding is a room temperature, dry process, this technique can be used to pattern metallic electrodes *on top of* organic semiconductors, which can easily be dissolved in solvents or degraded during a high-temperature process. In other technique, termed **nano-transfer printing**, a surface of a stamp and/or a substrate was modified by forming a SAM or by an oxygen plasma treatment to tailor the adhesion strength between the contacting surfaces [54, 55]. More recently, layers

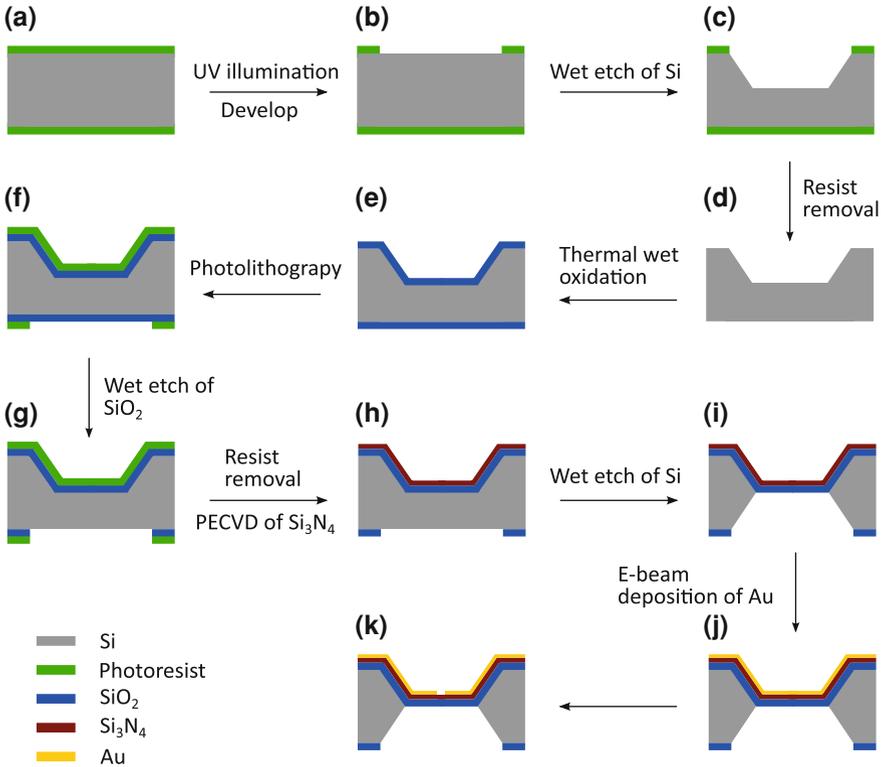


Fig. 8 Sequence of processes for the fabrication of nano-slit in Au, summarized in Table 1

of inorganic semiconductors were transferred onto otherwise incompatible substrates to demonstrate monocrystalline Si solar cells on PDMS [48] and curved glass [47], GaAs-based field effect transistors on glass [49] and near-infrared imager on Si [49], and inorganic light-emitting diodes based on AlInGaP quantum well structures on glass [56].

4.2 Metallic Nano-Slit for Terahertz Field Enhancement

The structure that we consider in this section is shown in Fig. 3, where a 70-nm-wide slit is defined in a 60-nm-thick Au layer on a thin membrane consisting of SiO₂/Si₃N₄. The fabrication process is briefly outlined in Ref. [32], whose details are described in Fig. 8 and Table 1. If you are familiar with processes discussed in Sect. 4.1, only Steps 3 (and 9), 5, and 11 need to be explained.

In Step 3 (Fig. 8b → c) and Step 9 (Fig. 8h → i), about a half the thickness of the Si wafer, $\sim 250 \mu\text{m}$, needs to be etched away, which requires a special

Table 1 Processing steps for the fabrication of nano-slit in Au

Step 1	Deposition of photoresist layers on both sides of a Si wafer by spin coating
Step 2, (a) → (b) ^a	Photolithography on the top photoresist to define a 3.7-mm-wide opening
Step 3, (b) → (c)	Anisotropic wet etching of Si in KOH to define a groove whose bottom face is 3.35-mm wide
Step 4, (c) → (d)	Removal of the photoresist layers by immersion in solvent
Step 5, (d) → (e)	Wet oxidation of Si to grow a 1.2- μm -thick SiO_2 layer on both sides
Step 6, (e) → (f)	Deposition of photoresist layers on both sides of the sample, and photolithography on the bottom side to define a 3.7-mm-wide opening like in Step 2
Step 7, (f) → (g)	Wet etching of the bottom SiO_2 layer in a diluted HF solution
Step 8, (g) → (h)	Photoresist removal and deposition of a 0.5- μm -thick Si_3N_4 layer on the top surface by PECVD
Step 9, (h) → (i)	Anisotropic wet etching of Si on the backside in KOH so that the bottom surface of the SiO_2 is exposed
Step 10, (i) → (j)	Deposition of a 60-nm-thick Au on the SiN by e-beam evaporation
Step 11, (j) → (k)	Focused ion beam milling of the Au layer to define the 70-nm-wide nanoslit

^a refers to figures in Fig. 8

consideration in selecting an etch process; (i) a process with high etch rate is desired; (ii) etch selectivity (\equiv etch rate of the material to be etched / that of the mask material) must be sufficiently high. Since the sidewall profile is not of a concern in this case, we use a wet etching in KOH with an appropriate mask layer (photoresist⁴ in Step 3, and SiO_2 in Step 9), which etches Si at different rates depending on crystallographic directions, leading to a tilted sidewall. Wet etching of Si in KOH is commonly used to form a v-groove on a Si surface or a free standing film whose thickness is less than 1 μm as in this example.

You may think that the formation of 1.2- μm -thick SiO_2 layers in Step 5 can be done using PECVD as in Sect. 4.1. Alternatively, they can be grown by wet thermal oxidation, where SiO_2 is formed via a chemical reaction involving Si and water vapor in a hot furnace: $\text{Si} + 2\text{H}_2\text{O} \rightarrow \text{SiO}_2 + 2\text{H}_2$. A SiO_2 layer formed in this way tends to be denser than that obtained by PECVD, and therefore it is a better etch mask for the KOH-etch of Si.

The free-standing film of $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{Au}$, whose total thickness is only 1,700 nm, is too fragile that e-beam lithography followed by dry etching of the Au may fracture the film. Therefore, a technique with a minimal number of processing steps that does not involve the use of a liquid is desired. In focused ion beam (FIB) lithography, which is conceptually similar to e-beam lithography, a focused beam of ions such as Ga^+ , Be^{2+} , and Si^{2+} is scanned across a resist layer, which is

⁴ Since the etch time is quite long (>1 h), the photoresist may delaminate from the substrate near the pattern boundary, leading to a larger opening than is defined by the photoresist pattern. This is acceptable since the lateral dimension is not very critical for the current application.

subsequently developed. In addition, the heavy mass of ions allows FIB lithography to be used to physically sputter off material from a sample surface. This process, called FIB milling, is a one-step process performed in a dry environment, which we use to define the nanogap in the Au (Step 11).

5 Conclusion

In this chapter, we presented discussions of micro- and nanofabrication techniques by going through fabrication processes to create nanostructures relevant to two recent publications in the field of THz science and technology. We did not intend our discussion to be comprehensive; instead, we aimed to provide readers with ideas on how various fabrication techniques are successively employed to create nanostructures required to realize novel devices or to experimentally seek answers to scientific questions. Important developments in nanofabrication, such as nanoimprint lithography, scanning probe lithography, and self assembly, are not covered in this chapter. For more thorough treatment of the subject, readers are encouraged to refer to review articles [36, 57] and monographs [37–39].

References

1. Painter, O.: Two-dimensional photonic band-gap defect mode laser. *Science* **284**(5421), 1819–1821 (1999)
2. Russell, P.: Photonic crystal fibers. *Science* **299**(5605), 358–362 (2003)
3. Craighead, H.: Future lab-on-a-chip technologies for interrogating individual molecules. *Nature* **442**(7101), 387–393 (2006)
4. Masmanidis, S.C., Karabalin, R.B., De Vlaminck, I., Borghs, G., Freeman, M.R., Roukes, M.L.: Multifunctional nanomechanical systems via tunably coupled piezoelectric actuation. *Science* **317**(5839), 780–783 (2007)
5. Ebbesen, T.W., Lezec, H.J., Ghaemi, H.F., Thio, T., Wolff, P.A.: Extraordinary optical transmission through sub-wavelength hole arrays. *Nature* **391**(6668), 667–669 (1998)
6. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Katsnelson, M.I., Grigorieva, I.V., Dubonos, S.V., Firsov, A.A.: Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**(7065), 197–200 (2005)
7. Park, H., Park, J., Lim, A.K., Anderson, E.H., Alivisatos, A.P., McEuen, P.L.: Nanomechanical oscillations in a single-C60 transistor. *Nature* **407**(6800), 57–60 (2000)
8. Ferguson, B., Zhang, X.-C.: Materials for terahertz science and technology. *Nat. Mater.* **1**(1), 26–33 (2002)
9. Tonouchi, M.: Cutting-edge terahertz technology. *Nat. Photonics* **1**(2), 97–105 (2007)
10. Knap, W., Nadar, S., Videlier, H., Boubanga-Tombet, S., Coquillat, D., Dyakonova, N., Teppe, F., Karpierz, K., Usakowski, J., Sakowicz, M., Kasalynas, I., Seliuta, D., Valusis, G., Otsuji, T., Meziani, Y., Fatimy, A., Vandenbrouk, S., Madjour, K., Théron, D., Gaquiere, C.: Field effect transistors for terahertz detection and emission. *J. Infrared Millim. Terahertz Waves* **32**(5), 618–628 (2011)
11. Komiyama, S., Astafiev, O., Antonov, V., Kutsuwa, T., Hirai, H.: A single-photon detector in the far-infrared range. *Nature* **403**(6768), 405–407 (2000)

12. Otsuji, T., Watanabe, T., Moutaouakil, A., Karasawa, H., Komori, T., Satou, A., Suemitsu, T., Suemitsu, M., Sano, E., Knap, W., Ryzhii, V.: Emission of terahertz radiation from two-dimensional electron systems in semiconductor nano- and hetero-structures. *J. Infrared Millim. Terahertz Waves* **32**(5), 629–645 (2011)
13. Köhler, R., Tredicucci, A., Beltram, F., Beere, H.E., Linfield, E.H., Davies, A.G., Ritchie, D.A., Iotti, R.C., Rossi, F.: Terahertz semiconductor heterostructure laser. *Nature* **417**(6885), 156–159 (2002)
14. Zhang, Y., Tan, Y.-W., Stormer, H.L., Kim, P.: Experimental observation of the quantum hall effect and Berry's phase in graphene. *Nature* **438**(7065), 201–204 (2005)
15. Castro Neto, A.H., Peres, N.M.R., Novoselov, K.S., Geim, A.K.: The electronic properties of graphene. *Rev. Mod. Phys.* **81**(1), 109–162 (2009)
16. Geim, A.K., Novoselov, K.S.: The rise of graphene. *Nat. Mater.* **6**(3), 183–191 (2007)
17. Schwierz, F.: Graphene transistors. *Nat. Nanotechnology* **5**(7), 487–496 (2010)
18. Bonaccorso, F., Sun, Z., Hasan, T., Ferrari, A.C.: Graphene photonics and optoelectronics. *Nat. Photonics* **4**(9), 611–622 (2010)
19. Vakil, A., Engheta, N.: Transformation optics using graphene. *Science* **332**(6035), 1291–1294 (2011)
20. Gusynin, V.P., Sharapov, S.G.: Transport of dirac quasiparticles in graphene: hall and optical conductivities. *Phys. Rev. B* **73**(24), (2006)
21. Gusynin, V.P., Sharapov, S.G., Carbotte, J.P.: Magneto-optical conductivity in graphene. *J. Phys. Condens. Matter* **19**(2), 026222 (2007)
22. Dionne, J., Sweatlock, L., Atwater, H., Polman, A.: Planar metal plasmon waveguides: frequency-dependent dispersion, propagation, localization, and loss beyond the free electron model. *Phys. Rev. B* **72**(7), 1–11 (2005)
23. Genet, C., Ebbesen, T.W.: Light in tiny holes. *Nature* **445**(7123), 39–46 (2007)
24. Weiner, J.: The physics of light transmission through subwavelength apertures and aperture arrays. *Rep. Prog. Phys.* **72**(6), 064401 (2009)
25. Treacy, M.: Dynamical diffraction explanation of the anomalous transmission of light through metallic gratings. *Phys. Rev. B* **66**(19), 1–11 (2002)
26. Griffiths, D.J.: *Introduction to Electrodynamics*, 3rd edn. Benjamin Cummings (1999)
27. Chen, H., Chan, C.T., Sheng, P.: Transformation optics and metamaterials. *Nat. Mater.* **9**(5), 387–396 (2010)
28. Chen, H.-T., Padilla, W.J., Zide, J.M.O., Gossard, A.C., Taylor, A.J., Averitt, R.D.: Active terahertz metamaterial devices. *Nature* **444**(7119), 597–600 (2006)
29. Fang, N., Lee, H., Sun, C., Zhang, X.: Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308**(5721), 534–537 (2005)
30. Atwater, H.A., Polman, A.: Plasmonics for improved photovoltaic devices. *Nat. Mater.* **9**(3), 205–213 (2010)
31. Mühlischlegel, P., Eisler, H.-J., Martin, O.J.F., Hecht, B., Pohl, D.W.: Resonant optical antennas. *Science* **308**(5728), 1607–1609 (2005)
32. Seo, M.A., Park, H.R., Koo, S.M., Park, D.J., Kang, J.H., Suwal, O.K., Choi, S.S., Planken, P.C.M., Park, G.S., Park, N.K., Park, Q.H., Kim, D.S.: Terahertz field enhancement by a metallic nano slit operating beyond the skin-depth limit. *Nat. Photonics* **3**(March), (2009)
33. Choi, M., Lee, S.H., Kim, Y., Kang, S.B., Shin, J., Kwak, M.H., Kang, K.-Y., Lee, Y.-H., Park, N., Min, B.: A terahertz metamaterial with unnaturally high refractive index. *Nature* **470**(7334), 369–373 (2011)
34. Judy, J.W.: *Microelectromechanical systems (MEMS): fabrication, design and applications*. *Smart Mater. Struct.* **10**, 1115 (2001)
35. Senturia, S.D.: *Microsystem Design*. Springer, New York (2000)
36. Geissler, M., Xia, Y.: Patterning: principles and some new developments. *Adv. Mater.* **16**(15), 1249–1269 (2004)
37. Plummer, J.D., Deal, M.D., Griffin, P.B.: *Silicon VLSI Technology: Fundamentals, Practice and Modeling*, 1st edn. Prentice Hall (2000)
38. Cui, Z.: *Nanofabrication: Principles, Capabilities and Limits*, 1st edn. Springer (2010)

39. Köhler, M., Fritzsche, W.: *Nanotechnology: an introduction to nanostructuring techniques*, 2nd edn. Wiley-VCH (2007)
40. Hoppe, H., Sariciftci, N.S.: Organic solar cells: an overview. *J. Mater. Res.* **19**(7), 1924–1945 (2004)
41. Baldo, M.A., Forrest, S.R.: Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395**(September), 151–154 (1998)
42. Tang, C.W., VanSlyke, S.A.: Organic electroluminescent diodes. *Appl. Phys. Lett.* **51**(12), 913 (1987)
43. Ohta, T., Bostwick, A., Seyller, T., Horn, K., Rotenberg, E.: Controlling the electronic structure of bilayer graphene. *Science* **313**(5789), 951–954 (2006)
44. Eda, G., Lin, Y.-Y., Miller, S., Chen, C.-W., Su, W.-F., Chhowalla, M.: Transparent and conducting electrodes for organic electronics from reduced graphene oxide. *Appl. Phys. Lett.* **92**(23), 233305 (2008)
45. Wang, X., Zhi, L., Müllen, K.: Transparent, conductive graphene electrodes for dye-sensitized solar cells. *Nano Lett.* **8**(1), 323–327 (2008)
46. Guo, H.L., Wang, X.F., Qian, Q.Y., Wang, F.B.: A green approach to the synthesis of graphene nanosheets. *ACS Nano* **3**(9), 2653–2659 (2009)
47. Meitl, M.A., Zhu, Z.-T., Kumar, V., Lee, K.J., Feng, X., Huang, Y.Y., Adesida, I., Nuzzo, R.G., Rogers, J.A.: Transfer printing by kinetic control of adhesion to an elastomeric stamp. *Nat. Mater.* **5**(1), 33–38 (2005)
48. Yoon, J., Baca, A.J., Park, S.-I., Elvikis, P., Geddes, J.B., Li, L., Kim, R.H., Xiao, J., Wang, S., Kim, T.-H., Motala, M.J., Ahn, B.Y., Duoss, E.B., Lewis, J.A., Nuzzo, R.G., Ferreira, P.M., Huang, Y., Rockett, A., Rogers, J.A.: Ultrathin silicon solar microcells for semitransparent, mechanically flexible and microconcentrator module designs. *Nat. Mater.* **7**(11), 907–915 (2008)
49. Yoon, J., Jo, S., Chun, I.S., Jung, I., Kim, H.-S., Meitl, M., Menard, E., Li, X., Coleman, J.J., Paik, U., Rogers, J.A.: GaAs photovoltaics and optoelectronics using releasable multilayer epitaxial assemblies. *Nature* **465**(7296), 329–333 (2010)
50. Briseno, A.L., Mannsfeld, S.C.B., Ling, M.M., Liu, S., Tseng, R.J., Reese, C., Roberts, M.E., Yang, Y., Wudl, F., Bao, Z.: Patterning organic single-crystal transistor arrays. *Nature* **444**(7121), 913–917 (2006)
51. Kim, C., Burrows, P.E., Forrest, S.R.: Micropatterning of organic electronic devices by cold-welding. *Science* **288**(5467), 831–833 (2000)
52. Kim, C., Shtein, M., Forrest, S.R.: Nanolithography based on patterned metal transfer and its application to organic electronic devices. *Appl. Phys. Lett.* **80**(21), 4051 (2002)
53. Whitesides, G.M., Ostuni, E., Takayama, S., Jiang, X., Ingber, D.E.: Soft lithography in biology and biochemistry. *Annu. Rev. Biomedical Eng.* **3**, 335–373 (2001)
54. Loo, Y.-L., Willett, R.L., Baldwin, K.W., Rogers, J.A.: Additive, nanoscale patterning of metal films with a stamp and a surface chemistry mediated transfer process: applications in plastic electronics. *Appl. Phys. Lett.* **81**(3), 562 (2002)
55. Loo, Y.-L., Willett, R.L., Baldwin, K.W., Rogers, J.A.: Interfacial chemistries for nanoscale transfer printing. *J. Am. Chem. Soc.* **124**(26), 7654–7655 (2002)
56. Park, S.-I., Xiong, Y., Kim, R.-H., Elvikis, P., Meitl, M., Kim, D.-H., Wu, J., Yoon, J., Yu, C.-J., Liu, Z., Huang, Y., Hwang, K.-C., Ferreira, P., Li, X., Choquette, K., Rogers, J.A.: Printed assemblies of inorganic light-emitting diodes for deformable and semitransparent displays. *Science* **325**(5943), 977–981 (2009)
57. Gates, B.D., Xu, Q., Stewart, M., Ryan, D., Willson, C.G., Whitesides, G.M.: New approaches to nanofabrication: molding, printing, and other techniques. *Chem. Rev.* **105**(4), 1171–1196 (2005)

Sub-single Cycle Pulses of Electromagnetic Radiation

Minwoo Yi and Jaewook Ahn

Abstract Sub-single cycle electromagnetic radiation has been invented a couple of decades ago in an experiment for the interaction of ultrafast optical pulses with semiconductor materials. During the surge of optically created photocarriers in an electrically-biased switch, a highly asymmetric electric pulse with a peak electric amplitude as high as 100 kV/cm has been produced. Also, it has been known that an optical nonlinear process of difference frequency mixing also generates various shapes of THz pulses. A typical THz pulse is composed of a sharp electric pulse with a given polarity determined by the direction of the current surge followed by a long tail of opposite polarity. As the duration of the sharp part of the pulse is significantly shorter than the duration of the tail part, the shape of the pulse is nearly a half-cycle pulse. In this part, we present a short discussion of the generation, detection, and propagation of the sub-single-cycle THz pulses.

1 Introduction

The generation of electromagnetic radiation from a large-area photo-conductive antenna is illustrated in Fig. 1. The metallic structure of an electrically biased antenna is drawn on a semiconducting substrate material of which the energy band gap is below the photon energy of an illuminating light source. Then, the electrical surge is triggered by light, producing a unipolar electric pulse. That is the THz

M. Yi · J. Ahn (✉)

Department of Physics, KAIST, Guseong-dong, Yuseong-gu 373-1,
Daejeon, 305-701 Korea
e-mail: jwahn@kaist.ac.kr

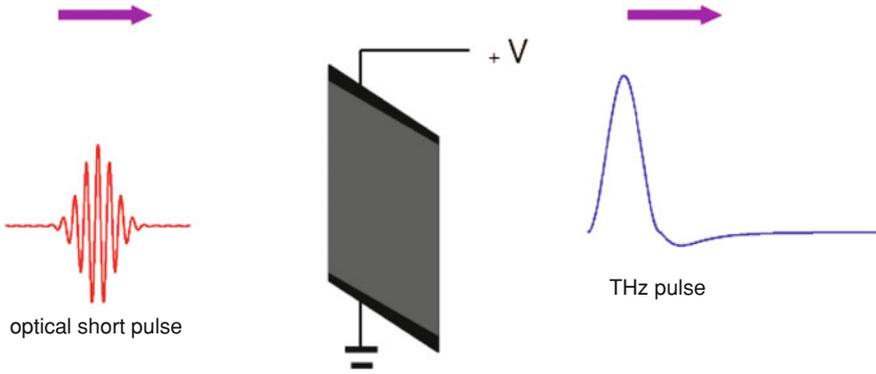


Fig. 1 Half-cycle electric pulse generation from a photoconductive antenna

pulse, or the electrical short pulse in the terahertz frequency range, which is the subject of the discussion in this part.

The generated THz pulse is temporally short and spectrally broad: The temporal width is 150 fs and the frequency bandwidth is $\Delta f_{\text{THz}} = 2$ THz in a typical experiment. Especially the spectral feature of the THz pulse may be considered as “extremely broadband,” because, in a THz pulse, the frequency bandwidth may be equivalent to or often exceeds the carrier frequency (f_{THz}), or

$$\Delta f_{\text{THz}}/f_{\text{THz}} \sim 1. \quad (1)$$

Therefore, the THz pulse is not simply described by its carrier frequency, which is 1 THz to say, while other electric waves are often described in terms of the carrier frequency and the envelope function.

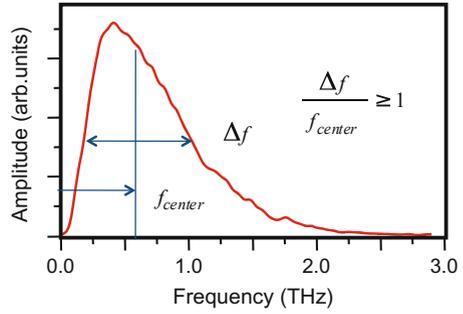
The Fourier spectrum of the generated electric pulse spans from zero up to a few THz, which is equivalent to the frequency bandwidth of the optical short pulse that is used for the switch triggering. For a 100-fs short optical pulse, the frequency bandwidth of the optical pulse is 5 THz, because the time-frequency bandwidth product is given by

$$\tau \times \Delta f = \frac{1}{2}. \quad (2)$$

for a gaussian function, where τ and Δf are the temporal width and the spectral bandwidth, respectively, of the optical pulse. Equivalently, a 10-fs ultrashort optical pulse, which has 50-THz frequency bandwidth, may generate an extremely broadband THz pulse. Indeed, a THz pulse of extremely broad bandwidth up to $\Delta f = 20$ THz has been reported [1].

Having the relationship between the temporal width and the frequency bandwidth in Eq. 2, it is necessary for a sub-single-cycle pulse $\tau < 1/f_{\text{THz}}$ to have an enough frequency bandwidth satisfying

Fig. 2 Electric field amplitude spectrum of a typical THz pulse



$$\Delta f_{\text{THz}} > \frac{1}{2} f_{\text{THz}}. \tag{3}$$

The spectrum of a typical THz pulse shown in Fig. 1, the requirement for a sub-single-cycle pulse in Eq. 3 is well satisfied. However, for a wave in optical frequency domain, this requirement is not easy to meet. This extreme broadband nature of a sub-single-cycle pulse is rather uniquely found in THz pulses, and the generation mechanism of the THz pulses may explain the reasons for this in the following section. (Fig. 2)

2 Generation of Sub-single Cycle Pulses

One may wonder what determines the direction of the unipolar main lobe and may think that the nature of unipolarity is based on the fact that the current in the photoconductive antenna surges along one direction: For example, in Fig. 1, the generated THz wave is polarized along the positive \hat{z} direction because the electric current surges along the same direction. But then think again! To understand the mechanism of THz wave generation from the electrically-biased antenna, we start from the Maxwell's equations [2], given by

$$\nabla \cdot \mathbf{E} = \frac{\rho_f(\mathbf{r}, t)}{\epsilon_0}, \tag{4}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{5}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \tag{6}$$

$$\nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}_f(\mathbf{r}, t), \tag{7}$$

where the differential equations for the electric field \mathbf{E} and the magnetic field \mathbf{B} are given in a free space where the charge and current distributions are $\rho_f(\mathbf{r}, t)$ and

$\mathbf{J}_f(\mathbf{r}, t)$, respectively. From these equations, the solutions for $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ can be formally obtained as

$$\mathbf{E}(\mathbf{r}, t) = -\nabla V(\mathbf{r}, t) - \frac{\partial \mathbf{A}(\mathbf{r}, t)}{\partial t}, \quad (8)$$

$$\mathbf{B}(\mathbf{r}, t) = \nabla \times \mathbf{A}(\mathbf{r}, t), \quad (9)$$

in terms of the scalar potential $V(\mathbf{r}, t)$ and the vector potential $\mathbf{A}(\mathbf{r}, t)$, and they are obtained as a function of the charge and current distributions, ρ_f and \mathbf{J}_f , respectively, as

$$V(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int \frac{[\rho_f]_c}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}', \quad (10)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{[\mathbf{J}_f]_c}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}', \quad (11)$$

where the retarded time notation (i.e., $[f(\mathbf{r}, t)]_c = f(\mathbf{r}, t = |\mathbf{r} - \mathbf{r}'|/c)$) is used.

So, the radiated electric field from a time varying current distribution is given as

$$\mathbf{E}(\mathbf{r}, t) = \frac{\mu_0}{4\pi} \int \frac{1}{|\mathbf{r} - \mathbf{r}'|} \frac{\partial [\mathbf{J}]_c}{\partial t} d^3\mathbf{r}', \quad (12)$$

for an electrically neutral media, $\rho_f(t) = 0$, which is the case under consideration. If we assume a homogeneity of the medium and also far-field measurement of the electric field, then we can neglect the spatial dependence and the time-retardation effect, and the electric field becomes simplified as

$$\mathbf{E}(t) \simeq \frac{\mu_0}{4\pi} \frac{V}{r} \frac{d\mathbf{J}(t)}{dt} \propto -\frac{d}{dt}(env)\hat{z}, \quad (13)$$

where n is the density of charge carriers, which are electrons in this case, and v is their drift velocity. The generated THz pulse is determined by the temporal behavior of the current density $\mathbf{J}(t)$, and, after the space integration, the current $I(t)$. So, $\mathbf{E}(t)$ is simply determined by the time dependence of the current $I(t)$. In other words, the emitted THz wave is determined by the electric current surge in the optically switched antenna.

Therefore, based on this ‘‘current surge model,’’ the polarity of the generated electric field, or of the THz pulse, is determined by the time derivative of the product of $n(t)$ and $v(t)$, as evident from Eq. 13. For a constant $n(t)$, if $v(t)$ is increased \mathbf{E} turns out parallel with the \hat{z} -direction and if $v(t)$ is decreased $\mathbf{E}(t)$ becomes anti-parallel. Especially during, and right after, the optical-short-pulse interaction, the carrier density $n(t)$ changes quite rapidly, compared with the change of $v(t)$, so the change of $n(t)$ plays a crucial role in determining the sign and magnitude of $\mathbf{E}(t)$.

In order to test the current surge model for the THz pulse generation, we investigate the time dependence of the current density $n(t)$. As we are considering the situation where an optical short pulse illuminates the surface of a semiconducting material, the charge carriers of the semiconductor are created by the bombardment of the photons in the optical pulse. The number of charge carriers created in a given time interval is directly proportional to the light intensity, for the light intensity is below the saturation intensity. Then, the governing equation for the time dependence of the density function $n(t)$ is given, including the relaxation of the electrons with relaxation rate γ , by

$$\frac{dn}{dt} = I(t) \frac{A}{h\nu} - \gamma n(t), \quad (14)$$

where I is the laser intensity, A the absorption coefficient, h the Planck constant, and ν the laser frequency. For a gaussian pulse with pulse duration τ , this above equation becomes,

$$\frac{dn}{dt} = n_0 e^{-t^2/\tau^2} - \gamma n(t), \quad (15)$$

where n_0 is the carrier density determined at the peak of the optical pulse, given as $n_0 = AI(0)/h\nu$. Then, the differential equation given in Eq. 15 is straight forward to solve and the solution is obtained as

$$n(t) = n_0 e^{-\gamma t} \int_0^t e^{-t'^2/\tau^2 + \gamma t'} dt'. \quad (16)$$

Now we put a set of typical numbers from an experiment. In the experiment illustrated in Fig. 1, $\tau = 50 - 100$ fs and $\gamma = 1/(1 - 10$ ps), so the relaxation is significantly slower than the duration of the absorption process. Then the result in given in Eq. 16 is little different from the following equation:

$$n(t) \simeq n_0 e^{-\gamma t} \int_0^t e^{-t'^2/\tau^2} dt'. \quad (17)$$

So, we have obtained the time dependent carrier density function $n(t)$ as

$$n(t) = n_0 e^{-\gamma t} \int_0^t \frac{I(t')}{I_0} dt', \quad (18)$$

and, therefore, the increase of $n(t)$ is contributed from the photon absorption and the decrease of $n(t)$ is from the carrier relaxation.

Now we consider to obtain the velocity $v(t)$ of the charge carrier. The Newton's second law for the charge carrier, which is electron in this case, can be expressed with two kinds of forces in this case: the external electric field along the biased antenna and the damping term due to elastic collisions. Then, the differential equation

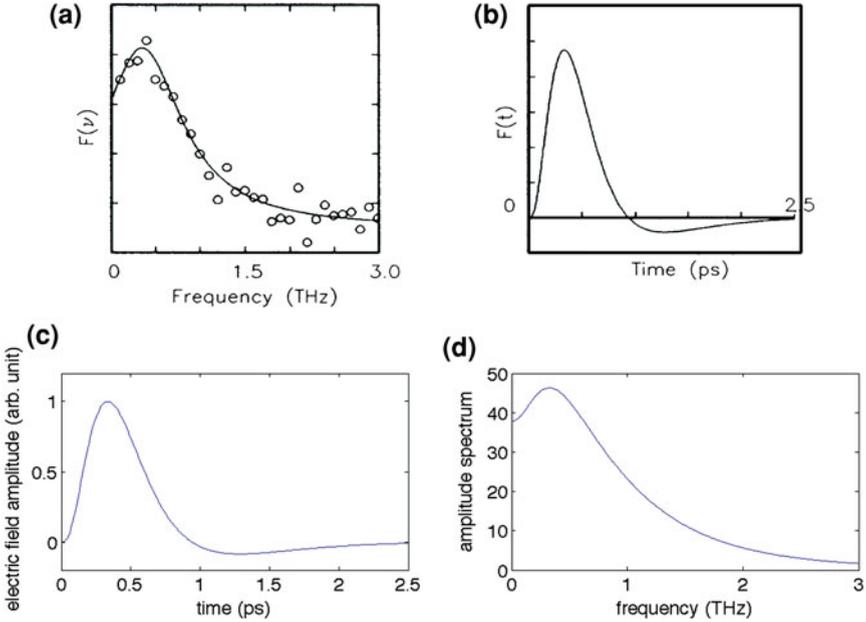


Fig. 3 Experimental measurement of (a) the temporal waveform of sub-single-cycle pulse and (b) its spectrum. (c) and (d) show the model calculations for the waveform and the spectrum, respectively

$$m \frac{d^2x}{dt^2} + b \frac{dx}{dt} = eE_{\text{ext}}, \quad (19)$$

gives the solution becomes

$$v(t) = v_t(1 - e^{-(eE_{\text{ext}}/mv_t)t}) = v_t(1 - e^{-\lambda_v t}), \quad (20)$$

where v_t is the electron mobility of the given semiconducting material, and the characteristic relaxation rate for the velocity is given by $\lambda_v = eE_{\text{ext}}/mv_t = (0.03 \text{ fs})^{-1}$. In the given experiment, $E_{\text{ext}} = 10^{6-7} \text{ V/m}$ and $v_t = 10^4 \text{ m/s}$, and, as the given numbers are used, the λ_t is given as $\lambda_t = 1/(3 \times 10^{-15} \text{ fs})$. The velocity relaxation is significantly faster than the pulse duration τ . So the velocity of the charge carriers is considered nearly constant during the THz pulse emission process, i.e.,

$$v(t) \simeq v_t. \quad (21)$$

Figure 3 shows the experimentally measured waveform of a sub-single-cycle pulse and its spectrum, and the results are compared with the model calculations based on Eqs. 13, 18, and 21. It has become clear that the current surge is mainly

contributed by the time dependent change of the carrier density $n(t)$ in THz pulse generation. Therefore, the electric field $\mathbf{E}(t)$ that is given as the time derivative of the current has a positive polarity during the light-interaction and a negative polarity during the relaxation process.

3 Difference Frequency Mixing for THz Wave Emission

Having considered the THz pulse generation from a photoconductive antenna in the previous section, we also briefly review the THz wave generation from an optoelectronic crystal such as ZnTe.

The generation of THz waves can be modeled using the nonlinear Maxwell equation expressed in the spatial and frequency domain for the difference frequency mixing process of the optical laser pulse, that is given as [3],

$$\frac{\partial^2}{\partial z^2} E^{\text{THz}}(\Omega, z) + \epsilon(\Omega) \frac{\Omega^2}{c^2} E^{\text{THz}}(\Omega, z) = -\frac{4\pi\Omega^2}{c^2} P(\Omega, z), \quad (22)$$

where $\Omega = 2\pi f_{\text{THz}}$. The polarization $P(\Omega, z)$ is the second-order nonlinear susceptibility $\chi^{(2)}(\Omega)$ multiplied by the Fourier components of the intensity of the pump laser pulse:

$$P(\Omega, z) = \chi^{(2)}(\Omega) I(\Omega) \exp[i\Omega z/v_g], \quad (23)$$

where, v_g is the group velocity of the optical pulse, and if the pump laser pulse is assumed to be a gaussian pulse with a pulse width τ_{opt} , the Fourier component of the intensity of the pump laser pulse may be expressed as $I(\Omega) = I_0 \exp[-(\tau_{\text{opt}}\Omega)^2/2]$.

The solution for this nonlinear Maxwell equation is obtained at each point in the crystals as [3]

$$E^{\text{THz}}(\Omega, z) \propto \Omega^2 \chi^{(2)}(\Omega) I(\Omega) z \text{sinc} \left[\frac{\Delta k z}{2} \right]. \quad (24)$$

In order to obtain terahertz spectral shapes appropriate for some particular purpose and to maximize $E^{\text{THz}}(\Omega, z)$ for a terahertz frequency range, one must choose the wavelength λ_{opt} and the duration τ_{opt} of the ultrashort optical pulse. So, we consider the requirements for the crystal thickness and the pulse width.

The efficiency of the terahertz pulse generation strongly depends on the momentum mismatch between the optical and the terahertz pulses in the nonlinear medium. The momentum mismatch condition for the wavelength at different frequencies is

$$\Delta k = k(w_{\text{opt}} + \Omega) - k(w_{\text{opt}}) - k(\Omega) = 0, \quad (25)$$

where the angular frequency of the optical wave is given $\omega_{\text{opt}} = 2\pi\nu$. The momentum mismatch Δk influences the inverse of the coherence length l_c for the optical rectification process in a material with dispersion at optical frequencies. The coherence length l_c , which depends on the efficiency of terahertz generation, is given by

$$l_c(\Omega; \lambda_{\text{opt}}) = \frac{\pi}{\Delta k} = \frac{\pi c}{\Omega |N_{\text{eff}}(\lambda_{\text{opt}}) - n_{\text{THz}}(\Omega)|}, \quad (26)$$

where $N_{\text{eff}}(\lambda_{\text{opt}})$ is the effective refractive index for an optical pump beam with a wavelength of λ_{opt} [4],

$$N_{\text{eff}}(\lambda_{\text{opt}}) = n_{\text{opt}}(\lambda_{\text{opt}}) - \lambda_{\text{opt}} \left. \frac{dn_{\text{opt}}(\lambda)}{d\lambda} \right|_{\lambda_{\text{opt}}}. \quad (27)$$

The spectrum for the generated THz pulse propagating through a ZnTe is obtained by integrating Eq. 24 over the crystal thickness z_o . Once the THz pulse travels a distance equivalent to the coherence length of this non-linear frequency mixing process, the spectral shape of the THz pulse is greatly altered. The calculated results are shown in the left panel of Fig. 4, where the spectral shapes are primarily governed by the coherence length of the thick crystal. As the pump laser wavelength is tuned below 790 nm, most of the low frequency parts of THz spectra becomes phase-mismatched as shown in the left panel of Fig. 4, and the obtained THz spectra are quite narrow. For comparison, the experimentally measured spectra and pulse shapes are shown in Fig. 4. Here the thickness of the ZnTe crystal is chosen to separate the coherent band of the THz spectrum into two, one at near zero frequency and the other at higher frequencies. This results in two distinct spectra humps as shown in the right panel of Fig. 4b and c. The obtained spectrum for the generated THz pulse is strongly dependent on the pump laser wavelength and the temporal shapes vary from a single-cycle pulse with pulse duration of 250 fs to an oscillating pulse stretched out to over 50 ps in time as in Fig. 4c.

4 Measurement of Sub-single Cycle Pulses

The first measurement of the waveform of a THz pulse has been obtained via interferometry. As shown in Fig. 5a Michelson interfereometer was setup to measure the autocorrelation of THz pulses. The THz pulse was generated as in Fig. 1 via short pulse photoexcitation of a semi-conducting InP. An off-axis paraboloidal mirror, wiregrid polarizers, metallic retroreflectors were used to collect the freely propagating broadband THz pulse without a significant chromatic aberration, and the signal was collected with a helium-cooled silicon composite bolometer [5]. The autocorrelation signal defined by

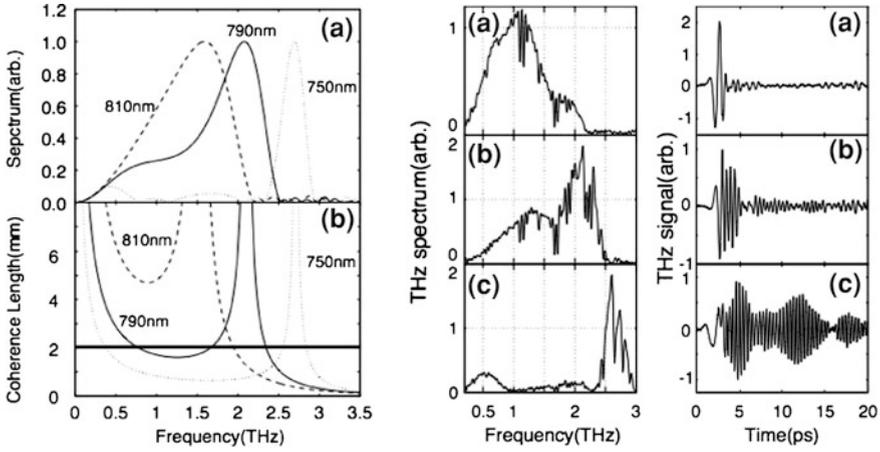


Fig. 4 Generation of THz waves via difference frequency mixing in ZnTe. The left panel shows the computed THz spectra and coherence length for various laser wavelengths. The *right* panel shows the measured THz waves (from top to bottom, 810, 790, 750 nm.) Reprinted from Ref. [3]

$$I_{\text{auto}}(t) = \int_{-\infty}^{\infty} E(t')E(t+t')dt' \tag{28}$$

As the Wiener-Khintchine theorem tells, the intensity spectrum is given as the Fourier transform of the autocorrelation signal $I_{\text{auto}}(t)$ [6].

$$S(\nu_{\text{THz}}) = \int_{-\infty}^{\infty} I_{\text{auto}}e^{-2\pi\nu_{\text{THz}}t} dt, \tag{29}$$

where $S(\nu_{\text{THz}})$ is the intensity spectrum which is the square of the amplitude spectrum. Based on an assumption that the amplitude spectrum is real-valued, the electric field $E(t)$ is given simply as the inverse Fourier transform of the amplitude spectrum, or

$$E(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sqrt{S(\nu_{\text{THz}})}e^{2\pi\nu_{\text{THz}}t} df. \tag{30}$$

The measured result in Fig. 5 shows that the THz pulse is a nearly half-cycle pulse.

Later, a more direct method, for example, the free-space electro-optic sampling has measured the actual shape of THz pulses directly [8]. The underlying mechanism of the electro-optic sampling is the Pockels effect, which utilizes the polarization rotation of propagating light through an electro-optic crystal, such as ZnTe, in the presence of an electric field. As an optic probe beam co-propagates with the THz pulse, which is the electric field, with a given time-delay between them, the polarization rotation of the optic probe beam is given proportional to the THz field amplitude. Therefore, the THz field amplitude is obtained as a function of the time-delay as

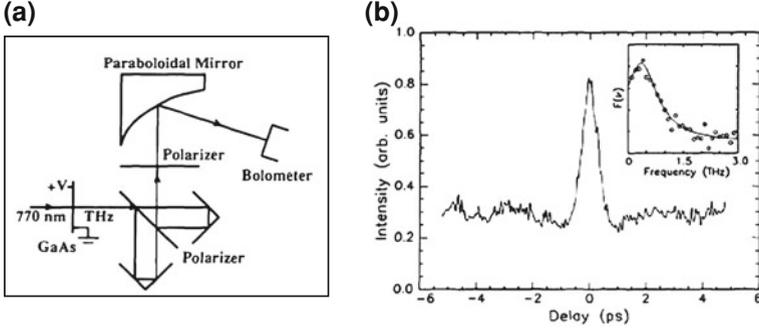


Fig. 5 (a) Michelson interferometry for THz pulse characterization. (b) The measured intensity autocorrelation. The inset shows the electric field amplitude. Reprinted from Ref. [7]

$$E_{\text{THz}}(t) = \Delta\phi(t) \left(\frac{2\pi\nu_{\text{THz}}L}{\sqrt{2}c} n_0^3 r_{41} \right)^{-1}, \quad (31)$$

where ϕ is the polarization angle change, L is crystal thickness, n_0 is the refractive index of optic beam, and r_{41} is the electro-optic coefficient. For example, with ZnTe ($r_{41} = 4 \text{ pm/V}$) of $L = 1 \text{ mm}$ -thickness an electric field of $E_{\text{THz}}^{\text{peak}} = 10 \text{ V}/\mu\text{m}$ leads to the probe polarization change by $\phi \approx 0.01$ degree.

5 Propagation of Sub-single Cycle Pulses

The THz pulse shapes in the near-field and in the far-field differ significantly with respect to each other due to diffraction and propagation. The near-field criterion for the far-field diffraction region required for the Fraunhofer formula is given by

$$|\mathbf{r} - \mathbf{r}'| \gg \pi \frac{d^2}{\lambda}, \quad (32)$$

where \mathbf{r}' and \mathbf{r} denote the source and measurement points, respectively, and d is the emission antenna size. And for a large-area antenna, as described in this article for the generation of THz sub-single-cycle pulses, that is of a size $d = 3 \text{ cm}$, after a propagation over a distance of a few meters, the criterion in Eq. 32 is satisfied. Therefore, instead of the famous Fraunhofer diffraction formula, the Huygens-Fresnel diffraction integral formula may be used to model the THz electric field at an arbitrary point beyond the antenna aperture, that is given as

$$E(\mathbf{r}, t) = \int_{\Sigma} \frac{\cos(\hat{\mathbf{n}}, \mathbf{r} - \mathbf{r}')}{2\pi c |\mathbf{r} - \mathbf{r}'|} \frac{d}{dt} E(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}) d^2 r', \quad (33)$$

where $\Sigma(\mathbf{r}')$ is the THz emission surface and \mathbf{n} is its surface normal [9].

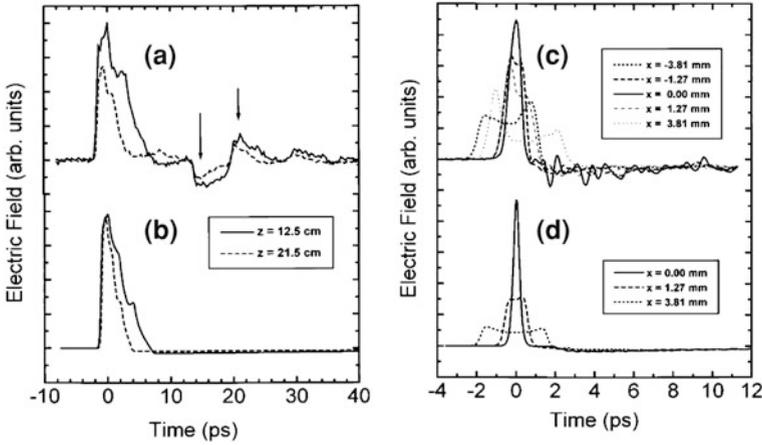


Fig. 6 Spatiotemporal profile of THz shapes of (a) experimental and (b) simulation results at a distance z directly measured behind the antenna, and (c) experimental and (d) simulation results at the focal point of a parabolic mirror. Reprinted from Ref. [9]

Using this formula, the spatiotemporal profile of THz pulses at the near- and far-field regions are calculated and the results are compared with experimental measurements in Fig. 6. It is noted that the low-frequency components of the near-field THz shapes are significantly damped compared to the ones of the far-field.

The near-field shape of the THz pulse is more likely a quarter-cycle pulse, while the far-field one is a half-cycle pulse. Such a significant loss of the low-frequency components of THz pulse is caused by the diffraction. As an illustrative example, we consider the wave diffraction from a narrow slit as shown in Fig. 7. When a wave is penetrate through or emitted from the narrow aperture, the intensity of the diffracted wave is given at a angle θ off from the normal direction \mathbf{n} as

$$I(\theta) = I(0) \left(\frac{\sin \beta}{\beta} \right)^2, \tag{34}$$

where the parameter β is defined with the slit width b as

$$\beta = v_{\text{THz}} \frac{\pi b}{c} \sin \theta. \tag{35}$$

Then, the electric field amplitude that is collected for a finite numerical aperture ($\sin \theta_c$) may be given by

$$E(v_{\text{THz}}) = \int_{-\theta_c}^{\theta_c} \frac{dE(v_{\text{THz}})}{d\theta} d\theta \approx E(v_{\text{THz}}) \frac{v_{\text{THz}}}{v_{\text{cutoff}}}, \tag{36}$$

for v_{THz} smaller than v_{cutoff} . The last approximation is valid, as θ_c is small for a typical example with $b = 1$ cm because

$$\theta_c = \sin^{-1} \frac{c/\pi b}{v_{\text{THz}}} \approx \sin^{-1} \frac{10^{10}(\text{Hz})}{v_{\text{THz}}}. \quad (37)$$

It is clear from Eq. 35 that the spectrum of the diffracted pulse is damped as a linear function of the THz frequency v_{THz} compared to the spectrum of the incident pulse. The smaller the size of the antenna is used, the more sharply the low frequency components of the spectrum is removed after the diffraction. Therefore the temporal shape of THz pulses from a small-size antenna ($b \approx 100 \mu\text{m}$) is further reshaped, and full-cycle pulses are produced in the far-field region in general.

6 Applications of Sub-Cycle Cycle Pulses

As we have considered the generation, measurement, and propagation of the sub-single-cycle THz pulses, we now begin to discuss the applications of such pulses in this section. In general, a broadband nature of a pulse can be used in various ways. One of the non-trivial ways of using the broadband nature may be the use of the spectral degree of freedom of a THz wave as a replacement for one of the spatial degrees of freedom in optical imaging [10]. A more scientific use of THz pulses has been found in the coherent manipulation of quantum systems [7, 11–13]. Especially with a species of atoms in highly excited states, called Rydberg atoms, there have been many fundamental studies for atom-field interactions.

For a numerical study, an analytical form of the temporal profile of the THz wave may be necessary, such as the one shown below, which is obtained from a numerical fit to the measured temporal waveform of the THz wave:

$$E(t)/E_{\text{max}} = 29.56 \left[17.75 \left(\frac{t}{\tau} \right)^3 e^{-8.87t/\tau} - 0.412 \left(\frac{t}{\tau} \right)^5 e^{-4.734/\tau} \right], \quad (38)$$

where $\tau = 1$ ps is used [11].

The photon energy of THz pulses induces the transition to and from adjacent energy levels, and, at the same time, as a result of the interaction with a THz pulse with unipolar shape Rydberg atom gets a directional momentum change along the THz polarization direction. The momentum is directly delivered to the quantum system at a precise time, e.g., kicking an electron out of an atom. So the wave function changes along the kick, so does the momentum distribution. In this case, the ionization probability of the Rydberg atom initially at a quantum state $|\psi_{lmm}\rangle$ is given by

$$P(Q) = \int d^3k |\langle \psi_k^- | e^{i\mathbf{Q}\mathbf{r}} | \psi_{lmm} \rangle|^2, \quad (39)$$

where \mathbf{Q} is the impulse, or a momentum kick, delivered to the atom as a result of the interaction of the THz pulse, especially from the main lobe of the pulse, defined as

$$\mathbf{Q} = - \int_{t-\epsilon}^{t+\epsilon} e\mathbf{E}(t')dt. \quad (40)$$

In conjunction with electron velocity-map imaging, the asymptotic velocity distribution of electrons ionized by the sub-single-cycle-pulse excitation can be studied, for an analysis of the ionization dynamics of the Ryberg atoms or the tomographic information of the initial atomic states.

7 Sub-single Cycle Pulses in Other Frequency Ranges

In the previous sections, we have found that an extreme broadness of the spectral bandwidth of a THz pulse has led to the temporal feature of the sub-single-cycle nature. Then, we can ask a question if there are sub-single cycle pulses in other frequency range of electromagnetic wave or in other kinds of waves than electromagnetic waves? In the longer wavelength region of electromagnetic wave than the wavelength of THz pulses, ways of making such pulses are trivial, for example, an electric pulse in radio frequency, so we don't discuss them here. However, in the shorter wavelength region, in particular in the optical frequency range, devising a method to produce sub-single-cycle pulses has been of great interest. We will discuss a couple of research efforts towards the generation of optical sub-single-cycle pulses in the below. Also, there are other kinds of short pulses that can be considered as sub-single-cycle pulses found in nature that may include the shock waves in solid-state materials and solitons generated on liquid surface.

In optical frequency region, producing a coherent waveform of such a broad bandwidth enough to generate a single-cycle pulse is not easy. In ultrafast optics, the high harmonic generation of an ultrashort infrared pulse in atoms produces a sub-femtosecond pulses that encompasses the whole spectral range of visible and UV/XUV [14]. Another method is the supercontinuum formed as a result of cascading nonlinear processes of IR short pulses, as seen, for example, in a photonic bandgap fiber, and its spectrum spans several optical octaves with a spectral range extending from UV to beyond mid-IR [15]. However, with these two methods, such an enough bandwidth comparable to the optical frequency that the optical pulses can have the temporal width less than, or comparable to, the optical cycle has not yet been achieved.

In order to make a single-cycle pulse, as discussed in earlier sections, it is required to have the bandwidth at least a half of the frequency itself, i.e.

$$\Delta\nu \geq \frac{1}{2}\nu. \quad (41)$$

Fig. 7 Illustration of wave diffraction through a narrow slit

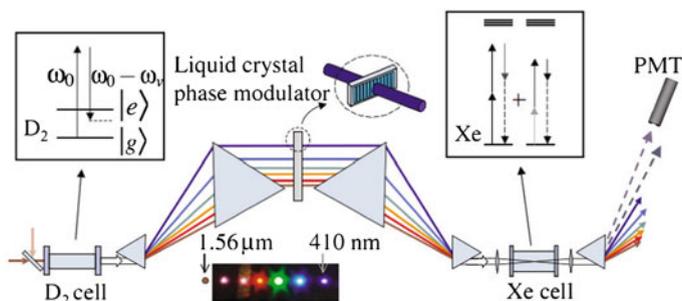
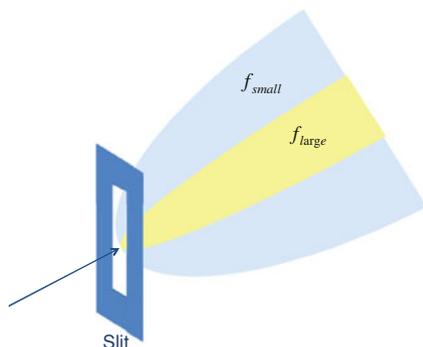
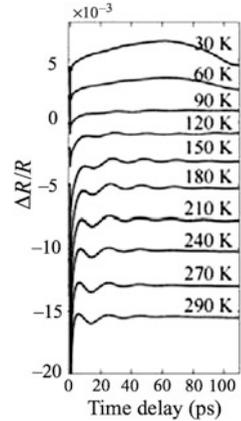


Fig. 8 Experimental setup for temporal synthesis of Raman sidebands of D_2 . Seven sidebands to produce a train of optical single-cycle pulses. Reprinted from Ref. [16]

If it is necessary to fill the huge range of continuous spectrum to make an isolated short pulse. However, one may instead use an idea of using a discrete set of frequency components that are distributed with a net spectral width of the half of the frequency. Of course, then the produced waveform is a train of pulses. In experiments with the vibrational transition in D_2 , it has been attempted to add the mutually coherent Raman sidebands [16]. Although they are separated from each other by the difference of the driving laser frequencies, the sidebands extend from $2.94 \mu\text{m}$ to 410 nm . As shown in Fig. 8, the seven Raman sidebands of D_2 were coherently added to produce a train of optical sub-single-cycle pulses. Temporally 11 fs that corresponds to 0.8 cycle for 650-nm light has been achieved (Fig. 8).

Pushing further the idea of coherently combining optical waves, now people have coherently put together two non-overlapping spectra of separate pulse trains. For this method to work, the timing jitter between the two separate lasers needs to be precisely controlled under a fraction of the optical cycle. In experiments in Ref. [17, 18], two erbium-doped fiber amplifiers that produce separate spectral windows were coherently combined with a timing jitter of less than 50 attoseconds, and the results shows a single-cycle pulse train of infrared light.

Fig. 9 Strain pulse propagation in multiferroic LuMnO_3 probed by femtosecond optical time-domain spectroscopy. Reprinted from Ref. [21]



Sub-single-cycle pulses, or more generally a non-oscillatory propagating waveform, are found in other kinds of waves as well. The acoustic strain pulses generated in solid with ultrashort laser pulse interaction shows a non-oscillatory but propagating wave. It is also referred to as the picosecond ultrasonics and the pulse propagate a fairly long distance as long as the relaxation process allows, without a significant deformation [19]. The propagation of the acoustic wave can be probed by an optical probe pulse because the wave equation is deformed with a small index change δn from the strain. Then the Maxwell's equations for the probe electric field $E(z, t)$ in the slowly varying envelope approximation are given as [19, 20]

$$\frac{\partial^2 E(z, t)}{\partial z^2} + \frac{4\pi^2 v_{\text{probe}}^2}{c^2} [n_0 + \delta n_0(z, t)] E(z, t) = 0, \tag{42}$$

where the v_{probe} is the frequency of the optical probe pulse, n_0 is the index of refraction, and δn is the index change due to the strain. The shape of the strain is merely a bump propagating through the medium and its non-oscillatory propagating pulse shape is resulted from a coherent superposition of a wide range of acoustic phonon spectrum [21]. (Fig. 9)

Lastly, a non-oscillatory propagating wave is in general a kind of solitons. The soliton wavepacket is a self-reinforcing solitary wave found as a result of a nonlinear effect of dispersion cancellation in a medium. As famously known with the solitary wave found travelling a long distance with a constant speed along a canal in Scotland, there have been various forms of soliton have been studied. But lately the soliton in optical fiber has resurrected its importance because of its potential use for a long distance data transmission. For example, Korteweg-de Vries equation models waves on water surface and it is given as a partial differential equation for a function $\phi(x, t)$ for a given time t and position x as

$$\frac{\partial \phi}{\partial t} + \frac{\partial^3 \phi}{\partial x^3} + 6\phi \frac{\partial \phi}{\partial x} = 0, \tag{43}$$

and this equation becomes for a traveling wave $\phi(x, t) = f(x - ct)$

$$-c \frac{\partial f}{\partial t} + \frac{\partial^3 f}{\partial x^3} + 6\phi \frac{\partial f}{\partial x} = 0, \quad (44)$$

or, after the integration with respect to x ,

$$-cf + \partial_x^2 f + 3f^2 = A. \quad (45)$$

where A is a constant. Then, the solution of the above equation is given as

$$\phi(x, t) = \frac{1}{2} c \operatorname{sech}^2 \left[\frac{\sqrt{c}}{2} (x - ct) - A' \right], \quad (46)$$

a solitary wave of which the shape does not change and moves with a constant speed of c . It is noted that in the solitons in an optical fiber the solitary waveform is maintained in the intensity envelope, that is in general much longer than an optical cycle. However, in the solitons in water surface, on the other hand, the surface wave propagates with a spatial stretch less than a characteristic wavelength of the corresponding wave equation.

Acknowledgements This research was supported by Basic Science Research and Mid-career Researcher Programs through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant Nos. 2009-0090843 and 2010-0013899).

References

1. Huber, R., Brodschelm, A., Tauser, F., Leitenstorfer, A.: Generation and field-resolved detection of femtosecond electromagnetic pulses tunable up to 41 THz. *Appl. Phys. Lett.* **76**, 3191–3193 (2000)
2. Jackson, J.D.: *Classical Electrodynamics*, 2nd ed. John Wiley & Sons, Inc, New York (1975)
3. Ahn, J., Efimov, A.V., Averitt, R.D., Taylor, A.J.: Terahertz waveform synthesis via optical rectification of shaped ultrafast laser pulses. *Opt. Express* **11**, 2486 (2003)
4. Nahata, A., Weling, A.S., Heinz, T.F.: A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling. *Appl. Phys. Lett.* **69**, 2321 (1996)
5. Greene, B.L., Federici, J.F., Dykaar, D.R., Jones, R.R., Bucksbaum, P.H.: Interferometric characterization of 160 fs far-infrared light pulses. *Appl. Phys. Lett.* **59**, 893 (1991)
6. Hecht, E.: *Optics*, 4th ed. Addison Wesley, Reading (2002)
7. Jones, R.R., You, D., Bucksbaum, P.H.: Ionization of Rydberg atoms by subpicosecond half-cycle electromagnetic pulses. *Phys. Rev. Lett.* **70**, 1236 (1993)
8. Winnewisser, C., Uhd Jepsen, P., Schall, M., Schyja, V., Helm, H.: Electro-optic detection of THz radiation in LiTaO₃, LiNbO₃ and ZnTe. *Appl. Phys. Lett.* **70**, 3069 (1997)
9. Budiarto, E., Pu, N.-W., Jeong, S., Boker, J.: Near-field propagation of terahertz pulses from a largeaperture antenna. *Opt. Lett.* **23**, 213 (1998)
10. Lee, K., Ahn, J.: Single-pixel coherent diffraction imaging. *Appl. Phys. Lett.* **97**, 241101 (2010)
11. Bugacov, A., Piraux, B., Pont, M., Shakeshaft, R.: Ionization of Rydberg hydrogen by a half-cycle pulse. *Phys. Rev. A* **51**, 1490 (1995)

12. Ahn, J., Hutchinson, D.N., Rangan, C., Bucksbaum, P.H.: Quantum phase retrieval of a Rydberg wave packet using a half-cycle pulse. *Phys. Rev. Lett.* **86**, 1179–1182 (2001)
13. Wetzels, A., Gurtler, A., Noordam, L.D., Robicheaux, F., Dinu, C., Muller, H.G., Vrakking, M.J.J., van der Zande, W.J.: Rydberg State Ionization by half-cycle-pulse excitation: strong kicks create slow electrons. *Phys. Rev. Lett.* **89**, 273003 (2002)
14. Li, X.F., L’Huillier, A., Ferray, M., Lompré, L.A., Mainfray, G.: Multiple-harmonic generation in rare gases at high laser intensity. *Phys. Rev. A* **39**, 5751 (1989)
15. Domachuk, P., Wolchover, N.A., Cronin-Golomb, M., Wang, A., George, A.K., Cordeiro, C.M.B., Knight, J.C., Omenetto, F.G.: Over 4000 nm bandwidth of mid-IR supercontinuum generation in sub-centimeter segments of highly nonlinear tellurite PCFs. *Opt. Express* **16**, 7161–7168 (2008)
16. Shverdin, M.Y., Walker, D.R., Yavuz, D.D., Yin, G.Y., Harris, S.E.: Generation of a single-cycle optical pulse. *Phys. Rev. Lett.* **94**, 033904 (2005)
17. Morgner, U.: Ultrafast optics: Single-cycle pulse generation. *Nat. Photon.* **4**, 14 (2010)
18. Krauss, G., Lohss, S., Hanke, T., Sell, A., Eggert, S., Huber, R., Leitenstorfer, A.: Synthesis of a single cycle of light with compact erbium-doped fibre technology. *Nat. Photon.* **4**, 33–36 (2010)
19. Thomsen, C., Grahn, H.T., Maris, H.J., Tauc, J.: Surface generation and detection of phonons by picosecond light pulses. *Phys. Rev. B* **34**, 4129 (1986)
20. Liu, R., Sanders, G.D., Stanton, C.J., Kim, C.S., Yahng, J.S., Jho, Y.D., Yee, K.J., Oh, E., Kim, D.S.: Femtosecond pump-probe spectroscopy of propagating coherent acoustic phonons in InxGa1-xN/GaN heterostructures. *Phys. Rev. B* **72**, 195335 (2005)
21. Jang, K.-J., Lim, J., Ahn, J., Kim, J., Yee, K.-J., Ahn, J.S., Cheong, S.-W.: Ultrafast IR spectroscopic study of coherent phonons and dynamic spin-lattice coupling in multiferroic LuMnO_3 . *New J. Phys.* **12**, 023017 (2012)

Resonant Antennas on Semi-Infinite and Lens Substrates at Terahertz Frequency

Truong Khang Nguyen and Ikmo Park

Abstract This part is concerned with four different types of resonant antennas on semi-infinite and lens substrates made of high-permittivity dielectric materials at terahertz frequency. Full-wavelength single-dipole, full-wavelength dual-dipole, half-wavelength single-slot, and full-wavelength four-leaf-clover-shaped dipole antennas are studied at a frequency of around 1.0 THz. The resonance characteristics of the four antennas are first investigated on a semi-infinite substrate, using the method of moments simulator FEKO. Based on these characteristics, the antennas are placed on an extended hemispherical lens to examine and compare the radiation effects of a substrate lens on each antenna, using the finite-integration time-domain simulator CST Microwave Studio. The shape of an extended hemispherical lens with optimum thickness and quarter-wavelength matching layer plays an important role in maximizing the directivity, while the radiation efficiency is generally affected by the antenna structure.

1 Introduction

The idea of using a planar antenna that has been lithographically integrated on the surface of a dielectric lens to couple electromagnetic radiation to active semiconductor devices is attractive because of its potential for eliminating substrate modes and increasing the Gaussian coupling efficiency (Gaussicity). In addition, the fabrication procedures are well suited to monolithic or hybrid integrated circuit technology, and offer greater dimensional accuracy and durability, and reduced

T. K. Nguyen · I. Park (✉)

School of Electrical and Computer Engineering, Ajou University, 5 Woncheon-dong, Youngtong-gu, Suwon 443-749, Republic of Korea
e-mail: ipark@ajou.ac.kr

cost. Since the lens is electrically large compared to the antenna, the antenna elements act as if they are at the interface of an air–dielectric half-space. Moreover, aside from the radiation patterns, an antenna exhibits similar resonant characteristics on both a lens substrate and a semi-infinite substrate [1]. Hence, an antenna designed on a semi-infinite substrate can be used in the initial investigation of its resonant characteristics, which saves a significant amount of computational time.

A comparison of dipole, bow-tie, spiral, and log-periodic antennas on a thin substrate at infrared frequencies has recently been reported [2]. However, planar antennas built on a substrate are quite different from ordinary antennas in free space or on a thin substrate, primarily because antennas tend to radiate most of their energy on the substrate side, especially on high-permittivity substrates at high frequencies. The properties of antennas on semi-infinite substrates have been studied extensively for many years, from substrates made of lossy dielectrics (such as water or earth) [3–5] to substrates made of lossless dielectrics (such as GaAs or silicon) [1, 5–9]. However, there have been no detailed studies comparing several types of resonant antenna on lossless semi-infinite and lens substrates made of high-permittivity materials at terahertz (THz) frequency.

For terahertz receivers, planar integrated quasi-optical technology, implemented by mounting a planar antenna on a substrate lens, is expected to be a preferable alternative to waveguide-based front ends. Recent progress in nanostructuring and micro-machining enables reliable production and alignment of planar antenna structures on dielectric lenses with sufficient accuracy for high-frequency operation. Furthermore, it has been shown that this type of integrated lens antenna makes a THz heterodyne receiver perform better when combined with a hot electron bolometer (HEB) [10–12]. Therefore, a planar integrated lens antenna with optimum substrate and antenna structure plays a critical role in optimizing an output power design. Single- and dual-antenna elements, such as dipoles and slots, have been successfully used on quartz and silicon dielectric lenses at millimeter-wave and terahertz-regime frequencies [13–17]. The conventional design technique for a planar integrated lens antenna is generally based on the hybrid complementation of geometrical optics (GO) inside the lens and physical optics (PO) outside the lens to characterize the field irradiated by the feeds and the Gaussian coupling efficiency. However, there remains a need for full characterization and comparison of several types of resonant antennas, particularly at THz frequencies.

In this part, we examine and compare the overall performance of several types of resonant antenna in terms of input impedance, radiation patterns, and radiation efficiency. The four antenna types selected for this research are full-wavelength single-dipole, full-wavelength dual-dipole, half-wavelength single-slot, and full-wavelength four-leaf-clover-shaped dipole, all designed to resonate at around 1.0 THz. Two radiation environments are used to investigate the characteristics of each antenna: a semi-infinite Si substrate ($\epsilon_r = 11.7$) and an extended hemispherical lens made entirely of Si. This study should be very helpful to optimize several types of resonant antenna on a substrate lens for THz photomixer designs.

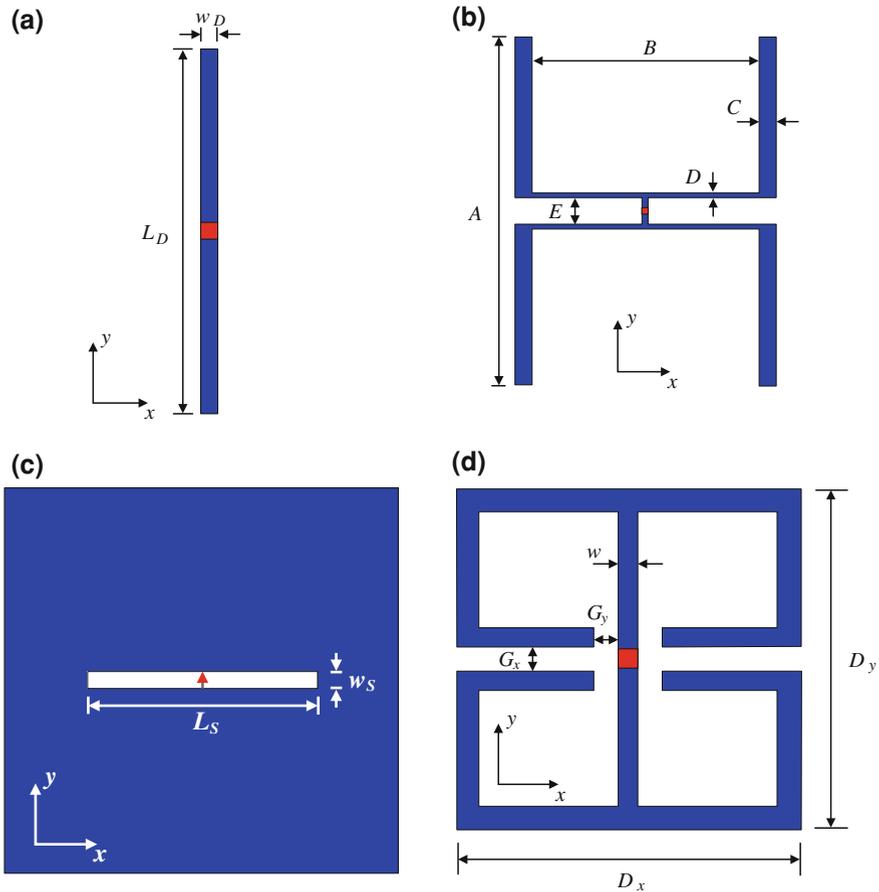
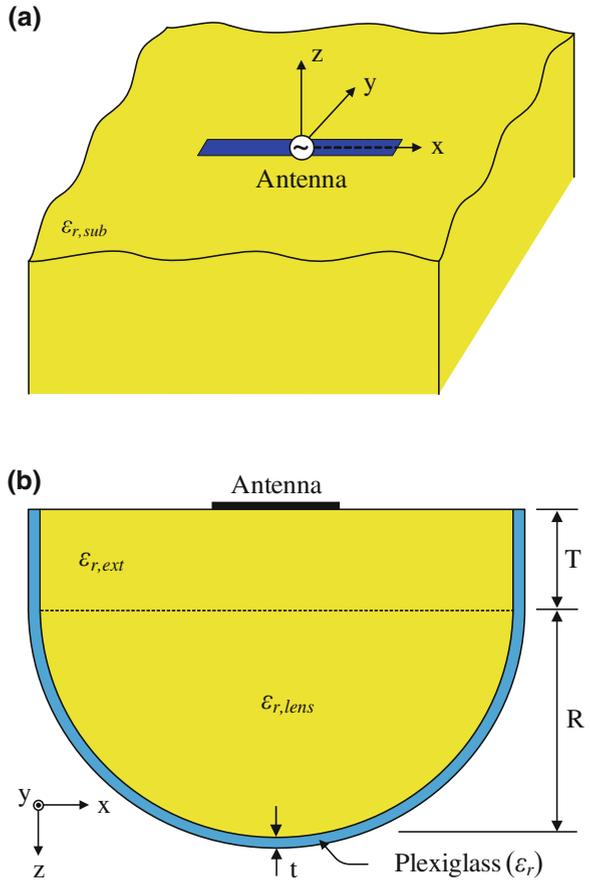


Fig. 1 Geometries of the four resonant antennas: **a** full-wavelength single-dipole, **b** full-wavelength dual-dipole, **c** half-wavelength single-slot, and **d** full-wavelength four-leaf-clover-shaped antennas

2 Geometries of the Four Antennas and Substrate Structures

Figure 1 shows the four types of antenna under consideration. The antennas were all designed to resonate at around 1.0 THz for the sake of comparison. The metal layer had a conductivity of 1.6×10^7 S/m and a thickness of $0.35 \mu\text{m}$. The single-dipole, dual-dipole, and four-leaf-clover-shaped antennas all had a total length of approximately 1λ at about 1.0 THz on the Si substrate, in order to provide maximum input resistance. However, the single-slot antenna provided maximum input resistance at a total length of approximately $\lambda/2$, since a slot antenna is complementary to a dipole antenna (in accordance with Babinet’s principle). Accordingly, these four resonant antennas exhibited high input resistance

Fig. 2 **a** Antenna on a semi-infinite substrate; **b** antenna on an extended hemispherical lens substrate structure



characteristics suitable for THz photomixer designs. The initial dimensions of the antennas were as follows: $L_D = 94 \mu\text{m}$ and $w_D = 3 \mu\text{m}$ for the single-dipole antenna (Fig. 1a); $A = 70 \mu\text{m}$, $B = 49 \mu\text{m}$, $C = 3 \mu\text{m}$, $D = 1 \mu\text{m}$, and $E = 5 \mu\text{m}$ for the dual-dipole antenna (Fig. 1b); $L_S = 46 \mu\text{m}$ and $w_S = 3 \mu\text{m}$ for the single-slot antenna (Fig. 1c); $D_x = D_y = 37 \mu\text{m}$, $G_x = G_y = 2 \mu\text{m}$, and $w = 3 \mu\text{m}$ for the four-leaf-clover-shaped antenna (Fig. 1d). Note that the right and left halves of the four-leaf-clover-shaped antenna each formed a full-wavelength dipole with a total length of approximately $2[(D_x + D_y) - (G_x + G_y) - 1.5w]$, corresponding to 1λ at about 1.0 THz on the Si substrate [18]. Figure 2a and b show prototypes of an antenna on a semi-infinite Si substrate ($\epsilon_{r, \text{sub}} = 11.7$) and an extended hemispherical Si lens ($\epsilon_{r, \text{ext}} = \epsilon_{r, \text{lens}} = 11.7$), respectively. The extension layer and radius of the lens are denoted by T and R , respectively. A $\lambda/4$ -thick matching layer, made of plexiglass ($\epsilon_r = 3.4$), was used to suppress reflected waves from the internal lens surface.

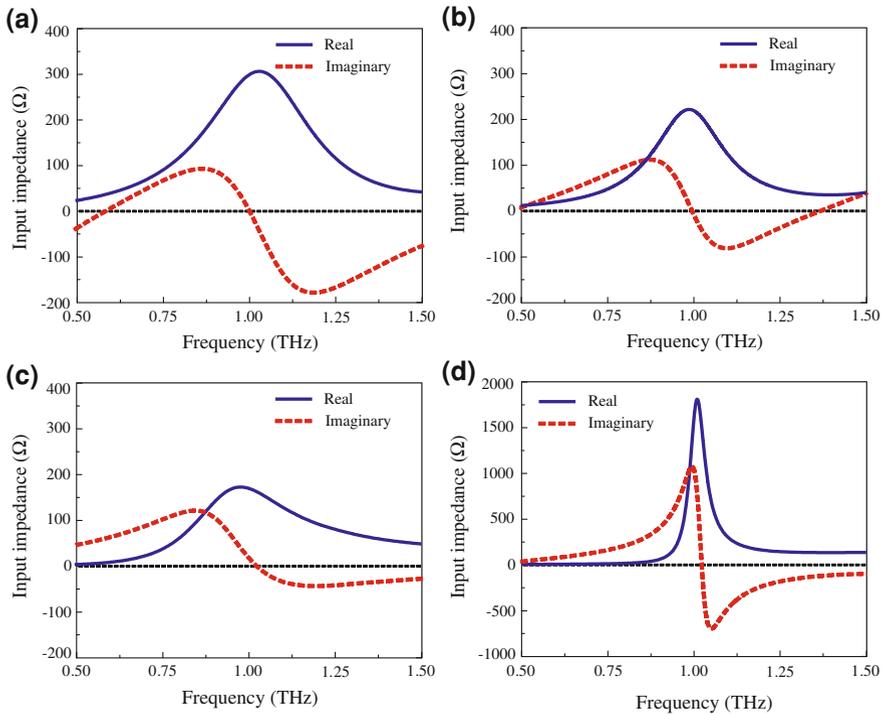


Fig. 3 Input impedance of the **a** single-dipole, **b** dual-dipole, **c** single-slot, and **d** four-leaf-clover-shaped antennas on a semi-infinite Si substrate

3 Antenna Characteristics

3.1 Semi-Infinite Si Substrate

The four antennas were simulated via the FEKO software package featuring the method of moments (MoM), the Si substrate being approximated with a semi-infinite Green's function layer. The single-slot antenna was fed by a wire port, while the single-dipole, dual-dipole, and four-leaf-clover-shaped antennas were all fed by an edge port with the same voltage source (of magnitude 1 V). The single-slot antenna was oriented in the x -direction, whereas the other three antennas were oriented in the y -direction, in order to provide a clear comparison of the antenna radiation patterns.

The input impedance characteristics of the single-dipole, dual-dipole, single-slot, and four-leaf-clover-shaped antennas are shown in Fig. 3a–d, respectively. The single- and dual-dipole antennas resonated at around 1.0 THz, and had input resistances of 302 and 220 Ω , respectively. The single-slot antenna exhibited a slightly lower input resistance than the two dipole antennas, and had an input

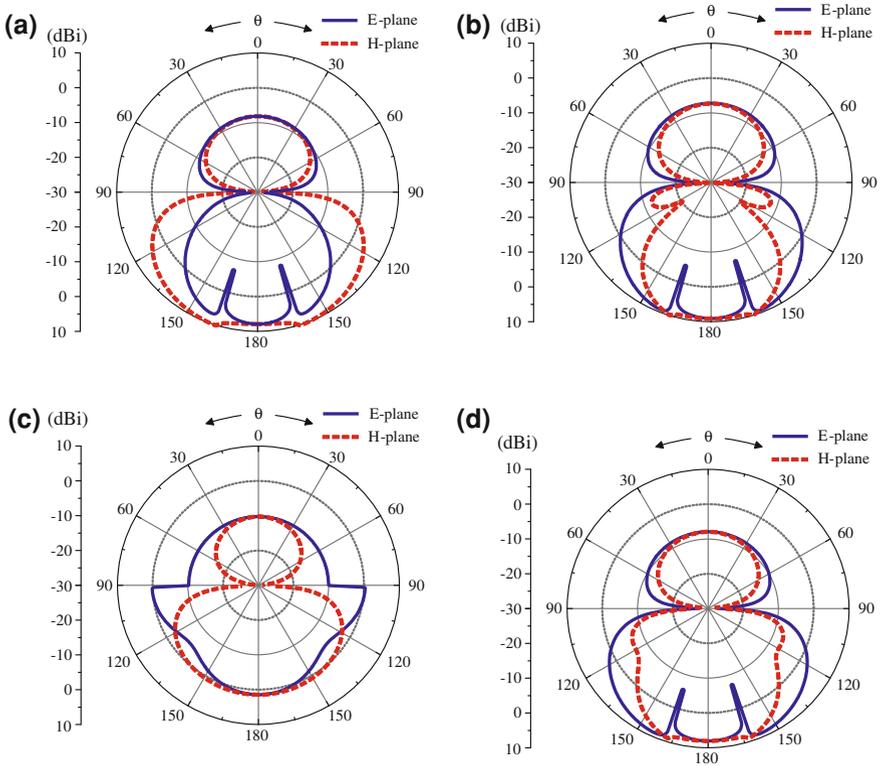


Fig. 4 Radiation patterns of the antennas on a semi-infinite Si substrate: **a** single-dipole, **b** dual-dipole, **c** single-slot, and **d** four-leaf-clover-shaped antennas at the resonant frequency of around 1.0 THz

resistance of 170Ω at the resonant frequency of 1.0 THz. The highly resonant four-leaf-clover-shaped antenna had a high input resistance of about 1800Ω at the resonant frequency of 1.0 THz. The high input resistance of this antenna was due to the high Q and narrow band characteristics of the design.

Figure 4a–d shows the radiation patterns of the four antennas plotted in the *E*- and *H*-planes, which respectively correspond to the *yz*- and *xz*-planes of the antennas' orientation. As the figure indicates, all of the antennas exhibited radiation patterns with a minimum in the *E*-plane and a maximum in the *H*-plane at the critical angle $\theta_c = \pi - \sin^{-1} [(\epsilon_r)^{-1/2}]$ on the dielectric side, which is about 163° for the Si substrate ($\epsilon_r = 11.7$). Both the *E*- and *H*-plane patterns of the four antennas had a null value at the air–dielectric interface. For the single-dipole antenna, the *E*-plane beam pattern was narrower than the *H*-plane beam pattern due to the nature of current distribution on the dipole. However, for the dual-dipole antenna, the *E*-plane beam pattern was wider than the *H*-plane beam pattern because of the array effect. Hence, a small side lobe appeared near the interface in

the H -plane pattern of the dual-dipole antenna. For the four-leaf-clover-shaped antenna, the E -plane and H -plane patterns were almost identical to those of the single-dipole antenna, but the E -plane beam pattern was wider than the H -plane pattern. This beam angle behavior was similar to that of the dual-dipole antenna. The radiation patterns on the air sides of the single-dipole, dual-dipole, and four-leaf-clover-shaped antennas were almost identical. However, the E -plane and H -plane radiation patterns of the single-slot antenna were quite different from those of the other three antennas, as can be seen in Fig. 4c. Since the patch was modeled to be infinite in extent, the two media (air and dielectric) were effectively isolated. Minimum and maximum values did not occur at the critical angle for the single-slot antenna. The E -plane had no null value at the air–dielectric interface, but the H -plane did, as noted in [5].

From the radiation efficiency and directivity perspectives, the four antennas behaved differently. The radiation efficiencies of the single-dipole, dual-dipole, and four-leaf-clover-shaped antennas were 95, 82.5, and 51%, respectively. The maximum directivities at the critical angle were 9.9, 9.8, and 8.9 dBi for the single-dipole, dual-dipole, and four-leaf-clover-shaped antennas, respectively. The directivities at $\theta = 180^\circ$ were 7.7, 8.7, and 7.9 dBi, corresponding to directivity differences (compared to the maximum values at the critical angles) of 2.2, 1.1, and 1.0 dBi for the single-dipole, dual-dipole, and four-leaf-clover-shaped antennas, respectively. For the single slot antenna, the radiation efficiency was 78.8% and the directivity at $\theta = 180^\circ$ was only 1.2 dBi. This behavior variation was also noted in [5], in which it was reported that the radiation of a slot in an infinite ground does not concentrate in a cone, as with a current dipole.

3.2 *Extended Hemispherical Si Lens Substrate*

The four antennas were simulated on a substrate lens using the commercially available electromagnetic simulator CST Microwave Studio, which is based on the finite-integration time-domain technique. Except for the radiation patterns, antennas exhibit similar resonant characteristics on either a lens substrate or a semi-infinite substrate [1]. Accordingly, we investigated the radiation of the four antennas in general, and the directivity of the main beam on the z -axis in particular ($\theta = 180^\circ$), by varying the ratio of the extension layer thickness to the radius of the hemisphere (T/R). The results are shown in Fig. 5a–d. We used hemisphere radii of 2.5, 3.5, and 4.5 λ at a resonant wavelength of 300 μm in free space. The ratio T/R was varied from 0.25 to 0.5 for each value of the hemisphere radius, and the results were plotted separately for each antenna. It should be noted that all directivities and radiation efficiencies shown in the figure were deduced from the radiation patterns of the antennas at the resonant frequency of around 1.0 THz. In general, higher directivity of the main beam was achieved with a larger hemisphere radius for all four antennas. In addition, the extension layer thickness

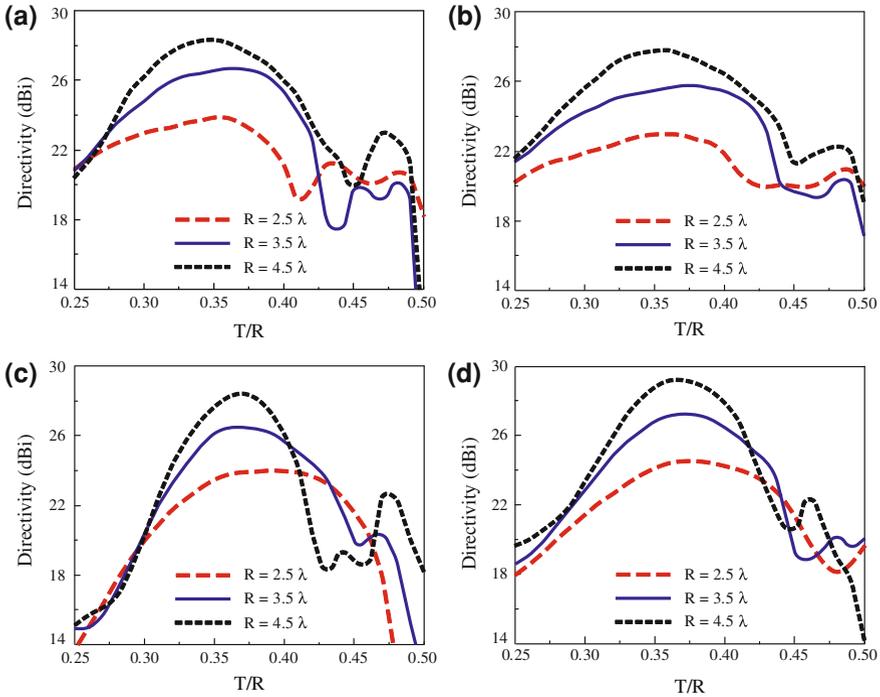


Fig. 5 Directivities of the **a** single-dipole, **b** dual-dipole, **c** single-slot, and **d** four-leaf-clover-shaped antennas on a lens at the resonant frequency of around 1.0 THz, for varying values of the ratio of the extension layer thickness to the radius of the hemisphere (T/R)

for maximum directivity became more distinct as the radius of the hemisphere increased.

For the single-dipole antenna (Fig. 5a), the maximum directivity of the radiation pattern was 24.0 dBi at a ratio of $T/R = 0.36$, 26.8 dBi at $T/R = 0.37$, and 28.4 dBi at $T/R = 0.35$ for hemisphere radii of 2.5, 3.5, and 4.5λ , respectively. Beam-splitting phenomena (i.e., maximum not at $\theta = 180^\circ$) occurred at ratios of $T/R = 0.41$, 0.44, and 0.46 for hemisphere radii of 2.5, 3.5, and 4.5λ , respectively. For the dual-dipole antenna (Fig. 5b), the maximum directivity of the radiation pattern was 23.6 dBi at a ratio of $T/R = 0.36$, 26.0 dBi at $T/R = 0.37$, and 28.0 dBi at $T/R = 0.36$ for hemisphere radii of 2.5, 3.5, and 4.5λ , respectively. Beam-splitting phenomena also occurred in this case, at ratios of $T/R = 0.42$, 0.44, and 0.45 for hemisphere radii of 2.5, 3.5, and 4.5λ , respectively, but with a degree of suffering far less than that of the single-dipole antenna. For the single-slot antenna (Fig. 5c), the maximum directivity of the radiation pattern was 24.0 dBi at a ratio of $T/R = 0.38$, 26.6 dBi at $T/R = 0.37$, and 28.5 dBi at $T/R = 0.37$ for hemisphere radii of 2.5, 3.5, and 4.5λ , respectively. The directivity was almost constant in the range of T/R values from 0.35 to 0.43 for the small

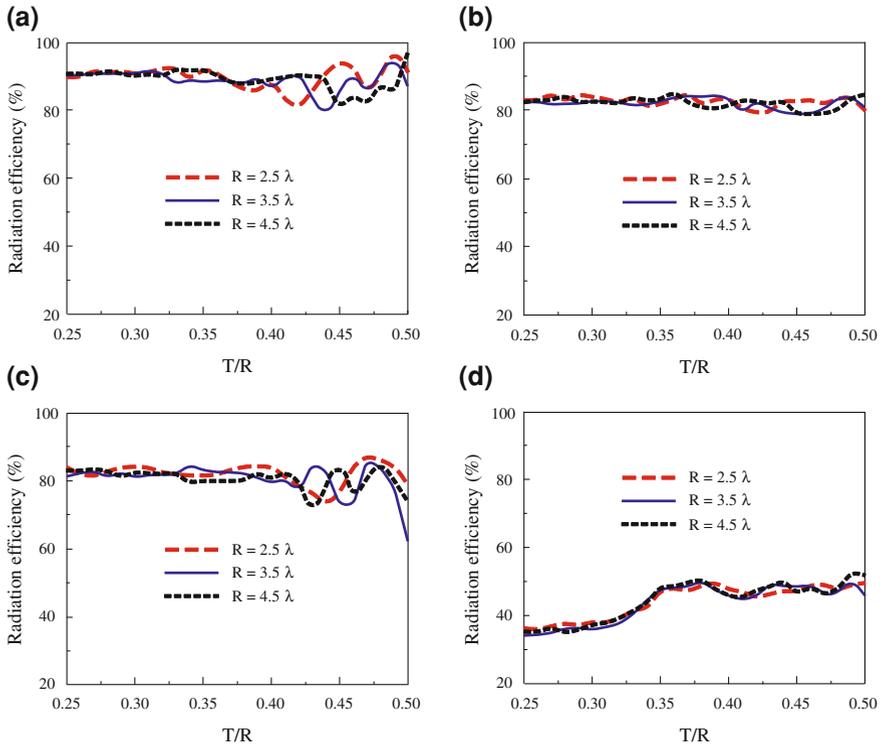


Fig. 6 Radiation efficiencies of the antennas on a lens as functions of the ratio of the extension layer thickness to the radius of the hemisphere (T/R): **a** single-dipole, **d** dual-dipole, **c** single-slot, and **d** four-leaf-clover-shaped antennas at the resonant frequency of around 1.0 THz

radius case $R = 2.5 \lambda$. For the larger hemisphere radii, main-beam splitting occurred (at ratios of $T/R = 0.45$ and 0.43 for hemisphere radii of $R = 3.5$ and 4.5λ , respectively). For the four-leaf-clover-shaped antenna (Fig. 5d), the maximum directivity of the radiation pattern was 24.6 dBi at a ratio of $T/R = 0.38$, 27.3 dBi at $T/R = 0.37$, and 29.3 dBi at $T/R = 0.37$ for hemisphere radii of 2.5, 3.5, and 4.5λ , respectively. Beam-splitting phenomena occurred at ratios of $T/R = 0.48$, 0.45, and 0.44 for hemisphere radii of 2.5, 3.5, and 4.5λ , respectively. For the single-dipole and dual-dipole antennas, the ratios at which main-beam splitting occurred increased slightly with the hemisphere radius, resulting in broad, gently sloping directivity curves. In contrast, the directivity curves for the single-slot and four-leaf-clover-shaped antennas became progressively sharper with increasing hemisphere radius, since the ratio at which main-beam splitting occurred decreased. Generally speaking, the four antennas exhibited almost the same maximum directivity for the same hemispherical lens substrate structure, but the optimum extension layer thicknesses varied according to antenna type.

In Fig. 6, the variation in radiation efficiency of each of the four antennas is plotted as a function of the ratio T/R for hemisphere radii of 2.5, 3.5, and 4.5λ . For

the single-dipole and single-slot antennas, the radiation efficiency curves showed stable variation only in the range of T/R from 0.25 to 0.42, irrespective of the hemisphere radius. These upper bounds were related to the ratios at which the antennas exhibited main-beam splitting, as was previously mentioned. The radiation efficiencies of the single-dipole antenna at optimum extension layer thickness were 92% at a ratio of $T/R = 0.36$, 90% at a ratio of $T/R = 0.37$, and 92% at a ratio of $T/R = 0.35$ for hemisphere radii of 2.5, 3.5, and 4.5 λ , respectively. The radiation efficiencies of the single-slot antenna at optimum extension layer thickness were 83% at a ratio of $T/R = 0.38$, 83% at a ratio of $T/R = 0.37$, and 80% at a ratio of $T/R = 0.37$ for hemisphere radii of 2.5, 3.5, and 4.5 λ , respectively. Obviously, the radiation efficiencies of the single-dipole and single-slot antennas showed fluctuations at high T/R ratios, i.e., $T/R > 0.4$, while those of the dual-dipole and four-leaf-clover-shaped antennas showed nearly stable variations over the entire range of the ratio T/R , regardless of the hemisphere radius. This indicates that the dual-element antennas provided more stable radiation efficiency characteristic than the single-element antennas with respect to the hemisphere lens geometry. The radiation efficiencies of the dual-dipole antenna at optimum extension layer thickness were 83% at a ratio of $T/R = 0.36$, 85% at a ratio of $T/R = 0.37$, and 85% at a ratio of $T/R = 0.36$ for hemisphere radii of 2.5, 3.5, and 4.5 λ , respectively. The radiation efficiencies of the four-leaf-clover-shaped antenna at optimum extension layer thickness were 48% at a ratio of $T/R = 0.37$, 49% at a ratio of $T/R = 0.37$, and 50% at a ratio of $T/R = 0.37$ for hemisphere radii of 2.5, 3.5, and 4.5 λ , respectively.

Figure 7a–d shows the radiation patterns of the four antennas at the resonant frequency of around 1.0 THz for the extended hemispherical lens with radius of $R = 4.5 \lambda$ and plotted at the optimum ratios T/R . The radiation patterns of the single-dipole, dual-dipole, and four-leaf-clover-shaped antennas were very similar with small side lobes in both xz - and yz -planes. However, the single-slot antenna exhibited several and high level of side lobes in the yz -plane. Table 1 gives a comparison in the overall characteristics, i.e., resonant frequency (f_{res}), input resistance (R_{in}), directivity (D), radiation efficiency (η_{rad}), half-power beamwidth (HPBW), sidelobe level (SLL), and front-to-back (F/B) ratio, of the four antennas on the extended hemispherical lens with radius of $R = 4.5 \lambda$. The four-leaf-clover-shaped antenna achieved the highest directivity of 29.3 dBi but requiring a slightly thicker extension layer than the other three antennas. Additionally, F/B ratio of the four-leaf-clover-shaped antenna was 25.2 dB which was the highest as compared to the single-dipole antenna (21.3 dB), dual-dipole antenna (24.3 dB), and single-slot antenna (21.8 dB). However, the four-leaf-clover-shaped antenna had a relatively low radiation efficiency of only 50% in comparison with those of the remains, i.e., single-dipole antenna (92%), dual-dipole antenna (85%), and single-slot antenna (80%). Obviously, the single-dipole antenna was outstanding among the four antennas in term of radiation efficiency. Therefore, single-dipole antenna is mostly suitable for applications with high radiation efficiency requirement while four-leaf-clover-shaped antenna could be the best candidate in using with a lens

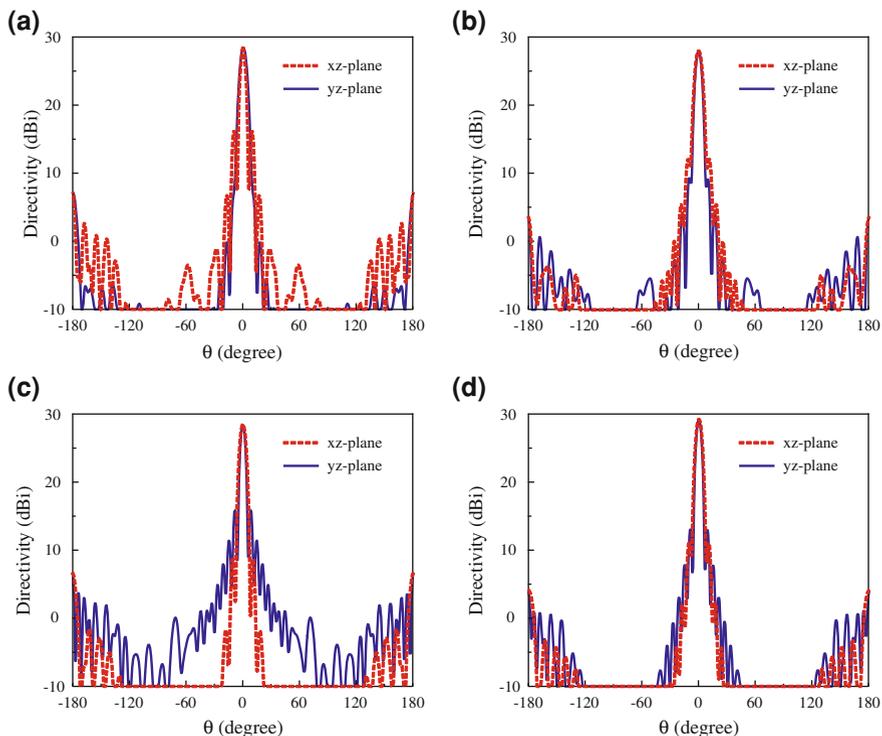


Fig. 7 Radiation patterns of the four antennas on the extended hemispherical lens with radius of $R = 4.5 \lambda$ ($\lambda = 300 \mu\text{m}$) at the resonant frequency of around 1.0 THz, plotted at the optimum T/R ; **a** single-dipole, **b** dual-dipole, **c** single-slot, and **d** four-leaf-clover-shaped antennas

Table 1 Comparison of characteristics of the four antennas on the extended hemispherical lens with radius of $R = 4.5 \lambda$ ($\lambda = 300 \mu\text{m}$)

	Single-dipole	Dual-dipole	Single-slot	Four-leaf-clover-shaped dipole	
T/R	0.35	0.36	0.37	0.37	
f_{res} (THz)	0.99	0.99	1.00	1.01	
R_{in} (Ω)	295	190	167	1705	
D (dBi)	28.4	28.0	28.5	29.3	
η_{rad} (%)	92	85	80	50	
HPBW (degree)	xz-plane	5.7	7.5	6.5	6.6
	yz-plane	7.2	6.3	5.0	5.5
SLL (dB)	xz-plane	-12.3	-15.8	-19.7	-18.2
	yz-plane	-21.3	-18.5	-12.6	-15.8
F/B (dB)	xz-plane	21.3	24.3	21.8	25.2
	yz-plane	21.3	24.3	21.8	25.2

substrate for applications demanding particularly high antenna input resistance, such as a THz photomixer design [18].

4 Conclusions

We have compared the overall performances of four resonant antennas designed on two different substrate structures (a semi-infinite Si substrate and an extended hemispherical Si lens substrate) in terms of input impedance and radiation characteristics. For the semi-infinite substrate, the input impedances of the four antennas exhibited similar behaviors, but the radiation characteristics varied, especially in the case of the single-slot antenna. The four-leaf-clover-shaped antenna had the highest input resistance among the four antennas, and the single-dipole antenna had good overall radiation characteristics (radiation efficiency, clean main beam, and directivity). However, when a lens substrate was used, the four antennas showed similar radiation patterns and maximum directivities for the same hemispherical lens substrate structure, but at different optimum extension layer thicknesses. From a radiation efficiency perspective, the dual-dipole and four-leaf-clover-shaped antennas (which can be classified as dual-element antennas) exhibited stable variation with respect to the ratio T/R , irrespective of the hemisphere radius. The single-dipole antenna exhibited the highest radiation efficiency among four representative antennas. The four-leaf-clover-shaped antenna with its highly resonant characteristic produces the highest directivity but the lowest radiation efficiency. Nevertheless, this antenna would be most preferred in a THz photomixer application that demand high antenna input resistance. These studies provide useful information for choosing and optimizing the antenna and substrate lens structure for certain applications in THz frequency regime.

References

1. Kominami, M., Pozar, D.M., Schaubert, D.H.: Dipole and slot elements and arrays on semi-infinite substrate. *IEEE Trans. Antennas Propagat.* AP-33(6), 600–607 (1985)
2. González, F.J., Boreman, G.D.: Comparison of dipole, bowtie, spiral and log-periodic IR antennas. *Infra. Phys. Tech.* **46**, 418–428 (2005)
3. Banos, A.: *Dipole Radiation in the Presence of a Conducting Half-Space*. Pergamon Press, New York (1966)
4. King, R.W.P., Smith, G.S.: *Antennas in Matter: Fundamentals, Theory, and Applications*. MIT Press, Cambridge (1981)
5. Brewitt-Taylor, C.R., Gunton, D.J., Rees, H.D.: Planar antennas on a dielectric surface. *Electron. Lett.* **17**(20), 729–731 (1981)
6. Smith, G.S.: Directive properties of antennas for transmission into a material half-space. *IEEE Trans. Antennas Propagat.* AP-32, 232–246 (1984)
7. Elefetheriades, G.V., Rebeiz, G.M.: Self and mutual admittance of slot antennas on a dielectric half-space. *Int. J. Infrared Millim. Waves* **14**(10), 1925–1946 (1993)

8. Tong, C., Blundell, R.: An annular slot antenna on a dielectric half-space. *IEEE Trans. Antennas Propag.* **42**, 967–974 (1994)
9. Weller, T.M., Katehi, L.P.B., Rebeiz, G.M.: Single and double folded-slot antennas on semi-infinite substrates. *IEEE Trans. Microw. Theory Tech.* **43**, 1423–1428 (1995)
10. Gerecht, E., Musante, C.F., Zhuang, Y., Yngvesson, K.S., Gol'tsman, G.N., Voronov, B.M., Gershenzon, E.M.: NbN hot electron bolometric mixer—A new technology for low-noise THz receivers. *IEEE Trans. Microw. Theory Tech.* **47**(12), 2519–2527 (1999)
11. Semenov, A.D., Hübers, H.-W., Richter, H., Birk, M., Krocka, M., Mair, U., Smirnov, K., Gol'tsman, G.N., Voronov, B.M.: 2.5 THz heterodyne receiver with NbN hot-electron-bolometer mixer. *Physica C* **372–376**(pt. 1), 448–453 (2002)
12. Semenov, A.D., Richter, H., Hübers, H.-W., Günther, B., Smirnov, A., Il'in, K.S., Siegel, M., Karamarkovic, J.P.: Terahertz performance of integrated lens antennas with a hot-electron bolometer. *IEEE Trans. Microw. Theory Tech.* **55**(2), 239–247 (2007)
13. Filipovic, D.F., Gearhart, S.S., Rebeiz, G.M.: Double-slot antennas on extended hemispherical and elliptical silicon dielectric lenses. *IEEE Trans. Microw. Theory Tech.* **41**(10), 1738–1749 (1993)
14. Zmuidzinas, J., Leduc, H.G.: Quasi-optical slot antenna SIS mixer. *IEEE Trans. Microw. Theory Tech.* **40**(9), 1797–1804 (1992)
15. Ganzevles, W.F.M., Swart, L.K., Gao, J.R., de Korte, P.A.J., Klapwijk, T.M.: Direct response of twin-slot antenna-coupled hot-electron bolometer mixer designed for 2.5 THz radiation detection. *Appl. Phys. Lett.* **76**(22), 3304–3306 (2000)
16. Filipovic, D.F., Ali-Ahmad, W.Z., Rebeiz, G.M.: Millimeter-wave double-dipole antennas for high-gain integrated reflector illumination. *IEEE Trans. Microw. Theory Tech.* **40**(5), 962–967 (1992)
17. Duffy, S.M., Verghese, S., McIntosh, K.A., Jackson, A., Gossard, A.C., Matsuura, S.: Accurate modeling of dual dipole and slot elements used with photomixers for coherent terahertz output power. *IEEE Trans. Microw. Theory Tech.* **49**(6), 1032–1038 (2001)
18. Woo, I., Nguyen, T.K., Han, H., Lim, H., Park, I.: Four-leaf-clover-shaped antenna for a THz photomixer. *Opt. Express* **18**(18), 18532–18542 (2010)

The Basic Principle of the Near-Field Superlens

Kyoungsik Kim

To overcome the diffraction limit in the optical imaging, the superlens with superresolution are developed using negative index materials. In this part, the basic principle of the near-field superlens is explained in details.

1 Diffraction Limit

Due to the diffraction limit, the imaging resolution cannot be better than approximately half of the incident light's wavelength based on the conventional optical lenses. As the diffraction limit is one of the fundamental principles in optical physics, the subwavelength imaging with superresolution has not been possible based on typical lenses with positive refractive-index (RI) materials.

Following Fourier optics, an arbitrary form of light can be represented as a superposition of harmonic functions of space (x, y, z) and time (t),

$$f(x, y, z) = \int \int_{-\infty}^{\infty} F(k_x, k_y) e^{i(k_x x + k_y y)} e^{ik_z z} dk_x dk_y,$$

with a relation of $k_x^2 + k_y^2 + k_z^2 = \epsilon\mu \frac{\omega^2}{c^2} = k_0^2$

$F(k_x, k_y)$ is the Fourier transform of $f(x, y)$ and k_x, k_y are the spatial frequencies in the x and y directions, respectively.

Since an arbitrary form of light wave is able to be represented using Fourier analysis of a superposition integral of spatial frequency harmonic waves, the light propagated through an optical element may be written as a superposition of spatial

K. Kim (✉)
Yonsei University, 50 Yonsei-ro,
Seodaemun-gu, Seoul 120-749, Korea
e-mail: kks@yonsei.ac.kr

frequency plane waves. The physical meaning of this spatial spectral analysis is as follows: The image of an object is formed by the linear summation of waves scattered by the object, which have Fourier component spatial frequency of (k_x, k_y) and the propagation wave vector k_z . The transverse wave vectors are k_x, k_y and the propagation wave vector is k_z .

The propagating waves appear for low spatial frequency satisfying

$$k_x^2 + k_y^2 < \varepsilon\mu \frac{\omega^2}{c^2},$$

because the real value of k_z generates harmonically propagating term in z direction. Therefore, the waves of lower spatial frequency than the illuminating wavelength can propagate in materials without significant intensity loss. On the other hand, for the waves of higher spatial frequency than the illuminating wavelength,

$$k_x^2 + k_y^2 > \varepsilon\mu \frac{\omega^2}{c^2},$$

k_z is an imaginary value, hence, the harmonic wave component exponentially decays in the propagating direction. The subwavelength features of an object can be transferred and imaged by use of these evanescent waves of higher spatial frequency components. However, these components are confined only to the near-field of the object because of the exponential decay in the propagation direction. In the far-field imaging method using conventional lens, there are no contributions from the waves of high spatial frequency components and the subwavelength imaging of superresolution has not been possible with typical positive RI materials.

2 Negative Index Materials

In 1968, a Russian scientist, V. G. Veselago suggested the possibility of negative index materials (NIM) because there is no violation on fundamental physical laws even if the permittivity and permeability of materials are negative simultaneously [1]. From the Maxwell's wave equations, the refractive index is expressed in terms of relative permittivity and permeability, $n = \sqrt{\varepsilon\mu}$. We define a left-handed (LH) material as a material which has both negative permittivity ε and negative permeability μ at the same time. These are also called as double negative index materials (DNG).

In electromagnetism, two important parameters of permittivity ε and permeability μ determine and describe the most of electromagnetic phenomena of the materials. For the linear and homogenous medium, from the relation of followings,

$$\begin{aligned}
\vec{D} &= \varepsilon_0 \vec{E} + \vec{P} = \varepsilon_0 \varepsilon \vec{E} \\
\vec{B} &= \mu_0 \vec{H} + \vec{M} = \mu_0 \mu \vec{H} \\
\vec{P} &= \varepsilon_0 \chi_e \vec{E} \\
\vec{M} &= \chi_m \vec{H}
\end{aligned}$$

the Maxwell's equations are written as following,

$$\begin{aligned}
\nabla \cdot \vec{D} &= \rho_f \\
\nabla \cdot \vec{B} &= 0 \\
\nabla \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\
\nabla \times \vec{H} &= \vec{J}_f + \frac{\partial \vec{D}}{\partial t}
\end{aligned}$$

In general, the permittivity and permeability can be written as

$$\begin{aligned}
\varepsilon &= \varepsilon' + i\varepsilon'' \\
\mu &= \mu' + i\mu''
\end{aligned}$$

The imaginary parts of ε and μ are attributed to the losses in materials. For example, in the cases of a dielectric or a perfect electric conductor, the imaginary part of ε is zero or infinity.

If we consider the Maxwell's equation for a plane harmonic wave, $e^{i(\vec{k}\cdot\vec{r}-\omega t)}$,

$$\begin{aligned}
\vec{k} \times \vec{E} &= \omega \mu_0 \mu \vec{H}, \\
\vec{k} \times \vec{H} &= -\omega \varepsilon_0 \varepsilon \vec{E}.
\end{aligned}$$

For simplicity, if we consider lossless mediums without sources ($\vec{M} = \vec{J} = 0$), the imaginary parts of ε and μ are all zeroes. From the above relation, \vec{E} , \vec{H} , and propagation vector \vec{k} follow the right-hand rule for the typical right-handed (RH) materials. On the other hand, for the case of a LH material, $\varepsilon, \mu < 0$, hence, they can be represented as,

$$\begin{aligned}
\varepsilon &= -|\varepsilon| < 0, \\
\mu &= -|\mu| < 0.
\end{aligned}$$

Applying these relations into the above equations, we can get

$$\begin{aligned}
\vec{k} \times \vec{E} &= -\omega \mu_0 |\mu| \vec{H}, \\
\vec{k} \times \vec{H} &= +\omega \varepsilon_0 |\varepsilon| \vec{E}.
\end{aligned}$$

From these relations, \vec{E} , \vec{H} , and propagation vector \vec{k} follow the left-hand rule, differing from the right-hand rule in the case of positive index materials. As the phase velocity is parallel with propagation vector and the frequency is always a positive number, the phase velocity in a LH material is opposite direction of that in a RH material.

The propagation wave number k is positive in a RH medium and negative in a LH medium. Therefore, we can write a common relation as following [2],

$$\begin{aligned}\vec{k} \times \vec{E} &= s\omega\mu_0|\mu|\vec{H}, \\ \vec{k} \times \vec{H} &= -s\omega\epsilon_0|\epsilon|\vec{E},\end{aligned}$$

where s is a handedness sign parameter defined as

$$s = \begin{cases} +1 & \text{for RH material} \\ -1 & \text{for LH material.} \end{cases}$$

$$k = nk_0 = n\frac{\omega}{c}$$

and the refractive index is defined as $n = \pm\sqrt{\epsilon\mu}$.

In a LH medium, $k < 0$, and we have a negative value for the refractive index ($n < 0$). In generalized form, the refractive index can be represented as $n = s\sqrt{\epsilon\mu}$. Finally, a medium with negative permittivity and negative permeability has a negative index of refraction. In contrast, the power flow is related with the Poynting vector \vec{S} , defined as $\vec{S} = \vec{E} \times \vec{H}^*$.

The Poynting vector \vec{S} is in the same direction of energy propagation over time and the same magnitude of power flow. Therefore, \vec{S} is parallel with the group velocity, \vec{V}_g . The Poynting vector and the group velocity depend only on \vec{E} and \vec{H} , not on the signs of ϵ and μ . Also, the impedance of a medium,

$$Z = \sqrt{\frac{\mu\mu_0}{\epsilon\epsilon_0}},$$

still has positive sign even when $\epsilon < 0$ and $\mu < 0$ independently of s value. Therefore, at the interface between a NIM and a positive index material, a light can transmit perfectly if the impedance is matched. Finally, in RH medium, $v_p > 0$, $k > 0$, $v_g > 0$, $Z > 0$, and in LH medium, $v_p < 0$, $k < 0$, $v_g > 0$, $Z > 0$.

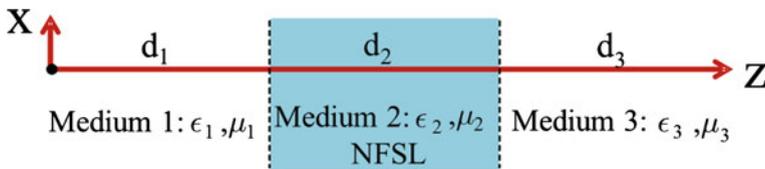


Fig. 1 Schematic of a slab superlens

3 Calculation of the Slab Near-Field Superlens Imaging System

Based on NIM, a lot of interesting phenomena are possible, such as inverse Doppler effect, inverse Snell’s law, inverse Cerenkov effect, optically perfect lens, and so on. In 2000, Pendry suggested a near-field superlens (NFSL) made of silver slab, which has superresolution beyond the diffraction limit [3].

Firstly, we calculate the transmission and reflection for a plane wave illuminating on a NIM slab suggested by Pendry, as shown in Fig. 1. Fresnel coefficients for the p-polarized light are represented as reflection coefficient (r_{ij} , incident light from i to j medium) and transmission coefficient (t_{ij} , incident light from i to j medium) at the interfaces.

$$r_{12} = \frac{k_z^{(1)}/\varepsilon_1 - k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2}$$

$$r_{23} = \frac{k_z^{(2)}/\varepsilon_2 - k_z^{(3)}/\varepsilon_3}{k_z^{(2)}/\varepsilon_2 + k_z^{(3)}/\varepsilon_3}$$

$$t_{12} = \frac{2k_z^{(1)}/\varepsilon_1}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2}$$

$$t_{23} = \frac{2k_z^{(2)}/\varepsilon_2}{k_z^{(2)}/\varepsilon_2 + k_z^{(3)}/\varepsilon_3}$$

where

$$k_z^{(1)} = \sqrt{\varepsilon_1\mu_1 \frac{\omega^2}{c^2} - k_x^2 - k_y^2}$$

$$k_z^{(3)} = \sqrt{\varepsilon_3\mu_3 \frac{\omega^2}{c^2} - k_x^2 - k_y^2}$$

$$k_z^{(2)} = \pm \sqrt{\varepsilon_2\mu_2 \frac{\omega^2}{c^2} - k_x^2 - k_y^2}.$$

To compute the total transmission and reflection of the slab with thickness d_2 ($\equiv d$), we summate the multiple reflections as following,

$$\begin{aligned}
T_P &= t_{12}t_{23}e^{ik_z^{(2)}d} + t_{12}r_{23}r_{21}t_{23}e^{3ik_z^{(2)}d} + t_{12}r_{23}r_{21}r_{23}r_{21}t_{23}e^{5ik_z^{(2)}d} + \dots \\
&= t_{12}t_{23}e^{ik_z^{(2)}d} + t_{12}(r_{23}r_{21})t_{23}e^{3ik_z^{(2)}d} + t_{12}(r_{23}r_{21})^2t_{23}e^{5ik_z^{(2)}d} + \dots \\
&= \frac{t_{12}t_{23}e^{ik_z^{(2)}d}}{1 - r_{23}r_{21}e^{2ik_z^{(2)}d}}
\end{aligned}$$

$$\begin{aligned}
R_P &= r_{12} + t_{12}r_{23}t_{21}e^{ik_z^{(2)}d} + t_{12}r_{23}r_{21}r_{23}t_{21}e^{3ik_z^{(2)}d} + t_{12}r_{23}r_{21}r_{23}r_{21}r_{23}t_{21}e^{5ik_z^{(2)}d} + \dots \\
&= r_{12} + t_{12}r_{23}t_{21}e^{ik_z^{(2)}d} + t_{12}r_{23}t_{21}(r_{23}r_{21})e^{3ik_z^{(2)}d} + t_{12}r_{23}t_{21}(r_{23}r_{21})^2e^{5ik_z^{(2)}d} + \dots \\
&= r_{12} + \frac{t_{12}r_{23}t_{21}e^{ik_z^{(2)}d}}{1 - r_{23}r_{21}e^{2ik_z^{(2)}d}}
\end{aligned}$$

To quantify the image quality of the superlens imaging system, the conventional optical transfer function (OTF) in optics, $H_{\text{img}}/H_{\text{obj}}$, the ratio of image field to the object field is introduced. The OTF from an object to an image in the slab superlens system is given by $OTF(k_x) = \frac{H_{\text{img}}}{H_{\text{obj}}} = e^{ik_z^{(3)}d_3} T_P(k_x) e^{ik_z^{(1)}d_1}$.

For a symmetric superlens, the host materials media 1 and 3 are the same.

$$\begin{aligned}
r_{12} &= \frac{k_z^{(1)}/\varepsilon_1 - k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \\
r_{23} = r_{21} &= \frac{k_z^{(2)}/\varepsilon_2 - k_z^{(1)}/\varepsilon_1}{k_z^{(2)}/\varepsilon_2 + k_z^{(1)}/\varepsilon_1} = -r_{12} \\
t_{12} &= \frac{2k_z^{(1)}/\varepsilon_1}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \\
t_{23} = t_{21} &= \frac{2k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2}
\end{aligned}$$

$$\begin{aligned}
T_P &= \frac{t_{12}t_{23}e^{ik_z^{(2)}d}}{1 - r_{23}r_{21}e^{2ik_z^{(2)}d}} = \frac{t_{12}t_{21}e^{ik_z^{(2)}d}}{1 - r_{21}^2e^{2ik_z^{(2)}d}} \\
&= \left(\frac{2k_z^{(1)}/\varepsilon_1}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \right) \times \left(\frac{2k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \right) \times \frac{e^{ik_z^{(2)}d}}{1 - \left(\frac{k_z^{(2)}/\varepsilon_2 - k_z^{(1)}/\varepsilon_1}{k_z^{(2)}/\varepsilon_2 + k_z^{(1)}/\varepsilon_1} \right)^2 e^{2ik_z^{(2)}d}} \\
&= \frac{2\varepsilon_2k_z^{(1)}}{\varepsilon_2k_z^{(1)} + \varepsilon_1k_z^{(2)}} \frac{2\varepsilon_1k_z^{(2)}}{\varepsilon_2k_z^{(1)} + \varepsilon_1k_z^{(2)}} \times \frac{e^{ik_z^{(2)}d}}{1 - \left(\frac{\varepsilon_1k_z^{(2)} - \varepsilon_2k_z^{(1)}}{\varepsilon_1k_z^{(2)} + \varepsilon_2k_z^{(1)}} \right)^2 e^{2ik_z^{(2)}d}}
\end{aligned}$$

$$R_P = r_{12} + \frac{t_{12}r_{23}t_{21}e^{ik_z^{(2)}d}}{1 - r_{23}r_{21}e^{2ik_z^{(2)}d}} = r_{12} - \frac{t_{12}r_{12}t_{21}e^{ik_z^{(2)}d}}{1 - r_{12}^2e^{2ik_z^{(2)}d}} = r_{12} \left(1 - \frac{t_{12}t_{21}e^{ik_z^{(2)}d}}{1 - r_{12}^2e^{2ik_z^{(2)}d}} \right)$$

$$= \frac{k_z^{(1)}/\varepsilon_1 - k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \left\{ 1 - \frac{\left(\frac{2k_z^{(1)}/\varepsilon_1}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \right) \left(\frac{2k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \right) e^{ik_z^{(2)}d}}{1 - \left(\frac{k_z^{(1)}/\varepsilon_1 - k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \right)^2 e^{2ik_z^{(2)}d}} \right\}$$

In the high spatial frequency limit of $\omega \ll c_0\sqrt{k_x^2}$, the wave numbers in the z direction inside medium 1, 2, 3 are approximated as $k_z^{(1)} \approx k_z^{(2)} \approx k_z^{(3)} \approx ik_x$ and r_{12} , r_{23} , t_{12} , t_{23} yield to

$$r_{12} = \frac{k_z^{(1)}/\varepsilon_1 - k_z^{(2)}/\varepsilon_2}{k_z^{(1)}/\varepsilon_1 + k_z^{(2)}/\varepsilon_2} \approx -\frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2} = -r_{21}$$

$$t_{12} = \frac{2\varepsilon_2}{\varepsilon_1 + \varepsilon_2} = 1 + r_{12}$$

$$t_{23} = t_{21} = \frac{2\varepsilon_1}{\varepsilon_1 + \varepsilon_2} = 1 + r_{21} = 1 - r_{12}.$$

Recalling $d_2 = d_1 + d_3 (= d)$, the OTF in this high spatial frequency limit, $k_x \gg k_0$, approximately becomes

$$OTF(k_x) \approx \frac{1 - r_{12}^2}{e^{2k_x d_2} - r_{12}^2},$$

$$r_{12}^2 \approx \left[\frac{(\varepsilon_1'^2 - \varepsilon_2'^2 + \varepsilon_1''^2 - \varepsilon_2''^2) + 2i(\varepsilon_1''\varepsilon_2' - \varepsilon_1'\varepsilon_2'')}{(\varepsilon_1' + \varepsilon_2')^2 + (\varepsilon_1'' + \varepsilon_2'')^2} \right]^2.$$

The field distribution at the image plane, $E_{img}(x)$, is obtained by the inverse Fourier transform of $E_{obj}(k_x)$ as follows:

$$E_{img}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} OTF(k_x)E_{obj}(k_x)e^{ik_x x} dk_x.$$

References

1. Veselago, VG.: Sov. Phys. Usp. **10**, 509 (1968)
2. Caloz, C., Itoh, T.: Electromagnetic metamaterials: Transmission line theory and microwave applications. Wiley-Interscience, New Jersey (2006)
3. Pendry, JB.: Phys. Rev. Lett. **85**, 3966 (2000)

Effective Material Parameter Retrieval for Terahertz Metamaterials

Teun-Teun Kim, Muhan Choi, Yushin Kim and Bumki Min

Abstract Metamaterials, which are generally composed of subwavelength scale metallic structures, have been the subject of intensive research in recent years. Because their effective electromagnetic properties can be engineered by designing subwavelength scale metallic structures, called ‘meta-atoms’, these artificially constructed materials are expected to lead to many new developments in the field of photonics. Furthermore, the terahertz (THz) frequency range has many important applications such as security detection, sensing, and biomedical imaging. Because many natural materials are inherently unresponsive to THz radiation, the natural materials that can be applied in devices in order to manipulate THz waves are very limited. Accordingly, the development of metamaterials with unusual optical properties in the THz frequency range has generated intense interest among researchers. In this part, design methods for metamaterials in the terahertz frequencies are introduced. This method is based on the unit cell design and *S*-parameter retrieval technique. Following a brief introduction to the method, some examples of terahertz metamaterial design will be presented in the last section.

Keywords Terahertz · Metamaterials · Unit cell · Effective medium theory · Drude model · *S*-parameter retrieval method · High refractive index · Negative permeability · Chirality

T.-T. Kim · Y. Kim · B. Min (✉)

Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-751, Republic of Korea
e-mail: bmin@kaist.ac.kr

M. Choi

Creative Research Center for Graphene Electronics, Electronics and Telecommunications Research Institute (ETRI), Daejeon 305-700, Republic of Korea

1 Introduction

Since metamaterials are typically periodic subwavelength scale metallic structures, the electromagnetic properties of entire structures can be determined using the analysis of a unit cell. The geometric size of the artificial atoms should be carefully selected in order to implement the desired material parameters with considerations of the functionality of the metamaterial and the frequency range of interest [1, 2]. For terahertz metamaterials, the operating frequencies are generally between 0.1 and 10.0 THz, which correspond to a wavelength range from 30.0 μm (far infrared) to 3.0 mm (microwave). In order that effective parameters are defined, metamaterials should be treated as effectively homogeneous media; for this reason, the size of the unit cell (L) must be much smaller than the wavelength of light (λ_0). Although there are no clear guidelines for applying a homogeneous medium approximation, it is generally accepted that the unit cell size should be smaller than a quarter wavelength, $L < \lambda_0/4$ (for 1 THz, $L < 75 \mu\text{m}$) so that the refractive effects dominate scattering and diffraction [3]. If the condition of effective homogeneity is satisfied, the microscopic detail of each individual structure and the corresponding microscopic field ‘distribution’ inside the unit cell becomes less important when describing the wave propagation through the metamaterial. In this case, it is the fields that are averaged over the unit cell that determine the effective material properties. Under the homogeneous assumption, metamaterials are usually characterized using two constitutive parameters: effective (relative) permittivity ε and (relative) permeability μ .

2 S-Parameter Retrieval Method

The scattering parameter (S -parameter) retrieval method [4, 5] is one of the most important techniques for assigning effective electromagnetic parameters to metamaterials. The S -parameter retrieval method has been widely used in material characterizations in microwave frequency ranges. With its simplicity and powerfulness, it has now become the most commonly used method for characterizing metamaterials over the entire range of electromagnetic spectra. The S -parameter retrieval method uses the transmission and reflection coefficients calculated by the finite difference time domain (FDTD) method or the finite element method (FEM). As the size of a unit cell is much smaller than the excitation wavelength, it is generally sufficient to include only the fundamental eigenmodes in the S -parameter retrieval method. From numerically calculated S -parameter sets, effective material parameters can be obtained by inverting the Fresnel equation. In this section, the S -parameter retrieval methods are reviewed for two distinct classes of metamaterials: metamaterials with linearly polarized eigenwaves [6, 7] and chiral metamaterials with circularly polarized eigenwaves [8–10].

2.1 Metamaterials with Linearly Polarized Eigenwaves

Elements of the scattering matrix (2×2) are given by S_{ij} . When a linearly polarized plane wave is normally incident on a homogeneous (meta) material slab with a thickness d , S_{11} is equal to the reflection coefficient R while S_{21} is related to the transmission coefficient T according to $S_{21} = T e^{i k_0 d}$, where k_0 represents the wave number of the incident wave in free space. These S -parameters are related to the complex refractive index n and impedance z using the Fresnel equations:

$$S_{11} = \frac{R_{01}(1 - e^{i2nk_0d})}{1 - R_{01}^2 e^{i2nk_0d}}, \quad (1)$$

$$S_{21} = \frac{(1 - R_{01}^2) e^{i nk_0 d}}{1 - R_{01}^2 e^{i2nk_0d}}, \quad (2)$$

where $R_{01} = z - 1/z + 1$.

The refractive index n and impedance z can be found by inverting Eqs. (1) and (2), yielding:

$$z = \pm \sqrt{\frac{(1 + S_{11})^2 - S_{21}^2}{(1 - S_{11})^2 - S_{21}^2}}, \quad (3)$$

$$e^{i nk_0 d} = X \pm i \sqrt{1 - X^2}, \quad (4)$$

where $X = 1/2S_{21}(1 - S_{11}^2 + S_{21}^2)$. Because the metamaterial under consideration is a passive medium, the signs in Eqs. (3) and (4) are determined by the requirements that $\text{Re}(z) \geq 0$ and $\text{Im}(n) \geq 0$ [6]. Using Eq. (4), the value of the refractive index n can be determined as:

$$n = \frac{1}{k_0 d} \{ [\text{Im}[\ln(e^{i nk_0 d})] + 2m\pi] - i \text{Re}[\ln(e^{i nk_0 d})] \}, \quad (5)$$

where m is an integer related to the branch index of $\text{Re}(n)$. The imaginary part of n is uniquely determined, but the real part is ambiguous due to the logarithm function branches. Therefore, some physical conditions must be applied in order to determine the branch number m . One such condition is that the branch number at zero frequency should be zero because there is no phase shift in the transmitted wave in a quasi-static regime. Another condition arises from the requirement that the branch number must be increased or decreased to prohibit discontinuities in the refractive index for the frequency range of interest (for which the homogeneous assumption is valid). The permittivity ε and permeability μ are then directly obtained from the calculated refractive index n and impedance z , using the relationships of $\varepsilon = n/z$ and $\mu = nz$.

2.2 Chiral Metamaterials with Circularly Polarized Eigenwaves

Typical chiral metamaterials are composed of subwavelength scale metallic structures that cannot be superposed on their mirror images. Since chiral metamaterials exhibit interesting electromagnetic phenomena such as large circular birefringence and dichroism, they are attracting much attention for potential application in polarizing devices. Chiral media can be characterized using the following constitutive relations [11]:

$$\mathbf{D} = \varepsilon_0 \varepsilon \mathbf{E} + i\kappa \sqrt{\mu_0 \varepsilon_0} \mathbf{H}, \quad (6)$$

$$\mathbf{B} = \mu_0 \mu \mathbf{H} - i\kappa \sqrt{\mu_0 \varepsilon_0} \mathbf{E}, \quad (7)$$

where κ is the dimensionless chirality parameter that describes the coupling between the electric fields and magnetic fields in the material. As can be seen from Eqs. (6) and (7), an electric dipole can also be induced by the magnetizing field (as well as the electric field), whereas a magnetic dipole can be induced by the electric field. Due to the rotational symmetry of the chiral metamaterials in the plane perpendicular to the axis of wave propagation, the eigenwaves of the propagation are circularly polarized waves rather than linearly polarized waves. For a slab of chiral metamaterial, the retrieval process is similar to that introduced earlier (Sect. 2.1) for conventional metamaterials, except for a slight complication that stems from the cross coupling terms (related to κ).

For chiral materials, two separate refractive indices for right (n_+) and left (n_-) circularly polarized eigenwaves need to be obtained [9, 10]. If T and R are denoted as the transmission and reflection coefficients, respectively, the impedance and refractive indices can be extracted, respectively.

$$z = \pm \sqrt{\frac{(1+R)^2 - T_+ T_-}{(1-R)^2 - T_+ T_-}}, \quad (8)$$

$$n_{\pm} = \frac{i}{k_0 d} \left\{ \log \left[\frac{1}{T_{\pm}} \left(1 - \frac{z-1}{z+1} R \right) \right] \pm 2m\pi \right\}, \quad (9)$$

where $R_{\pm} = R$, $T_{\pm} = T e^{i\kappa k_0 d}$, and m is an integer that can be determined using the branch considerations. The sign and branch integer in Eqs. (8) and (9) must be chosen carefully by imposing the conditions of $\text{Re}(z) \geq 0$ and $\text{Im}(n) \geq 0$ for passive media. Once z and n_{\pm} are obtained, the remaining parameters can be determined from the following relations: $n = (n_+ + n_-)/2$, $\kappa = (n_+ - n_-)/2$, $\varepsilon = n/z$, and $\mu = nz$.

3 Examples of Parameter Retrieval

Three representative examples of THz metamaterials will be introduced and analyzed using the S -parameter retrieval method that was introduced in previous sections: I-shaped metallic patch metamaterials for high refractive index, parallel metal plate metamaterials for negative permeability, and conjugated gammadion metamaterial for chirality.

3.1 I-Shaped Metallic Patch Metamaterials for High Refractive Index

Metamaterial composed of I-shaped metallic patch structures can be used to provide a wide range of refractive indices by varying the geometric dimensions of the I-shaped metallic ‘meta-atoms’. For this reason, I-shaped patch structures have been used previously for the demonstration of a ground plane cloak with a moderate range of refractive index variations [12]. Recently, Choi et al. [13] explored a new regime of unnaturally high refractive indices using similar I-shaped metamaterials at THz frequencies. The high refractive index THz metamaterials that are considered in this section have an identical structure to that presented by Choi et al. [13].

The effective permittivity of a metamaterial can be expressed as $\varepsilon = 1 + (P/\varepsilon_0 E)$ and the effective permeability can be defined by $\mu = 1 + (M/H)$, where E , H , P , and M represent the electric field, magnetizing field, polarization, and magnetization, respectively. In order to achieve high effective indices, Choi et al. [13] maximized the effective permittivity by placing the I-shaped meta-atoms sufficiently close to each other in order to enhance the capacitive coupling and minimize the diamagnetic effect by decreasing the central beam width and the thickness of the metallic patches.

In Fig. 1, an I-shaped gold patch embedded in a polyimide substrate ($n_{\text{poly}} = 1.8 + 0.04i$) is schematically shown with definitions of the geometrical parameters. The complex dielectric constants of gold for the frequency range of interest (~ 1 THz) can be fitted using the Drude model with a plasma frequency of $\omega_p = 1.37 \times 10^{16} \text{ s}^{-1}$ and collision frequency of $\gamma = 4.07 \times 10^{13} \text{ s}^{-1}$ [14]. The direction of the propagation and polarization of the incident THz waves are indicated by the arrows in Fig. 1. The geometric parameters used for the simulation of a high index I-shaped metamaterial are: $L = 60 \text{ }\mu\text{m}$, $a = 58.5 \text{ }\mu\text{m}$, $w = 3 \text{ }\mu\text{m}$, $t = 0.1 \text{ }\mu\text{m}$, and $d = 2.5 \text{ }\mu\text{m}$. The gap width between the metallic patches is defined as $g = L - a$.

In the unit cell calculations, the periodic boundary conditions are usually applied in the transverse plane. After running the frequency domain or time domain simulations with the appropriate boundary conditions, the S -parameters can be extracted from the simulation results. The effective material parameters are

Fig. 1 Schematic view of I-shaped high index metamaterial unit cell

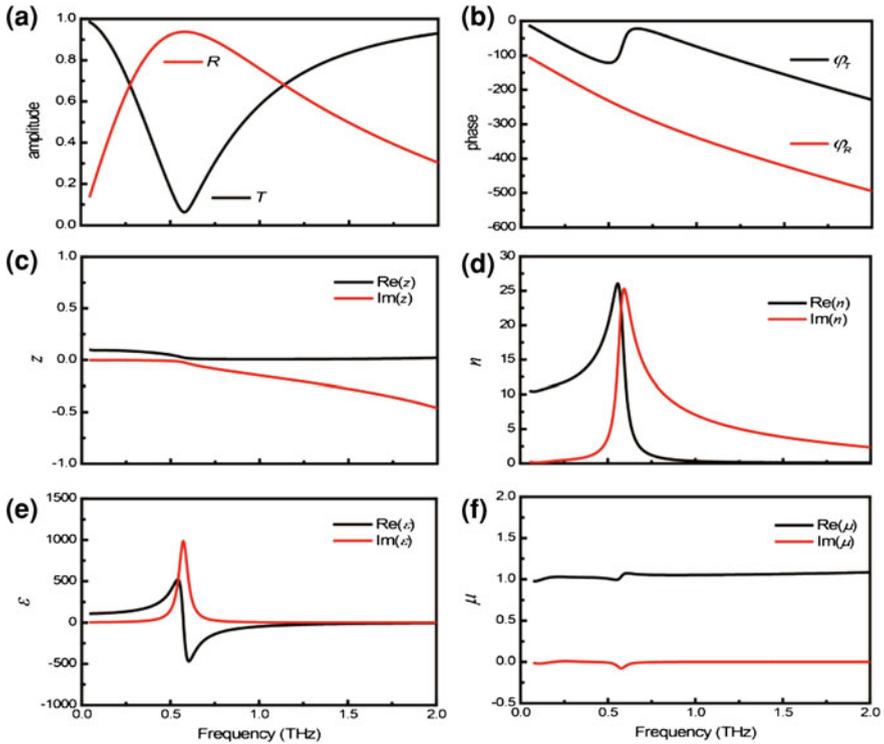
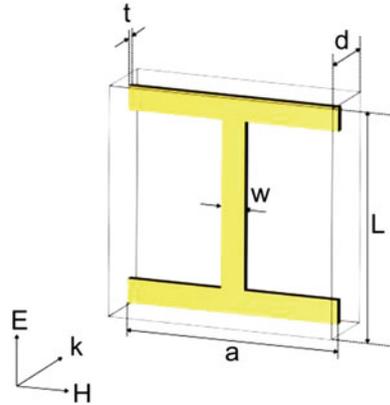


Fig. 2 Frequency dependent THz transmission (*black curves*) and reflection (*red curves*) **a** amplitude, **b** phases for I-shaped metamaterials, and the retrieved effective parameters for **c** impedance, **d** refractive index, **e** permittivity, and **f** permeability

then obtained using the *S*-parameter retrieval method introduced earlier in [Sect. 2.1](#). Figure. 2a and b show the simulated amplitude and phase of transmission/reflection plotted as a function of frequency; as can be seen in the figure, an

electric resonance and the corresponding dip in the transmission are observed near 0.5 THz. The refractive index n (Fig. 2c) and the impedance z (Fig. 2d) can be obtained from Eqs. (3) and (5), respectively. It can be seen that the real part of z and the imaginary part of n satisfy the law of energy conservation for passive mediums. Then, the permittivity ϵ and permeability μ are obtained directly using the relations $\epsilon = n/z$ and $\mu = nz$ from the calculated n and z (Figs. 2e and f). The peak value (~ 542) of the relative permittivity ϵ at a resonance frequency (0.5 THz) is significantly higher than that of bare polyimide substrate films ($\text{Re}(\epsilon)$, ~ 3.24).

In order to quantify the effect of the gap width variation between the I-shaped metallic patches, the effective parameters are plotted in Fig. 3 as a function of the gap width g . In this plot, the effective permeability does not significantly change over the frequency range of interest, but the effective permittivity and index increase gradually as the gap size g decreases. In order to confirm the previously mentioned physical origin of the extremely high refractive index, the electric fields around the metallic patches with $g = 30 \mu\text{m}$ and $g = 1.5 \mu\text{m}$ were calculated for the resonance frequencies corresponding to 1.75 and 1.5 THz, respectively. It is clear that the electric field is strongly concentrated in the gap, especially when the gap was very narrow (Fig. 4b).

The diamagnetic effect can be reduced effectively by employing an I-shaped metallic patch structure with a small central beam width (w_c) and thickness (t). In order to verify the effect of the metallic volume fraction on the magnetic permeability, the effective permeability is plotted in Fig. 5a as a function of w_c . The thickness of the metallic patch and the gap size were fixed at 1.0 and 1.5 μm , respectively. The plot illustrates that the effective permeability increases gradually and the value approaches unity over the frequency band of interest as the central beam width w_c decreases. Additionally, the effective permeability is plotted as a function of the thickness of the metallic patch t in Fig. 5b. As can be seen from the plot, the thickness of the metallic patch is also important in decreasing the diamagnetic effect.

3.2 Parallel Metal Plate Metamaterials for Negative Permeability

Split ring resonators (SRRs), which are typically composed of two metallic rings with split openings facing opposite directions, are a well known design for strong artificial magnetism [15]. From the viewpoint of circuit theory, SRRs are equivalent to a subwavelength scale LC resonator. When an external magnetizing field is coupled to an SRR, a resonant current is induced in each split ring, thereby creating a magnetic dipole moment. If the magnetic response is sufficiently strong, effective magnetic permeability μ can be negative near the resonance frequency. For this range of frequencies, the magnetization is in the opposite direction to the magnetizing field and its magnitude is larger than that of the magnetizing field. In the following example, it will be demonstrated that the negative permeability can

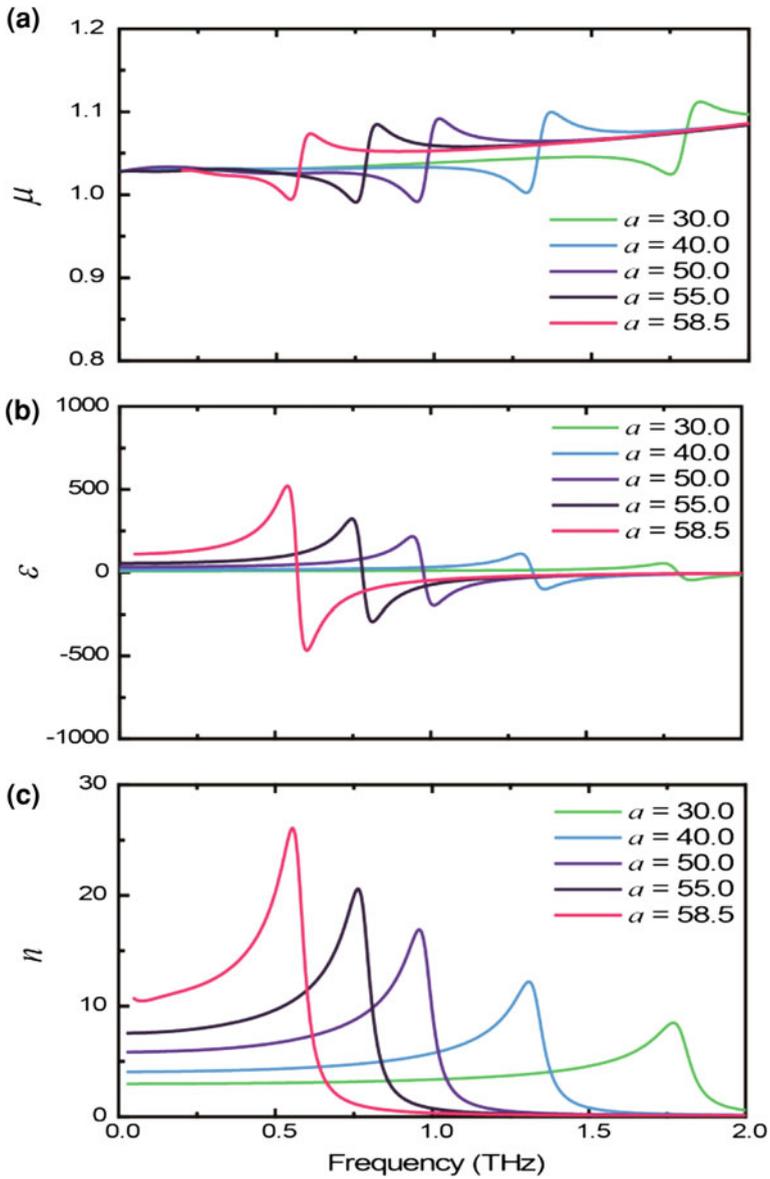


Fig. 3 Gap width dependent behavior of effective **a** permittivity, **b** permeability, and **c** index

be obtained by applying the *S*-parameter retrieval method to two parallel metal plates that have a topology which is equivalent to that of SRRs.

Figure 6 shows the schematic view of the parallel metal plate metamaterial. The parallel metal plates were embedded in a polyimide substrate with a complex refractive index of $1.8 + 0.04i$. The geometric parameters were given as $L = 40$,

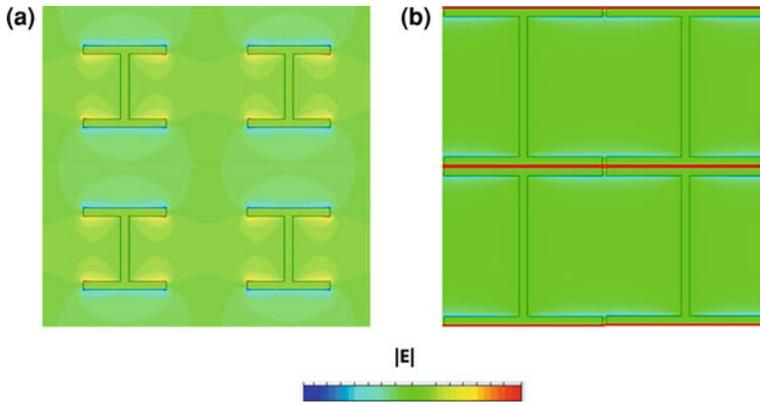


Fig. 4 Electric field plots for I-shaped metamaterials with **a** $g = 30 \mu\text{m}$ and **b** $g = 1.5 \mu\text{m}$

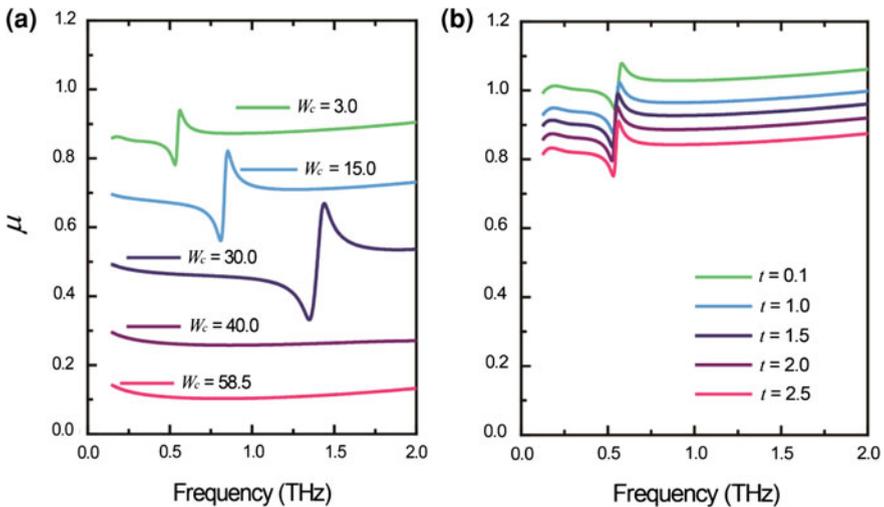


Fig. 5 Effective permeability of I-shaped high index metamaterials with variations in the metallic volume fraction

$a = 38.5$, and $t = 0.1 \mu\text{m}$. The amplitude/phase of the transmission/reflection for the two parallel metal plates obtained from the simulation is shown in Figs. 7a and b. Using the transmission and reflection coefficients, the retrieved effective parameters can be obtained using the *S*-parameter retrieval method (Fig. 7). As predicted by the topological analogy with the SRRs, it can be seen that the real part of the permeability μ becomes negative at approximately 2.0 THz (magnetic resonance frequency).

When an electromagnetic wave is incident on the parallel metal plates, an effective magnetic dipole moment is induced using the anti-parallel currents

Fig. 6 Unit cell structure of two parallel metal plates

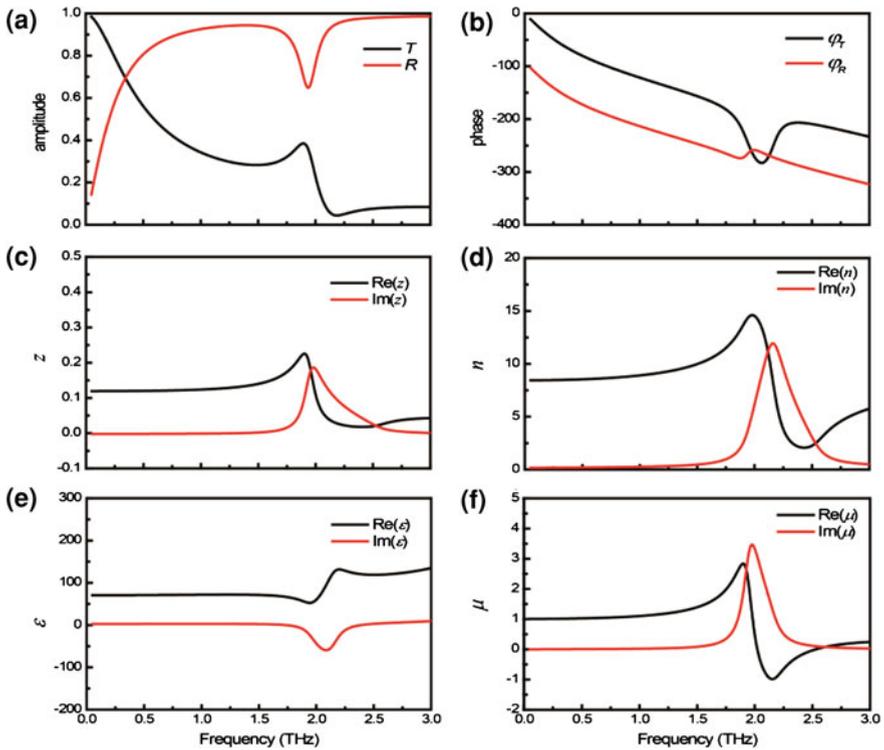
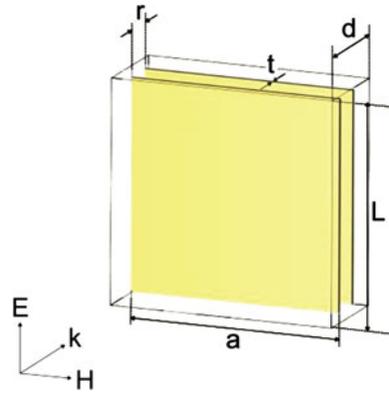


Fig. 7 Frequency dependent THz transmission (*black curves*) and reflection (*red curves*) **a** amplitude and **b** phases for parallel metal plate metamaterial, and the retrieved effective parameters of **c** impedance, **d** refractive index, **e** permittivity, and **f** permeability

flowing in each metallic plate (as shown in Fig. 8). For the parallel metal plate metamaterials, the magnetic permeability is strongly dependent on the distance (r) between the metal plates as well as other parameters. For simplicity, the gap

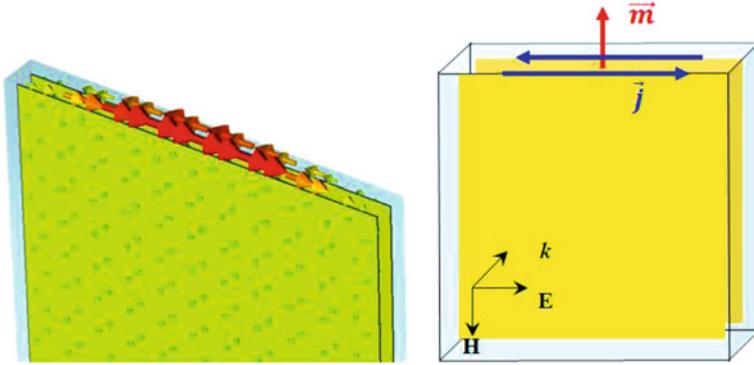


Fig. 8 Current distribution in a double layered metallic structure; the magnetic dipole moment (\vec{m}) is induced by a current loop on top of the parallel plates

width between adjacent metallic plates ($g = L - a$) is fixed and only the distance variation is considered here. If two parallel metallic plates are sufficiently close to each other, the induced current can be sufficiently large due to the strong capacitive effect. This effect leads to the observation of negative effective permeability near the resonance frequency.

In order to verify the dependency of the magnetic permeability on the distance between the plates, the effective permeability extracted using the *S*-parameter retrieval method is plotted as a function of distance in Fig. 9. The plot clearly demonstrates that the negative effective permeability can be obtained when the metallic plates are sufficiently close to each other (in the case of $r = 1$ and $2 \mu\text{m}$).

3.3 Conjugated Gammadion Metamaterial for Chirality

In this section, the *S*-parameter retrieval method is applied to specific THz chiral metamaterials. For this purpose, the conjugated gammadion structure was chosen as a chiral meta-atom. This structure was proposed for operation at microwave frequencies by Zhao et al. [16]; however, in this section, the size of unit cell was modified in order to shift the operating frequency into the THz regime. Figure. 10 shows the detailed structure of the conjugated gammadion with the definitions of the geometrical parameters. The geometric parameters were $a = 50 \mu\text{m}$, $l = 40 \mu\text{m}$, $w = 4 \mu\text{m}$, $t = 0.1 \mu\text{m}$, $d = 5 \mu\text{m}$, and $r = 5 \mu\text{m}$.

The transmission coefficients for the linearly polarized waves are expressed as T_{xx} , T_{xy} , T_{yx} , and T_{yy} , where the first subscript indicates the polarization (x or y) of the transmitted wave and the second subscript indicates the polarization of the incident wave. From these transmission coefficients calculated in a linear basis, the

Fig. 9 Effective permeability as a function of distance between two plates (r)

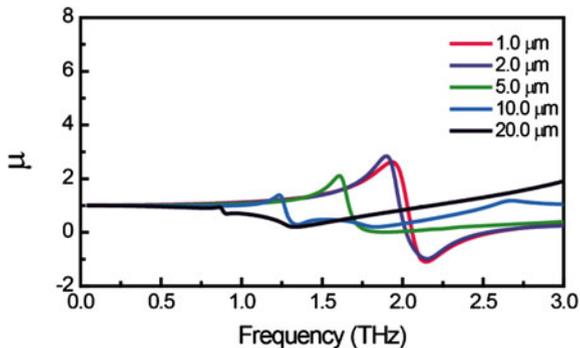
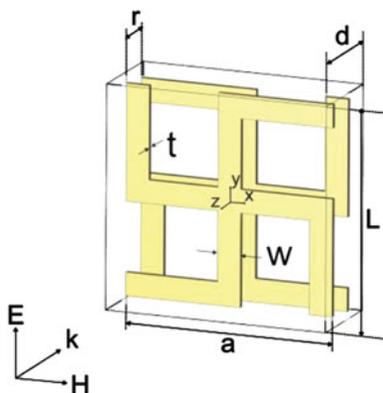


Fig. 10 Unit cell structure of the chiral metamaterial composed of conjugated gammadions



right and left circularly polarized transmission (RCP/LCP) coefficients (T_{++} , T_{+-} , T_{-+} , and T_{--}) can be obtained as:

$$\begin{pmatrix} T_{++} & T_{+-} \\ T_{-+} & T_{--} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (T_{xx} + T_{yy}) + i(T_{xy} - T_{yx}) & (T_{xx} - T_{yy}) - i(T_{xy} + T_{yx}) \\ (T_{xx} - T_{yy}) + i(T_{xy} + T_{yx}) & (T_{xx} + T_{yy}) - i(T_{xy} - T_{yx}) \end{pmatrix}. \quad (10)$$

Figures 11a and b illustrate the numerically obtained amplitudes and phases of the transmission/reflection coefficients in a linear basis. Since the gammadion structure possesses C_4 symmetry in the z axis, the transmission coefficients in the linear basis satisfy the relations $T_{xy} = T_{yx} = T_{\parallel}$ and $T_{xy} = -T_{yx} = T_{\perp}$. Similarly, the reflection coefficients are expressed as $R = R_{\parallel}$ and $R_{\perp} = 0$. As required by the C_4 symmetry, the cross-coupling transmission coefficients T_{-+} and T_{+-} should be zero and Eq. (10) can be reduced to:

$$T_{++} = T_{+} = T_{\parallel} + iT_{\perp}, \quad T_{--} = T_{-} = T_{\parallel} + iT_{\perp}, \quad R_{++} = R_{--} = R. \quad (11)$$

Figure 12a shows the simulated transmission coefficients $|T_{+}|$ and $|T_{-}|$ for the circularly polarized waves estimated from the linear basis using Eq. (11). The

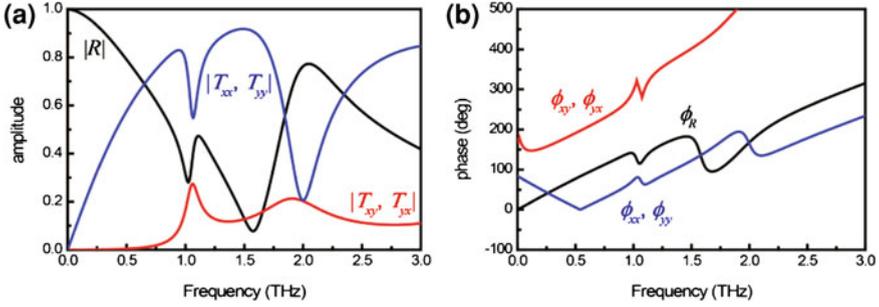


Fig. 11 a Amplitude and b phase of the transmission and reflection coefficients for the linearly polarized waves

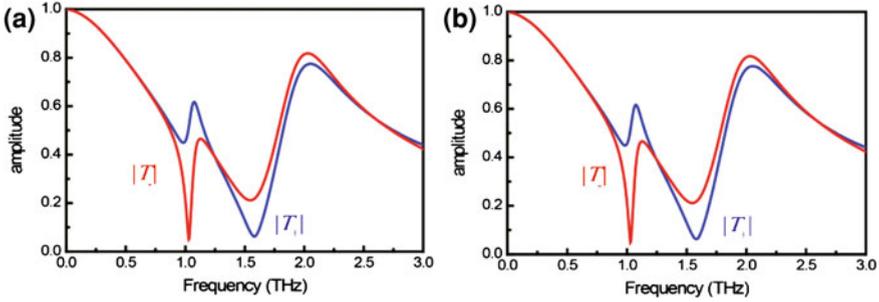


Fig. 12 Comparison between the transmission coefficients for circularly polarized waves obtained indirectly from the linear basis and those simulated with circularly polarized waves

transmission amplitudes have two resonances at the frequencies near 1 and 1.6 THz. As can be seen clearly in the plot, there are obvious differences between $|T_+|$ and $|T_-|$ near these resonances. In order to validate the conversion of the transmission coefficients, a conjugated gammadion structure with circularly polarized waves (Fig. 12b) was simulated directly. Excellent agreement was confirmed between the values obtained in the linear basis and circular basis. With the transmission coefficients ($|T_+|$ and $|T_-|$), and the reflection coefficient ($|R|$), Eqs. (8) and (9) can be applied in order to retrieve the impedance, refractive index, and remaining parameters (Fig. 13). As the refractive indices (n_+ and n_-) are modified from the linear refractive index by the chirality ($n_{\pm} = n \pm \kappa$), the real part of the refractive indices for the RCP and LCP waves can reach negative values near the resonance frequencies, even without negative refractive index (n). In particular in this example, the chirality and negative refractive index for the LCP waves were very large at the resonance frequency of 1.5 THz ($\kappa = 7.3$, $n_- = -9.8$).

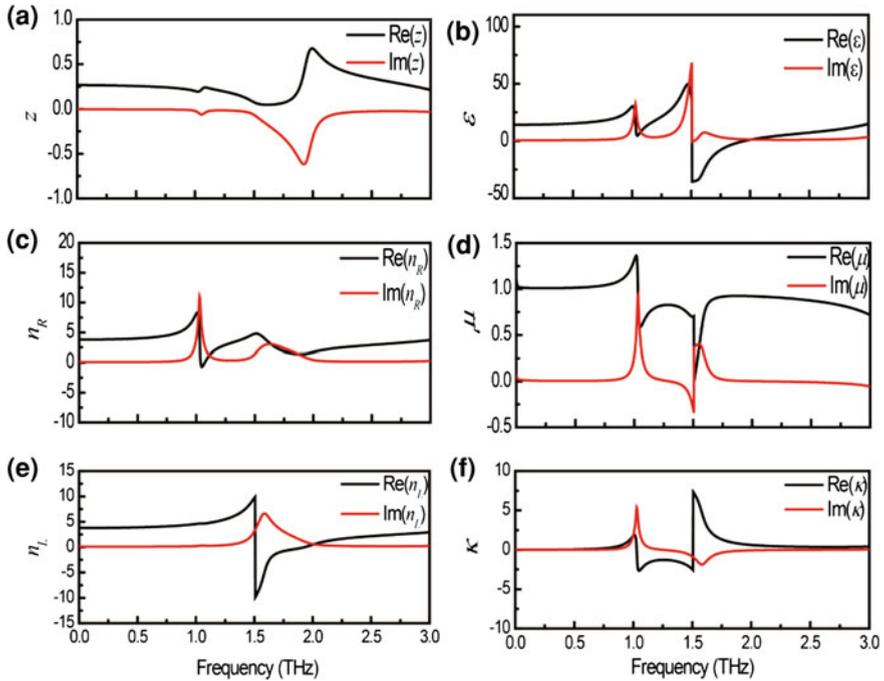


Fig. 13 Retrieved effective parameters of conjugated gammadion chiral metamaterial

4 Conclusion

In this part, the S -parameter retrieval method has been reviewed and retrieval techniques were applied in order to obtain effective material parameters for three different types of THz metamaterials. These examples demonstrated that the metamaterial design can be facilitated using these retrieval methods. With the help of physical intuition, this retrieval method is a powerful tool for the prediction of metamaterial functionalities.

References

1. Smith, D.R., Pendry, J.B., Wiltshire, M.C.K.: Metamaterials and negative refractive index. *Science* **305**, 788–792 (2004)
2. Ramakrishna, S.A.: Physics of negative refractive index materials. *Rep. Prog. Phys.* **68**, 449–521 (2005)
3. Caloz, C., Itoh, T.: *Electromagnetic Metamaterials Transmission Line Theory and Microwave Applications: the Engineering Approach*. Wiley-Interscience, New Jersey (2006)
4. Weir, W.B.: Automatic measurement of complex dielectric constant and permeability at microwave frequencies. *Proc. IEEE* **62**, 33–36 (1974)

5. Nicholson, A.M., Ross, G.F.: Measurement of the intrinsic properties of materials by time-domain techniques. *IEEE Trans. Instrum. Meas.* **19**, 377–382 (1970)
6. Chen, X., Grzegorzczak, T.M., Wu, B.-I., Pacheco, J., Kong, J.A.: Robust method to retrieve the constitutive effective parameters of metamaterials. *Phys. Rev. E* **70**, 016608 (2004)
7. Smith, D.R., Vier, D.C., Koschny, Th, Soukoulis, C.M.: Electromagnetic parameter retrieval from inhomogeneous metamaterials. *Phys. Rev. E* **71**, 036617 (2005)
8. Wang, B., Zhou, J., Koschny, T., Kafesaki, M., Soukoulis, C.M.: Chiral metamaterials: simulations and experiments. *J. Opt. A: Pure Appl. Opt.* **11**, 114003 (2009)
9. Kwon, D.-H., Werner, D.H., Kildishev, A.V., Shalaev, V.M.: Material parameter retrieval procedure for general bi-isotropic metamaterials and its application to optical chiral negative-index metamaterial design. *Opt. Express* **16**, 11822–11829 (2008)
10. Zhao, R., Koschny, T., Soukoulis, C.M.: Chiral metamaterials: retrieval of the effective parameters with and without substrate opt. *Express* **18**, 14553–14567 (2010)
11. Lindell, I.V., Sihvola, A.H., Tretakov, S.A., Vitanen, A.J.: *Electromagnetic Waves in Chiral and Bi Isotropic Media*. Artech House, Boston (1994)
12. Liu, R., Ji, C., Mock, J.J., Chin, J.Y., Cui, T.J., Smith, D.R.: Broadband Ground-Plane Clock. *Science* **323**, 366–369 (2009)
13. Choi, M., Lee, S.H., Kim, Y., Kang, S.B., Shin, J., Kwak, M.H., Kang, K.-Y., Lee, Y.-H., Park, N., Min, B.: A terahertz metamaterial with unnaturally high refractive index. *Nature* **470**, 369–373 (2011)
14. Ordal, M.A.: Optical properties of the metals Al, Co, Cu, Au, Fe, Pb, Ni, Pd, Pt, Ag, Ti, and W in the infrared and far infrared. *Appl. Opt.* **22**, 1099–1119 (1983)
15. Pendry, J.B., Holden, A.J., Robbins, D.J., Stewart, W.J.: Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Trans. Microw. Theory Tech.* **47**, 2075–2084 (1999)
16. Zhao, R., Zhang, L., Zhou, J., Koschny, Th, Soukoulis, C.M.: Conjugated gammadion chiral metamaterial with uniaxial optical activity and negative refractive index. *Phys. Rev. B* **83**, 035105 (2011)

Ultrabroadband Terahertz Spectroscopy

Inhee Maeng, Jaeseok Kim and Hyunyong Choi

1 Introduction

Low energy region correspond to the Boltzmann thermal energy ($k_B T$) show condensed matter properties and electric phenomena. Terahertz (THz) spectroscopy technique enables us to access this low energy level excitation including the conductivity of itinerant charges, plasmons and polarons [1, 2]. THz spectroscopy measures material's ground state at finite temperature and its dynamics as a result of electromagnetic or electronic perturbation.

The importance of THz spectroscopic information has made the researchers keep finding the efficient source and sensitive detectors. THz region which is called THz Gap, resides between the microwave electronics and infrared optoelectronics region. Upscaling the speed of electronic circuit is limited by the scattering and downscaling optics is also limited by the thermal excitation. There are many recent developments to overcome these electronic and photonic limits. In the electronic technique, the solid state oscillators such as Gunn and impact avalanche and transit time (IMPATT) diode and the Schottky diode multipliers are used to generate sub-THz radiation. In the photonic technique, quantum cascade laser (QCL) for specific THz frequency region has been developed.

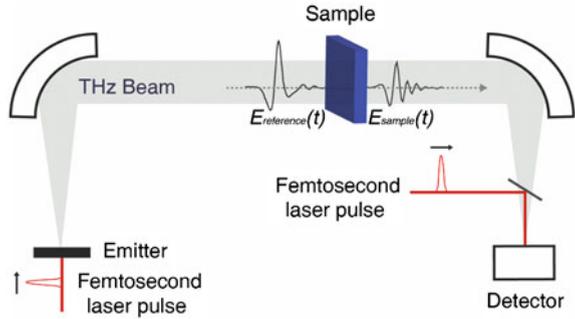
In fact, a method based on the ultrafast pulse laser is known to be the most efficient way to generate and detect a wide range of THz spectrum. In this part, we will review the generation and detection methods of coherent few-cycle THz pulse with table-top ultrafast lasers and their THz spectrum.

The general scheme of THz time domain spectroscopy (THz-TDS) is shown in Fig. 1. The femtosecond near-IR laser pulses are used to generate and detect of the coherent few-cycle THz pulse. THz-TDS measures the amplitude and phase

I. Maeng · J. Kim · H. Choi (✉)

School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
e-mail: hychoi@yonsei.ac.kr

Fig. 1 A schematic diagram of terahertz time domain spectroscopy using femtosecond laser pulse to generate and detect coherent terahertz field



information of the THz field. The THz properties of sample are measured with the difference between the generated THz pulse $E_{reference}(t)$ and the transmitted sample THz pulse $E_{sample}(t)$. Using Fast Fourier Transform, the reference and sample signals provide complex-valued optical properties such as refractive indices, dielectric constants and conductivities [3].

In the Sect. 2, we will discuss the generation and detection with photoconductive switching technique. In the Sects 3 and 4, optical rectification and electro-optic sampling for ultrabroadband THz generation and detection will be presented.

2 Photoconductive Switching Technique

Photoconductive antennas are extensively used for THz generation and detection. The photoconductive technique uses photoconductive antennas which are electrically shortened by incident laser beam. The antennas are made of a pair of metal striplines with a μm gap between them on the photoconductive semiconductor surface as shown in Fig. 2a [4, 5]. The electrodes, the antenna striplines, are biased across the gap. The bias current creates the electric field on the edge of the dipole gap as shown in Fig. 3a and the electrons and holes are accumulated. When the femtosecond laser pulse illuminates on the semiconducting gap, photo-generated electrons and holes are accelerated by the bias field as shown in Fig. 3b. The current density is expressed as

$$J(t) = N(t)e\mu E_{dc} \quad (1)$$

where $N(t)$ is density of photocarriers, e is the electronic charge, μ is the carrier mobility, and E_{dc} is the bias electric field. A fast current pulse $J(t)$ is set by the laser pulse duration and RC time constant of the antenna structure [5, 6]. As decay time is influenced by the carrier recombination, short decay time enables to cover broadband THz region, for which a low-temperature grown GaAs or a radiation-damaged silicon-on-sapphire are used for the antenna substrate. The generated THz is proportional to the differential of photocurrent as follows

$$E_{THz} \propto \frac{\partial}{\partial t} J(t) \quad (2)$$

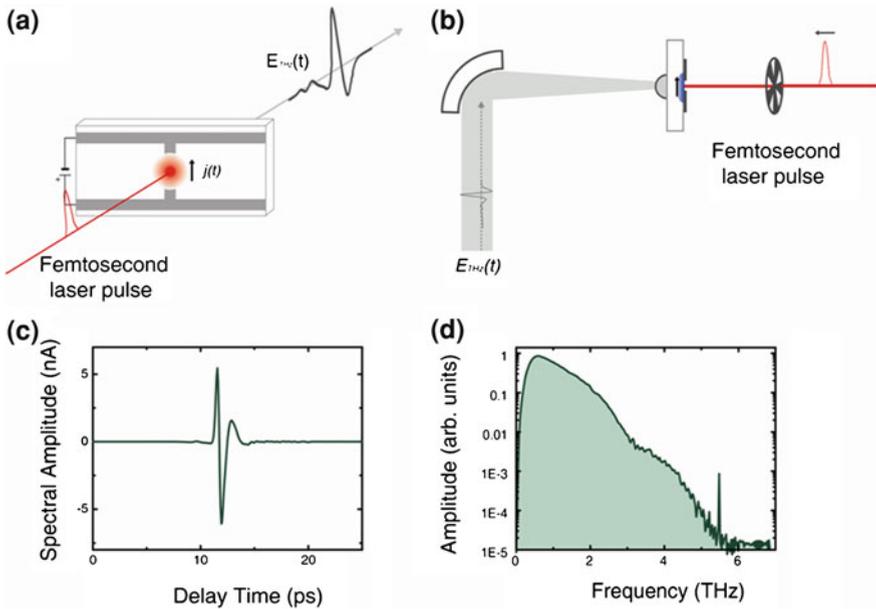
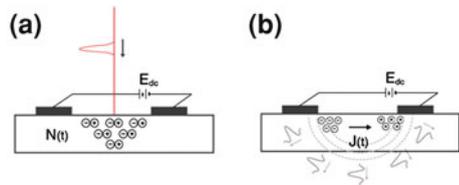


Fig. 2 Schematic diagram of generation **a** and detection **b** using photoconductive switching technique. Temporal waveform generated and detected via photoconductive switching method **c** and corresponding THz spectra **d**

Fig. 3 The photo-generated electron and hole by femtosecond laser pulse on the semiconducting gap **a** and THz generated by photocurrent when the photo-generated electron and hole are accelerated by the bias field **b**



The emitted THz by the photoconductive switching technique covers several THz.

THz cross detection in a photoconductive antenna is the almost same as that of the THz generation in PC antenna. The photoconductive antennas induce the DC photocurrent depending on cross-correlation between the THz pulses and the photo carriers which are generated by the incident femtosecond laser pulses (Fig. 2b). The THz field induced photocurrent $J(t)$ is expressed as

$$\bar{J}(\tau) = \bar{N}e\mu E(\tau) \tag{3}$$

where \bar{N} is the average electron density and τ is the time delay between the THz electromagnetic pulses and the probe laser pulses. Current amplifier reads the photo current gating by detecting the laser pulse.

Small width of the dipole gap and the distance between the antenna striplines are able to generate wide spectrum range, but the generation power is reduced with the dipole gap antenna and the distance between the coplanar lines. The width of the dipole gap and the distance between the coplanar lines are affected as the trade-off between the generation power and the spectrum width.

Photoconductive sampling has high signal-to-noise ratio (over 1:100,000), and low background noise caused by low dark current. Measured temporal waveform generated and detected by photoconductive switching technique and its spectral range are shown in Fig. 2c, d. Photoconductive detection has bandwidth limit of about 5 THz by decay time and laser pulse duration. The decay time is an intrinsic property of the substrate material used, however the laser pulse duration is tunable. The laser pulse duration mostly affect the high frequency region, and short laser pulses provide boarder spectral range. 10 femtosecond laser enables to demonstrate 95 THz ultrabroadband THz detection without the effect of decay time [7].

3 Optical Rectification Technique

In this section, we will review the THz generation by optical rectification which has benefit of ultrabroadband spectral bandwidth. Typical photoconductive antenna and femtosecond laser pulse generate THz pulse up to 3 THz bandwidth. Compared to the THz generation by the photoconductive antenna, the optical rectification technique easily obtains ultrabroadband THz spectrum of about 30 THz.

Optical rectification is a second-order nonlinear effect in nonlinear crystals. It is basically part of different frequency mixing used for generation of THz or mid-IR pulses with spectrally separated pulses. Optical rectification explains the process which has very small frequency difference, close to zero of the incidence pulses.

In the optical rectification, a near-IR femtosecond laser pulse is used as a source for second-order nonlinear effect. The applied electric field \mathbf{E} affects on a polarization of nonlinear crystals such as Zinc Telluride (ZnTe), Gallium Phosphide (GaP), or Gallium Selenide (GaSe). The polarization \mathbf{P} induced by electric field \mathbf{E} is expressed as [8]

$$\mathbf{P} = \epsilon_0 \left(\chi^{(1)} \mathbf{E} + \chi^{(2)} \mathbf{E} \mathbf{E} + \chi^{(3)} \mathbf{E} \mathbf{E} \mathbf{E} + \dots \right) \quad (4)$$

where $\chi^{(n)}$ is the n th order nonlinear susceptibility tensor. Optical rectification uses the nonlinear crystal which has nonzero second-order susceptibility tensor $\chi^{(2)}$. If incoming electric field has only two frequency components ω_1 and ω_2 then \mathbf{E}

Table 1 Properties of widely used nonlinear crystals for THz generation and detection with corresponding values from Refs. [12–14]

Material	Structure	Spectral range in thick crystals(THz)	Electro-optic coefficients (pm/V)	IR group Index n_g @ 800 nm	Refractive index n @ 1THz
ZnTe	Zincblende ($\bar{4}3m$)	0.1–3	$d_{14} = 4$ ($\lambda = 633$ nm)	3.24	3.17
GaP	Zincblende ($\bar{4}3m$)	0.1–7	$d_{14} = 1$ ($\lambda = 633$ nm)	3.55	3.32
GaSe	Hexagonal ($\bar{6}2m$)	14–43	$d_{22} = 54$ ($\lambda = 10.6$ μm)	3.13 (0) 2.86 (e)	3.26

can be expressed by $\mathbf{E} = E_1 e^{-i\omega_1 t} + E_2 e^{-i\omega_2 t} + c.c.$ where c.c. stands for complex conjugate. The second-order polarization is given by

$$\begin{aligned} \mathbf{P} = \varepsilon_0 \chi^{(2)} \mathbf{E} \mathbf{E} = \varepsilon_0 \chi^{(2)} [2(|E_1| + |E_2|)e^0 + (|E_1|^2 e^{-i2\omega_1 t} + |E_2|^2 e^{-i2\omega_2 t}) \\ + 2E_1 E_2 e^{-i(\omega_1 + \omega_2)t} + 2E_1 E_2^* e^{-i(\omega_2 - \omega_1)t}] \end{aligned} \quad (5)$$

The polarization has frequency components of 0, $2\omega_1$, $2\omega_2$, $\omega_1 + \omega_2$, $\omega_2 - \omega_1$, which are known as optical rectification, second harmonic generation, sum frequency and difference frequency generation, respectively. As generated far field, THz pulse is related to the second derivative of the polarization with respect to the time t , $E_{THz} \propto \frac{\partial^2}{\partial t^2} \mathbf{P}$, and the difference frequency generation is the main part of the THz pulse generation. The spectrum of generated THz pulse corresponds to the difference frequency $\omega_1 - \omega_2$ of incident pulses. Thus, the generated THz bandwidth is determined by the pulse width of the incident near-IR femtosecond laser pulse. Thus, if the second-order susceptibility tensor $\chi^{(2)}$ and the temporal width of the incident pulse are known, the field strength and bandwidth of the generated THz pulse can be estimated.

In practical operations, crystal orientation, thickness, dispersion, absorption, diffraction, phase matching should be considered. These properties of widely used nonlinear crystals are summarized in Table 1. Several hundreds of μm thick ZnTe is used for THz generation of 0.1–3 THz via optical rectification. TO phonon mode of ZnTe at 5.3 THz restricts the generated THz bandwidth and dispersion. Thus ZnTe generation has limit of 3 THz [9, 10]. Aside from the bandwidth limitation, there are two important factors that determine THz pulse shape generated in nonlinear crystals.

First, the crystal orientation of nonlinear susceptibility tensor associated with the incident pulse strongly affects the generated THz pulse. Considering the crystal orientation, we can rewrite the second order nonlinear polarization as following,

$$P_x^{(2)} = \sum_{y,z} \varepsilon_0 \chi_{xyz}^{(2)} E_y(\omega) E_z^*(\omega) \quad (6)$$

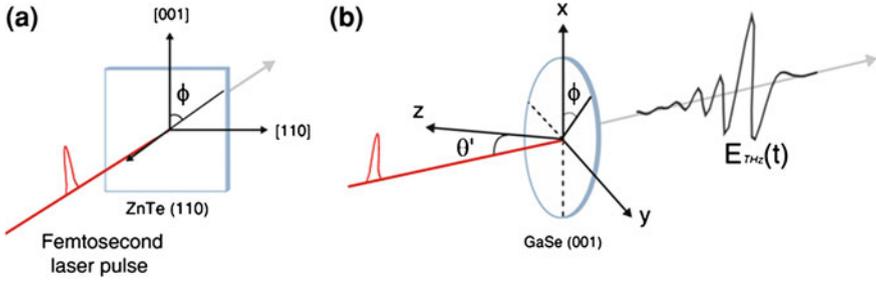


Fig. 4 Schematic diagrams of phase-matched THz generation using $\langle 110 \rangle$ cut ZnTe crystal **a** and $\langle 001 \rangle$ cut GaSe crystal **b**

The notation of x, y, z are indicated in a Cartesian-coordinate direction of the incident pulse E . The 3-dimensional second order nonlinear susceptibility tensor can be written in compact form after considering the directional symmetry

$$d_{xl} = \frac{1}{2} \chi_{xyz}^{(2)} \tag{7}$$

Then, each nonlinear tensor element is expressed as

$$d_{11} = \frac{1}{2} \chi_{111}^{(2)}, d_{12} = \frac{1}{2} \chi_{122}^{(2)}, d_{13} = \frac{1}{2} \chi_{133}^{(2)}, d_{14} = \frac{1}{2} \chi_{123}^{(2)}, d_{15} = \frac{1}{2} \chi_{131}^{(2)}, d_{16} = \frac{1}{2} \chi_{112}^{(2)}, d_{21} = \frac{1}{2} \chi_{131}^{(2)}, d_{22} = \frac{1}{2} \chi_{131}^{(2)}, d_{23} = \frac{1}{2} \chi_{131}^{(2)}, d_{24} = \frac{1}{2} \chi_{131}^{(2)}, d_{25} = \frac{1}{2} \chi_{112}^{(2)}, d_{26} = \frac{1}{2} \chi_{121}^{(2)} \tag{8}$$

The induced second order nonlinear polarizations by the susceptibility tensor above become,

$$\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} = 2\epsilon_0 \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{pmatrix} \begin{pmatrix} E_x^2 \\ E_y^2 \\ E_z^2 \\ 2E_y E_z \\ 2E_z E_x \\ 2E_x E_y \end{pmatrix} \tag{9}$$

Because most of the crystal structure are highly symmetric, it contains parts of 18 susceptibility tensors. In the case of ZnTe, it has symmetry lattice of zinc blende structure with non-zero d -matrix components of $d_{14} = d_{25} = d_{36}$. The directional interaction between optical pulse and crystal structure is also a crucial factor. For normal incident of the optical field, only $\langle 110 \rangle$ cut ZnTe orientation can generate the THz field. THz field amplitude also depends on an angle ϕ between the optical field and the crystal axes. Figure 4a describes an incident field of optical pulse on ZnTe. According to a study reported by Chen et al. the maximum THz electric field can be generated with the angle of $\phi = 54.7^\circ$ [11].

The other important property is the phase matching between optical pulse and generated THz in the nonlinear crystal. If an optical pulse coherently generates the THz field in the crystal, the amplitude of the THz field can be enhanced by propagating through the crystal. This phase-matching condition has the relation of $\Delta k = (k_2 - k_1) - k_{THz} \approx 0$, where k_1 and k_2 are the wave vectors of two optical pulses and k_{THz} is the wave vector of generated THz pulse. With relation $\frac{2\pi}{\lambda} = \frac{\omega_n(\omega)}{c_0}$ the phase matching condition is given by the following relation. [12].

$$n(\omega_2)\omega_2 - n(\omega_1)\omega_1 = n(\omega_{THz})\omega_{THz} \quad (10)$$

When the interaction length with the coherence length L_C reaches π , i.e. $\Delta k L_C = \pi$, then, the generated THz is called phase-mismatched. If the propagation distance is longer than the coherence length, then the radiation efficiency is significantly reduced showing sinc-function dependence. Thus the thickness of the nonlinear crystal should be shorter than the coherent length in order to achieve a maximum THz field strength. In addition, since the THz frequency is much lower than that of the optical laser pulse, the group velocity $v_g(\omega) = \frac{\partial \omega}{\partial k}$ of femtosecond laser pulse and the phase velocity $v_p(\omega) = \frac{\omega}{k}$ of THz pulse should be considered in the phase-matching condition.

GaSe used for generation of higher THz frequency range, up to the mid-IR. GaSe crystal is fragile and only has cleaved form along $\langle 001 \rangle$ plane due to its stacking structure. However, it has outstanding nonlinear coefficient $d_{22} = 54$ pm/V and relatively high damage threshold. Another advantage of the GaSe is its small internal phase-matching angles of around 10° .

When the orientation of femtosecond laser pulse and the $\langle 001 \rangle$ GaSe crystal are set as shown in Fig.4b, two types of phase matching conditions are given

$$\text{type 1 : } n_o(\omega_{THz})\omega_{THz} = n_{eo}(\omega_2, \theta)\omega_2 + n_o(\omega_1)\omega_1 \quad (11)$$

$$\text{type 2 : } n_{eo}(\omega_{THz}, \theta)\omega_{THz} = n_{eo}(\omega_2, \theta)\omega_2 + n_o(\omega_1)\omega_1 \quad (12)$$

where n_o and n_{eo} denote the ordinary (o) and extraordinary (eo) refractive indices. Θ is the angle between direction of incident pulse inside the GaSe and the optical z axis which has the relation of $\theta = \arcsin\left[\frac{\sin(\theta')}{n}\right]$ [15, 16]. To adjust the polarization of incidence femtosecond laser pulse, a half-wave plate can be inserted before the GaSe crystal. When propagating the GaSe, an input pulse simultaneously generates both ordinary and extraordinary polarized waves [15]. Adjusting the incidence angle θ' , the generated THz spectral range can be easily changed by the angled-tuned phase matching. Thus, GaSe crystal enables us to generate quite broad THz frequency, typically from 15 to 50 THz [17].

For GaSe crystal which has thickness over 200 μm , the interaction length is affected not only by the phase difference of optical pulse and generated THz, but also by the phase difference between the ordinary and extraordinary polarized optical pulses, since these two polarizations have different group velocities. A typical number of the group velocity mismatch is around 100–200 fs/mm and about 1 ps/mm for the optical pulse and generated THz pulse, respectively.

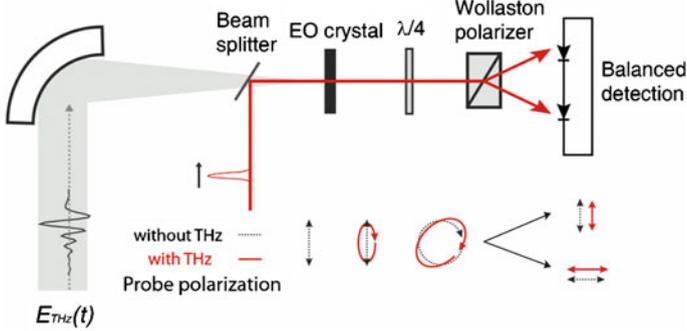


Fig. 5 The schematic diagram of electro-optic sampling and polarization of probe beam with and without THz

Ultrabroadband THz pulses have been generated up to 30 THz by non-phase-matched optical rectification technique with 10 femtosecond laser pulse on nonlinear crystal. But, the very short (≈ 10 femtosecond) laser sources do not provide enough power for spectroscopic applications. Instead of the non-phase-matching method, thin GaSe crystal could be used. Thinner GaSe crystal with thickness of 30–150 μm reduces the power, which also relaxes the phase-matching condition and increases transmittance below 15 THz [18, 19]. Several examples of a broadband THz generation with thin GaSe crystal is shown in Fig. 5c, f representing the THz spectrum of 10–25 THz.

Depending on the nonlinear crystal, optical rectification technique generates and detects different THz spectrum. Several nonlinear crystals such as ZnTe, GaP and GaSe are combined to cover the full range from below 1 to 50 THz. Using ZnTe and GaSe, the frequency region are covered within 0.1–3 THz and 10–50 THz, respectively. GaP can partially cover this intermediate gap but its field strength is much lower than ZnTe because of its small d coefficient.

Recently, even more broadband spectra were generated by using phase-matched difference frequency mixing with 7 femtosecond laser pulses in LiIO₃, resulted in ultrabroadband 50–130 THz pulses [20]. Poled polymers and organic crystals such as 4-dimethylamino-N-methylstilbazolium tosylate (DAST) have large nonlinear coefficients. These crystals have been used for generation of ultrabroadband spectrum [21–25] of up to several tens of THz.

4 Electro-Optic Sampling

The electro-optic (EO) sampling technique allows us to accomplish broadband THz detection. It measures the actual THz electric field in time domain, and its detection principle is basically the same as the photoconductive antenna detection mentioned earlier. The EO sampling uses so called Pockels effect of the EO

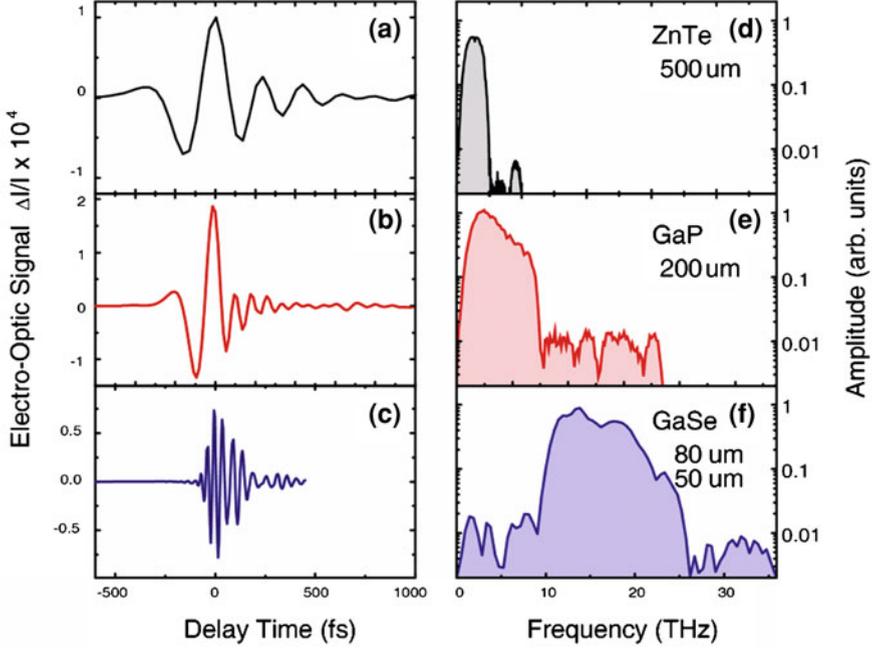


Fig. 6 **a–c** Temporal waveform measured via electro-optic sampling of the THz radiation generated in different crystals with a 250 kHz amplifier system. **d–f** corresponding THz spectra. **a, d** 500 μm thick ZnTe emitter and detector, **b, e** 200 μm thick GaP emitter and detector, **c, f** 80 μm thick GaSe emitter and 50 μm thick GaSe detector

crystals which is related to the optical rectification. When static electric field is applied, it induces birefringence in the EO crystal. In this case, THz pulse acts as static electric gating and the femtosecond laser pulse is used to detect their temporal changes of the induced birefringence. As shown in Fig. 5, a linearly polarized femtosecond laser pulse and a THz pulse are incident on the EO crystal together. As they are propagating through the crystal, the polarization of the laser pulse is changed into slightly elliptical polarization. The following $\lambda/4$ -plate induces quarter-wave phase shift of the incident laser pulse. Without the THz pulse, the polarization state of the linearly polarized pulse is changed into a circular polarization. On the other hand, with the THz pulse, the optical pulse is more elliptically polarized pulse. Then, a Wollaston prism divides the circularly polarized beams into two orthogonal linearly polarized beams. These two pulses are incident on a balanced detector, which the intensity differences of the two pulses are measured [26]. The measured signal can be expressed as

$$\frac{\Delta I}{I} = \frac{\omega_{THz} n^3}{c} dr E_{THz} \quad (13)$$

where n is the near-IR refractive index of the EO crystal, d is the crystal thickness, and r represents the EO coefficient (see the Table 1). The balanced photodiode circuit can be easily fabricated with two photodiodes as shown in Fig. 5. ΔI can be measured by a current preamplifier and followed by the lock-in detection technique. Using this method, 250 kHz amplifier-based system has a signal-to noise (S/N) ratio over 1:100000, which is an order-of-magnitude higher than the oscillator-based system.

The Pockels effect can be understood as sum-frequency generation process of the second-order nonlinear effect where the optical pulse and THz pulse are the applied electric field. This is exactly the reverse process of the optical rectification. By the same manner, EO sampling is limited by the interaction length and the specific phase-matching condition. The thickness of crystal is a trade-off factor which determine the detection bandwidth and the sensitivity. Nevertheless, considering the dispersion effect of THz and optical pulse, thinner crystals are preferred (Fig. 6).

Optical rectification and EO sampling enable us to cover a spectral range up to 100 THz with combinations of ZnTe, GaP, and GaSe crystals. Thin ZnTe which has the thickness of 30 μm enables to detect up to 37 THz. Ultrathin ZnTe and GaP crystals demonstrated that their detectable signals can be reached 70 THz [27]. In order to prevent the multiple reflections within the crystal, typically appeared in the measurable time window, the EO crystals can be attached onto a thick EO-inactive material [14, 27, 28].

References

1. Dressel, M., Gruner, G.: *Electrodynamics of Solids*. Cambridge University Press, Cambridge (2002)
2. Basov, D.N., Timusk, T.: *Electrodynamics of high- T_C superconductors*. *Rev. Mod. Phys.* **77**, 721–779 (2005)
3. Nuss, M.C., Orenstein, J.: *Millimeter and Submillimeter Wave Spectroscopy of Solids*. Springer, Berlin (1998)
4. Auston, D.H.: Picosecond optoelectronic switching and gating in silicon. *Appl. Phys. Lett.* **26**, 101–103 (1975)
5. Smith, P.R., Auston, D.H., Nuss, M.C.: Subpicosecond photoconducting dipole antennas. *IEEE J. Quantum Electron.* **24**, 255–260 (1988)
6. Abdullaev, G.B., Kulevskii, L.A., Prokhorov, A.M., Savel'ev, A.D., Salaev, E.Y., Smirnov, V.V.: GaSe, a new effective material for nonlinear optics. *JETP Lett.* **16**, 90–92 (1972)
7. Ashida, M.: Ultra-broadband terahertz wave detection using photoconductive antennas. *Jpn. J. Appl. Phys.* **47**, 8221–8225 (2008)
8. Boyd, R.W.: *Nonlinear Optics*. Academic Press, San Diego (2003)
9. Gallot, G., Zhang, J., McGowan, R.W., Jeon, T.-I., Grischkowsky, D.: Measurements of the THz absorption and dispersion of ZnTe and their relevance to the electro-optic detection of THz radiation. *Appl. Phys. Lett.* **74**, 3450–3452 (1999)
10. Brazis, R., Nausewicz, D.: Far-infrared photon absorption by phonons in ZnTe crystals. *Opt. Mater.* **30**, 789–791 (2008)
11. Chen, Q., Tani, M., Jiang, Z., Zhang, X.C.: Electro-optic transceivers for terahertz-wave applications. *J. Opt. Soc. Am. B* **18**, 823–831 (2001)

12. Nahata, A., Weling, A.S., Heinz, T.F.: A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling. *Appl. Phys. Lett.* **69**, 2321–2323 (1996)
13. Dmitriev, V.G., Gurzadyan, G.G., Nikogosyan, D.N.: Handbook of nonlinear optical crystals, Springer series in optical sciences, vol. 64. Springer, Heidelberg (1997)
14. Wuand, Q., Zhang, X.C.: Terahertz broadband GaP electro-optic sensor. *Appl. Phys. Lett.* **70**, 1784–1786 (1997)
15. Kaindl, R.A., Smith, D.C., Joschko, M., Hasselbeck, M.P., Woerner, M., Elsaesser, T.: Femtosecond infrared pulses tunable from 9 to 18 μm at an 88-MHz repetition rate. *Opt. Lett.* **23**, 861–863 (1998)
16. Vodopyanov, K.L., Kulevskii, L.A.: New dispersion relationships for GaSe in the 0.65–18 μm spectral region. *Opt. Commun.* **118**, 375–378 (1995)
17. Kaindl, R.A., Eickemeyer, F., Woerner, M., Elsaesser, T.: Broadband phase-matched difference frequency mixing of femtosecond pulses in GaSe: experiment and theory. *Appl. Phys. Lett.* **75**, 1060–1062 (1999)
18. Huber, R., et al.: Generation and field-resolved detection of femtosecond electromagnetic pulses tunable up to 41 THz. *Appl. Phys. Lett.* **76**, 3191–3193 (2000)
19. Kubler, C., Huber, R., Leitenstorfer, A.: Ultrabroadband terahertz pulses: generation and field-resolved detection. *Semicond. Sci. Technol.* **20**, S128–S133 (2005)
20. Zentgraf, T., et al.: Ultrabroadband 50–130 THz pulses generated via phase-matched difference frequency mixing in LiIO_3 . *Optics Express* **15**, 5775–5781 (2007)
21. Auston, D.H., Nuss, M.C.: Electrooptic generation and detection of femtosecond electrical transients. *IEEE J. Quant. Electron.* **24**, 18–196 (1988)
22. Yeh, K.L., Hoffmann, M.C., Hebling, J., Nelson, K.A.: Generation of 10 μJ ultrashort terahertz pulses by optical rectification. *Appl. Phys. Lett.* **90**, 171121–171123 (2007)
23. Zhang, X.C., Jin, Y., Yang, K., Schowalter, L.J.: Resonant nonlinear susceptibility near the GaAs band gap. *Phys. Rev. Lett.* **69**, 2303–2306 (1992)
24. Lee, Y.-S., Meade, T., Perlin, V., Winful, H., Norris, T.B., Galvanauskas, A.: Generation of narrow-band terahertz radiation via optical rectification of femtosecond pulses in periodically poled lithium niobate. *Appl. Phys. Lett.* **76**, 2505–2507 (2000)
25. Nahata, A., Auston, D.H., Wu, C., Yardley, J.T.: Generation of terahertz radiation from a poled polymer. *Appl. Phys. Lett.* **67**, 1358–1360 (1995)
26. Planken, P.C.M.: Measurement and calculation of the orientation dependence of terahertz pulse detection in ZnTe. *J. Opt. Soc. Am. B* **18**, 313–317 (2001)
27. Leitenstorfer, A., Hunsche, S., Shah, J., Nuss, M.C., Knox, W.H.: Detectors and sources for ultrabroadband electro optic sampling: experiment and theory. *Appl. Phys. Lett.* **74**, 1516–1518 (1999)
28. Liu, K., Xu, J., Zhang, X.C.: GaSe crystals for broadband terahertz wave detection. *Appl. Phys. Lett.* **85**, 863–865 (2004)

Part II
Terahertz-Bio Interaction

Terahertz Pulse Near-Field Microscopes

Kiwon Moon, Meehyun Lim, Youngwoong Do
and Haewook Han

Abstract We review the recent progress in the terahertz (THz) apertureless near-field microscopes. We demonstrate quantitative analysis and measurements of the THz near-fields interactions in the probe-sample system. We also present a self-consistent line dipole image method for the quantitative analysis of the near-field interaction. The measurements of approach curves and relative contrasts on gold and silicon substrates were in excellent agreement with calculations based on the self-consistent line dipole image method.

1 Introduction

Nanoscale near-field imaging in the terahertz (THz) spectral range provides a powerful means for studying intriguing phenomena such as intermolecular vibrational spectroscopy and dynamic charge transport in a variety of quantum-confined nanostructures. Conventional THz time-domain spectroscopy (TDS) can provide macroscopic imaging averaged over an ensemble of such nanostructures, which inevitably suffers from inhomogeneous spectral broadening. Moreover, its spatial resolution is limited to $\sim \lambda/2$ by diffraction. Therefore, several types of THz pulse scanning near-field optical microscopes (SNOMs) have been developed to achieve sub-wavelength resolutions [1–9].

In contrast to visible or IR SNOMs based on frequency-domain spectroscopy systems, most THz SNOMs have been based on THz pulse TDS systems [1–8],

K. Moon · M. Lim · Y. Do · H. Han (✉)
Department of Electrical and Computer Engineering,
National Laboratory for Nano-THz Photonics, POSTECH,
Pohang, Gyeongbuk 790-784, Korea
e-mail: hhan@postech.ac.kr

making it possible to perform broadband coherent THz spectroscopy. Among the THz SNOM systems, the scattering-type SNOM (s-SNOM) has been the most successful technique so far in terms of spatial resolution and image quality [6–13]. In the THz s-SNOM, the scattered field from the tip apex is measured in the far-field region. Sub-micrometer resolutions are enabled by the strongly localized near-field around the probe tip [14]. Thus, it is essential to understand the near-field interaction in the tip-substrate system, and there have been several analytic models [15–18] and also numerical simulations [19, 20] to solve the problems. The most popular approach has been the point dipole image method (PDIM) [8–12] where the probe tip is replaced by a polarizable point dipole [10]. The PDIM has been widely used to analyze experimental data [8–12], and has provided qualitative understanding on s-SNOMs, including resolution [10] and optical phase contrast [11]. However, because the electromagnetic boundary conditions are not fully matched on the surface of the probe sphere, the PDIM becomes incorrect as the probe sphere approaches the substrate [15–18].

In this part, we analyze the THz near-field scattering signals of a THz pulse s-SNOM by using a self-consistent line dipole image method (LDIM), closely following our recent publications [21, 22]. We also present a line scan across the edge of a gold film on a GaAs substrate, showing that the THz s-SNOM has a nanoscale resolution of ~ 80 nm. The measurements of the THz scattering signals on gold and GaAs surfaces were in reasonably good agreement with the LDIM calculations.

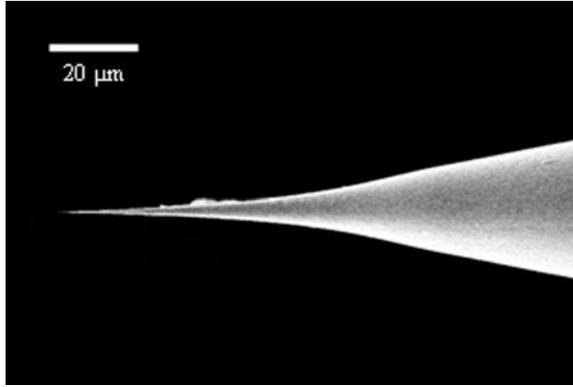
2 Instrumentations

The THz s-SNOM system was based on a THz TDS system combined with an atomic force microscope (AFM). The THz TDS system generates the THz pulses incident on the probe tip and detects the THz scattering signals. The sample scanning and the distance between the s-SNOM and the sample surface were precisely controlled by the AFM. The first nanoscale THz near-field imaging was based on a scanning tunneling microscope (STM) where the tunneling current was used to control the probe-substrate gap distance [5]. Although the STM can provide the highest spatial resolution among the scanning probe microscopes (SPMs), the principal disadvantage is that the tunneling current always requires a conducting substrate. Since many materials used in THz s-SNOMs are not conducting, we used the AFM which can operate on both conducting and insulating substrates.

2.1 THz TDS System

Sub-picosecond THz pulses were generated from an InAs substrate pumped by a femtosecond laser, and measured by a photoconductive antenna on a low-temperature-grown (LTG) GaAs substrate using optical gate detection techniques

Fig. 1 Scanning electron micrograph of a typical probe tip fabricated by electro-chemical etching method. The tip size can be controlled by the etching condition



[23–25]. This optical gating detection makes it possible to directly measure the transient electric fields of the THz pulses, providing both amplitude and phase spectra, whereas the conventional THz frequency-domain spectroscopy can usually measure only the amplitude spectrum.

2.2 AFM System and Background Suppression

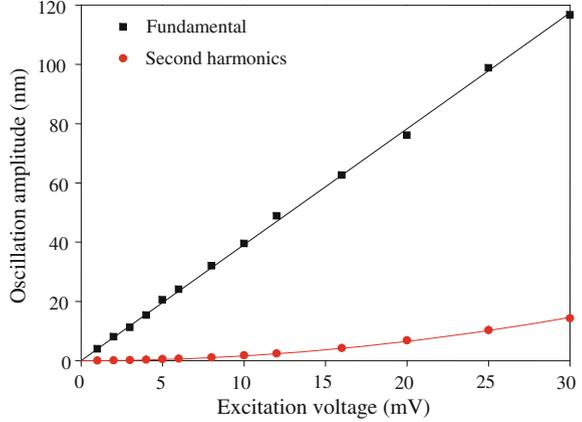
The AFM tip was fabricated by an electro-chemical etching method [28] and glued to a quartz tuning fork. Figure 1 shows the scanning electron micrograph of a tungsten probe tip. The output current from the tuning fork is amplified and demodulated by a preamplifier and a lock-in amplifier, respectively. The extracted output current of the tuning fork is proportional to the oscillation amplitude of the tuning fork, and there is a relationship between the tip-to-sample distance and the oscillation amplitude. Therefore, the measured output current is used to control the tip-to-sample distance by a PID controller which controls the sample height. By scanning the sample-mounting nanostage, the topography of the sample can be measured by recording the sample height.

In the THz *s*-SNOM system, the AFM tip oscillates at the mechanical resonance frequency Ω of a quartz tuning fork. The tip-sample gap distance is then given by

$$g(t) = g_0 + g_1(V) \cos \Omega t \quad (1)$$

where t denotes the time coordinate for the tip oscillation, and the oscillation amplitude g_1 is a function of the excitation voltage V applied to the tuning fork. When scanning the time delay of the THz TDS system, the scattered field \mathbf{E}_{sc} was measured in the far-field region. Since the near-field is a highly nonlinear function of $g(t)$, the scattered field can be expressed as

Fig. 2 Measured oscillation amplitude of the quartz tuning fork



$$\mathbf{E}_{sc}(t) = \mathbf{E}_0 + \sum_{m=1}^{\infty} \mathbf{E}_m \cos(m\Omega t). \quad (2)$$

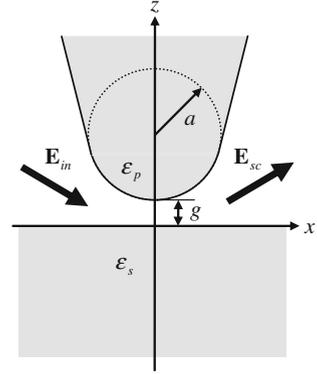
The large background signal \mathbf{E}_0 can be suppressed by a lock-in demodulation technique to extract harmonic component \mathbf{E}_m [8–13]. For the analysis and measurement of the near-field interaction, a precise regulation of the oscillation amplitude is essential. Figure 2 shows that the measured oscillation amplitude has a linear dependence on the applied voltage with an electromechanical responsivity of $\zeta_1 = g_1/V = 3.913 \text{ nm/mV}$.

3 Theory

The schematic of a THz s-SNOM is shown in Fig. 3 where the incident THz pulse (\mathbf{E}_{in}) induces dipole moments in both the probe and the substrate, and the probe is strongly coupled with the substrate, generating the scattered THz pulse (\mathbf{E}_{sc}). In general, the analytic theory for the near-field interaction in the probe-substrate system is very involved because the electromagnetic boundary conditions should be matched at the surfaces of the probe and the substrate. Moreover, numerical simulations are practically impossible since the metallic probes requires large mesh sizes and the numerical accuracies are usually not good enough for quantitative comparison with experiments.

For an analytic theory of s-SNOMs, the probe tip has usually been replaced by a polarizable sphere of radius a and complex relative permittivity ϵ_p , as shown in Fig. 3. Because the probe sphere is much smaller than the wavelength of the incident field \mathbf{E}_{in} , especially at THz frequencies, quasi-static image methods are applicable to the analysis of the near-field interactions. In the quasi-static image methods, the near-field interaction and the subsequent scattering field \mathbf{E}_{sc} can be

Fig. 3 Schematic diagram of a probe-substrate system in THz s-SNOM. The incident field \mathbf{E}_{in} is p-polarized, and the scattered field \mathbf{E}_{sc} is measured in the far-field region. In the image analysis, the probe is replaced by a polarizable dielectric sphere. Reprinted from [21]



calculated by the dipoles induced in the probe and substrate. In the far-field region, the scattering field is given by

$$\mathbf{E}_{sc}(t) = \frac{k^2[\mathbf{r} \times \mathbf{d}(t)] \times \mathbf{r}}{4\pi\epsilon_0 r^3} \quad (3)$$

where \mathbf{d} is the total induced dipole in the tip-substrate system, \mathbf{r} is the far-field position vector ($r = |\mathbf{r}|$), k and ϵ_0 are the wave number and permittivity in vacuum, respectively. In this section we review the point dipole image model and the line dipole image model [21, 22, 26].

3.1 Point Dipole Image Method

In the conventional PDIM [8–12], the incident field polarizes the probe sphere, resulting in a point image dipole $\mathbf{d}_p^{PD} = \alpha\mathbf{E}_{in}$ at the center of the sphere. The sphere polarizability is given by $\alpha = 4\pi\epsilon_0 a^3(\epsilon_p - 1)/(\epsilon_p + 2)$ where ϵ_p is the probe permittivity. The probe dipole induces a substrate image dipole, $\mathbf{d}_s^{PD} = \beta_T \cdot \mathbf{d}_p^{PD}$ at $z = -h = -(a + g)$. The probe-substrate image generation dyadic is given by $\beta_T = \beta(\mathbf{I} - \mathbf{u}_x\mathbf{u}_x)$ where $\beta = (\epsilon_s - 1)/(\epsilon_s + 1)$, ϵ_s is the substrate permittivity, \mathbf{u}_x (\mathbf{u}_z) are the unit vectors along the x (z) axes, and $\mathbf{I} = \mathbf{u}_x\mathbf{u}_x + \mathbf{u}_z\mathbf{u}_z$. By taking into account the recursive generation of the point image dipoles [10, 11], the total image dipole in the probe-substrate system is given by $\mathbf{d}^{PD} = \alpha^{PD} \cdot \mathbf{E}_{in}$ with the effective polarizability tensor,

$$\alpha^{PD} = \frac{\alpha(1 - \beta)}{1 - \frac{\alpha\beta}{32\pi h^3}} \mathbf{u}_x\mathbf{u}_x + \frac{\alpha(1 + \beta)}{1 - \frac{\alpha\beta}{16\pi h^3}} \mathbf{u}_z\mathbf{u}_z. \quad (4)$$

In the PDIM, the continuous dipole distributions are usually replaced by point image dipoles. This point dipole approximation, however, works well only for a

large tip-substrate distance. In principle, the electric field from the substrate dipole induces not only a single point dipole but also a continuous dipole distribution in the probe sphere, as we shall see in the LDIM.

3.2 Line Dipole Image Method

In contrast to the PDIM, the LDIM applies exact electrostatic boundary conditions to all dielectric surfaces of the sphere-substrate system through iteration processes, which results in continuous line image dipole densities. This means that the LDIM is self-consistent within the framework of the polarizable sphere model. Here we describe the LDIM and then present quantitative analyses for the THz near-field interaction in the probe-substrate system [21, 22, 26].

In the first iteration, the initial dipole moment at the probe sphere is induced by the incident field, resulting in the probe dipole given by

$$\mathbf{d}_{p(1)}^{LD} = \alpha(1+r)\mathbf{E}_{in} \quad (5)$$

where we include the reflection coefficient r at the substrate surface because the actual excitation field is the sum of the incident and reflected fields. It should be noted that we have to consider an image dipole of the radiation source for the incident field to match the exact boundary conditions at the substrate surface. The probe dipole in turn induces the substrate image dipole at $z = -h$ given as

$$\mathbf{d}_{s(1)}^{LD} = \boldsymbol{\beta}_T \cdot \mathbf{d}_{p(1)}^{LD} \quad (6)$$

In the second iteration, the substrate point dipole $\mathbf{d}_{s(1)}^{LD}$ induces a line dipole density $\mathbf{p}_{s(2)}^{LD}(z, h) = \gamma(z, h, 2h) \cdot \mathbf{d}_{s(1)}^{LD}$. From the exact image theory [21, 22, 26], we can derive the substrate-probe image generation tensor defined as

$$\begin{aligned} \gamma(z, h, L) = & \frac{\epsilon_p - 1}{\epsilon_p + 1} \left(\frac{a}{L}\right)^3 \delta(z - h + z_K) (\mathbf{I} - 2\mathbf{u}_x \mathbf{u}_x) \\ & + \frac{\epsilon_p - 1}{(\epsilon_p + 1)^2} \frac{a}{L^2} \left(\frac{h-z}{z_K}\right) \frac{1}{\epsilon_p + 1} \left[U(z - h + z_K) - U(z - h) \right] \left[\mathbf{I} + (\epsilon_p - 1)\mathbf{u}_z \mathbf{u}_z \right] \end{aligned} \quad (7)$$

where $z_K = a^2/L$ is the Kelvin distance [26], and $\delta(z)$ and $U(z)$ are the Dirac delta and Heaviside unit step functions, respectively. From the third iteration, we must consider not only the point image dipoles but also the line image dipole densities in both the probe and the substrate. For the n th iteration, the probe ($n \geq 3$) and substrate ($n \geq 2$) dipole densities are given by convolution integrals,

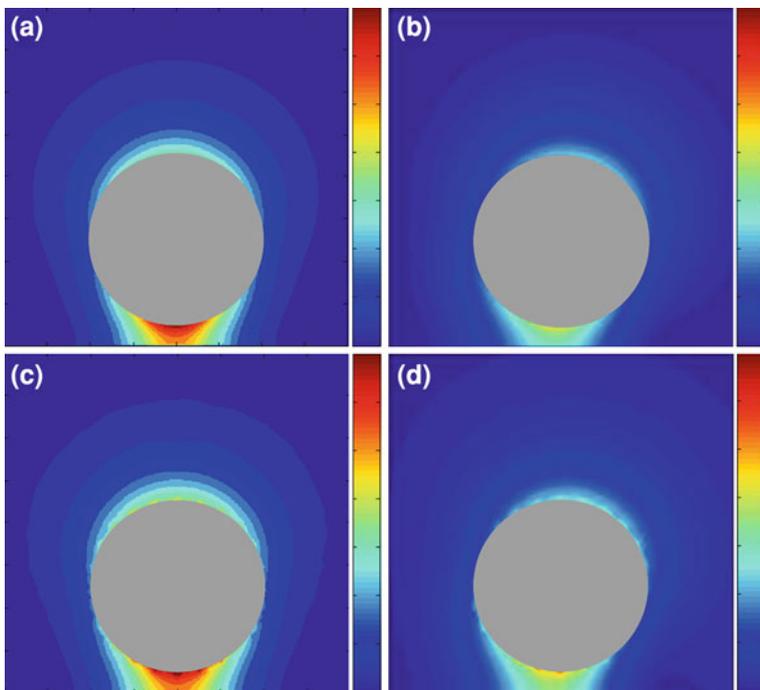


Fig. 4 Near-field distributions (80×80 nm) of the sphere-substrate system. LDIM calculations on (a) Au substrate and (b) FZ-Si substrate, and FEM simulations on (c) Au substrate and (d) FZ-Si substrate. The diameter of the tungsten sphere is 40 nm and the gap distance is 5 nm

$$\begin{aligned} \mathbf{p}_{p(n)}^{LD}(z, h) &= \int_{-h}^{a-h} dz' \gamma(z, h, h - z') \cdot \mathbf{p}_{s(n-1)}^{LD}(z', h) \\ \mathbf{p}_{s(n)}^{LD}(z, h) &= \beta_T \cdot \mathbf{p}_{p(n-1)}^{LD}(-z, h) \end{aligned} \quad (8)$$

and the total dipole moment in the sphere-substrate system is then given by

$$\mathbf{d}^{LD}(h) = \sum_{n=1}^{\infty} \left[\int_{h-a}^h dz \mathbf{p}_{p(n)}^{LD}(z, h) + \int_{-h}^{a-h} dz \mathbf{p}_{s(n)}^{LD}(z, h) \right] \quad (9)$$

where the dipole densities in the first iteration are given by

$$\begin{aligned} \mathbf{p}_{p(1)}^{LD}(z, h) &= \mathbf{d}_{p(1)}^{LD} \delta(z - h) \\ \mathbf{p}_{s(1)}^{LD}(z, h) &= \beta_T \cdot \mathbf{d}_{p(1)}^{LD} \delta(z + h). \end{aligned} \quad (10)$$

Figure 4 depicts the accuracy of the near-field interactions calculated by the LDIM, where the incident angle of p-polarized \mathbf{E}_{in} is 60° and FEM simulations using a commercial software (HFSS) are in excellent agreement with the LDIM calculations on gold and Si substrates.

4 Analyses and Measurements

The samples for the THz line scan were prepared by evaporating a 35 nm thick gold film on the half surface of a semi-insulating GaAs substrate. The etched tip radius of a tungsten wire was $a \sim 20$ nm. The probe tip was mounted to a quartz tuning fork such that the tip oscillates perpendicularly to the sample surface. As the tip approaches the sample surface, the amplitude, phase, and resonant frequency of the tuning fork's oscillation change. In our experiments, the amplitude was used as a feedback signal for tip-to-sample distance regulation, maintaining a constant tip-substrate distance by controlling the piezo actuator. The incident THz pulses were p-polarized and focused upon an AFM tip with an incident angle of 60° . The THz scattering signal was detected by a THz photoconductive antenna in the far-field region, and simultaneously the AFM topography of the sample surface was obtained from the feedback signals for the piezo actuators of the nanostages.

4.1 Line Dipole Image Analysis

The LDIM calculations of the dipole distributions for the gold and GaAs substrates are shown in Fig. 5, where the point dipoles and continuous dipole distribution were discretized into piecewise-constant dipole distributions. The red and blue curves represent calculations for gold and GaAs substrates, respectively. As expected from the standard electrostatic image theory, we can neglect the tangential component (d_x) that is much smaller than the vertical component (d_z). For the LDIM calculation, we used the Drude model for gold and tungsten, where the permittivity is given by

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i\omega_\gamma)}. \quad (11)$$

The plasma (damping) frequencies, ω_p (ω_γ) of gold and tungsten were 2,184 THz (1,551 THz) and 6.45 THz (14.61 THz), respectively [27]. For the GaAs substrate, we used a frequency-independent permittivity $\varepsilon_s = 12.918$ [25].

Figure 6 shows the LDIM calculations of the first harmonic scattering field ratios for gold and GaAs substrates, $E_{1,\text{Au}}/E_{1,\text{GaAs}}$. The scattering ratios increase rapidly for small tip-sample distances, and become saturated for large oscillation amplitudes ($g_1 > 15$ nm). This means that the precise control of the tip-substrate distance and oscillation amplitude is crucial for quantitative analysis and measurements. In general, the LDIM calculations result in larger induced dipole moments and also steeper increase of the dipole moments than the PDIM as the tip approaches the substrate. Consequently, the LDIM produces larger harmonic components than the PDIM [22].

Fig. 5 LDIM calculation of dipole distribution on (a) gold and (b) semi-insulating GaAs substrates. The point dipoles and continuous dipole distribution were discretized into piecewise-constant dipole distributions. The tungsten tip has a radius of $a = 20$ nm. The incident field \mathbf{E}_{in} is p-polarized, and the incident angle is 60° . Reprinted from [21]

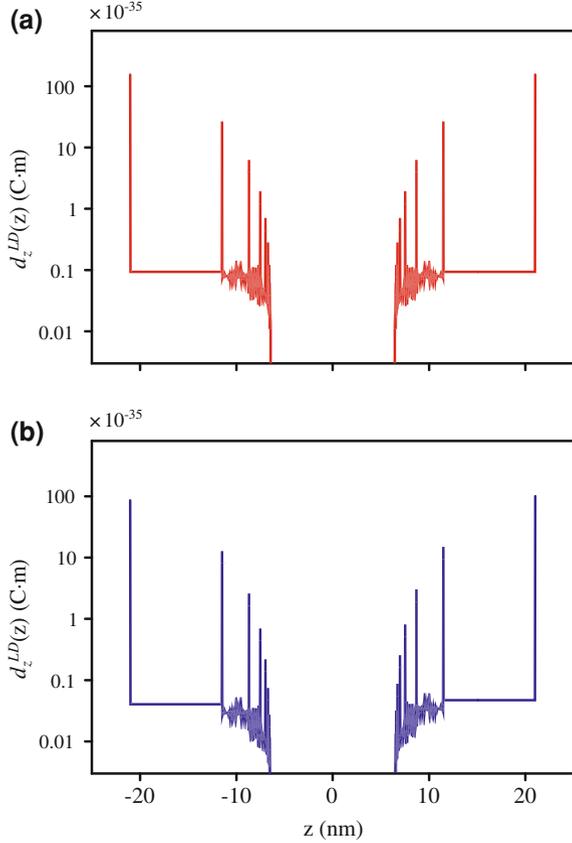


Fig. 6 LDIM calculation of the first harmonic scattering field ratios of gold and GaAs substrates. The tungsten tip has a radius of $a = 20$ nm. The incident field \mathbf{E}_{in} is p-polarized, and the incident angle is 60° . Reprinted from [21]

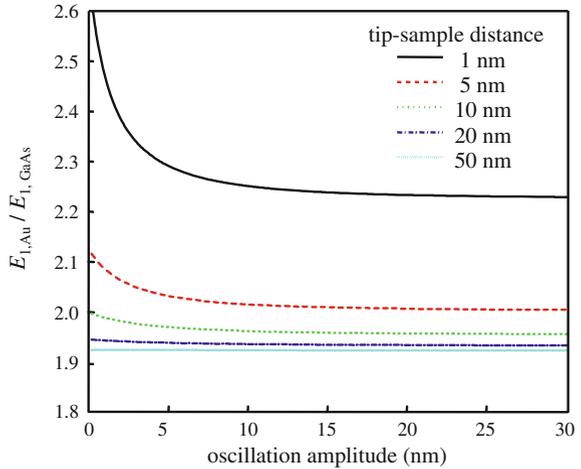
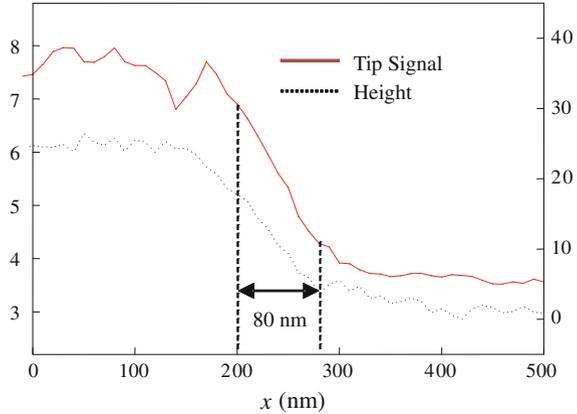


Fig. 7 Line scan across the edge of a gold film on a GaAs substrate. The red and black lines represent the THz scattering signal and the AFM topography, respectively. The 10–90% resolution of the THz s-SNOM was ~ 80 nm



4.2 THz Line Scan

The THz line scans across the edge of the gold film is shown in Fig. 7, where the first harmonic components (E_1) of the THz scattering signals were used. During the measurements, an oscillation amplitude of $g_1 = 45$ nm was maintained by a proportional-integral-derivative (PID) controller, and the mechanical delay line was fixed at the positive peaks of the THz scattering pulses. The minimum tip-to-sample distance was less than 1 nm, which is the typical operation condition for our AFM system.

The measured tip signal from the Au surface was stronger than that from the GaAs surface as expected from the image theory since the vertical dipole on a metallic surface is enhanced by the image dipole. The THz scattering signal is proportional to the total dipole moment induced in the tip-substrate system, and a metal surface induces larger dipoles than a dielectric surface. The spatial resolution of the THz s-SNOM was ~ 80 nm, estimated from the 10 to 90% transition distance in the line scan. The measured scattering field ratio ($E_{1,Au}/E_{1,GaAs}$) was 2.36 ± 0.3 which is in good agreement with the calculated scattering ratio of ~ 2.23 , corresponding to the saturation value of scattering ratio for large oscillation amplitude (45 nm) and small tip-sample distance (≤ 1 nm) used in our measurements.

5 Conclusion

We have used the line dipole image method to analyze the scattering signals in THz pulse s-SNOMs. We have also demonstrated that the THz s-SNOM has a nanoscale resolution of ~ 80 nm. The relative ratios of the scattering signals measured on gold and GaAs substrates were in good agreement with calculations by the line dipole image theory. We believe that this is an important step toward quantitative imaging contrast in THz near-field microscopy.

References

1. Hunsche, S., Koch, M., Brener, I., Nuss, M.C.: THz near-field imaging. *Opt. Comm.* **150**, 22–26 (1998)
2. Chen, Q., Zhang, X.-C.: Semiconductor dynamic aperture for near-field terahertz wave imaging. *IEEE J. Sel. Top. Quantum Electron.* **7**(4), 608–614 (2001)
3. van der Valk, N.C.J., Planken, P.C.M.: Electro-optic detection of subwavelength terahertz spot sizes in the near field of a metal tip. *Appl. Phys. Lett.* **81**(9), 1558–1560 (2002)
4. Yuan, T., Park, H., Xu, J., Han, H., Zhang, X.-C.: Field induced THz wave emission with nanometer resolution. *Proc. SPIE* **5649**, 1–8 (2005)
5. Chen, H.T., Kersting, R., Cho, G.C.: Terahertz imaging with nanometer resolution. *Appl. Phys. Lett.* **83**(15), 3009–3011 (2003)
6. Park, H., Kim, J., Han, H.: THz pulse near-field microscope with nanometer resolution, presented at the 35th workshop: physics and technology of THz photonics 2005, pp. 20–26. Erice, Italy (July 2005)
7. Park, H., Kim, J., Kim, M., Han, H., Park, I.: Terahertz near-field microscope, presented at the Joint 31st International Conference on Infrared Millimeter Waves and 14th International Conference on Terahertz Electronics (IRMMW-THz 2006), pp. 18–22. Shanghai, China (September 2006)
8. von Ribbeck, H.-G., Brehm, M., van der Weide, D.W., Winnerl, S., Drachenko, O., Helm, M., Keilmann, F.: Spectroscopic THz near-field microscope. *Opt. Express* **16**(5), 3430–3438 (2008)
9. Huber, A.J., Keilmann, F., Wittborn, J., Aizpurua, J., Hillenbrand, R.: Terahertz near-field nanoscopy of mobile carriers in single semiconductor nanodevices. *Nano. Lett.* **8**(11), 3766–3770 (2008)
10. Knoll, B., Keilmann, F.: Enhanced dielectric contrast in scattering-type scanning near-field optical microscopy. *Opt. Comm.* **182**, 321–328 (2000)
11. Hillenbrand, R., Keilmann, F.: Complex optical constants on a subwavelength scale. *Phys. Rev. Lett.* **85**(14), 3029–3032 (2000)
12. Taubner, T., Hillenbrand, R., Keilmann, F.: Nanoscale polymer recognition by spectral signature in scattered infrared near-field microscopy. *Appl. Phys. Lett.* **85**(21), 5064–5066 (2004)
13. Labardi, M., Patanè, S., Allegrini, M.: Artifact-free near-field optical imaging by apertureless microscopy. *Appl. Phys. Lett.* **77**(5), 621–623 (2000)
14. Novotny, L., Bian, R.X., Xie, X.S.: Theory of nanometric optical tweezers. *Phys. Rev. Lett.* **79**(4), 645–648 (1997)
15. Aravind, P.K., Metiu, H.: The effects of the interaction between resonances in the electromagnetic response of a sphere-plane structure; applications to surface enhanced spectroscopy. *Surf. Sci.* **124**, 506–528 (1983)
16. Suknov, S.V.: Role of multipole moment of the probe in apertureless near-field optical microscopy. *Ultramicroscopy* **101**, 111–122 (2004)
17. Porto, J.A., Johansson, P., Apell, S.P., López-Ríos, T.: Resonance shift effects in apertureless scanning near-field optical microscopy. *Phys. Rev. B* **67**, 085409 (2003)
18. Cvitkovic, A., Ocelic, N., Hillenbrand, R.: Analytical model for quantitative prediction of material contrast in scattering-type near-field optical microscopy. *Opt. Express* **15**(14), 8550–8565 (2007)
19. Esteban, R., Vogelgesang, R., Kern, K.: Tip-substrate interaction in optical near-field microscopy. *Phys. Rev. B* **75**, 195410 (2007)
20. Brehm, M., Schliesser, A., Čajko, F., Tsukerman, I., Keilmann, F.: Antenna-mediated back-scattering efficiency in infrared near-field microscopy. *Opt. Express* **16**(15), 11203–11215 (2008)
21. Moon, K., Jung, E., Lim, M., Do, Y., Han, H.: Terahertz near-field microscope analysis and measurements of scattering signals. *IEEE Trans. Terahertz Sci. Technol.* **1**, 164–168 (2011)

22. Moon, K., Jung, E., Lim, M., Do, Y., Han, H.: Quantitative analysis and measurements of near-field interactions in terahertz microscopes. *Opt. Express* **19**(12), 11539–11544 (2011)
23. van Exter, M., Fattinger, Ch., Grischkowsky, D.: Terahertz time-domain spectroscopy of water vapor. *Opt. Lett.* **34**(20), 1128–1130 (1989)
24. van Exter, M., Fattinger, Ch., Grischkowsky, D.: High-brightness terahertz beams characterized with an ultrafast detector. *Appl. Phys. Lett.* **55**(4), 337–339 (1989)
25. Grischkowsky, D., Keiding, S., van Exter, M., Fattinger, Ch.: Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors. *J. Opt. Soc. Am. B* **7**(10), 2006–2015 (1990)
26. Lindell, I.V., Sten, J.C.-E., Nikoskinen, K.I.: Electrostatic image method for the interaction of two dielectric spheres. *Radio. Sci.* **28**(3), 319–329 (1993)
27. Ordal, M.A., Bell, R.J., Alexander, R.W.Jr., Long, L.L and Querry, M.R.: Optical properties of fourteen metals in the infrared and far infrared: Al, Co, Cu, Au, Fe, Pb, Mo, Ni, Pd, Pt, Ag, Ti, V, and W. *Appl. Opt.* **24**(24), 4493–4499 (1985)
28. Kulawik, M., Nowicki, M., Thielsch, G., Cramer, L., Rust, H.-P., Freund, H.-J., Pearl, T.P., and Weiss, P.S.: A double lamellae dropoff etching procedure for tungsten tips attached to tuning fork atomic force microscopy/scanning tunneling microscopy sensors. *Rev. Sci. Instrum.* **74**(2), 1027–1030 (2003)

Optical Imaging and Control of Neurons

Yoon-Kyu Song

1 Introduction

Although remarkable progress has been made in our understanding of the function, organization, and development of the brain by various approaches of modern science and technology, how the brain performs its marvelous function remains unsolved or incompletely understood [1–3]. This is mainly attributed to the insufficient capability of currently available research tools and conceptual frameworks to deal with enormous complexity of the brain. Hence, in the last couple of decades, a significant effort has been made to crack the complexity of brain by utilizing research tools from diverse scientific areas [4–6]. The research tools include the optical neurotechnology which incorporates the exquisite characteristics of optics, such as multi-parallel access and non-invasiveness, in sensing and stimulating the excitable membrane of a neuron, the basic functional unit of the brain [7]. This part is aimed to serve as a short introduction to the optical neurotechnology for those who wish to use optical techniques as one of their brain research tools.

The part is organized as follows; in Sect. 2, a neuron is introduced as the basic functional unit of the brain. Ion channels and their functional importance for the neurons are addressed as well. In Sect. 3, optical characteristics of the neuronal membrane due to its electrochemical activity are briefly described. In Sect. 4, some methods for functional optical imaging of the neurons are introduced as new research tools for recording neuronal activity. In Sect. 5, various techniques to manipulate neuronal activity are addressed, including recently developed opto-genetic techniques. Finally, in Sect. 6, the future direction of research is suggested, including potential application of optical techniques in THz domain.

Y.-K. Song (✉)
Department of Nano Science and Technology,
Seoul National University, Seoul, Republic of Korea
e-mail: songyk@snu.ac.kr

2 Neuron: Basic Functional Unit of the Brain

The neuron is the basic working unit of the brain, processing sensory, motor, cognitive information through their network and transmitting them to other neuronal or non-neuronal cells. The enormous complexity of the brain comes from the extremely complex structural organization of neurons in the brain. There are about 100 billion neurons in human brain, and they make rich interconnections (on average 10,000 per neuron) among themselves through a special interface called synapse [8]. Figure 1 shows the basic structure of a neuron with its synaptic connections to other neurons [9].

The neuron consist of a *cell body* with nucleus, *dendrites* receiving signals from other neurons, and an *axon* transmitting electrical signals to other neurons through an interface called “synapse,” as shown in Fig. 1. Neurons transmit signals as electrical impulses along their axons, often wrapped with myelin sheath to speed up the transmission. The electrical impulses in neurons, called *action potentials*, involve the activity of *ion channels*—tiny pores imbedded in the cell membrane, allowing ions and other small molecules to enter or leave the cell. The sequential gating of ion channels creates a pattern of voltage change across the neuronal membrane, which passes along the axon until reaching the axon terminals at the end of the axon. At the axon terminal, the voltage change triggers the release of neurotransmitters, chemical messengers in neurons, which diffuse across the intra-synaptic space and bind to the specific receptors, switching on-and-off the post synaptic cells (for review of the mechanism in detail, see references [10]).

To date, electrodes have been a valuable tool to record or stimulate electrical activity of neurons, mainly due to their excellent temporal resolution on the millisecond time scale in recording and stimulation, as well as unprecedented signal to noise ratio in recording. There are a number of experimental challenges in investigation of neuronal activity with electrodes, particularly when studying the intricate functional network of multiple neurons [11]. Investigation of neuronal activity with electrodes has limited spatial resolution due to invasive nature of the technique, and it is impractical to register multiple specific targets due to mechanical instability. Hence, there is clearly a need for research tools better equipped to investigate the complex network of a large number of interacting neurons, and the solutions have begun to emerge from active adoption of various techniques in the diverse scientific areas.

3 Optical Properties of Neurons in Action

There have been numerous reports to date that optical properties of neurons were found to change during their electro-chemical activities [7]. The optical properties of the cells are closely related to the structural properties based on the molecular orientation of membrane, volumetric dimension, and organization of membrane bound molecules. There is an excellent review article written by L. B. Cohen at

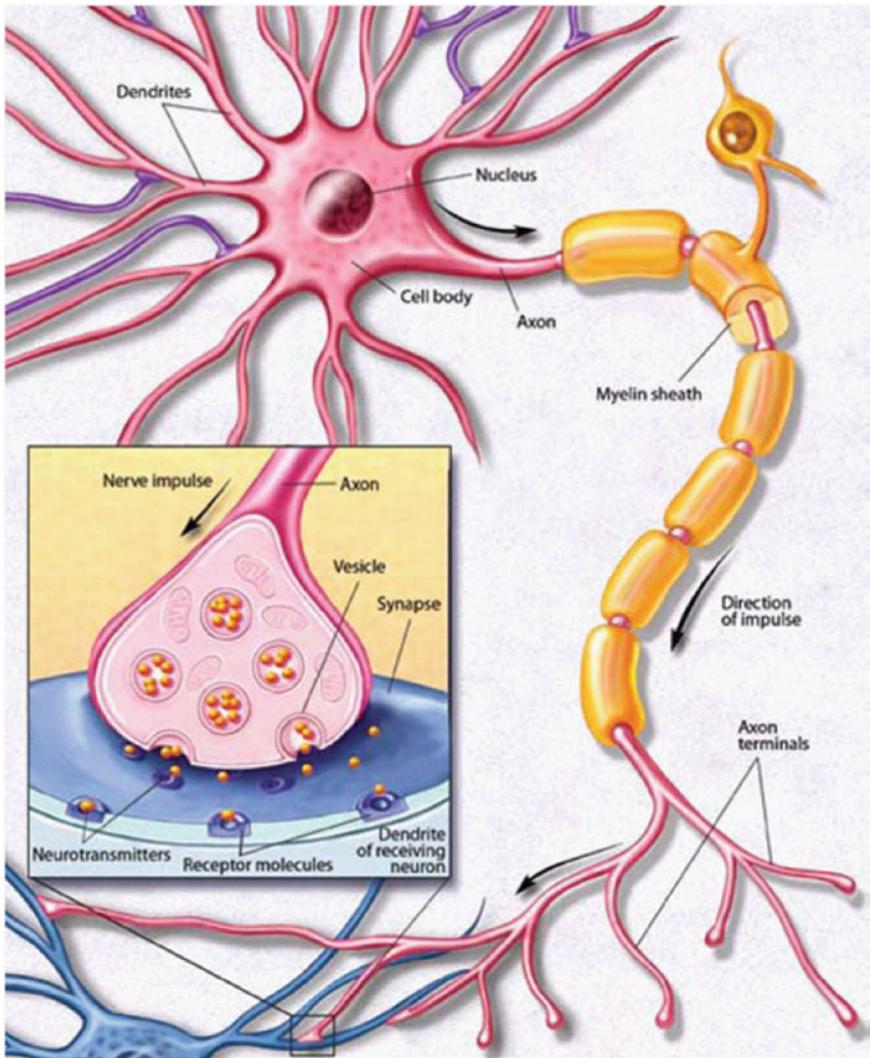


Fig. 1 Basic structure of a neuron with a synapse at the axon terminal (*inset*) [9]

Yale University [12], which comprehensively compiled a significant amount of early experimental works related to the structural changes in neuron (including optical properties) during action potential propagation.

Electro-chemical activity of neuron triggers instantaneous reversal of polarity across the cellular membrane due to changes in permeability of ion channels, before returning to the resting state in the millisecond time scale. Change in polarity causes corresponding change in optical properties, such as birefringence or retardation [13], intrinsic fluorescence [14], light scattering [15], ultraviolet/

Table 1 Physical origins of change in optical properties of neurons in action [18]

Technique	Birefringence	Fluorescence	Scattering	Absorption	Volume changes
Physical origin	Kerr effect electrostriction	NADH proteins with Trp/Tyr	Water of hydration transport number	Not clearly understood	Ion exchange chloride flux

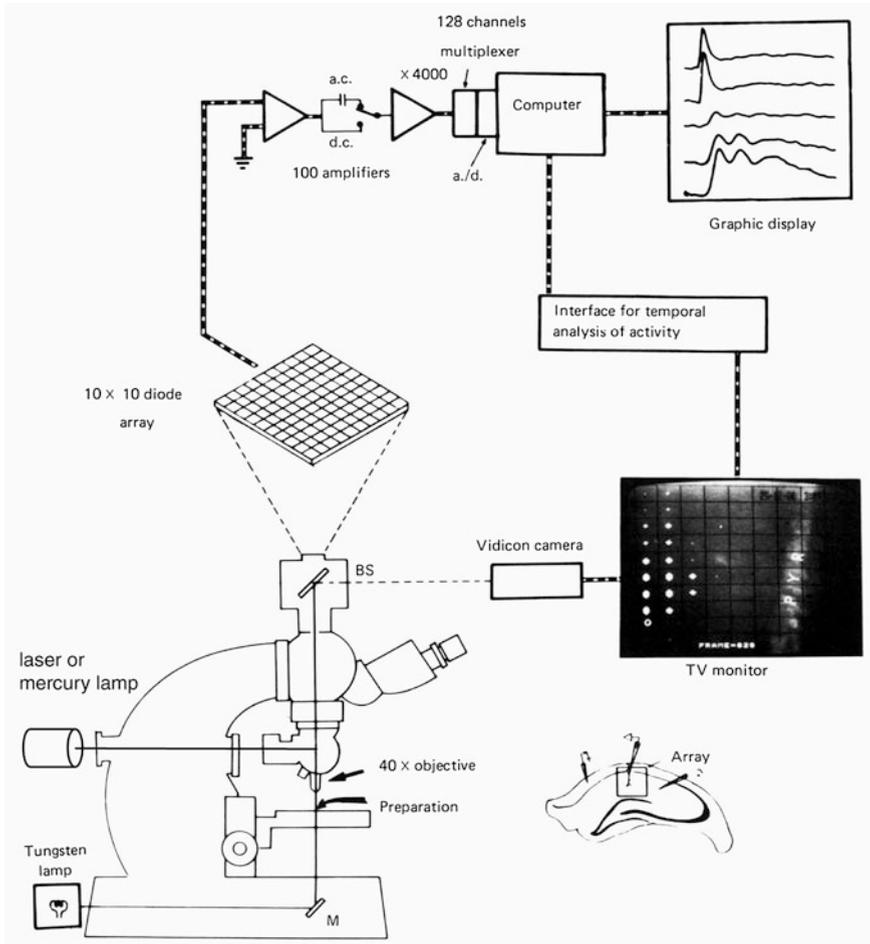


Fig. 2 Imaging apparatus equipped with upright microscope, bottom illuminating a tungsten lamp for transmission measurement, a laser or a mercury lamp as an excitation source for fluorescence imaging, 10 × 10 photodiode array for the real time imaging (modified from Grinvald et al. [19])

infrared absorption [16], and cell volume changes [17]. Physical origins of the changes in optical properties based on the Ref. [18] are summarized in Table 1.

Although physical mechanism of change in optical properties of neuron has been understood to significant extent, more detailed investigations in the specific areas remain to be done using advanced optical tools that bring us to the unexplored experimental spaces. Figure 2 shows a typical optical apparatus for real-time imaging of neuronal activity [19].

4 Optical Recording of Neuron with Voltage Indicators

In the last few decades, ionic bases of action potential propagation and synaptic transmission has been revealed in considerable detail, thanks to the impressive progress in the knowledge about ion channels related to the neuronal activity using various research techniques. Among those techniques, the patch clamp recording sets a gold standard in determining temporal behavior of membrane potential with an unprecedented signal-to-noise ratio [20]. However, patch clamp technique poses a few problems as a research tool for investigating network activity of the neurons, even in a small neural circuitry with a few neuronal cells. First, patch clamp technique disrupts the neuronal membrane so much that long-term recording with the technique is significantly compromised. Second, the simultaneous recording capacity is severely limited due to the bulkiness of the apparatus, typically up to 6 parallel recordings maximum. Third, in the case of large animal models, the usage of the technique is often limited to the *in vitro* recordings, since the recorded signal is extremely sensitive to the movement (cardiac or others) of the cell under patch clamp due to the structural weakness of the recording interface [21].

As described in the Sect. 3, membrane potential dependent changes in optical properties of neurons may prove to be extremely useful for studying the dynamics of a complex neuronal network with a large number of neurons. Intrinsic optical recordings, however, generally demonstrate poor signal-to-noise ratio, presumably due to low optical density of excitable membrane in neurons [22]. Also, intrinsic optical signal may not be temporally resolved during single action potential propagation due to very short integration time, typically in the millisecond time scale. To overcome these problems, voltage indicators have been developed based on small molecules [23, 24], fluorescent proteins [25, 26], and hybrid protein–dye systems [27, 28]. Recently, Adam Cohen and coworkers have developed a couple of fast and sensitive voltage indicators, proteorhodopsin optical proton sensor (PROPS) and Archaeorhodopsin 3 (Arch) [29, 30], and successfully expressed Arch in mammalian neurons [30].

Four different categories with examples are shown in Fig. 3, and temporal resolution and sensitivity ($\Delta F/F$ per 100 mV) for several different indicators are compared in Fig. 4.

Action potentials have been routinely recorded from mammalian neurons using voltage sensitive dye (VSD), but this approach requires intracellular injection of

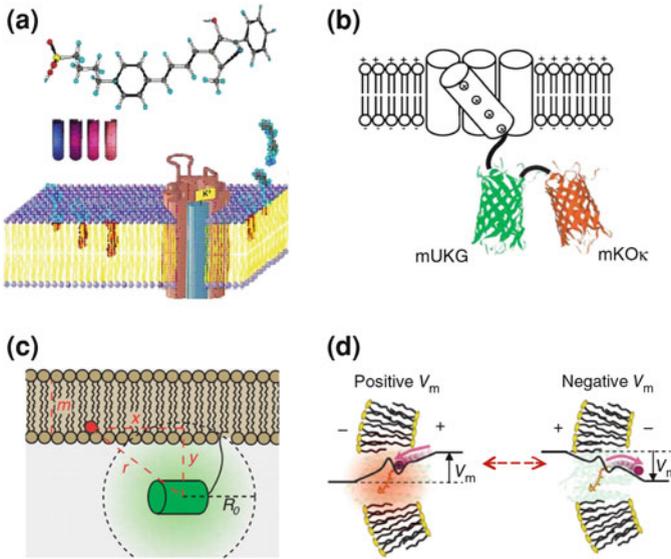


Fig. 3 Examples representing four different voltage indicator categories. **a** di-4-ANEPPS: small molecule voltage sensitive dye (VSD). **b** Mermaid: voltage sensitive fluorescent protein (VSFP) [26]. **c** eGFP-DAP: Hybrid Voltage Sensor (hVOS) [27]. **d** Arch: voltage sensitive microbial rhodopsin (VSMR) [30]

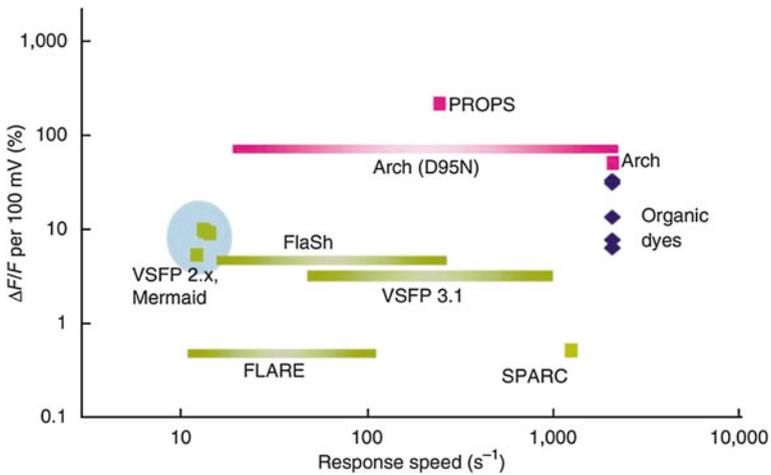


Fig. 4 Characteristics of various optical indicators. *Green* indicators based on fusions of GFP homologs to membrane proteins. *Pink* indicators based on microbial rhodopsins. *Blue* hybrid dye-protein indicators. *Extended bars* indicators where two time constants have been reported [30]

Table 2 Properties of various voltage indicators [31]

Voltage indicator	Classification	Binding	Sensitivity (%)	Temporal resolution
Voltage sensitive dye (di-4-ANEPPS)	Small molecule	Organic probe	0.02–2	<1 μ s
SHG chromophore (FM 4-64)	Small molecule	Organic probe	10–15	<1 μ s
VSNP (Mermaid)	Fluorescent protein	Genetic encoding	40	5–20 ms ^a
VSNP (FlaSh)	Fluorescent protein	Genetic encoding	5	85 ms/ 160 ms ^b
hVSNP (DPA-GFP)	Hybrid sensor system	Organic/genetic hybrid	34	~0.5 ms
hVSNP (DPA-DiO)	Hybrid sensor system	Organic hybrid	Up to 60	~0.12 ms
VSNP (PROPS)	Microbial rhodopsin	Genetic encoding	150	~5 ms
VSNP (Arch)	Microbial rhodopsin	Genetic encoding	N/A	~0.12 ms

^a Measured at T = 33°C

^b τ_{on}/τ_{off}

dye. Photo-toxicity could be another hurdle for VSD as voltage indicator. Action potentials in mammalian neurons have been recorded *in vivo* using a genetically encoded voltage indicator, but it requires averaging over multiple trials due to low sensitivity. Calcium imaging is a popular alternative only in the case that temporal resolution is not required. Although an impressive single-trial based action potential recording has been made using microbial rhodopsin protein, the sensitivity is still too low for any practical application. All the methods have one or more serious limitations, including slow response, low sensitivity, difficulties in targeting, and photo-toxicity. Table 2 shows the properties of various voltage indicators.

Currently, the focus of research in optical imaging of neurons has been shifted to combining with new research tools from other scientific areas, particularly with modern genetics/genetic engineering as shown in this section.

5 Optical Control of Neuronal Activity

Since the membrane of the neuronal cells is electrically excitable, the neuron's activity could be easily controlled by external electrical stimulation. The excitability of the neurons has been used for many important applications, spanning from the basic research in neuroscience such as mapping of neuronal circuitry in the brain [32], to the device applications in biomedical industry, e.g. a deep brain stimulator (DBS) for Parkinson's disease [33]. However, electrical stimulation of neurons with electrodes has some drawbacks difficult to overcome, as discussed in Sect. 2. Since the first experimental attempt by Richard Fork at Bell Laboratories using intense light to evoke action potentials in *Aplysia* ganglion cells [34],

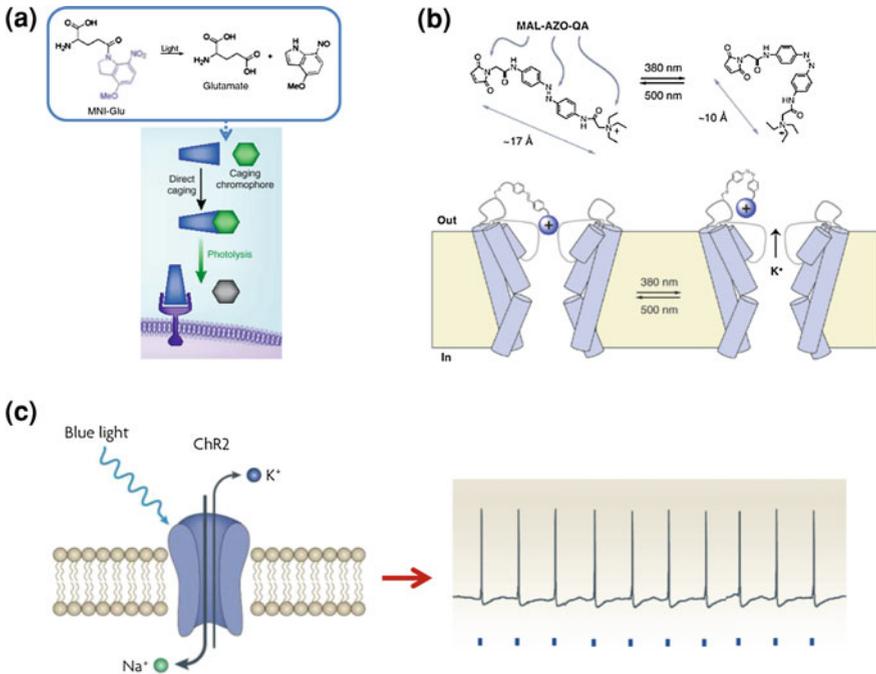


Fig. 5 Different photostimulation approaches: **a** photo-release of signaling molecules (caged glutamate [38]), **b** chemically modification of ion-channels/receptors (MAL-AZO-QA [41]), and **c** genetic introduction of light sensitive channel proteins (channelrhodopsin-2 [44]) to neuronal membrane (left), enabling blue light pulses to trigger electrical impulses (right)

a number of research groups have developed photostimulation techniques as a versatile alternative to the conventional electrical stimulation, demonstrating simultaneous access to many neurons with relaxed mechanical stability requirement [35–37].

So far, photostimulation techniques have been developed in three categories: light-mediated release of chemically modified signaling molecules [38], chemically modified photo-responsive ion channels and receptors [39–41], and genetically introduced photo-sensitive channel proteins [42–44], as shown in Fig. 5. Also, temporal, spatial, and technical properties of different photostimulation techniques are summarized in Table 3.

As noted in the Table 3, *optogenetics*, the method involving genetically expressed light-sensitive ion channels, channelrhodopsin-2 (ChR2), provides multi-parallel access to the network of neurons with temporal resolution in the millisecond time scale. Moreover, a specific type of neurons in the intricate neuronal network may be easily targeted through genetic and viral engineering procedures necessary for ChR2 transfection. It is one of the reasons that

Table 3 Comparison of different photostimulation techniques [47]

Technique	Voltage actuation	Excitation wavelength (nm)	Temporal precision	Exogenous chemicals	Spatial resolution
Glutamate uncaging	Depolarizing	355	1–3 ms	Caged glutamate	5 μ m
ChARGE	Depolarizing	400–600	Seconds to minutes	Retinal	Genetically targetable
ChR2	Depolarizing	480	1–3 ms	Not needed for mammalian cells	Genetically targetable
Modified glutamate receptor	Depolarizing	380 (act.) 580 (inact.)	100 m	Azobenzene-tethered agonist	Genetically targetable
Ion channel ligand uncaging	Depolarizing	355	<1 s	Caged ATP, capsaicin, etc.	Genetically targetable
Modified potassium channels	Hyperpolarizing	380 (act.) 580 (inact.)	1–3 s	Azobenzene-tethered blocker	Genetically targetable
Vertebrate rhodopsin	Hyperpolarizing	475	1–3 s	Retinal	Genetically targetable

optogenetics has emerged as a tool of choice among neuroscientists whose research area involves functional analysis of neuronal circuitry [45], fast mapping of neuronal connectivity [46], and selective stimulation of excitatory/inhibitory neurons [47]. Optogenetics is believed to make a key role as versatile new technology in many areas of the neuroscience, ranging from neurophysiology to behavioral neuroscience, for years to come.

6 Future Research Direction

In this overview, we have attempted to provide a brief introduction to basic characteristics of neurons as well as optical properties of neuronal membrane under stimulation. It is only recently that electrical activity has been converted to optical signal, thanks to the tremendous technological contribution from modern genetics/genetic engineering. In our view, such a technological trend will be maintained in the near future, until the solution for all optical recording and stimulation become available, with non-invasiveness, high temporal resolution (sub-millisecond time scale), high sensitivity, specifically targetable (cell-type specific and/or physical location specific), multi-parallel optical access.

Although optogenetics has been defined in a scientific area related to the neuroscience, there would be an enormous opportunity for optogenetics when exploring other area of research for application. THz spectrum may create such a space for optogenetics since conformation change in the channel protein during pore activation/deactivation has energy scale similar to the THz level. Also, photo cycle of microbial rhodopsin will likely have transitions at the energy level of THz. There has been very limited amount of work done in this research area though, with only a few publications available for references [48, 49].

References

1. Bear, M. F., Connors, B. W., Paradiso, M. A.: Neuroscience: Exploring the Brain (3rd ed., Lippincott Williams & Wilkins, 2007)
2. Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., Lamantia, A.-S.: Neuroscience (5th ed., Sinauer Associates, Inc., 2011)
3. Bressler, S. L., Menon, V.: Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences* **14**, 277–290 (2010)
4. Luo, L., Callaway, E. M., Svoboda, K.: Genetic Dissection of Neural Circuits. *Neuron* **57**, 634–660 (2008)
5. Guo, Z. V., Hart, A. C., Ramanathan, S.: Optical interrogation of neural circuits in *Caenorhabditis elegans*. *Nat. Methods* **6**, 891–898 (2009)
6. Eisenstein, M.: Neural circuits: Putting neurons on the map. *Nature* **461**, 1149–1152 (2009)
7. Cohen, L. B. and Salzberg, B. M.: Optical measurement of membrane potential. *Rev. Physiol. Biochem. Pharmacol.*, **83**, 35–88 (1978).
8. Zimmer, C.: 100 trillion connections: New efforts probe and map the brain’s detailed architecture. *Sci. Am. Mag.* **1**, 51–63 (2011)
9. Brain Facts: A Primer on the Brain and Nervous System (6th ed.), Society for Neuroscience [Online]. Available: <http://www.sfn.org/brainfacts>
10. Voglis, G., Tavernarakis, N.: The role of synaptic ion channels in synaptic plasticity. *EMBO reports* **7**, 1104–1110 (2006)
11. Margrie, T. W., Meyer, A. H., Caputi, A., Monyer, H., Hasan, M. T., Schaefer, A. T., Denk, W., Brecht, M. Targeted whole-cell recordings in the mammalian brain in vivo. *Neuron* **39**, 911–918 (2003)
12. Cohen, L.B.: Changes in neuron structure during action potential propagation and synaptic transmission. *Physiological Reviews* **53**, 373–418 (1973)
13. Cohen, L.B, Keynes, R.D, Hille, B.: Light scattering and birefringence changes during nerve activity. *Nature* **218**, 438–441 (1968)
14. Ritchie, J.M.: Energetic aspects of nerve conduction: The relationships between heat production, electrical activity and metabolism. *Progr. Biophys. Mol. Biol.* **26**, 147–187 (1973)
15. Landowne, D., COHEN, L. B. Changes in light scattering during synaptic activity in the electric organ of the skate, *Raia erinacea*. *Biol. Bull.* **137**, 407–408 (1969).
16. Howarth, J.V., Keynes, R.D., Ritchie, J.M. The origin of the initial heat associated with a single impulse in mammalian non-myelinated nerve fibres. *J. Physiol. (London)* **194**, 745–793 (1968)
17. Hill, D.K.: The volume change resulting from stimulation of a giant nerve fibre. *J. Physiol. (London)* **111**, 304–327 (1950)
18. Ross, W. N., Salzberg, B. M., Cohen, L. B., Grinvald, A., Davila, H. V., Waggoner, A. S., Wang, C. H.: Changes in absorption, fluorescence, dichroism, and birefringence in stained

- giant axons : optical measurement of membrane potential. *J. Membr. Biol.* **33**, 141–183 (1977)
19. Grinvald, A., Manker, A., Segal, M.: Visualization of the spread of electrical activity in rat hippocampal slices by voltage-sensitive optical probes. *J. Physiol. (London)* **333**, 269–291 (1982)
 20. Sakmann, B., Neher, E.: Patch clamp techniques for studying ion channels in excitable membranes. *Ann. Rev. Physiol.* **46**, 455–472 (1984)
 21. Miesenböck, G., Kevrekidis, I.G.: Optical imaging and control of genetically designated neurons in functioning circuits. *Annu. Rev. Neurosci.* **28**, 533–563 (2005)
 22. Kerr, J. N. D., Denk, W.: Imaging in vivo: watching the brain in action. *Nat. Rev. Neurosci.* **9**, 195–205 (2008)
 23. Bashford, C. L.: The measurement of membrane potential using optical indicators. *Bioscience Reports* **1**, 183–196 (1981)
 24. Carlson, G.C., Coulter, D.A.: In vitro functional imaging in brain slices using fast voltage-sensitive dye imaging combined with whole-cell patch recording. *Nature Protocols* **3**, 249–255 (2008)
 25. Siegel, M.S., Isacoff, E.Y.: A genetically encoded optical probe of membrane voltage. *Neuron* **19**, 735–741 (1997)
 26. Tsutsui, H., Karasawa, S., Okamura, Y., Miyawaki, A.: Improving membrane voltage measurements using FRET with new fluorescent proteins. *Nat. Methods* **5**, 683–685 (2008)
 27. Chanda, B., Blunk, R., Faria, L. C., Schweizer, F. E., Mody, I., Bezanilla, F.: A hybrid approach to measuring electrical activity in genetically specified neurons. *Nat. Neurosci.* **8**, 1619–1626 (2005)
 28. Sjulson, L., Miesenböck, G.: Rational optimization and imaging in vivo of a genetically encoded optical voltage reporter. *J. Neurosci.* **28**, 5582–5593 (2008)
 29. Kralj, J.M., Hochbaum, D.R., Douglass, A.D., Cohen, A.E.: Electrical spiking in *Escherichia coli* probed with a fluorescent voltage indicating protein. *Science* **333**, 345–348 (2011)
 30. Kralj, J. M., Douglass, A. D., Hochbaum, D. R., Maclaurin, D., Cohen, A.E. Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. *Nature Methods* **9**, 90–95 (2012)
 31. Scanzani, M., Häusser, M.: Electrophysiology in the age of light. *Nature* **461**, 930–939 (2009)
 32. Yoshimura, Y., Dantzker, J.L., Callaway, E.M.: Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005)
 33. Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardouin, C., Hoffmann, D., Benabid, A.-L.: Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease. *N. Engl. J. Med.* **339**, 1105–1111 (1998)
 34. Fork, R. L.: Laser stimulation of nerve cells in *Aplysia*. *Science* **171**, 907–908 (1971)
 35. Melyan, Z., Tattelin, E. E., Bellingham, J., Lucas, R. J., Hankins, M. W.: Addition of human melanopsin renders mammalian cells photoresponsive. *Nature* **433**, 741–745 (2005)
 36. Miesenböck, G.: Optogenetic control of cells and circuits. *Annu. Rev. Cell Dev. Biol.* **27**, 731–758 (2011)
 37. Shoham, S., O'Connor, D. H., Sarkisov, D. V., Wang, S. S.: Rapid neurotransmitter uncaging in spatially defined patterns. *Nat. Methods* **2**, 837–843 (2005)
 38. Ellis-Davies, G.C.R.: Caged compounds: Photorelease technology for control of cellular chemistry and physiology. *Nat. Methods* **4**, 619–628 (2007)
 39. Zemelman, B. V., Lee, G. A., Ng, M., Miesenböck, G.: Selective photostimulation of genetically chARGed neurons. *Neuron* **33**, 15–22 (2002)
 40. Volgraf, M., Gorostiza, P., Numano, R., Kramer, R. H., Isacoff, E. Y., Trauner, D. Allosteric control of an ionotropic glutamate receptor with an optical switch. *Nat. Chem. Biol.* **2**, 47–52 (2006)
 41. Banghart, M., Borges, K., Isacoff, E., Trauner, D., Kramer, R. H. Lightactivated ion channels for remote control of neuronal firing. *Nat. Neurosci.* **7**, 1381–1386 (2004)

42. Nagel, G., Szellas, T., Huhn, W., Kateriya, S., Adeishvili, N., Berthold, P., Ollig, D., Hegemann, P., Bamberg, E.: Channelrhodopsin-2, a directly lightgated cation-selective membrane channel. *Proc. Natl. Acad. Sci. USA* **100**, 13940–13945 (2003)
43. Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., Deisseroth, K.: Millisecond timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
44. Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L., Deisseroth, K.: Circuit-breakers: Optical technologies for probing neural signals and systems. *Nat. Rev. Neurosci.* **8**, 577–581 (2007).
45. Zhang, F., Wang, L.-P., Brauner, M., Liewald, J. F., Kay, K., Watzke, N., Wood, P. G., Bamberg, E., Nagel, G., Gottschalk, A., Deisseroth, K.: Multimodal fast optical interrogation of neural circuitry. *Nature* **446**, 633–639 (2007)
46. Petreanu, L., Huber, D., Sobczyk, A., Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nat. Neurosci.* **10**, 663–668 (2007)
47. Zhang, F., Wang, L.P., Boyden, E.S., Deisseroth, K.: Channelrhodopsin-2 and optical control of excitable cells. *Nat. Methods* **3**, 785–792 (2006)
48. Balu, R., Zhang, H., Zukowski, E., Chen, J.Y., Markelz, A.G., Gregurick, S.K.: Terahertz spectroscopy of bacteriorhodopsin and rhodopsin: similarities and differences. *Biophys. J.* **94**(8), 3217–3226 (2008)
49. Siegel, P. H., Pikov, V.: THz in biology and medicine: towards quantifying and understanding the interaction of millimeter- and submillimeter-waves with cells and cell processes. *SPIE Photonics West: BiOS*, San Francisco, CA, USA Jan 23-28, 2010

Genome-Wide Analysis of THz-Bio Interaction

Kyu-Tae Kim and Woong-Yang Park

1 Introduction

We now can predict the biological effects of electromagnetic field (EMF) exposure owing to advances in genomic technologies. Keeping pace with the development of available radiation sources, studies on biological interactions with specific frequency region has been required to precede the fundamental understanding and evaluation of bio-safety. Experimentally verified molecular mechanisms of biological systems with EMF exposure can provide powerful evidence to detect significant microscopic changes in living organism. Accumulated research results ultimately can be contributed to guide and regulate the safe application of EMF sources in diverse fields.

Conjecture for EMF-induced toxicity to human was brought up in 1960s. Workers in substation complained of poor sleep, headache, dyspnea and so on. Firstly led by Soviet Government, epidemiological investigation for relation between EMF and those symptoms were initiated and the result was reported no evidence in a general CIGRE (Conseil International des Grands Reseaux Electrique, [1]). Worldwide research to investigate how EMF affects biological systems has been followed and supported to set the criteria of dosimetry. IARC (International Agency for Research on Cancer) announced officially in May 2011 a newly carcinogenic classification guideline (Table 1) and radiofrequency (RF) was upwardly assessed in group 2B [2] along with ELF (extremely low frequency) magnetic fields based on overall related laboratory evidences [3].

Advent of practical sources associated with THz-band, which finally bridged so-called “THz-Gap”, has prompted scientific interests in diverse application of

K.-T. Kim · W.-Y. Park (✉)

Department of Biomedical Sciences, Seoul National University College of Medicine,
Seoul, 110-799, Korea
e-mail: wypark@snu.ac.kr

Table 1 IARC monographs on the evaluation of carcinogenic risk to humans (2011)

Group	Carcinogenicity (to human)	Scientific evidence	Physical/chemical agents
1	“ <i>Certainly</i> ” carcinogenic	• Sufficient (humans)	(107 agents) Asbest, benzene, cigarette, coltar and etc.
2	2A “ <i>Probably</i> ” carcinogenic	• Limited (humans) • Sufficient (experimental animals)	(59 agents) Formaldehyde, IR, PCB, diesel gas, solar radiation and etc.
	2B “ <i>Possibly</i> ” carcinogenic	• Limited (humans) • Inadequate (experimental animals)	(267 agents) Coffee, DDT, lead, gasoline gas, gathered wild herbs, ELF magnetic fields, RF and etc.
3	“ <i>Not classifiable</i> ” as to its carcinogenicity	• Inadequate (humans) • Inadequate (experimental animals)	(508 agents) Cholesterol, ink, tea, saccharin, ELF electric fields and etc.
4	“ <i>Probably not</i> ” carcinogenic	• Lack of evidence	(1 agent) Caprolactam (nylon)

“T-rays” for imaging and sensing technologies [4]. THz-wave simultaneously takes both merits in light wave and electric wave; the straightness away from electromagnetic perturbation and the non-ionizing penetrability without damage [5]. To cite application instances of THz-region, it has been attempted in various areas including medical fields for pathologic diagnosis [6, 7] and security systems for screening improper things [8]. Furthermore, terahertz wave is also expected to be applied in telecommunication technologies with ultrahigh speed, only if overcoming the noteworthy characteristic for vapor absorption in frequency range from 400 GHz to 10 THz [9].

It is obvious to perform preceding studies to investigate how THz-irradiation affects biological systems, and to make public criteria corresponding to limiting exposure time and dose, before widespread practical utilization. This part highlights what sort of genomic methodologies have been applied to examine the biological effects with EMF-exposure and how cutting-edge genomic technologies could meet for studies with promising THz research fields.

2 Overview of Genomic Backgrounds for THz Research

The compound term “Genome” describes the whole cluster of genes placed in a chromosome of eukaryotic cells. Genome is a structural and functional unit of communication in organisms, and genes are defined as an expressed part of genome and a critical factor to destine the nature of distinct entity. Genetic information is delivered to daughter cells by replication and also applied to specific functional protein synthesis. Approximately 100,000 genes are contained in the mammalian genome and 12,000 ~ 15,000 transcripts are estimated as to hold a function in a cell. Certain represented features of a cell ultimately come from a variety of genetic interactions. Quantitative comparison of specific functional gene

expression enables to indicate and predict the state of cell. Furthermore, phenotype of physical symptoms can be explained with analysis of differentially expressed genes (DEGs), so appropriate relevant reactions can be suggested.

Traditional biological study generally followed a “Hypothesis-Driven” way; (1) Development of a hypothesis explaining observed biological phenomenon. (2) Validation of the hypothesis with designed experiments. (3) Conclusion for the hypothesis. Even if researchers design well for an experiment based on the hypothesis, with this approach, it would be difficult to control various other factors affecting focused phenomenon and afterward interpretation of the experiment results would not be perfect to support the hypothesis. Otherwise, genomic research can introduce another approach which is called as a “Data-Driven” way. Various plausible hypotheses are producible from cumulative massive database relevant to various profiling sets of DEGs based on microarray. More detailed explanation about microarray is dealt in next section.

For understanding how a wide scope of genetic world could be applied in THz-study with biological effects, this section covers basic backgrounds on biology. We can comprehend the flow of biological interactions occurred in organisms in regard to any stimuli including THz-exposure, with the knowledge of how, where and when the genetic expression is differentially responsive.

2.1 Genetic Flow: Expression from DNA to RNA and Proteins

Fundamental hierarchy of genetic flow is initiated with biomolecules composed of nucleic acids. Simply, chemical components of DNA contain ribose, phosphate and four sorts of nucleotides which are adenine (A), guanine (G), thymine (T) and cytosine(C). In 1944, Oswald Avery cleared positive evidence of DNA (deoxy-ribonucleic acid) as the genetic material [10]. From the innovative suggestion for a double-helix model of DNA structure by Watson and Crick [11], Erwin Chargaff established a so-called “Chargaff’s Rules”; four bases of A, G, T and C consisting DNA strand have a complementary base pair (e.g., $A = T$, $G \equiv C$) by hydrogen bonds [12].

Represented letters of bases (A, G, T, C) can standardize the genetic flow as a coding system. With this concept of the so-called codon, we can easily understand and handle the large amount of genetic information. In 1958, Francis Crick designated the process of delivery of genetic information as “Central Dogma”; genetic instruction stored in DNA ultimately synthesizes proteins through the series of processes as follows; DNA replication, RNA transcription and protein synthesis (Fig. 1). With the significant signal in a cellular nucleus, replication of genomic DNA is initiated and RNA is synthesized complementary to sequence of DNA. RNA transcripts move to cytoplasm out of the nucleus through nucleus pore that involves in the movement of biomolecules between nucleoplasm and cytoplasm. According to the arranged sequence of transcripts, translation is occurred in ribosome where twenty sorts of amino acids are linked to be a specific functional

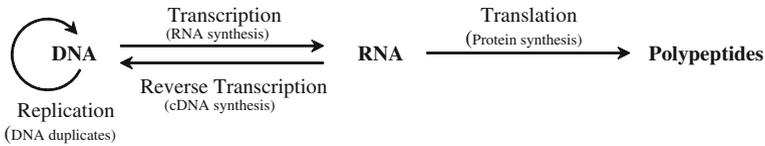


Fig. 1 The central dogma of molecular biology

protein. Primary structure of a protein corresponds to the sequential combinations of three bases having a distinct genetic code for an amino acid. Structural assembly of polypeptides forms tertiary structure of a protein that has an intrinsic function at last.

Genetic sequence in a higher organism is divided into coding region and non-coding region literally unrelated to protein synthesis. We regard coding region as exon separately from intron relevant to non-coding region. Processed mRNA is comprised only of exons and acts as the template for encoded protein synthesis in ribosome. David Baltimore and Howard Temin David Baltimore discovered mRNA could transcribe reversely to complementary DNA (cDNA) with reverse transcription enzyme [13, 14]. This innovative paradigm has rendered stable molecular biological works with cDNA instead of mRNA which is more fragile. Overall steps of reverse transcription are shown in Fig. 2. Isolated mRNA from samples acts as a template to synthesize cDNA with supplemented poly (T) primer, mixture of nucleotides, and reverse transcriptase. Degraded mRNA by RNase H is applied as a primer aligned to poly T region in pre-synthesized cDNA. DNA polymerase completes the synthesis of a newly complementary DNA strand. Compared to genomic DNA, this form of double-stranded cDNA has no intron sequence and is ready to be applied in a wide variety of genomic functional studies.

In response to any types of ceaseless external stimuli, cells adopt a secure regulation system to maintain cellular homeostasis between rapid growth and drastic cell death. Intrinsic quality and quantity of various biomolecules in a living organism are constantly sensed and tightly controlled to manage ordinary survival. If this system fails to cope with cellular stresses, a typical instance can be examined in cancer cells exhibiting mutated forms of DNA sequence especially relevant to genes involved in cell cycle regulation. In regard to THz-research in biological effects, various attempts to detect any abnormal symptoms by validation of gene expression are remained to perform.

2.2 Previous Evidence for EMF or THz-Induced Effects on Biological System

THz-study with biological effects is at the beginning stage keeping pace with the development of feasible radiation sources. Considering the frequency portion of

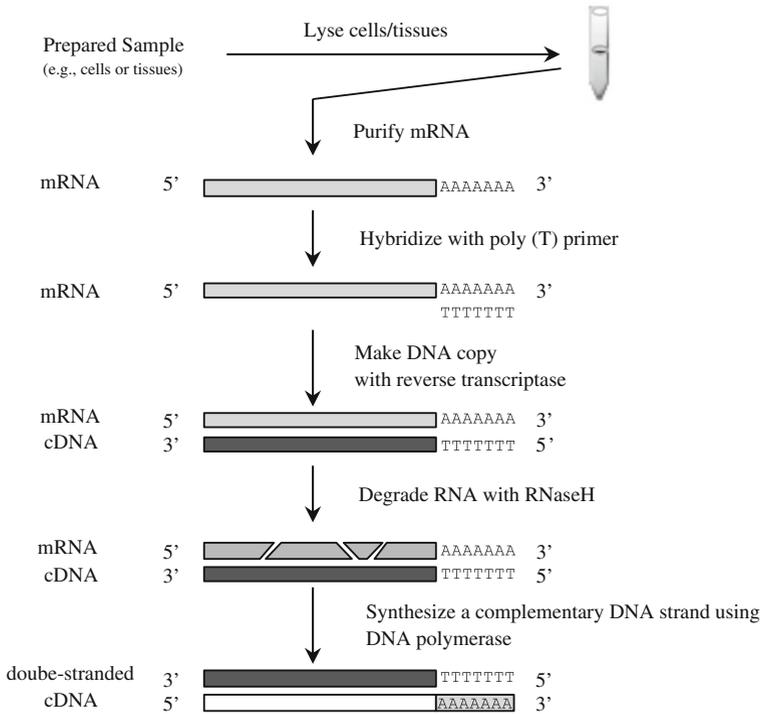


Fig. 2 Process of synthesis of cDNA with mRNA

THz between the IR (infrared) and microwave regions, cumulative abundant prior scientific evidences on the bioelectromagnetic studies for neighboring EMF spectrum of the THz region can suggest the general features of biological response to EMF irradiation. Ahead of dealing with various phenotypes caused by EMF exposure, looking into physical mechanisms of EMF interaction with biological systems can enlighten how initial stimulation on biomolecules leads to biological reactions by and large.

Incidental electromagnetic wave arriving at an intentional spot of samples undergoes the attenuation of the intensity traveling through several materials. The transformational state (*I*) of the incidence irradiance (*I*₀) in the spatial direction of +*z* propagating through a path length δ_p in the material is predictable based on the processing of formulas starting with the Beer-Lambert law:

$$I(z) = I_0 \cdot e^{-\alpha\delta_p}$$

where α refers to the attenuation coefficient corresponding to the material. The intensity of the initial signal (*I*₀) exponentially (e^{-1}) decreases as it reaches in the penetrated point (δ_p). Given the intensity of electro-magnetic wave in a given medium is proportionate to the square of the amplitude, the extent of the reduced intensity corresponds to (e^{-2}). The penetration depth (δ_e) of electromagnetic wave

in samples having distinct impedance (Im) reflected upon the complex refractive index (\tilde{n}) of incidental wave is predictable as follows:

$$\delta_e = 2\delta_p = \frac{\alpha}{2} = \frac{c}{\omega \times \text{Im}(\tilde{n}(\omega))}$$

As is shown above, the penetration depth (δ_e) of electromagnetic wave is inversely proportional to the frequency (ω) as well as the particular resistant properties of irradiated material. From this point, it is predictable that THz wave is relatively limited to penetrate deeply into materials mostly by absorption in the shallow surface. Corresponding to rising frequency, measured absorption coefficient indicates tendency of upward and the descending curve to the index of refraction for liquid water [15].

A variety of bio-molecules composing a living organism have distinct electric properties based on chemical conformation. Such components synthetically generate endogenous bioelectricity. From the point of that photon is the energy carrier for the electromagnetic force, EMF radiation can affect to the stable state of electrochemical energy dynamics in a cell by Coulomb forces. Imposed bundle of photons can cause the changes in existing resonant molecular oscillatory modes involving vibration, rotation and electronic reactions, or non-resonant effects resulting from fluctuating temperature, material concentration, mechanical stress and background electric fields [16]. Fröhlich et al. initially verified that micro-thermal or non-thermal effects with electromagnetic stimulus could make bio-molecules responsive to coherent electric vibration [17]. Recent studies with THz source confirmed again non thermal coupling mechanism could induce the oscillation of biomolecules [18, 19].

The level of photon energy is increasing corresponding to ascending frequency in electromagnetic spectrum. Taking various combinational experimental conditions into physical consideration on radiation source, applied sample, and so on, minimal change in temperature under EMF exposure could be detectable by empirical dosimetric measurement by an IR camera or computational simulation. Given examples in THz studies, temperature fluctuation was observed in different patterns depending on the exposure system. Bock et al. demonstrated cellular reprogramming could be induced by THz-radiation, and no temperature change was monitored by IR camera considering the measuring efficiency [20]. In case of exposure setup with high-power THz source by Air Force Laboratory, on the other hand, $\approx 3^\circ\text{C}$ increase was observed during THz exposure [21]. When focusing only on the electromagnetic influence, exclusion of thermal induced effect is available with accessory settings keeping the temperature uniformly [22].

A lot of fragmentary experiment results can elucidate to figure out general phenotype upon EMF exposure, though biological responses to specific EMF wave show irregular or sometimes conflicting patterns in similar experiment condition involving signal type, exposed intensity or exposure time. Either micro-thermal effect or resonant induction is regarded to date as potential triggers for biological alteration. A series of heat shock proteins (HSP) that are one of biomarkers

unregulated in response to increased temperature have been validated in a wide variety of bioelectromagnetic approaches. A few of laboratory results demonstrated induction of HSP70 in various exposure conditions [23–26]. In case of resonant mode to affect the stable state of cell membrane upon EMF radiation, ion fluctuation between cell membrane is proposed to excite serial cascades of interaction. Even subtle change of endogenous ion concentration could bring about vital fate of cells [27]. Linz et al. proposed that exposure to pulsed radiofrequency fields on heart muscle cells could disturb the stable state of Ca^{2+} dynamics [28]. Biological effects in condition of EMF exposure were observed in diverse phenotypes: DNA synthesis [29], ATP synthesis [30], proliferation and differentiation [31], inhibition of adherence [32], nerve cell damage [33], reduction of sperm motility [34] and so on. As most EMF researches have been on purpose for verification of inimical effects, it still remains ambiguous to make a clear conclusion about biosafety. Rather by taking single window focused on biological phenomenon under EMF signals, it may be available to explore positive approaches applicable in medical fields.

Recently, laboratory results associated THz are spurred to shed light on understanding physical and biological properties of THz source. Laman et al. successfully measured the line-narrowing of THz vibrational modes of biomolecules including thymine, deoxycytidine, adenosine, D-glucose, tryptophan, glycine and L-alanine. Each molecule showed different spectral amplitudes and indicated potential possibility to affect the resonant mode at specific frequency in THz range [35]. As exposed to terahertz field, partial denaturation of DNA, composed of adenine, guanine, thymine and cytosine (see Sect. 2.1), was analyzed by Alexandrov et al. [19] and meant that following genetic alteration caused from THz radiation would be elusive. Suggested biological phenotypes include: DNA synthesis [36], induction of stress responsive genes [21], fall of membrane potential [37], increased membrane permeability [38], stem cell reprogramming [20], reduced blood viscosity [39], platelet aggregation [40, 41], epithelialization and microbial dissemination [42], and etc. Emerging cutting-edge genomic technologies could enhance to discern what the THz-radiation means, and how to practically apply in clinical and industrial fields.

3 Design of Genomic Studies on THz Exposure

The persuasive research results can be originated from a well-designed experimental design. US Food and Drug Administration (FDA) [43] provided detailed guideline on good laboratory practice. Simply, experimental techniques, methods, and conditions should be as completely objective as possible. All data analyses should be completely objective, with no relevant data deleted from consideration and with uniform use of analytical methods.

Because bioelectromagnetic research is required for harmonized collaboration between biological and electromagnetic approaches, both relevant considerations

should be sustained for experiment design [44]. Establishment of appropriate exposure system for bioelectromagnetic research is prerequisite to derive the reproductive and clear experiment results. Owing to International EM Field Project established at World Health Organization (WHO) in 1996, design and characterization of exposure setups have been significantly improved; associated researches have different nonstandard customized exposure systems though. To prevent conflicting results, multidisciplinary collaboration between biological and electromagnetic research fields should satisfy following fundamental setup conditions (Table 2).

To test the potential effectiveness either of experimental results or epidemiological analysis, validation in another experimental model system can imply more powerful evidence. With one experimental model system, the reliability is relatively limited and further investigation of whether certain findings are also valid in another experimental model system can solidly enhance the effectiveness of such findings (Fig. 3). For example, increased social concern of whether mobile phone-use induces genotoxicity to human has triggered worldwide investigation. Either from *in vitro* or *in vivo* research model on brain damage from mobile phone exposure, laboratory evidences showed nerve cell damage [33, 45], increased blood–brain barrier permeability [46] and cognitive impairment [47]. Aside from experimental results, epidemiological analysis in relation between brain tumor risk and mobile telephone use by the INTERPHONE Study Group reported in 2010 that no link was found in the statistics [48]. It is still disputable to make a clear conclusion and further investigation is required to reveal potential effects.

As shown in Fig. 3, experiment models are organically associated each other and comprehensive understanding can lead to lucid explanation for research findings. In following two sections, we deal with *in vitro* and *in vivo* study models corresponding to molecular biological study tools.

3.1 In vitro Approaches

Experiments with cells are useful to seek cellular mechanisms with TH radiation. After THz exposure on cells, various biological assays are available as follows:

- Expression level of specific genes or proteins
- Perturbation of specific ions
- Cellular morphologies
- Molecular interaction dynamics
- Cell cycle analysis and cellular growth rate
- Sequence variation

Any subtle alterations in cellular level, among exemplified parameters above, could cause considerable changes in eventual phenotype through signal cascade transduction. Cells are vulnerable to external stimuli so cellular experiment

Table 2 Fundamental considerable factors for bioelectromagnetic research design

Biological considerations for exposure setup	
Sample preparation	Corresponding to purpose and hypothesis of the study, selectable range for type or species of biomolecules, tissue, cells and animals is varied. For further mechanism studies, a variety of approaches are feasible; treatment of drugs or manipulation of gene expression
Sample holder	Various size and sort of cell containers (e.g. petri dish, flask, plate, tube or cuvette) can be applied in experiments suitable for distinct character of cells and EMF signals. In case of in vivo approaches, multiform of carousels or restrainers are available. Sample holders should not interrupt the distribution of EMF signals
Exposure environment	To focus on the response of EMF exposure without any experimental noises, exposure environment should be kept and monitored perfectly. Basic environmental conditions for cell culture are strict maintenance of temperature ($\approx 37^{\circ}\text{C}$), CO_2 level ($\approx 5\%$), and isolated and sterilized incubation. For animal studies, any other stress factors to experiment subjects should be considered and required to be excluded in exposure setup
Mechanical and electromagnetic considerations for exposure setup	
Signal-generating apparatus	General characteristics of applied signal cover frequency, average power, peak power, band width, intermittency and transmission mode. Within scheduled course of EMF exposure, signal should be stably emitted and monitored to detect fluctuation of propagation. User-friendly interface in exposure system during experiment (from turn-on to turn-off) course can make users manage easier to perform the repetitive experiments
Exposure chamber	Unless exposure is conducted in ambient circumstances, bioelectromagnetic researches generally demand articulated exposure space delivering the EMF signal to prepared samples. Dimension, material and organized structure of exposure chambers depend on signal propagating type and control modality allowing adequate exposure environment. Shield system is needed to protect users from potential effects of EMF exposure
Dosimetry	Practical applied EMF on samples should be defined by computational simulation. Reflecting various physical and electromagnetic features (e.g. average power, conductivity, current density, electric field strength, magnetic field strength, sample mass, sample density, measured temperature increase and etc.), exposed dose can be calculated as quantity of SAR (specific energy absorption rate, W kg^{-1}) and power density (W m^{-2}). SAR distribution can show where and how the EMF is absorbed in samples

condition including temperature, composition of air, and sterilization should be thoroughly controlled to monitor single response under THz irradiation.

To culture cells out of a living tissue, optimal conditions similar to original environment in a living body should be prepared. In physical properties of culture media, automatic controlled incubators are generally used for regulate the concentration of CO_2 and temperature. Optimal pH (commonly, ≈ 7.4) is maintained with bicarbonate system keeping the level of CO_2 , which of solubility depends on

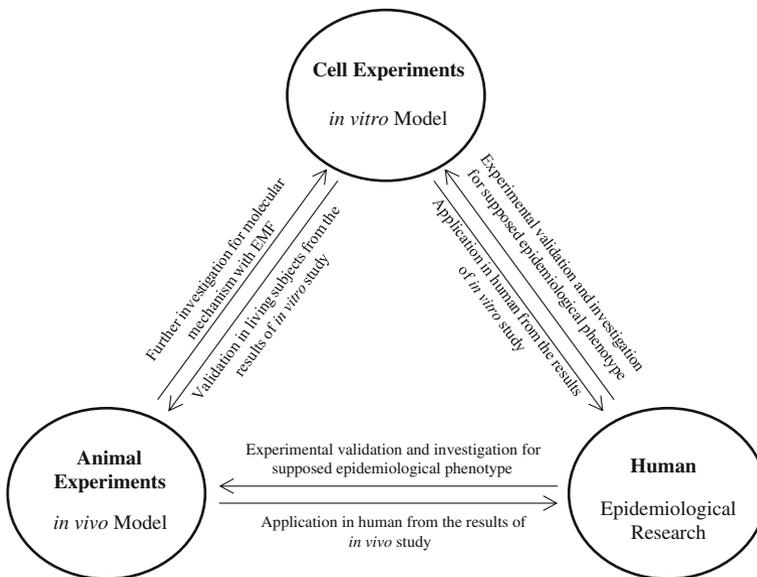


Fig. 3 Scheme of experimental procedure

the temperature. Cell culture media with various supplements provide artificial physiological environment suitable for cells. Considerable composition of supplements in culture media is essential amino acids, vitamins, glucose, antibiotics and various salts (e.g. Na^+ , K^+ , Mg^{2+} , Ca^{2+} , Cl^- , SO_4^{2-} , PO_4^{3-} , HCO_3^- and etc.). Salts are used in order to maintain electric membrane potential and osmotic pressure, and to regulate the physiological environment according to sorts of cells. Nutrient ingredients in culture media are varied in cell types, and various additives, as inhibitors or enhancers, can be applied for a purpose of experiment design. When it comes to EMF researches, meanwhile, the physical properties, including the refractive index and dielectric constant, should be measured for whatever composition of culture media.

The process of homogenization is to lysis cells and extract intended specific molecules into pure RNA, DNA or proteins. Because cell components, including nucleus, mitochondria, lysosome and so on, are easily degraded in ambient environment, the procedure for isolation of RNA/DNA/proteins should be conducted in as low-temperature and RNase-free condition as possible. Then, isolated samples are available for works in molecular biology.

PCR (polymerase chain reaction) is an elementary technique in molecular biology to amplify an interested DNA segment in a complex mixture by multiple cycles of DNA synthesis from short oligonucleotide primers designed to target the specific sequence of an interested gene. In particular, RT-PCR (reverse transcription PCR) is utilized to analyze the quantitative expression of specific mRNA. The first step is to make cDNA from mRNA (see Sect. 1). The extent of amplified

copies of a particular cDNA sequence is detectable by electrophoresis or quantitative real-time PCR monitored with interacting computer system, and provides a clue for the expression level of a specific gene compared to comparison groups.

To separate according to the size of expression and hybridize biomolecules with interesting probes, blotting is a classical method using an electric force moving samples in agarose or polyacrylamide gels. Originated from Edwin Southern in 1975, the so-called southern blotting is to validate the expression level of DNA. In cardinal points, relevant to RNA, it is called northern blotting. As for western blotting using an antigen–antibody reaction, it is utilized to test the degree of expression in protein level for specific functional genes pre-validated in mRNA level. To give an example, Buttiglione et al. validated stress responsive markers both in gene and protein levels, to investigate the genotoxic effect of radiofrequency radiation in human neuroblastoma cells [49]. From checking induction of Egr-1 (early growth response protein 1), a tumor suppressor gene, by RT-PCR, validation for the phosphorylation of MAPK (mitogen-activated protein kinase) pathway-associated proteins including ERK 1/2 (extracellular signal-regulated kinases 1/2), JNK (c-Jun N-terminal kinases) and Elk-1 (E twenty-six-like transcription factor 1), were followed by western blotting analysis to support the activated apoptotic pathways. As genetic flow, such experiment flow from gene to protein validation can provide more persuasive evidence supporting for designed hypothesis.

3.2 In vivo Approaches

Experimental animals are artificially produced and modified on purposes of validation, diagnosis, education and so on. As living samples instead of human, experimental animals take a role for bio-measures in response to certain treatments. Considering its indispensable part in life science, it should be strictly managed to standardize environment condition of raising facility, and maintain high quality in heredity and pathogenic free. General environmental conditions to keep stably involve: temperature, humidity, ventilation and velocity of air, atmospheric pressure, illumination and noise. With such supervised animals, accuracy and reproducibility of experiments can be guaranteed. Animal studies should be performed based on an elaborate experiment design to prevent from dissipation, considering the dignity of life and welfare for animals.

Acquired laboratory results from in vitro experiments can be attempted to in vivo approach in order to confirm whether such findings are also happening in living animals. If so, theoretical justification could be more powerful. Furthermore, integrally verified evidence through in vivo test can be an indirect clue applicable practically in human body, as well. In bioelectromagnetic researches with animals, on the other hand, it is difficult to detect the time point of specific biological response in the inner body because of reversible reaction mechanism. Moreover, each animal has individual variation which obscures statistical process of given

data, even though species, strain, sex, age and raising environmental conditions are identical to all animal samples.

In case of THz researches applying in vivo model, only a few molecular biological literatures have been reported to date in exception of behavioral tests showing depression [41] and anxiety [50]. Kirichuk and his colleagues observed reduced blood viscosity and erythrocyte deformability upon THz exposure, using human blood with isoket [39]. Subsequently, they inquired into blood related responses including platelet aggregation, lipo-peroxidation, coagulation, and fibrinolysis, and demonstrated reacting with rats exposed to THz source (0.129 [51] and 0.15 THz [40, 41, 52]). Ensuing experiments are expected to support the results.

Overall procedure of the in vivo validation responsive for THz exposure is similar to in vitro model. According to interested part of animals and purpose of the studies, right after scheduled electromagnetic exposure, taken tissues from anesthetized animals should be stored in RNAlater solution not to be degraded for RNA works or in liquid nitrogen tank for preservation of proteins. Prepared samples go through the process of homogenization both by physical grinding and chemical lysis. Then, isolated RNA/DNA/proteins are ready to utilize in molecular biological approaches.

When studying further mechanism of specific gene and relevant phenotypes in animals, genetic modified mice are available. Largely, it can be divided into two types according to whether a certain exogenous gene is inserted into mice for overexpression or the function of a specific endogenous gene is turned off. Usually, the former is called transgenic mice, the latter knockout mice. Adequately for experimental intention, artificially genetic manipulation in mice is possible to apply. To give an example, HSP70.1-deficient mice indicating more prone to external physical stress were used for investigation of genetic and histological alteration upon RF exposure [53].

Meanwhile, histological analysis is achieved by microscopic observation. It can provide life-like visualized information from basic shape of tissues to the distribution of specific biomolecules. General process to make excised tissues observable in microscope begins with tissue fixation for effective preservation. After construction of tissue blocks transmittable for light, histological analysis are feasible with various types of staining according to interesting biomolecules (e.g. ions, DNA, RNA, polysaccharides, lipids or various enzymes.) composed of prepared tissue. Overall observation of how and where the interesting biomolecules are placed is possible in a field. Especially, hybridization histochemistry is a useful tool to validate the expression of mRNA in tissue. Compared to classical methods as northern blotting and PCR, this technique can provide positional information as well as the expression intensity. In response to bio-electrical stimulus, for example, histological alteration indicating increased permeability of BBB (blood-brain barrier) under radiofrequency exposure was verified by the increased intensity of stained albumin and shrunken morphology [45].

4 Current Methodologies Applied in Bioelectromagnetic Researches

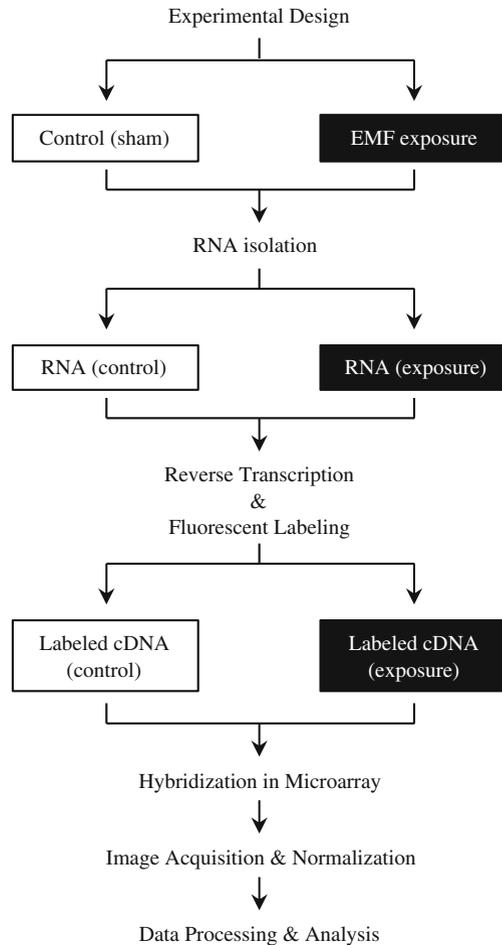
A huge body of biological phenomena is now understandable in the context of methodical ensemble with data processing by computer, referred as bioinformatics. Proteins have an interdependent interacting relationship and develop functional organic networks. Aggregated networks, at last, indicate phenotype in cell units. If the altered expression of particular gene in a certain network is experimentally observed, compared to control group, we could suppose abnormal cellular signal and subsequently validate related works to support the hypothesis. Beyond classical molecular biology deployed as a methodology for bioelectromagnetic studies, advanced biological technologies originated from molecular biology can implement seminal opportunities to delve into potential effects of any treatment including THz irradiation. In this section, we cover three of representative techniques, involving expression analysis, proteomics and epigenetics, in biological studies searching for biomarkers responsive to electromagnetic exposure.

4.1 Expression Analysis

With a purpose of the investigation for overall gene expression pattern, DNA chip technology was firstly invented by Pat Brown in 1995. DNA microarray is an improved technique from northern blotting. Main concept of this, in other words, is to harness the complementary hybridization of DNA probes in a chip to mRNA. Prepared mRNA samples labeling with fluorescent dye hybridize with tens of thousands of gene probes put in a microarray. According to the extent of expression, the luminous intensity from respective gene probes is detected by a fluorescent scanner. Acquired intensity image subsequently converted into corresponding numeric information through data processing and normalization. Overall diagram to get microarray data is shown in Fig. 4.

Noteworthy part in microarray experiment is how to process and analyze the acquired expression data. We can access a vast amount of microarray data sets via public genomic database repository web site, GEO (The Gene Expression Omnibus, URL: <http://www.ncbi.nlm.nih.gov/geo/>). As searching for interested data, we can take advantage of processed data to design experiment. Depending on approaches of data mining, it can be applied to diverse fields: gene discovery, tumor classification, risk assessment, disease diagnosis, prognosis prediction, drug discovery, toxicological research, and more attempts using microarray analysis are expected. In case of bioelectromagnetic researches, characterization of responsive gene expression has been widely utilized and Bock et al. performed microarray experiment with THz radiation on mouse embryo fibroblast and filtered statistically significant responsive genes from 21,000 annotated genes [20].

Fig. 4 Overall procedure to acquire microarray data



From global gene expression data, the first step is to filter out statistically non-significant genes. There may be unexpected error factors, such as sample variability, mechanical errors of scanner, or experimental errors, affecting the characterization of differentially expressed genes between the control group and the treatment group. Filtered candidate genes can be determined generally by how expression values between the two groups differ to a great extent and also to a statistical significance satisfying the P-values below 0.05 as a two-tailed student's t-test. As shown in Fig. 5, for example with microarray data corresponding to 1,763 MHz radiofrequency radiation for 24 h at SAR level of 60 W/kg on human lung fibroblast cells [54], statistically filtered genes are represented in a heat map with hierarchical clustering. In a common way to show a heat map of gene expression with a color key relevant to the extent of expression level, genes are arrayed vertically and

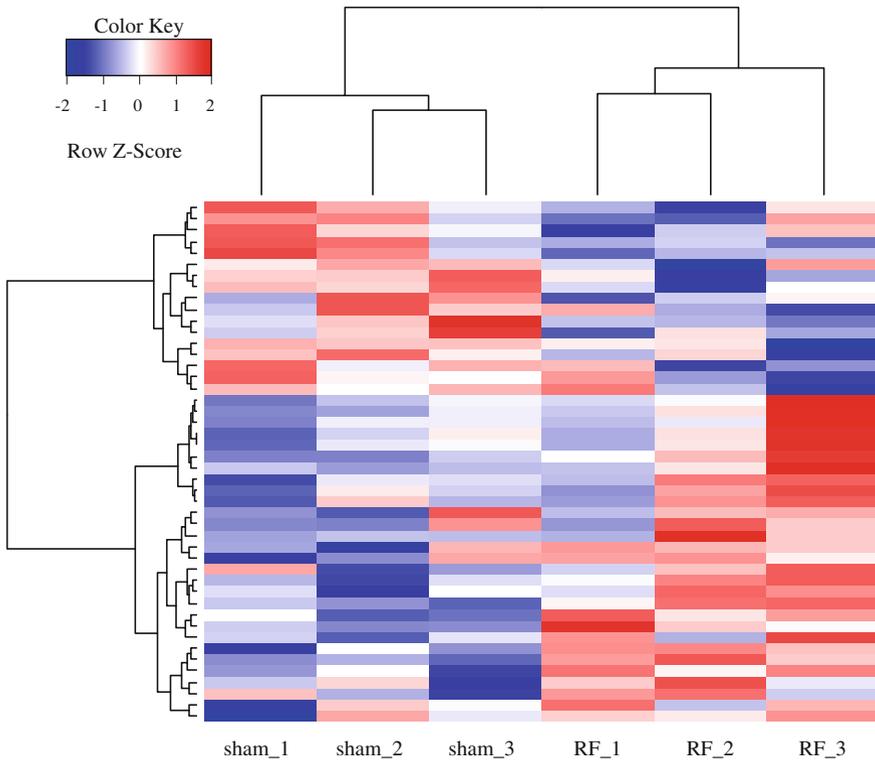


Fig. 5 Filtered gene expression heat map with hierarchical clustering (modified from Im et al. [54])

sample sets are placed in the vertical. A variety of similarity metric can be applied to categorize genes having similar expression pattern of high correlation.

4.2 Proteomics

Proteomics focus on the ‘functional’ proteins considering all translated proteins from mRNA do not have particular functions and some go through the process of modification. Proteome refers to the composition of several thousand proteins expressed by the genome, and is ultimately the last expressed medium reacting in a living body. Further investigation for protein folding and posttranslational modification is essentially required to understand the proteome as much as the genome. Though whole proteome map is not yet identified, it is one of the promising approaches to elucidate the pathway of metabolic activities and identify possible drug targets.

Two-dimensional gel electrophoresis (2-D) is the basis of proteomic experiments. Proteins have particular electric properties according to the combination of comprising amino acids. Immobilized pH gradients are applied for the one-dimensional separation of protein mixtures. Subsequent two-dimensional SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) sorts them into molecular weight. With 2-D gel, up or down regulated proteins of interest are trackable based on corresponding reference data. For more detailed identification of proteins, peptide mass fingerprinting and amino acid sequence analysis are possible to achieve. MALDI ToF instruments are usually used to acquire peptide mass spectrum and it can be compared to the theoretical tryptic peptides by genomic database search. Because structure of proteins is highly complex originated from disulfide bonds, phosphorylation, truncation, acylation and glycosylation, tandem mass spectrometry is applied to characterize structural information of proteins.

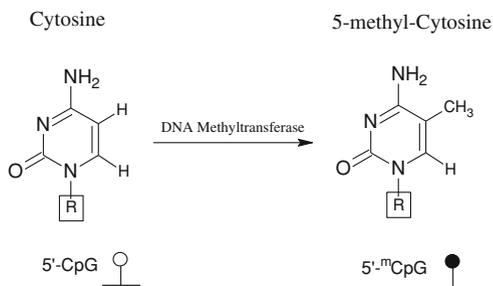
A few proteomic approaches have been reported in bioelectromagnetic researches. Zeng et al. investigated altered gene and protein expression in human breast cancer cell upon 1,800 MHz radiofrequency radiation [55]. Systemic proteome analysis validated distinct RF EMF-responsive proteins which were not detected in genomic level by RT-PCR assay. Meanwhile, Sinclair et al. applied Sty1p deleted mutant *Schizosaccharomyces pombe*, more vulnerable to cellular stresses compared to wild-type, to identify no biological effects of ELF (extremely low-frequency) EMF exposure [56].

It is expected to take much more advantage of proteome analysis in upcoming bioelectromagnetic studies. Harmonized systemic development of techniques, which are involved in protein separation, protein functional interpretation, protein sequencing by mass analysis and bioinformatics, can support proteomic approaches beyond genomics.

4.3 Epigenetics

Even if two independent organisms have identified DNA sequence, phenotypes would be prominently different. This is difficult to explain with classical concept of genetics but possible with epigenetics. Specific expression pattern in particular cells is determined by how it is differentiated. Multiple interactions, like DNA methylation, histone modification or exogenous stimuli, are concerned to this. Epigenetics is the study of genetic functional change without DNA sequence variation.

DNA methylation occurs only at 5' carbon of cytosine (5mC) ahead of guanine through interaction of DNA-methyltransferase enzyme (Fig. 6). In general, its roles are control of gene expression, control of chromosomal integrity, control of recombinational events and protection host DNA from retro-virus or parasitic DNA fragments. Methylated genes are suppressed from expression, whilst unmethylated genes are free to be expressed. Corresponding to what genes are

Fig. 6 DNA CpG methylation

methylated or not, induced cellular function or phenotype would be changeable. There are several mechanisms to be suggested for how methylation inactivates gene expression. Such as, staking energy and polarity of 5mC [57], histone modification and structural change of chromatin [58], and reaction of methyl-CpG-binding proteins or histone deacetylases [59]. In case of the EMF study for biological effect, the investigation for sequence variation at CpG sites was performed in brains of Big Blue mice under 1.5 GHz EMF radiation exposure [60] and showed statistically no induction. Through validation of whether de novo methylation in specific genes is occurred, we can predict functional change from external stimulation including EMF radiation.

5 Promising Genomic Technologies on THz-Bio Interaction

In 1990, the decryption for three billion pairs of human genome was initiated mainly along by the United States, Japan and Europe. With technological advance in data processing and competition between a national institute, NIH (National Institutes of Health), and a private enterprise, Celera Genomics, the draft of human genome map was officially announced in February, 2001 and it was completed in April, 2004, which was earlier to original time schedule, 2005. Owing to the vast amount of whole genome sequence information, we now challenge to reveal the function and interaction of the genome for explanation for the reason of diseases and adequate preparation against them. From the future forecast by the Burrill Report based on current trend of technologies, the change of medical market and the desire of consumers [61], an era of “Tailored Medicine” is coming and a private health care system reflecting a personal genomic data, a present metabolic information and a lifestyle is possible to predict. Though there are still unsolved problems related with social, religious and ethical concerns, benefits from the cutting-edge biotechnologies are expected to successfully lead us to utopia of good health and longevity, beyond the Industrial Revolution and the info-communications revolution.

In this section, we cover three of prospective bio-technological approaches, including polymorphisms, CNV (copy number variation) and genome sequencing,

which are expected to be potential methodologies applicable in biological researches with a wide range of EMF frequencies including THz region.

5.1 Polymorphisms

SNP (Single nucleotide polymorphism) was emerged as an alternative approach to analyze correlation between phenotypes and sequence variation, instead of whole genome sequencing with astronomical costs. Because human share identified genome sequences above 99.5% of total three billion base pairs, the SNP study effectively can take attention only into sequence variation occurring at least 1%. In other words, approximately ten millions of SNPs have a crucial role to distinguish individual properties and the cause of diseases. Focusing on candidate SNP with insertion or deletion of DNA base, we can reveal where and how genetic variation is occurred for determination of specific phenotype by analysis of linkage disequilibrium between SNPs. As establishment of International HapMap Project in 2002 with the United States, the United Kingdom, Canada, China, Japan and Nigeria, the worldwide collaborative investigation was conducted to figure out common patterns of human genetic variation [62]. The interest of this project included how many SNPs are exhibited among individuals, how linkage disequilibrium blocks are formed in SNPs and how to sort representative tagging SNPs. The database for distinct tagging SNPs is now constructed and we can access to internet based searching program (URL: www.broad.mit.edu/mpg/haploview/). Due to remarkable development of high-throughput DNA chip technology, the genome-wide association study using cross analysis of SNP data and genetic profile is expected to disclose the mode of “CD-CV (common disease-common variant)” hypothesis which suggests the highly occurred diseases or specific phenotypes would be correlated with certain highly variant sequences.

5.2 Copy Number Variation

CNV (copy number variation) is the other concept of genetic polymorphism and indicates structural variation. With the verified reference for genome sequence in many species, it is possible to compare the copy number variation within and between species. To discover CNV, the utilization modality is based on array-CGH, which combines the principle of CGH (comparative genomic hybridization) with fundamental DNA chip technology. Scanned data shows where and how specific sequence region is amplified or in the reverse pattern in comparison to the reference. Whilst SNPs are found in the junk DNA region, CNVs are detected in specific functional gene region so correlated possibility with certain phenotypes are higher than the former polymorphism. For example, it was demonstrated that

the copy number of amylase, reacting in digestion of carbohydrate, was increased in people whose staple food is carbohydrate like rice rather than other peoples enjoying mainly creophagous diets, such as proteins or fats [63]. Various diseases, especially in cancer, also have indicated dynamic CNVs in gene loci associated with tumorigenesis [64]. In regard to that diseases like cancer could be occurred by specific cellular stresses as well as the hereditary mode, tracking for DNA polymorphisms including SNPs and CNVs can be applied for bioelectromagnetic researches.

5.3 Next Generation Sequencing

The Human Genome Project took an enormous sum of money about 2.5 billion dollars in 15 years only for sequencing nearly corresponding to one person. It was conducted based on conventional dideoxynucleotide termination method, as so-called Sanger sequencing, and mainly acquired information is summarized as follows:

1. Human genome is comprised of approximately three billion bases of DNA sequence and about 99.9% of DNA sequence is identified in human.
2. Single gene has averagely 3,000 bases of DNA sequence. (Maximum sized gene in human is dystrophin that supports structural maintenance of cellular cytoskeleton.)
3. In general, 2,968 genes are placed in a chromosome. (In cases of Y-chromosome, 231 genes.)
4. About 2% of total genome stores protein information in genetic codes and 50% of genome sequence is highly repetitive.
5. Specific genes are closely related with certain types of disorders like diabetes, rheumatism, cancer or cardiovascular diseases.

Facing with the era of ‘personalized medicine’ feasible with personal sequencing, totally new paradigms for sequencing methods were required to meet the demand. A biotechnology company called 454 Life Sciences launched a first ‘Next Generation Sequencing (NGS)’ developed from pyrosequencing and simultaneous analysis for hundreds of thousands sequence bases was archived. From this innovative device, the whole genome analysis for James Watson, a Nobel Prize laureate in 1962, was announced [65]. Estimated cost for sequencing one person with this method is approximately one million dollars which are relevant to 1/2500th portion of the amount for the Human Genome Project. Subsequent competitive development of NGS devices from Solexa, Applied Biosystems, Helicos and Pacific Bioscience has led the cost of sequencing lower and lower and now we can expect one thousand dollar-sequencing comes soon (Fig. 7) [66].

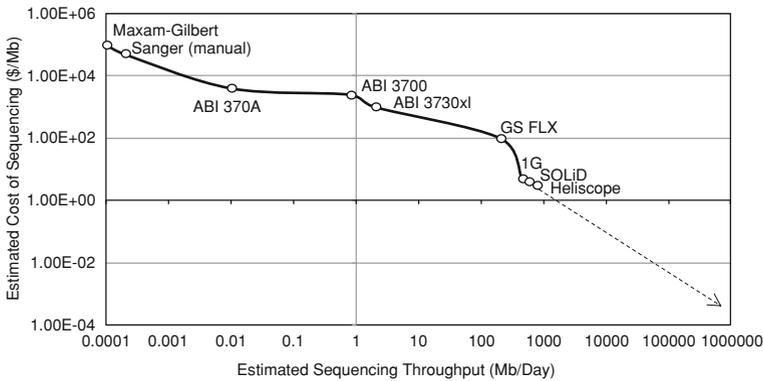


Fig. 7 Prospect of cost and time for DNA sequencing

6 Conclusion

We have presented an overview of how bioelectromagnetic researches have been performed and what we could apply the advanced biological technologies in Terahertz studies. Officially beginning from the first international terahertz biological research activity, the THz-BRIDGE project [67] funded in the EU “Quality of Life” program, recent works to investigate the interaction of THz radiation with biological systems are being performed by worldwide collaborative research groups including a Center for THz-Bio Application Systems, South Korea. A substantial effort to satisfy social concern for verification of biosafety is required to THz-bio related researchers. It is surely primary prerequisite to provide the definite guarantee before utilization of available EMF sources. In the end, we can go to the next stage to formulate how we could take infinite advantages of the sources.

Acknowledgments This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. 2011-0001291) and the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2011- 11911-01108).

References

1. A. Blair, A. Dowel, G. Godly: Environmental Aspects of High Voltage Substations. CIGRE **1**(23.05), 1–6 (1972)
2. IARC: IARC classifies radiofrequency electromagnetic fields as possibly carcinogenic to humans. www.iarc.fr/en/media-centre/pr/2011/pdfs/pr208_E.pdf (2011)
3. Ahlbom, A., et al.: Possible effects of electromagnetic fields (EMF) on human health—opinion of the scientific committee on emerging and newly identified health risks (SCENIHR). *Toxicology* **246**(2–3), 248–250 (2008)
4. Alfano, R.R., Shumyatsky, P.: Terahertz sources. *J. Biomed. Opt.* **16**(3), 033001 (2011)

5. Zhang, X.C., Dawes, D., Chen, Y.Q.: THz wave sensing and imaging. Abstr. pap. Am. Chem. Soc. **230**, U346–U346 (2005)
6. Shen, Y.C.: Terahertz pulsed spectroscopy and imaging for pharmaceutical applications: a review. *Int. J. Pharm.* **417**(1–2), 48–60 (2011)
7. Tewari, P., et al.: Terahertz imaging of biological tissues. *Stud. Health. Technol. Inform.* **163**, 653–657 (2011)
8. Grossman, E., et al.: Passive terahertz camera for standoff security screening. *Appl. Opt.* **49**(19), E106–E120 (2010)
9. Moller, L., et al.: Data encoding on terahertz signals for communication and sensing. *Opt. Lett.* **33**(4), 393–395 (2008)
10. Avery, O.T., Macleod, C.M., McCarty, M.: Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **79**(2), 137–158 (1944)
11. Watson, J.D., Crick, F.H.: Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**(4356), 737–738 (1953)
12. Chargaff, E.: Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6**(6), 201–209 (1950)
13. Baltimore, D.: RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**(5252), 1209–1211 (1970)
14. Temin, H.M., Mizutani, S.: RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**(5252), 1211–1213 (1970)
15. Thrane, L., Jacobsen, R.H., Jepsen, P.U., Keiding, S.R.: THz reflection spectroscopy of liquid water. *Chem. Phys. Lett.* **240**(4), 330–333 (1995)
16. Vaughan, T.E., Weaver, J.C.: Molecular change signal-to-noise criteria for interpreting experiments involving exposure of biological systems to weakly interacting electromagnetic fields. *Bioelectromagnetics* **26**(4), 305–322 (2005)
17. Fröhlich, H.: Long range coherence and action of enzymes. *Nature*. **228**(5276), 1093 (1970)
18. Reimers, J.R., et al.: Weak, strong, and coherent regimes of Frohlich condensation and their applications to terahertz medicine and quantum consciousness. *Proc. Natl. Acad. Sci. USA* **106**(11), 4219–4224 (2009)
19. Alexandrov, B.S., et al.: DNA breathing dynamics in the presence of a terahertz field. *Phys. Lett. A* **374**(10), 1214 (2010)
20. Bock, J., et al.: Mammalian stem cells reprogramming in response to terahertz radiation. *PLoS One* **5**(12), e15806 (2010)
21. Wilmlink, G.J., et al.: In vitro investigation of the biological effects associated with human dermal fibroblasts exposed to 2.52 THz radiation. *Lasers Surg. Med.* **43**(2), 152–163 (2011)
22. Huang, T.Q., et al.: Molecular responses of Jurkat T-cells to 1763 MHz radiofrequency radiation. *Int. J. Radiat. Biol.* **84**(9), 734–741 (2008)
23. Miyakoshi, J., et al.: Effects of exposure to a 1950 MHz radio frequency field on expression of Hsp70 and Hsp27 in human glioma cells. *Bioelectromagnetics* **26**(4), 251–257 (2005)
24. Czyz, J., et al.: High frequency electromagnetic fields (GSM signals) affect gene expression levels in tumor suppressor p53-deficient embryonic stem cells. *Bioelectromagnetics* **25**(4), 296–307 (2004)
25. Tian, F., et al.: Exposure to 2.45 GHz electromagnetic fields induces hsp70 at a high SAR of more than 20 W/kg but not at 5 W/kg in human glioma MO54 cells. *Int. J. Radiat. Biol.* **78**(5), 433–440 (2002)
26. Fritze, K., et al.: Effect of global system for mobile communication microwave exposure on the genomic response of the rat brain. *Neuroscience* **81**(3), 627–639 (1997)
27. Hess, P.: Calcium channels in vertebrate cells. *Annu. Rev. Neurosci.* **13**, 337–356 (1990)
28. Linz, K.W., et al.: Membrane potential and currents of isolated heart muscle cells exposed to pulsed radio frequency fields. *Bioelectromagnetics* **20**(8), 497–511 (1999)
29. Takahashi, K., et al.: Effect of pulsing electromagnetic fields on DNA synthesis in mammalian cells in culture. *Experientia* **42**(2), 185–186 (1986)

30. Zrimec, A., Jerman, I., Lahajnar, G.: Alternating electric fields stimulate ATP synthesis in *Escherichia coli*. *Cell. Mol. Biol. Lett.* **7**(1), 172–174 (2002)
31. Foletti, A., et al.: Cellular ELF signals as a possible tool in informative medicine. *Electromagn. Biol. Med.* **28**(1), 71–79 (2009)
32. Jandova, A., et al.: Effects of sinusoidal magnetic field on adherence inhibition of leucocytes: preliminary results. *Bioelectrochem. Bioenerg.* **48**(2), 317–319 (1999)
33. Grafstrom, G., et al.: Histopathological examinations of rat brains after long-term exposure to GSM-900 mobile phone radiation. *Brain Res. Bull.* **77**(5), 257–263 (2008)
34. Mailankot, M., et al.: Radio frequency electromagnetic radiation (RF-EMR) from GSM (0.9/1.8 GHz) mobile phones induces oxidative stress and reduces sperm motility in rats. *Clinics (Sao Paulo)* **64**(6), 561–565 (2009)
35. Laman, N., et al.: High-resolution waveguide THz spectroscopy of biological molecules. *Biophys. J.* **94**(3), 1010–1020 (2008)
36. Hadjiloucas, S., Chahal, M.S., Bowen, J.W.: Preliminary results on the non-thermal effects of 200–350 GHz radiation on the growth rate of *S. cerevisiae* cells in microcolonies. *Phys. Med. Biol.* **47**(21), 3831–3839 (2002)
37. Ol'shevskaia Iu, S.: Influence of terahertz (submillimeter) laser radiation on neurons in vitro. *Zh. Vyssh. Nerv. Deiat. Im. I. P. Pavlova* **59**(3), 353–359 (2009)
38. Ramundo-Orlando, A., et al.: Permeability changes induced by 130 GHz pulsed radiation on cationic liposomes loaded with carbonic anhydrase. *Bioelectromagnetics* **28**(8), 587–598 (2007)
39. Kirichuk, V.F., et al.: Use of terahertz electromagnetic radiation for correction of blood rheology parameters in patients with unstable angina under conditions of treatment with isoket, an NO donor. *Bull. Exp. Biol. Med.* **146**(3), 293–296 (2008)
40. Kirichuk, V.F., et al.: Sex-specific differences in changes of disturbed functional activity of platelets in albino rats under the effect of terahertz electromagnetic radiation at nitric oxide frequencies. *Bull. Exp. Biol. Med.* **145**(1), 75–77 (2008)
41. Kirichuk, V.F., Efimova, N.V., Andronov, E.V.: Effect of high power terahertz irradiation on platelet aggregation and behavioral reactions of albino rats. *Bull. Exp. Biol. Med.* **148**(5), 746–749 (2009)
42. Ostrovskiy, N.V., Nikituk, C.M., Kirichuk, V.F., Krenitskiy, A.P., Majborodin, A.V., Tupikin V.D., Shub, G.M.: Application of the terahertz waves in therapy of burn wounds. Infrared and millimeter waves and 13th international conference on terahertz electronics, IRMMW-THz (2005)
43. U.S. Food and Drug Administration: Good Laboratory Practice for Nonclinical Laboratory Studies. Code of Federal Regulations, FDA. Title 21, Part 58 (1993)
44. Valberg, P.A.: Designing EMF experiments: what is required to characterize “exposure”? *Bioelectromagnetics* **16**(6), 396–401 (1995), discussion 402–406
45. Salford, L.G., et al.: Nerve cell damage in mammalian brain after exposure to microwaves from GSM mobile phones. *Environ. Health Perspect.* **111**(7), 881–883 (2003), discussion A408
46. Nittby, H., et al.: Increased blood-brain barrier permeability in mammalian brain 7 days after exposure to the radiation from a GSM-900 mobile phone. *Pathophysiology* **16**(2–3), 103–112 (2009)
47. Nittby, H., et al.: Cognitive impairment in rats after long-term exposure to GSM-900 mobile phone radiation. *Bioelectromagnetics* **29**(3), 219–232 (2008)
48. INTERPHONE Study Group: Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. *Int. J. Epidemiol.* **39**(3), 675–694 (2010)
49. Buttiglione, M., et al.: Radiofrequency radiation (900 MHz) induces *Egr-1* gene expression and affects cell-cycle control in human neuroblastoma cells. *J. Cell. Physiol.* **213**(3), 759–767 (2007)
50. Bondar, N.P., et al.: Behavioral effect of terahertz waves in male mice. *Bull. Exp. Biol. Med.* **145**(4), 401–405 (2008)

51. Kirichuk, V.F., Tsymbal, A.A.: Use of terahertz electromagnetic waves for correcting the hemostasis functions. *Med. Tekh.* **1**, 12–16 (2010)
52. Kirichuk, V.F., Tsymbal, A.A.: Effects of terahertz irradiation at nitric oxide frequencies on intensity of lipoperoxidation and antioxidant properties of the blood under stress conditions. *Bull. Exp. Biol. Med.* **148**(2), 200–203 (2009)
53. Lee, J.S., et al.: Subchronic exposure of hsp70.1-deficient mice to radiofrequency radiation. *Int. J. Radiat. Biol.* **81**(10), 781–792 (2005)
54. Im, C.N., Kim, E.H., Park, A.K., Park, W.Y.: Classification of biological effect of 1,763 MHz radiofrequency radiation based on gene expression profiles. *Genomics Informat* **8**(1), 34–40 (2010)
55. Zeng, Q.L., et al.: Effects of GSM 1800 MHz radiofrequency electromagnetic fields on protein expression profile of human breast cancer cell MCF-7. *Zhonghua Yu Fang Yi Xue Za Zhi* **40**(3), 153–158 (2006)
56. Sinclair, J., et al.: Proteomic response of *Schizosaccharomyces pombe* to static and oscillating extremely low-frequency electromagnetic fields. *Proteomics* **6**(17), 4755–4764 (2006)
57. Breslauer, K.J., et al.: Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA* **83**(11), 3746–3750 (1986)
58. Englander, E.W., Howard, B.H.: Nucleosome positioning by human Alu elements in chromatin. *J. Biol. Chem.* **270**(17), 10091–10096 (1995)
59. Knoepfler, P.S., et al.: A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. *Nucleic Acids Res.* **27**(18), 3752–3761 (1999)
60. Takahashi, S., et al.: Lack of mutation induction with exposure to 1.5 GHz electromagnetic near fields used for cellular phones in brains of Big Blue mice. *Cancer Res.* **62**(7), 1956–1960 (2002)
61. Burrill-Report, T.: Life sciences: a 20/20 vision to 2020. www.burrillandco.com/content/BT08_execSum.pdf (2008)
62. Frazer, K.A., et al.: A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**(7164), 851–861 (2007)
63. Perry, G.H., et al.: Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**(10), 1256–1260 (2007)
64. Shlien, A., Malkin, D.: Copy number variations and cancer. *Genome Med.* **1**(6), 62 (2009)
65. Wheeler, D.A., et al.: The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**(7189), 872–876 (2008)
66. Coombs, A.: The sequencing shakeup. *Nat. Biotechnol.* **26**(10), 1109–1112 (2008)
67. Gallerano, G.P.: THz-BRIDGE: tera-hertz radiation in biological research, investigations on diagnostics and study on potential genotoxic effects. <http://www.frascati.enea.it/thz-bridge/reports/THz-BRIDGE%20Final%20Report.pdf> (2004)

Structure and Function of Skin: The Application of THz Radiation in Dermatology

Seong Jin Jo and Oh Sang Kwon

1 Introduction

Skin, the largest organ of human being, is a soft membrane covering the exterior of the body. It protects the host from mechanical injuries, toxic materials, pathogenic organisms, and so on. Although its basic function is protection from the environment like this, it is not a simple and static shield but a complex and dynamic organ which performs important roles in maintaining the homeostasis of the body. Skin controls evaporation to prevent massive water loss, and regulates body temperature by controlling the blood flow of skin and perspiration [1]. It is responsible for the synthesis of vitamin D and a storage center for lipid and water. In addition, skin contains nerve endings and provides sensation for temperature, touch, pressure, and vibration.

The range of terahertz (THz) radiation typically defined as the frequency range 0.1–10 THz, previously known as the terahertz gap due to the lack of sources and detectors [2]. THz Radiation is nonionizing but it excites intermolecular interactions including the librational and vibrational modes in liquid water [3]. Because of its sensitivity to water, terahertz may interact with biologic system but hardly penetrate deep tissue under the skin. Therefore, terahertz radiation may be particularly useful in dermatology field.

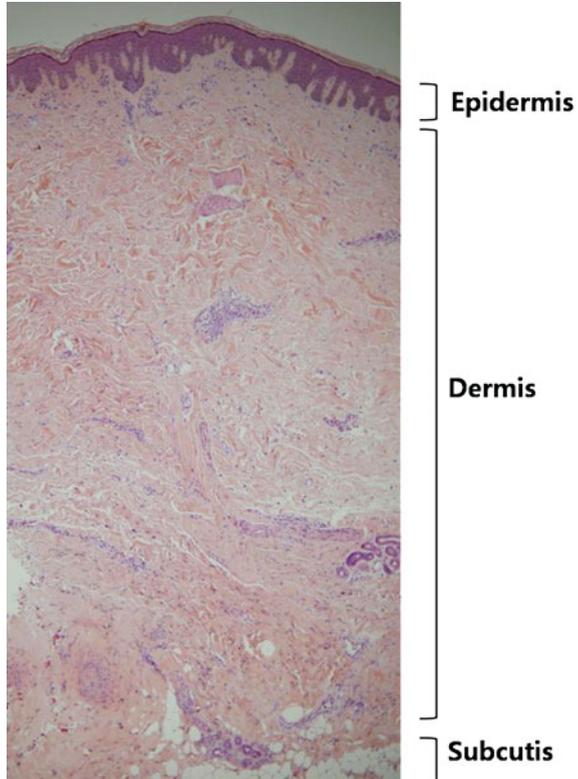
S. J. Jo · O. S. Kwon

Department of Dermatology, Seoul National University College of Medicine,
Institute of Dermatological Science, Seoul National University,
Seoul, Korea

O. S. Kwon (✉)

Department of Dermatology, Seoul National University Hospital,
101 Daehangno, Jongno-gu, Seoul, 110-744, Korea
e-mail: oskwon@snu.ac.kr

Fig. 1 The histology of normal skin stained with hematoxylin and eosin (H&E stain). The skin is a layered structure; the epidermis mainly composed of keratinocytes, the dermis mainly composed of collagen fibers, and the subcutis mainly composed of fat cells



2 The Structure of Skin

Skin is composed of three layers; the epidermis, dermis and subcutis from outside to inside (Fig. 1). The epidermis, the outermost layer of skin, is stratified squamous cornifying epithelium, which means it consists of flat and scale-like cells. Between epidermis and dermis is the basement membrane, a very thin connective tissue mainly consisting of type IV collagen [4]. The dermis, mid-layer of skin, is 15–40 times thicker than the epidermis. The major component of dermis is connective tissue such as collagen and elastic fiber and cells occupies only small parts. It contains nerve, blood vessels, lymphatic vessels, and folliculo-sebaceous unit. The subcutis, also called hypodermis, is the innermost layer and composed mainly of fat cells.

2.1 Epidermis

The epidermis is the thinnest layer among the divisions and ranges in thickness from 0.4 to 1.5 mm. The majority of cells in the epidermis are keratinocytes. Other

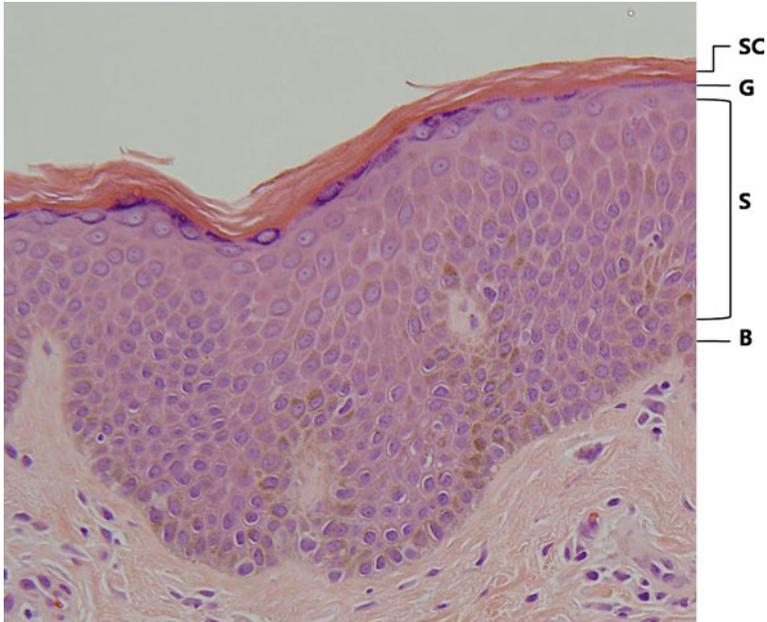


Fig. 2 The structure of the epidermis shown by H&E stain. The most of epidermis is filled with keratinocytes including purple color-stained nuclei. Note that any nuclei do not exist in SC that means full maturation of keratinocytes, so called keratinization. Granular layer are shown as dark purple color due to the cytoplasmic granules of keratinocytes, so called keratohyaline granules. (*B* basal layer, *S* spinous layer, *G* granular layer, *SC* stratum corneum)

cells in the epidermis, such as melanocytes, Langerhans cells, and Merkel cells, are very sparse and intercalates among the keratinocytes.

2.1.1 Keratinocytes

Keratinocytes is originated from ectoderm during embryogenesis and constitute about 90% of epidermis. In epidermis, they attach to each other at desmosome, a cell structure specialized for cell-to-cell adhesion. Keratinocytes contain fibrous structural proteins in their cytoplasm, named as keratin intermediate filaments. Keratin filaments play important roles in forming cytoskeleton. Not only in the skin but also in hair and nails, keratin filaments are produced by various types of keratinocytes [5].

In the epidermis of skin, keratinocytes are piled up into several layers. The shape and metabolic function of keratinocytes change as they progressively differentiate and move upward from the basal cells to the outermost layer. The differentiation of keratinocytes, so called keratinization, is complicated but well-controlled metabolic process. By their shape and differentiation stage, keratinocytes are organized into four layers; basal layer, spinous layer, granular layer, and stratum corneum (Fig. 2).

Basal layer is a single layer of column-shaped keratinocytes that attach to the basement membrane zone (BMZ) at hemidesmosome. Keratinocytes in basal layers are mitotically active and can proliferate. Spinous layer consists of 5 to 10 layers of keratinocytes above the basal layer. This mid-epidermal layer named for the spine-like appearance of the cell margins in histological sections. The shape of keratinocytes in the spinous layer correlates with their position in this layer. Keratinocytes just above the basal layer are polyhedral with round nucleus, but they become flatter as differentiating and moving upward. Therefore, keratinocytes in granular layer become very flat and polygonal. At the level of stratum corneum, the keratinocytes completely differentiate to the corneocytes, flattened cornified cells without nucleus, and are stacked in layers.

Stratum corneum (SC) provides mechanical protection and a barrier to water loss. Its structure is explained by 'two-compartment system' of corneocyte and extracellular lipids. In this explanation model, lipid-depleted and keratin-enriched corneocytes are surrounded by a continuous lipid layer like so called 'bricks and mortar'. The main components of this extracellular lipid layer are ceramide, cholesterol, and free fatty acid.

2.1.2 Melanocytes

Melanocytes are melanin-producing dendritic cells which are derived from neural crest during embryonic development. Melanin is a pigment that absorbs or scatters ultraviolet rays and a major determinant of human skin color. Skin melanocytes are located in the basal layer of epidermis [6] and constitute about 5% of epidermal cells. They transfer melanin-containing melanosomes to adjacent keratinocytes through their dendrites and thus almost keratinocytes through the epidermis have melanin in their cytoplasm. One melanocyte provides melanosomes to about 36 surrounding keratinocytes. This functional unit of a melanocyte and keratinocytes are called as 'epidermal melanin unit' [7].

2.1.3 Langerhans Cells

Langerhans cells are derived from bone marrow and move into the skin through blood vessels. They are dendritic and located in the supra-basal area of epidermis, occupying about 2% of epidermal cell population. Langerhans cells are dendritic and contain characteristic small tennis racket-shaped structure called Langerhans cell granules or 'Birbeck' granules [8]. Their basic function is to sample, process and present antigens to T cells of the epidermis [9]. Therefore, Langerhans cells, regarded as immune cells, play important roles especially in allergic skin diseases.

2.1.4 Merkel Cells

Merkel cells are slow-adapting type I mechanoreceptors. They are located in the basal layer of epidermis only in the site of high-tactile sensitivity and contact with nerve-ending to form the Merkel-neurite complex like synapse. Membrane-bound granules in Merkel cells are similar to neurosecretory granules in neurons and contain neurotransmitter-like substance.

2.2 Dermo-Epidermal Junction

The junction of epidermis and dermis is basement membrane zone (BMZ), which is detected as purple color by periodic acid-Schiff (PAS) stain. This zone is a thin sheet of fibers and consists of four components: plasma membrane of epidermal basal cells with the hemidesmosomes, lamina lucida, lamina densa, and fibrous component of dermis. Lamina lucida, an electron-lucent zone, contains laminin, integrin, and so on. Lamina densa, an electron-dense zone under lamina lucida, is mainly composed of type IV collagen. Anchoring fibrils, dermal microfibrils, and collagen fibers from dermis attach to BMZ.

Basically, BMZ provides structural support to epidermis and holds epidermis and dermis together. It gives resistance to skin against external shearing force. BMZ also serves as a semi-permeable filter, which permits cells and fluid exchange between epidermis and dermis. In addition, BMZ provides signals for epidermal keratinocytes to differentiate normally.

2.3 Dermis

Dermis is a connective tissue constituting most of the skin in regards to volume. It derived from mesoderm during embryogenesis. Dermis structurally supports epidermis and provides nutrients for epidermis. Dermal area beneath epidermis is called papillary dermis and the remaining lower part is reticular dermis. Major component of dermis is connective tissue composed of collagen, elastic fiber, and amorphous ground substance. Fibroblasts, macrophages, mast cells reside in dermis, but the numbers are small. Various structures are contained in dermis, such as nerves, blood and lymphatic vessels, and epidermal appendages.

2.3.1 Collagen Fiber

Collagen is major components of dermis and occupies about 75% of skin in dry weight. It provides skin with tensile strength. About 80–85% of collagen in dermis of adults is type I, and 10% is type III [10].

2.3.2 Elastic Fiber

Elastic fibers are responsible for the elasticity of skin. Mature elastic fiber is composed of elastin, a kind of amorphous protein, and microfibrillar protein.

2.3.3 Ground Substance

Ground substance is defined as heterogeneous amorphous extracellular material intercalating between the cells and fibrous component of dermis. Mucopolysaccharide is one of the major materials of ground substance, so mucin is used as synonym of ground substance. Although ground substance is only 0.2% in dry weight of dermis, it absorbs water up to 1,000 times of its volume.

2.3.4 Blood Vessel

Blood vessels in dermis form plexus, network of very small vessels named as arteriole and venule. There are two kinds of blood plexus in dermis; superficial and deep. Superficial plexus lies in the junction of papillary dermis and reticular dermis, and deep plexus lies in the lower reticular dermis. These two kinds of plexus are connected to each other by communicating vessels. Blood plexus parallels skin surface. Communicating vessels are connected to artery or vein in the septum of subcutis and run perpendicularly to the skin surface.

2.3.5 Nerve

The nerve networks of the skin contain somatic sensory and sympathetic autonomic fibers [11]. The sense of pain, itch, temperature, touch, pressure, and vibration is received by free nerve endings or specialized receptors like Meissner's or Pacinian corpuscle and is transmitted through sensory nerve network. Free nerve endings, the most widespread receptors, are particularly common in the papillary dermis and responsible for the sense of touch, pain, temperature and itch. Some of them are connected to the Merkel cells making Merkel cell-nerve complex. Meissner's corpuscles and Pacinian corpuscles are located in the papillary dermis of the fingers and in the deep dermis of weight-bearing sites, respectively. The sensory receptors are regionally variable in their types and density. In general, the receptors are dense in hairless area such as palm according to the site of body.

Sympathetic autonomic fibers are distributed together with the sensory nerves in the dermis until their branches innervate the sweat glands, vascular smooth muscle, and arrector pili muscle of follicle. They regulate perspiration, vasoconstriction, and erection of hair shaft.

2.4 Epidermal Appendage

Epidermal appendage consists of pilosebaceous unit, a combined structure of hair follicle and sebaceous unit, and eccrine and apocrine sweat unit. They are originated from ectoderm as down-growth of epithelium during embryonic period, and are therefore named as 'epidermal' appendages although most parts of them are included in dermis. In addition to their own specific functions of epidermal appendages, they also serve as reserve of epidermis. For example, epithelial stem cells in hair follicles serve as a source of reepithelization of superficial epidermis in the case of skin injury. The distribution of epidermal appendages varies according to the site of skin. Face and scalp contain pilosebaceous unit in abundance while skin site such as back has comparatively small number of appendages. In general, the sites containing numerous epidermal appendages are also abundant in nerves and blood vessels.

2.4.1 Hair Follicle

Hair is a solid cylindrical fiber made with hard keratin of hair epithelial cells. Only mammals have hair in their skin. Hairs in human are classified into lanugo hair, vellus hair, and terminal hair. Lanugo hair is fine and downy hair that covers the skin of fetus. Lanugo hair is normally shed at 33 to 36 weeks of gestational age and replaced by short (usually less than 2 cm long), fine (less than 30 μm in diameter), and light-colored hair, named vellus hair. Vellus hair is hardly noticeable but covers most of the host's body except lips, palms, soles, and some external genital areas such as glans penis. Vellus hair in body can be replaced with terminal hair by the increase of androgenic hormone during puberty. The terminal hair including hair in scalp skin is long, thick (more than 60 μm in diameter) and dark as compared with vellus hair.

The hair follicle is divided into three segments; inferior segment from the lowest base to the bulge to which arrector pili muscle attaches, isthmus from the bulge to the opening of sebaceous duct, follicular infundibulum from the opening of sebaceous duct to the surface of epidermis. Hair bulb is the root of the hair which is enlarged in diameter, and consists of dermal papilla as a core and enveloping hair matrix. The cells in dermal papilla are originated from mesenchyme in aspect of embryology, in contrast that most of other cells in follicle are epithelium-origin. The dermal papilla controls the hair growth by releasing many growth factors to the keratinocytes in hair matrix which proliferate and differentiate to form hair shaft by keratinization. Thus, the volume of dermal papilla correlates with the number of matrix keratinocytes and the size of hair shaft [12].

The cross-sectional shape of hair follicle is concentric circles of hair shaft, inner root sheath, and outer root sheath from inside to outside. Hair shaft in the center of follicle consists of medulla, cortex and cuticle covering hair shaft. Hair shaft arises from rapidly proliferating matrix keratinocytes in the bulb and hair shaft cells immediately above the matrix cells begin to keratinization. The inner root sheath

also arises from bulb and extends to the level of isthmus. While the hair shaft and inner root sheath grow together toward the skin surface, the outer root sheath is fixed in the position. The outer root sheath envelops the whole hair follicle from the level of bulb to the level infundibulum and is contiguous with the epidermis at the infundibulum.

Hair growth is not continuous, but cyclic through three phase; anagen phase of growth, catagen phase of involution, and telogen phase of rest. During anagen, the matrix cells actively proliferate and produce the growing hair. As the cells stop proliferation, the follicle enters the catagen and develops club-like shape due to incomplete keratinization of the cells. At this phase, the lower portion of the follicle is disappeared leaving a thin epithelial strand. The dermal papilla which was located in subcutaneous fat during anagen, begins to move to the bulge upward during catagen. During telogen, the epithelial strand shortens to the level of bulge and dermal papilla also comes to rest below the bulge. As described above, the upper part including isthmus and infundibular zone is permanent through hair cycle, but the lower segment below bulge totally replaced at each new hair cycle. This is because the hair follicle stem cells that govern the growth-cycle are located in the bulge [13–16]. Provided secondary germ, a transient structure that forms the lower portion of follicles, by stem cells in bulge, the telogen hair goes into the anagen and thus new hair cycle starts.

The duration of each phase during hair cycle varies according to the body site. In the case of human scalp hair, the average period of anagen is 2 to 8 years and the catagen and telogen phases last for approximately 3 months. In human, each individual hair follicle has its own hair-cycle as an independent unit, while most animals show synchronous cycle of hair growth that leads in the synchronous shedding of hair. Owing to the asynchronous hair cycle, approximately 90% of human scalp hairs are in anagen and the rest are in catagen or telogen [17].

2.4.2 Sebaceous Gland

The sebaceous gland is a major component of sebaceous unit. It is a uni- or multi-lobular structure of lipid-producing sebocytes and enveloped by basement membrane. Basal cells at the periphery of the lobule are small, cuboidal, and highly mitotic and give rise to the lipid-filled cells as differentiating and progressing toward the middle of the gland. The lipid produced by sebocytes, called sebum, is composed of squalene, cholesterol, cholesterol esters, was esters, and triglycerides. The sebum is exuded by disintegration of entire sebocytes, a process called holocrine secretion.

Embryologically, sebaceous glands are formed from bulge on the developing hair follicle. In anatomic and functional aspect, therefore, they are usually associated with hair follicles all over the body, except the eyelid, the buccal mucosa and vermilion border of the lip, the prepuce and the areola of women. Although there are no hair follicles in these listed sites of body, sebaceous glands are found. However, sebaceous glands are totally absent in the palms and soles. The size and

density of sebaceous gland is various according to the site of body. For example, the density is 400–900 glands/cm² in the scalp and face [18].

2.4.3 Eccrine Sweat Unit

The function of eccrine sweat unit is regulation of the body temperature by evaporative heat loss of sweat. Not only heat but also other emotional stress can stimulate sweating mediated by cholinergic innervations. Approximately 2 to 4 million eccrine sweat glands are distributed on all skin sites [19], but the density varies according to the site of the body. They are more abundant on the soles (620/cm²) than on the back (64/cm²) [20].

The eccrine sweat unit is composed of three segments, a secretory coil, an intra-dermal duct and an intra-epidermal duct. The secretory coil, which generates an isotonic precursor fluid, is found near the junction of dermis and subcutis and composed of three types of cells; clear, dark and myoepithelial cells. The clear cells are secretory, initiation the formation of sweat. The dark cells border most of the luminal surface of the secretory tubule, but their function is not well-known. Beneath these epithelial cells and above the basement membrane are myoepithelial cells.

During traveling through eccrine ducts, the isotonic fluid generated in secretory coil is modified into hypotonic skin surface sweat by the re-absorption of sodium and chloride. The eccrine sweat duct consists of an inner luminal cells and outer basal cells and is well designed for the effective absorption of sodium. While the intra-dermal duct is straight, the intra-epidermal duct is spiral and it opens directly onto the skin surface.

2.4.4 Apocrine Sweat Unit

Embryologically, apocrine sweat unit develop as outgrowth of the upper portion of the hair follicle and are therefore associated with pilar unit. In contrast to the eccrine sweat unit, apocrine sweat unit is confined to specific sites including the eyelid, the external auditory canal, axillae, areolae, and the anogenital area [21]. The apocrine glands do not begin to function until just before puberty. The secretory manner of apocrine gland is pulsatile by the adrenergic stimulation. Although some protein, carbohydrate, ammonia, and lipid are included in the secretion, the composition has not been known exactly because it is very difficult to obtain apocrine secretion without contamination of eccrine sweat or sebum. Apocrine secretion is odorless, milky, and viscid when first secreted, but it is altered in the skin surface by bacteria and becomes odoriferous. Its function in human is still unknown while it has protective and sexual function in other animals.

Like eccrine sweat gland, apocrine sweat gland also consists of three segments. The coiled secretory gland is located at the junction of dermis and subcutis and consists of inner layer of columnar cells and outer layer of myoepithelial cells. The

secretory process of the columnar cells, so called ‘decapitation secretion’, is distinct that the apexes of cells project into the lumen and are extruded. The apocrine duct is composed of double layer of cuboidal epithelial cells. Contrary to the intra-epidermal eccrine duct, however, the intra-epidermal or intra-follicular apocrine duct usually opens into the infundibulum of follicle and is straight.

2.5 Subcutis

Subcutis is tissue located under the dermis. Most of subcutis is constituted with fat cells namely lipocytes, in which lipid is synthesized and stored, so subcutis is also called subcutaneous fat. The functions of this fat tissue are protection of the skin, providing skin mobility over underlying structures, preventing heat loss, and molding body contours as well as reserve of energy. Lipocytes in subcutis hold together to form lobules [22, 23] which are separated by fibrous septa composed of collagen. Small blood vessels, lymphatics, and nerves run through the fibrous septa. The subcutaneous tissue varies in thickness according to the site of skin, sex, and the nutritional status of the individual.

2.6 Variation in Structure of Skin

There is marked variation in structure according to the site of skin, age of subjects, and so on (Fig. 3). For example, the dermis of back is thick compared with dermis of other sites and is characterized by broad parallel fascicles of collagen. Thick stratum corneum, abundance of eccrine sweat glands, but absence of pilosebaceous units are characteristics of the palms and soles. In scalp skin, a number of terminal hairs exist and their bulbs are located in subcutaneous tissue. In the skin of face, there are many vellus hairs with large sebaceous glands and this is very conspicuous at the skin of nose. At the mucosal aspect of lip, the epithelium does not normally keratinize and the keratinocytes often show glycogen-rich cytoplasm.

Ageing also affects the structure of skin. Old people have thinner epidermis and more flattened dermo-epidermal junction. The collagen fibers in dermis decrease as the host getting old.

3 The Dermatological Application of Terahertz and Millimeter Wave

The electromagnetic radiation in the range of terahertz has been known as ‘terahertz gap’. However, the developments of new sources and detectors accelerate the progress in the technology of terahertz wave such as spectroscopy and imaging

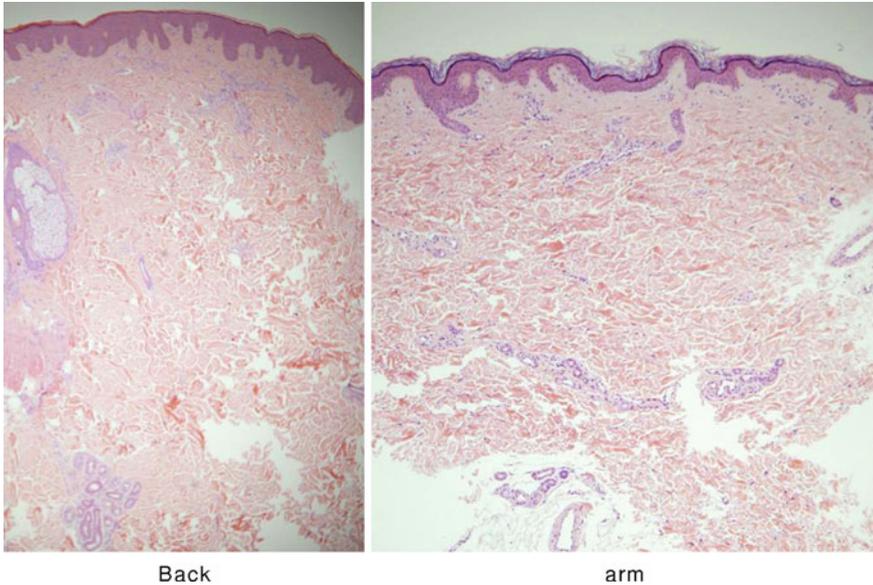


Fig. 3 Variation in the structure of skin according to the site of body. Dermis from the back skin is thick and featured by broad parallel fascicle of collagen. In the skin of arm, the epidermis is thin compared with that of back. Skin from the palm of hand is characterized by a thickened SC, in which spiral eccrine sweat duct is shown. Skin from the scalp and face is abundant in pilosebaceous units. Skin from the outer aspect of lip contains numerous vessels in its dermis and skeletal muscle fibers. (*SC* stratum corneum, *HF* hair follicle, *SG* sebaceous glands)

system [3, 24] and in the understating of the interaction between THz and biological molecules and tissue [25]. Particularly, many researches on THz-skin interaction has been carried out because the skin is the outermost organ enveloping our body and therefore firstly exposed to the THz radiation. In this section, we review the recent understandings of THz in dermatologic area. We also introduce the studies using millimeter waves (MMW) radiation whose wavelengths are in 1–10 mm because MMW radiation has similar characteristics to THz radiation that MMW are non-ionizing electromagnetic waves and sensitive to water. MMW radiation at power densities less than 30 mW/cm^2 , so called low-power MMW, has been widely used for the treatment of many diseases in the former Soviet Union [26–28]. When MMW is irradiated to skin, nearly all of the is absorbed in the epidermis and dermis [29, 30].

As explained earlier, skin is a kind of multilayered structure and each layer has different amount of water, which is one of the major determinants of THz-tissue interaction. The cutaneous water content has been measured by many methods [31–39]. The SC has the smallest amount of water (30–43%) while the total water content in the viable epidermis (spinous and basal layer) and dermis under SC is almost 70%. However, the amount of free water in the viable epidermis and outer dermis may be higher than that in the inner dermis [33, 39].

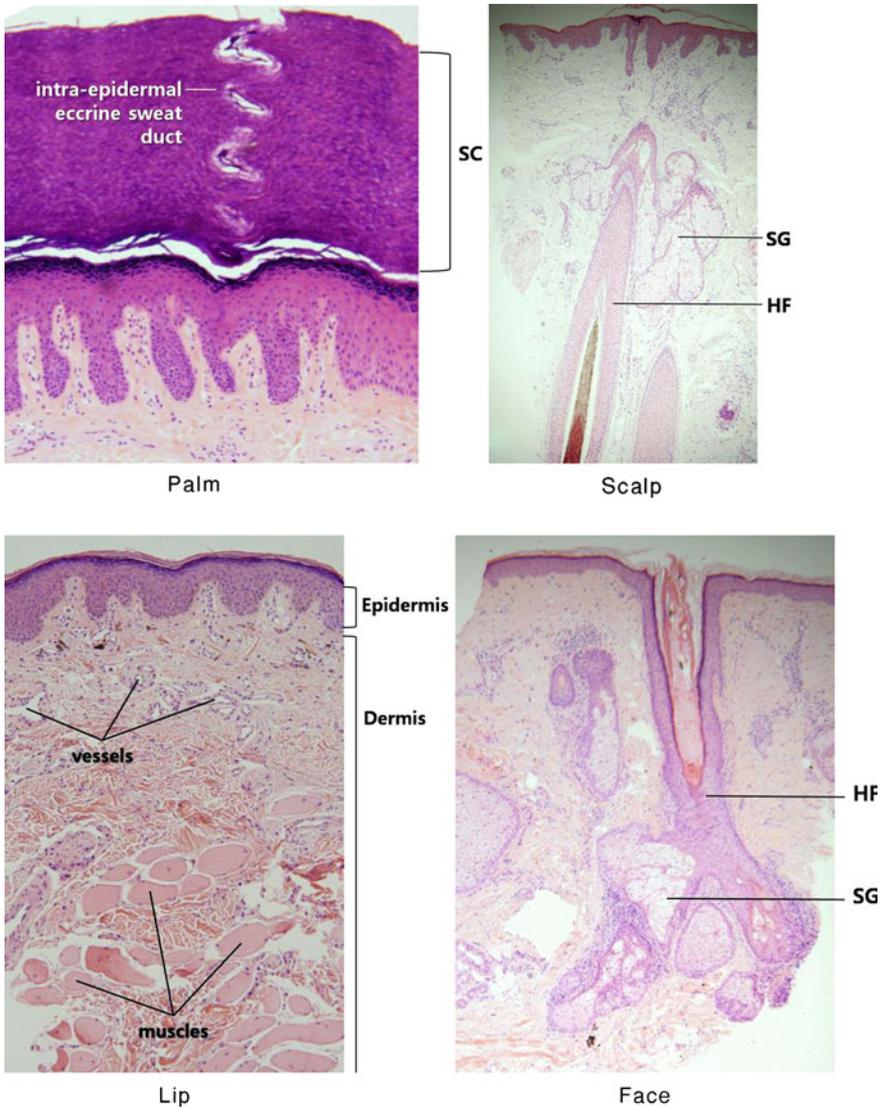


Fig. 3 (continued)

Variation in water content of the skin layer can cause non-uniformities in local electromagnetic fields with in the skin. Electromagnetically, therefore, a heterogeneous multilayered skin model would be more appropriate (Fig. 4) [40, 41], although the early study for the theoretical calculation of wave reflection and absorption in the skin was based on a homogeneous skin model. Alekseev et al. [42] reported that homogeneous uni-layer model is well fitted for the skin with

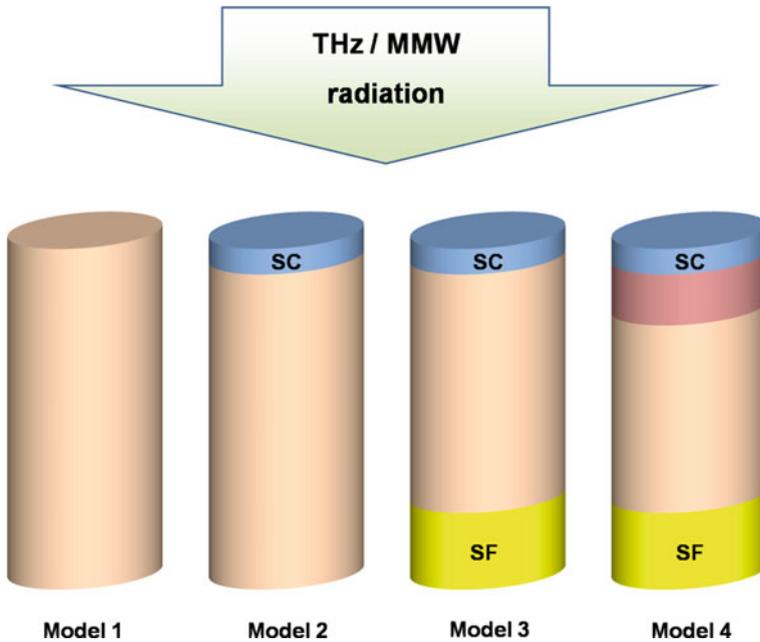
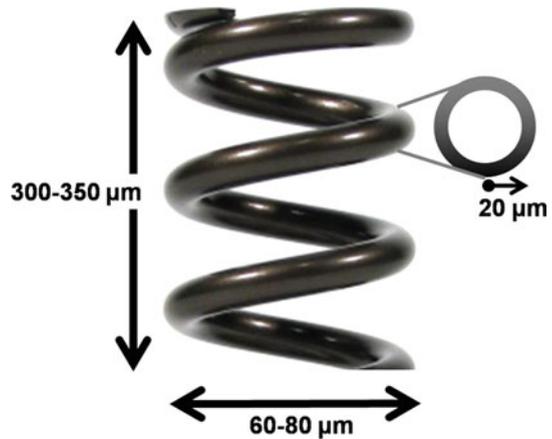


Fig. 4 Various skin models used for calculation of THz/MMW reflection [40, 41]. Model 1 is homogeneous skin model and model 2 is two-layered skin which consists of the upper stratum corneum (SC) and the lower homogeneous tissue. Model 3, three-layered model, contains SC, the rest viable epidermis and dermis, and the lower subcutaneous fat (SF). Model 4 contains four layers, each of which represents SC, viable epidermis, dermis, and SF, respectively

thin SC such as the skin of forearm while multilayer model is used for the skin with thick SC such as palm. [41].

Although the layered skin model is very useful for simulating the interaction between skin and THz radiation, numerous appendages are included within the skin and they may play roles the skin-THz radiation interaction. Here is an interesting example. As mentioned in the previous section, the intra-epidermal eccrine sweat ducts are helically shaped (or spiral) tubes, filled with a conductive aqueous solution. A computer simulation study suggested that the human skin may function as an array of low- Q helical antennas for MMW and sub-MMW bands owing to the characteristic structure of sweat duct (Fig. 5) [43, 44]. Experimental evidence is also presented that the spectral response in the sub-THz region is affected by the degree of perspiration with the minimum of reflectivity in the frequency band of 75–110 GHz [43]. In addition, the skin spectral response in the sub-THz region was found to be correlated to physiological stress as manifested by the pulse rate and the systolic blood pressure and the possibility of the remote sensing of the physiological parameters and the mental state of subjects were suggested [44]. Using local homogenization theory for the skin embedded with sweat ducts, however, Ney et al. [45] found that multiple interference effects from

Fig. 5 Structural model of the intra-epidermal eccrine sweat duct [43–45]. The possibility was suggested that this helical (or spiral) shaped sweat duct, filled with conducted aqueous solution containing ions including Na^+ and Cl^- , functions as low-Q helical antenna, but there is a controversy



the skin layers play the major role in determining the skin electromagnetic characteristics in the MMW and THz regions without the need for the assumption of the sweat ducts acting as low-Q helical antennae.

3.1 The Evaluation of Skin Hydration

Because dryness of the SC may cause and aggravate dermatitis and other skin diseases, many investigators tried to monitor the water content in certain conditions, using non-invasive methods including THz and MMW radiation. For example, the reflection of the millimeter waves in the 37–53 GHz, detected the changes of free water content in the SC by the treatment with various moisturizers [46]. Another study using time-domain THz spectroscopy (TDTS) was performed to determine the influence of hemodialysis on the content of water in human skin and found that the measured refractive index of skin is not related directly to the body hydration status alone [47].

3.2 Skin Imaging Using Terahertz Radiation

Recently, investigators tried to apply THz radiation to medical field. One of these application is Terahertz pulsed imaging (TPI), a noninvasive reflection imaging modality that explores the terahertz frequency range between 0.1 and 3 THz [48, 49]. Since TPI reconstitutes image from data of THz reflection, water content in the tissue is a source of image contrast [50]. Not only the surface information but also the depth information can be extracted because TPI is a time-domain technique. Considering that THz hardly penetrates the skin to reach deep tissues, the TPI may be applicable mainly for the diagnosis of skin diseases. For example,

several studies demonstrate that bound and free water content is different between normal and cancerous tissue [51–55], and tryptophan is also reported to increase in cancer [56]. These differences result in the contrast in TPI between normal and cancerous tissue.

Basal cell carcinoma (BCC), a malignant neoplasm derived from cells in the basal layer of the epidermis, is the most common form of cancer worldwide in white population and over 1 million new cases occur each year in the United States [57]. It can develop anywhere on the body, but usually occurs on sun-exposed areas such as head and neck. Characteristic clinical presentation of BCC includes ulceration, small dilated vessels called telangiectasia, and the presence of a rolled border. A significant difference between the response of THz radiation in normal skin and BCC is observed in several studies using TPI [48, 58, 59] and the increased interstitial water within BCC account for the contrast [58, 60]. Also, in vivo regions of contrast of BCC were seen well THz images and correlated well with histology [59, 61].

Another study suggested the usefulness of THz radiation for the diagnosis of melanoma, a malignant tumor of melanocytes. It is important to differentiate melanoma from pigmented nevus because the latter is benign and do not need rigorous treatment. Imaging using 0.19 THz electromagnetic wave successfully distinguished between the melanoma and nevus tissue, both of which were embedded in paraffin [62].

3.3 The Effects of Terahertz Radiation on the Various Skin Cells

Although THz radiation is non-ionizing, it can affect biological molecules by the interaction between molecular rotational and vibrational modes [25]. It is easily imaginable that the first and the most affected tissue is skin, when THz wave is irradiated to human beings. In fact, many studies using skin cells including keratinocytes were carried out to assess the effect of THz radiation on the skin.

Some studies suggested that the THz radiation is biologically safe. The THz irradiation in the range of 1–3 THz at up to 0.45 J/cm^2 did not show specific inhibition or stimulation of human keratinocytes in vitro and the differentiation of keratinocytes was normally occurred [63]. Besides, Szabo et al. studied the effect of low-power MMW radiation on the HaCaT human keratinocytes, which is a spontaneously immortalized human epithelial cell line from adult skin but maintain full epidermal differentiation capacity [64]. They found no significant change in spontaneous proliferation, adhesion to tissue culture plate, random migration, inducible IP-10 producing, constitutive RANTES and RNATES- and interleukin (IL)-8-induced chemotaxis [65, 66].

However, low-power MMW irradiation resulted in a significant increase of the intracellular level of IL- 1β , a cytokine mediating inflammatory response [65]. Moreover, higher power MMW radiation at 225 mW developed heat stress reaction and induced cellular damage [66]. In other study using HaCaT human

keratinocytes, MMW radiation ($1.0\text{--}3.5\text{ mW/cm}^2$) was reported to be able to reverse the suppression of gap junction intercellular communication by 12-O-tetradecanoylphorbol-13-acetate. Gap junction provides cells for intercellular communication and transfer of essential metabolite ions.

THz radiation can affect not only keratinocytes but also fibroblasts, which are normally located in dermis and produce connective tissue like collagen. In vitro exposure to the radiation of 2.52 THz at 84.8 mW/cm^2 induced heat shock protein in fibroblasts. However, the effects by THz irradiation are comparable with thermal effects in mammalian cells [67].

THz irradiation may have a potential on embryonic development or tissue regeneration by cellular reprogramming. The broad band ($\sim 10\text{THz}$) of THz radiation at the average density of 1 mW/cm^2 was irradiated to the mouse mesenchymal stem cells and certain genes responded to the THz irradiation, activated or repressed. Among the genes, the transcription factor peroxisome proliferator-activated receptor gamma (PPARG) was activated, which means THz irradiation accelerates cell differentiation toward adipose phenotype [68].

THz and MMW may be applicable for the therapeutic purpose. MMW irradiation generates a dose-dependent heating of skin [69, 70]. Szabo et al. reported that in vitro MMW irradiation of keratinocytes and melanoma cells revealed distinct susceptibility of melanoma cells to MMWV hyperthermia with higher thermotolerance of keratinocytes [70]. Additionally, in vivo irradiation of cutaneous melanoma in mice with MMW at the incidence power density of 1.25 W/cm^2 for 30 min resulted in a selective melanoma destruction [70].

3.4 The Future Perspectives of Thz-Application in Dermatology

Much of the biologic characteristics of THz radiation are still unknown because of the structural complexity and molecular heterogeneity of biologic samples. For example, the electromagnetic characteristics of skin in the THz-range are still under investigation although several skin models for the simulation of THz-skin interaction.

Although TPI has some value in diagnosing and evaluating the size of skin cancer including BCC, it is not practical yet due to its big size and low-resolution. However, we expect that the technical progress in THz sources and detectors will solve these problems and furthermore the THz imaging system may be applicable for the diagnosis of the inflammatory skin diseases because inflammatory diseases, like tumors, also show abnormal different water-, cellular-, and molecular profiles. In addition, the evaluation of skin hydration using THz radiation may be particularly useful for the patients with atopic dermatitis, xerotic dermatitis, and so on.

THz radiation may be able to be used for the treatment of skin diseases. For example, certain skin tumors may be able to be selectively removed due to the difference of thermotolerance between tumor cells and normal cells. In addition,

molecular biologic changes can be occurred even by low-powered THz radiation. Therefore, we can apply THz radiation more widely for the treatment of skin diseases if we know the mechanism of frequency-specific effects and delicately control the dose of THz radiation.

References

1. Cork, M.J., Danby, S.: Skin barrier breakdown: a renaissance in emollient therapy. *Br. J. Nurs.* **18**(14), 872, 874, 876–877 (2009)
2. Siegel, P.H.: Terahertz technology. *IEEE Trans. Microw. Theory Tech.* **50**(3), 910–928 (2002)
3. Pickwell-MacPherson, E., Wallace, V.P.: Terahertz pulsed imaging—a potential medical imaging modality? *Photodiag. Photodyn. Ther.* **6**(2), 128–134 (2009)
4. Chan, L.S.: Human skin basement membrane in health and in autoimmune diseases. *Front Biosci.* **2**, d343–d352 (1997)
5. Bragulla, H.H., Homberger, D.G.: Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. *J. Anat.* **214**(4), 516–559 (2009)
6. Passeron, T., Mantoux, F., Ortonne, J.P.: Genetic disorders of pigmentation. *Clin. Dermatol.* **23**(1), 56–67 (2005)
7. Nordlund, J.J.: The melanocyte and the epidermal melanin unit: an expanded concept. *Dermatol. Clin.* **25**(3), 271–281 (2007)
8. Wolff, K., Stingl, G.: The Langerhans cell. *J. Invest. Dermatol.* **80**, 17s–21s (1983)
9. Valladeau, J., Saeland, S.: Cutaneous dendritic cells. *Semin. Immunol.* **17**(4), 273–283 (2005)
10. Chu, D.H.: Development and structure of skin. In: Wolff, K., et al. (eds.) *Fitzpatrick's Dermatology in General Medicine*, pp. 739–749. McGraw-Hill, New York (2008)
11. Johansson, O.: The innervation of the human epidermis. *J. Neurol. Sci.* **130**(2), 184–190 (1995)
12. Sharov, A.A., et al.: Bone morphogenetic protein signaling regulates the size of hair follicles and modulates the expression of cell cycle-associated genes. *Proc. Natl. Acad. Sci.* **103**(48), 18166 (2006)
13. Cotsarelis, G.: Epithelial stem cells: a folliculocentric view. *J. Invest. Dermatol.* **126**(7), 1459–1468 (2006)
14. Ito, M., et al.: Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat. Med.* **11**(12), 1351–1354 (2005)
15. Morris, R.J., et al.: Capturing and profiling adult hair follicle stem cells. *Nat. Biotechnol.* **22**(4), 411–417 (2004)
16. Oshima, H., et al.: Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* **104**(2), 233–245 (2001)
17. Whiting, D.A.: Chronic telogen effluvium: increased scalp hair shedding in middle-aged women. *J. Am. Acad. Dermatol.* **35**(6), 899–906 (1996)
18. Downie, M., Guy, R., Kealey, T.: Advances in sebaceous gland research: potential new approaches to acne management. *Int. J. Cosmet. Sci.* **26**(6), 291–311 (2004)
19. Szabo, G.: The number of eccrine sweat glands in human skin. In: Montagna, W., Ellis, R., Silver, A. (eds.) *Advances in Biology of Skin*, p. 1. Pergamon, New York (1962)
20. Sato, K., Dobson, R.L.: Regional and individual variations in the function of the human eccrine sweat gland. *J. Invest. Dermatol.* **54**(6), 443–449 (1970)
21. Robertshaw, D.: Apocrine sweat glands. In: Goldsmith, L. (ed.) *Biochemistry and Molecular Biology of the Skin*, p. 763. Oxford University Press, Oxford (1991)
22. Koutnikova, H., Auwerx, J.: Regulation of adipocyte differentiation. *Ann. Med.* **33**(8), 556–561 (2001)

23. Holst, D., Grimaldi, P.A.: New factors in the regulation of adipose differentiation and metabolism. *Curr. Opin. Lipidol.* **13**(3), 241 (2002)
24. Plusquellic, D.F., et al.: Applications of terahertz spectroscopy in biosystems. *ChemPhysChem* **8**(17), 2412–2431 (2007)
25. Smye, S., et al.: The interaction between terahertz radiation and biological tissue. *Phys. Med. Biol.* **46**, R101 (2001)
26. Korpan, N.N., Saradeth, T.: Clinical effects of continuous microwave for postoperative septic wound treatment: a double-blind controlled trial. *Am. J. surg.* **170**(3), 271–276 (1995)
27. Rojavin, M., Ziskin, M.: Medical application of millimeter waves. *QJ Med* **91**(1), 57–66 (1998)
28. Radziewsky, A.A., et al.: Suppression of pain sensation caused by millimeter waves: a double-blinded, cross-over, prospective human volunteer study. *Anesth. Analg.* **88**(4), 836 (1999)
29. Alekseev, S.I., Ziskin, M.C.: Distortion of millimeter-wave absorption in biological media due to presence of thermocouples and other objects. *IEEE Trans. Biomed. Eng.* **48**(9), 1013–1019 (2001)
30. Alekseev, S., Ziskin, M.: Millimeter wave power density in aqueous biological samples. *Bioelectromagnetics* **22**(4), 288–291 (2001)
31. Warner, R.R., Myers, M.C., Taylor, D.A.: Electron probe analysis of human skin: determination of the water concentration profile. *J. Invest. Dermatol.* **90**(2), 218–224 (1988)
32. Richard, S., et al.: In vivo proton relaxation times analysis of the skin layers by magnetic resonance imaging. *J. Invest. Dermatol.* **97**(1), 120–125 (1991)
33. Mirrashed, F., Sharp, J.C.: In vivo morphological characterisation of skin by MRI micro imaging methods. *Ski. Res. Technol.* **10**(3), 149–160 (2004)
34. Mirrashed, F., Sharp, J.C.: In vivo quantitative analysis of the effect of hydration (immersion and Vaseline treatment) in skin layers using high resolution MRI and magnetisation transfer contrast*. *Ski. Res. Technol.* **10**(1), 14–22 (2004)
35. Imokawa, G., Kuno, H., Kawai, M.: Stratum corneum lipids serve as a bound-water modulator. *J. Invest. Dermatol.* **96**(6), 845–851 (1991)
36. Caspers, P., Lucassen, G., Puppels, G.: Combined in vivo confocal Raman spectroscopy and confocal microscopy of human skin. *Biophys. J.* **85**(1), 572–580 (2003)
37. Caspers, P.J., et al.: In vivo confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles. *J. Invest. Dermatol.* **116**(3), 434–442 (2001)
38. Mashimo, S., et al.: Dielectric relaxation time and structure of bound water in biological materials. *J. Phys. Chem.* **91**(25), 6337–6338 (1987)
39. Richard, S., et al.: Characterization of the skin in vivo by high resolution magnetic resonance imaging: water behavior and age-related effects. *J. Invest. Dermatol.* **100**(5), 705–709 (1993)
40. Alekseev, S., et al.: Millimeter wave dosimetry of human skin. *Bioelectromagnetics* **29**(1), 65–70 (2008)
41. Alekseev, S., Ziskin, M.: Human skin permittivity determined by millimeter wave reflection measurements. *Bioelectromagnetics* **28**(5), 331–339 (2007)
42. Gandhi, O.P., Riazi, A.: Absorption of millimeter waves by human beings and its biological implications. *IEEE Trans. Microw. Theory Tech.* **34**(2), 228–235 (1986)
43. Feldman, Y., et al.: Human skin as arrays of helical antennas in the millimeter and submillimeter wave range. *Phys. Rev. Lett.* **100**(12), 128102 (2008)
44. Feldman, Y., et al.: The electromagnetic response of human skin in the millimetre and submillimetre wave range. *Phys. Med. Biol.* **54**, 3341 (2009)
45. Ney, M., Abdulhalim, I.: Does human skin truly behave as an array of helical antennae in the millimeter and terahertz wave ranges? *Opt. Lett.* **35**(19), 3180–3182 (2010)
46. Alekseev, S., Szabo, I., Ziskin, M.: Millimeter wave reflectivity used for measurement of skin hydration with different moisturizers. *Ski. Res. Technol.* **14**(4), 390–396 (2008)
47. Kadlec, F., et al.: Assessing skin hydration status in haemodialysis patients using terahertz spectroscopy: a pilot/feasibility study. *Phys. Med. Biol.* **53**, 7063 (2008)

48. Woodward, R.M., et al.: Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue. *Phys. Med. Biol.* **47**, 3853 (2002)
49. Loffler, T., et al.: Visualization and classification in biomedical terahertz pulsed imaging. *Phys. Med. Biol.* **47**, 3847 (2002)
50. Pickwell, E., et al.: In vivo study of human skin using pulsed terahertz radiation. *Phys. Med. Biol.* **49**, 1595 (2004)
51. Ross, K., Gordon, R.: Water in malignant tissue, measured by cell refractometry and nuclear magnetic resonance. *J. Microsc.* **128**(Pt 1), 7 (1982)
52. Chen, J., et al.: In vivo relaxation times and hydrogen density at 0.063–4.85 T in rats with implanted mammary adenocarcinomas. *Radiology* **184**(2), 427 (1992)
53. Rofstad, E., et al.: Magnetic resonance imaging of human melanoma xenografts in vivo: proton spin-lattice and spin–spin relaxation times versus fractional tumour water content and fraction of necrotic tumour tissue. *Int. J. Radiat. Biol.* **65**(3), 387–401 (1994)
54. Gniadecka, M., Nielsen, O.F., Wulf, H.: Water content and structure in malignant and benign skin tumours. *J. Mol. Struct.* **661**, 405–410 (2003)
55. Joseph, C.S., et al.: Continuous wave terahertz transmission imaging of nonmelanoma skin cancers. *Lasers Surg. Med.* **43**(6), 457–462 (2011)
56. Brackxridge, C.: The tyrosine and tryptophan content of blood serum in malignant disease. *Clin. Chim. Acta* **5**(4), 539–543 (1960)
57. Crowson, A.N.: Basal cell carcinoma: biology, morphology and clinical implications. *Mod. pathol.* **19**, S127–S147 (2006)
58. Woodward, R.M., et al.: Terahertz pulse imaging of ex vivo basal cell carcinoma. *J. Invest. Dermatol.* **120**(1), 72–78 (2003)
59. Pickwell, E., et al.: Simulating the response of terahertz radiation to basal cell carcinoma using ex vivo spectroscopy measurements. *J. Biomed. Opt.* **10**, 064021 (2005)
60. Wallace, V.P., et al.: Terahertz pulsed spectroscopy of human basal cell carcinoma. *Appl. spectrosc.* **60**(10), 1127–1133 (2006)
61. Wallace, V.P., et al.: Terahertz pulsed imaging of basal cell carcinoma ex vivo and in vivo. *Br. J. Dermatol.* **151**(2), 424–432 (2004)
62. Mitobe, K., et al.: Imaging of epithelial cancer in sub-terahertz electromagnetic wave. *IEEE*, 2005
63. Clothier, R., Bourne, N.: Effects of THz exposure on human primary keratinocyte differentiation and viability. *J. Biol. Phys.* **29**(2), 179–185 (2003)
64. Boukamp, P., et al.: Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. cell biol.* **106**(3), 761 (1988)
65. Szabo, I., et al.: Reactions of keratinocytes to in vitro millimeter wave exposure. *Bioelectromagnetics* **22**(5), 358–364 (2001)
66. Szabo, I., et al.: Low power millimeter wave irradiation exerts no harmful effect on human keratinocytes in vitro. *Bioelectromagnetics* **24**(3), 165–173 (2003)
67. Wilmink, G.J., et al.: In vitro investigation of the biological effects associated with human dermal fibroblasts exposed to 2.52 THz radiation. *Lasers Surg. Med.* **43**(2), 152–163 (2011)
68. Bock, J., et al.: Mammalian Stem Cells Reprogramming in Response to Terahertz Radiation. *PLoS one* **5**(12), e15806 (2010)
69. Blick, D.W., et al.: Thresholds of microwave evoked warmth sensations in human skin. *Bioelectromagnetics* **18**(6), 403–409 (1997)
70. Szabo, I., et al.: Destruction of cutaneous melanoma with millimeter wave hyperthermia in mice. *IEEE Trans. Plasma Sci.* **32**(4), 1653–1660 (2004)

Science of Water at Nanoscale

Sung-Jin Chang and Wonho Jhe

1 Introduction

A water molecule contains one oxygen and two hydrogen atoms connected by covalent bonds. Despite its apparent simplicity, when a large number of molecules form a bulk liquid, it exhibits a number of anomalous properties [1]. These unusual properties are thought to arise from structure of water molecules binding to each other by hydrogen bonds. The conventional picture of bulk water posits that each water molecule is bonded on average to four others in a tetrahedral motif (see Fig. 1) [2].

Recently, however, this prevailing view has been questioned, and it has claimed that molecules in bulk water bind on average to just two others [3]. It implies that most molecules are arranged in strongly hydrogen bonded chains and rings embedded in disordered cluster network connected mainly by weak hydrogen bonds [3, 4]. However, it is not simple problem to accept this suggestion because such a two-state model based suggestion would fundamentally change our picture of how dissolved substances behave, which plays an important role in various situations including geochemistry, molecular biology, and industrial processes [5].

Meanwhile, there are many forms of water such as interfacial water and confined water as well as bulk water. Recent investigations have increased our understanding of interfacial water and confined water. When taken together with the recent investigations, their main message is that properties of interfacial water and confined water are not equivalent to those of bulk water, in general.

S.-J. Chang

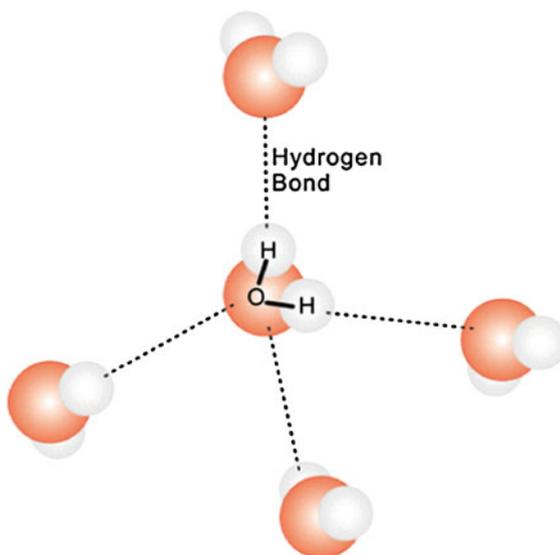
Division of Materials Science, Korea Basic Science Institute, Daejeon, 305-333, Korea

W. Jhe (✉)

Department of Physics and Astronomy, Center for Nano-liquid, Seoul National University, Seoul, 151-747, Korea

e-mail: whjhe@snu.ac.kr

Fig. 1 Schematic of water molecules forming a tetrahedral arrangement by hydrogen bonds



Here some recent understanding of interfacial water and confined water will be introduced in brief. Firstly, three examples of interfacial water will be provided: water molecules (i) at the air–water interface, (ii) at the mica–water interface, and (iii) around solutes or participated in hydration. Secondly, two examples of confined water will be also given: water molecules confined (i) in reverse micelles, and (ii) between two hydrophilic surfaces, which are compressed.

2 Interfacial Water

2.1 Air–Water Interface

The air–water interface is one of the most ubiquitous liquid interfaces in our planet. Its good understanding is important because its surface phenomena affect many disciplines ranging from oceanography to atmospheric and environmental chemistry, and because it is considered as a model system that allows rigorous tests of theoretical proposals. However, assignment of the structure of the water molecules at the air–water interface is still a difficult question to clearly answer.

When an oxygen–hydrogen (OH) group of the water molecule forms a hydrogen bond (HB) with the other water molecule, the OH stretch frequency of this group decreases by an amount determined by the HB strength. As such, the line-shape, including the frequency and line-width, of the OH stretch vibration of the water molecule provides a sensitive marker of the local environment of water molecules. Note that a donor HB, directly affecting the electron distribution around H, causes the OH stretch frequencies to shift much more than an acceptor HB [6].

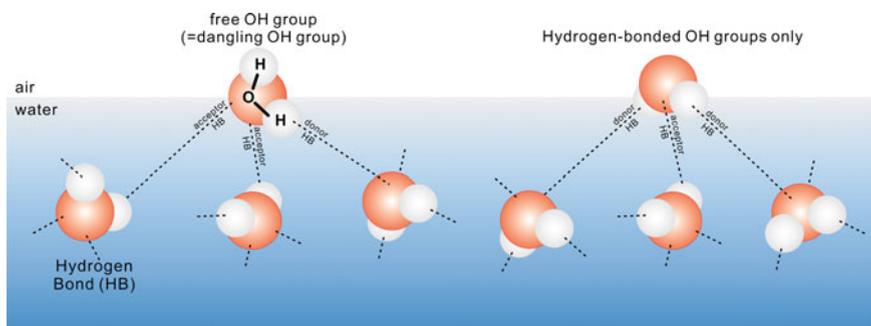


Fig. 2 Schematic of structure of water molecules at the air–water interface, largely assumed in previous works

To probe the line-shape of the OH stretch vibration of interfacial water, which is nearly one monolayer of water molecules, sum-frequency vibrational spectroscopy (SFVS) [7] and its advanced types, including phase-sensitive SFVS (PS-SFVS) [8] and heterodyne-detected SFVS (HD-SFVS) [9], have been used as a particularly versatile tool due to its inherent suppression of the isotropic bulk signal. The SFVS spectrum of the air–water (or heavy water, D_2O) interface exhibits spectral features originating from specific interfacial water molecules: the “free OH (or oxygen-deuterium, OD)” from water molecule with an OH (or OD) group protruding from the surface ($\sim 2740\text{ cm}^{-1}$ for D_2O [10] and $\sim 3690\text{ cm}^{-1}$ for H_2O [11]) (see Fig. 2) and two peaks in the hydrogen-bonded frequency region ($\sim 2200\text{--}2600\text{ cm}^{-1}$ for D_2O and $\sim 3000\text{--}3600\text{ cm}^{-1}$ for H_2O) (see Fig. 2). Although it is clear that such SFVS spectra reflect details of the interfacial water structure, the assignment of the two prominent peaks in the hydrogen-bonded frequency region has been much debated. In largely used explanation, the two peaks have been attributed to ice-like and liquid-like water species at the interface [12–21]. Recently, however, this explanation has been questioned, and it has instead been proposed that the two peaks have the same origin, being a single peak split by intramolecular coupling between the overtone of the bending mode and the stretch fundamental [10, 22]. This is completely distinguished from previous other suggestions [23, 24]. This debate is very important because it is strongly related the assignment of the structure of the water molecules at the air–water interface. The former implies that hydrogen bonding property of water molecules at the air–water interface is very different from that of water molecules in bulk water. On the other hand, the latter implies that the water molecules at the air–water interface are very similar to those in bulk water.

In recent years, instead of the OH (or OD) groups which are hydrogen-bonded to other water molecules (or heavy water molecules) (see Fig. 2), it has been probed the “free OD” bonds (see Fig. 2), which exist only at the surface, as sensitive reporters for detecting interactions of surface water molecules with molecules in the underlying subsurface water layer by means of HD-SFVS [9] combination with accurate theoretical approach [25]. And it has been also claimed

that water molecules at the air–water interface form hydrogen bonds that are only slightly weaker than those in bulk water [25]. It implies that only the surface layer is distinctly different in structure from the rest of the liquid due to existence of the free OH groups, and that the water surface has the thickness of just one layer of water molecules, which is about 0.3 nm.

2.2 Mica–Water Interface

Water molecules tend to be adsorbed on hydrophilic solid surfaces when the surfaces are exposed to humid environments. The adsorbed water molecules, which tend to form the first monolayer at the surface, play an essential role as a structural template in growth of ice, embodiment of the boundary condition for water transport, and mediation of aqueous interfacial chemistry. As such, it thus influences on many processes such as rain precipitation, migration rate of pollutants in rock and soil, corrosion, and catalysis. As a model system of the hydrophilic solid–water interface, (muscovite) mica [26] has been extensively used in various areas because it has importance not only as a prototype of clay minerals in fundamental research regarding clay swelling in geological science [27, 28] and cloud seeding in ecological science [29, 30] but also as a simply obtained atomically flat surface for a good understanding of phenomena at nanoscale [31, 32].

Since the successful use of scanning force microscopy (SFM) for direct imaging of the water molecules adsorbed at mica as the first monolayer [33] (see Fig. 3a), SFM has been also used to achieve two remarkable investigations for a good understanding of the mica–water interface in recent years. Firstly, distribution of the water molecules at the mica surface in aqueous environments has been precisely and three-dimensionally visualized with an atomic-scale resolution [34] (see Fig. 3b). The results in this study support the model proposed on the basis of X-ray scattering [35] and computer simulation [36], where the adsorbed and laterally distributed water molecules coexist at the interface. The coexistence of water molecules having a long relaxation time (adsorbed water) and laterally distributed disordered water molecules (two dimensional hydration layer) may reconcile the two opposing ideas of “ice-like” [37–39] and “liquid-like” [40, 41] water molecules at the mica–water interface. Secondly, height of the first water monolayer at mica exposed to air conditions as well as that of the second and thicker layers have been precisely measured with the help of graphene as a cover layer which does not destroy the original structural properties [42] (see Fig. 3c). Here it has been claimed that submonolayers form atomically flat, faceted islands of height nearly 0.37 nm, in agreement with the height of a monolayer of ice. The second adlayers, observed at higher relative humidity, also appear ice-like, and thicker layers appear liquid-like. It has also claimed that water adlayers grow epitaxially on the mica substrate in a layer-by-layer fashion, which is consistent with previous studies [39, 43]. It can also explain why water adsorption isotherms cannot be modeled with theories based on continuum models [44].

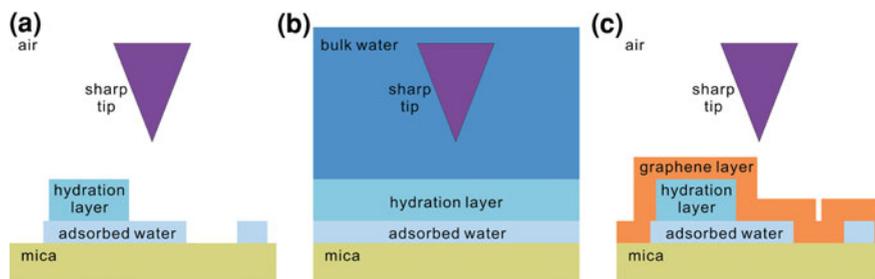
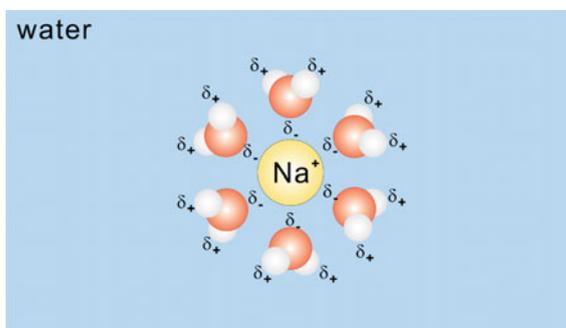


Fig. 3 Schematic of the mica-water interfaces: **a** mica surface directly exposed to air, **b** mica surface immersed in water, and **c** air-exposed mica covered by graphene monolayer

Fig. 4 Ions, considered as one of solute particles, in water are surrounded by a hydration shell (or hydration layer) of water molecules. White and red spheres mean hydrogen and oxygen atoms, respectively



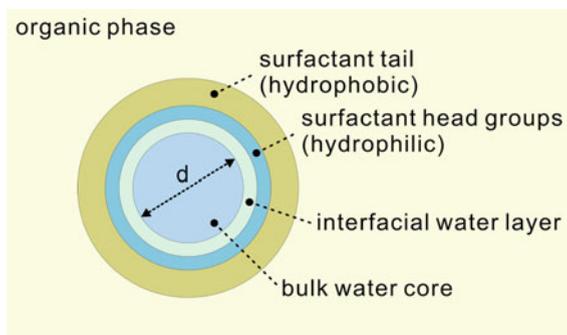
2.3 Hydration and THz Techniques

Clustering of solvent molecules around a solute particle (solvation) (see Fig. 4) is at the heart of a myriad of molecular phenomena. Water is well-known as the native solvent for biological processes. And recently its role has been revealed to be far more active than just being an inert solvent [45]. For example, the clustering of water molecules around protein (hydration) is a key feature for the protein function [46, 47], with the coupling between protein and water dynamics [48] as an important ingredient to understand protein folding and binding. Thus, water molecules used for hydration have been in the focus of numerous studies [45].

Various experimental techniques [49–52] as well as theoretical methods [53, 54] have been extensively used for a good understanding of characteristics of water molecules participating in hydration. For example, it has been revealed using femtosecond nonlinear spectroscopy that water molecules surrounding halogenic anions fluctuate with a characteristic time constant of 10–25 ps (depending on the ion), which is 20–50 times slower than water molecules fluctuations in bulk liquid water [49].

Recently, spectroscopy based on terahertz (THz) laser has been also successfully applied as a promising tool to study the hydration dynamics around solutes [55–57]. On the basis of the results from THz laser spectroscopy, it has been shown that water

Fig. 5 Schematics of a reverse micelle. Water molecules reside in the interior and surfactant tails reside in contact with the continuous, nonpolar, and organic supporting phase such as hexane



dynamics is influenced within a dynamical hydration shell of about 0.6–0.7 nm for carbohydrates (\sim small solutes) [55] by changes in the absolute absorption intensity as probed at 2.4 THz ($\sim 80 \text{ cm}^{-1}$). For proteins (\sim biomacromolecules) [57], this dynamical hydration shell extends even up to almost 2 nm. The long-ranged sensitivity of THz absorption to dynamical properties of water at interfaces results from the presence of numerous processes in water occurring on picosecond time scales, for example, hydrogen bond (HB) rearrangements and rotational relaxation. In recent theoretical study [58], it has been suggested that the prominent peak at about 6.7 THz ($\sim 200 \text{ cm}^{-1}$) is dominated by first hydration shell dynamics, whereas a concerted motion involving the second hydration shell contributes most significantly to the absorption at about 2.4 THz ($\sim 80 \text{ cm}^{-1}$). However, a full interpretation of the insights revealed by THz spectroscopy still remains a daunting task.

3 Confined Water

3.1 Reverse Micelles

In biology as well as in natural and synthetic materials, water is typically tucked away in tiny crevices inside proteins or in porous materials. It is important to understand well water molecules in confinement because these pockets of water are thought to be involved with various crucial processes such as folding and relaxation of proteins. As a tenable model for confined water, especially, in biological systems, reverse micelles have been typically used because their interior dimensions are similar to confined spaces in the biological systems (see Fig. 5). For example, reverse micelles exist in membrane lipid bilayers during endosome formation.

Reverse micelles (see Fig. 5) are spherical droplets of water sequestered from a nonpolar phase by a surfactant layer [59]. Its diameter can be well-characterized and monodispersely varied over a wide range [60, 61]. Such variation makes it possible to determine the influence of the size of a water nanopool on water dynamics and, for large reverse micelles, to determine the distinct dynamics at the interface. The

smallest reverse micelle (50–100 water molecules) have radii of <1 nm, whereas the largest ($\sim 400,000$ water molecules) have radii of up to 14 nm.

Reverse micelles whose diameter (d) is larger than ~ 5.5 nm can be separated into two parts. One comprises the core, which is well removed from the interface. The other includes water molecules at the interface. Water molecules in the core part undergo orientational relaxation with the decay time of 2.6 ps, which is the same with that of bulk water [61–66]. However, water at the interface undergoes orientational relaxation that is substantially slower than that of bulk water [61–66]. For large reverse micelles, the interfacial orientational relaxation time is independent of the size of the water nanopool. Furthermore, water at the interface of large AOT reverse micelles has the same orientational relaxation time as water at the interface of the AOT planar lamellar structures [65]: implying that the dynamics of water do not depend on the geometry of the interface.

Intermediate-size reverse micelles ($2.5 \text{ nm} < d < 6 \text{ nm}$) also exhibit two types of orientational relaxation. However, the core exhibits dynamics that are slower than those of bulk water, and the interfacial water dynamics are slower than those at the interface of large reverse micelles. It implies that the core does not consist of bulk-like water. For example, water molecules in the core and at the interface of the reverse micelle with a diameter of ~ 4 nm undergo orientational relaxation with the decay time of 4 and 26 ps, respectively [67].

For water molecules in reverse micelles whose diameter (d) is smaller than ~ 2.5 nm, orientational relaxation is always slower than that of bulk water [61, 62, 66]. Furthermore, the orientational relaxation time depends heavily on size and increases subsequently as the size of the water nanopool decreases [67]. For example, water molecules in the reverse micelles with a diameter of 2.5 and 1.7 nm undergo orientational relaxation with the decay time of 30 and 110 ps, respectively. It implies that there are so few water molecules that all of the water is substantially affected by interactions with the interface.

These experimental results show that the dynamics are very sensitive to the size of the confining structure for intermediate and small sizes. However, certain systems, namely those that are large enough that water in an interfacial layer interacts with so much water that it is effectively bulk water, display no size dependence.

3.2 Water Films Compressed to a Few Single Molecular Layers

When two materials are brought together, separated or moved with respect to one another, understanding the nanoscopic processes occurring at their interface is central to many technological problems, including adhesion, friction, wear, and lubrication [68]. Especially, water molecules confined between interfaces with nanoscopic separation are of critical importance in many phenomena, ranging from ageing in granular media [69] and swelling of layered clays [70] to molecular transports in nanolithography [71].

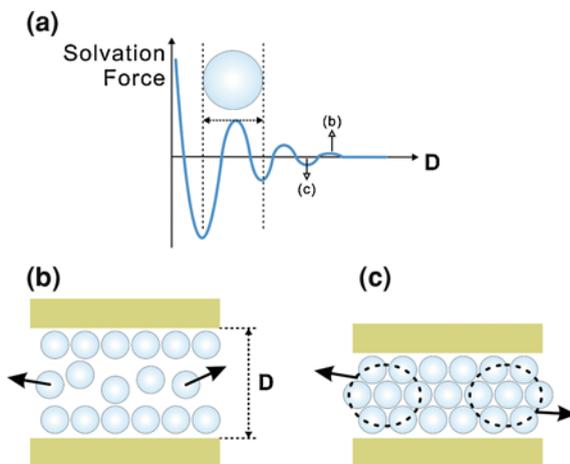


Fig. 6 Liquid molecules confined between the approaching two surfaces: (a) The short-range oscillatory solvation force (also known as the potential of mean force) between the two surfaces in a liquid varies between attraction and repulsion with a periodicity equal to some dimension of the confined liquid molecules [32]. This force is intimately related to the structure that the confined liquid is forced to adopt under confinement, as illustrated in (b) to (c); (b) The confined liquid layer is between three and four molecular layers thick and does not exhibit ordered layers. Molecules easily diffuse; (c) The liquid layer is further compressed to three molecular layers. Molecules are ordered in the vertical direction and cannot easily diffuse individually. Squeeze-out of a layer requires collective motion of many molecules

One of the most baffling aspects about confined water is whether it behaves in a special, even unique, manner or similarly to other confined fluids. A central property of liquids confined between solid boundaries which are smooth on the molecular scale as well as of films adsorbed on a solid surface is their tendency to organize into layered structures [32, 68] (see Fig. 6). For example, this tendency experimentally probed by means of surface force apparatus (SFA) techniques [31, 72] for octamethylcyclotetrasiloxane (OMCTS), one of simple liquids, as well as water.

When SFA techniques were also used to measure OMCTS in confinement, it was consistently concluded that the effective viscosity dramatically increased comparing with that of OMCTS in bulk liquid phase [73, 74]. Even though the same SFA technique were also used to measure water in confinement, very different magnitude of the effective viscosity of the confined water were concluded [75, 76]. That is, one conclusion is that the effective viscosity of the confined water is nearly same as bulk water [75], and the other is that the effective viscosity of water exhibits like that of simple liquids under nanoscopic confinement [76]. There has been a sharp conflict of claim on this issue [77–83] until the elastic and viscous response of nanoconfined water layers were precisely, carefully measured under different dynamic conditions [84]. On the basis of the recent measurement, now it can newly be concluded that the two opposing claims are originated from

two different experimental conditions for measurements such as compression rate [85]. This comprehensive understanding is very essential and important for optimizing quality of new devices at nanoscale.

4 Conclusions

Some recent understanding of interfacial water and confined water has been introduced. There has been a continuous attempt to better understand interfacial water and confined water with new techniques. The new techniques have been increased our understanding of interfacial water and confined water by probing deeper into the molecular architecture of water molecules. Sometime, however, their results have been throwing up more puzzles under debate. As happens so commonly when scientists disagree, the problem lies not in the measurements but in how they are interpreted as introduced above.

References

1. <http://www.lsbu.ac.uk/water/index2.html>
2. Franks, F. (ed.): *Water: A Comprehensive Treatise*, vol. 1. Plenum Press, New York (1975)
3. Wernet, Ph, et al.: *Science* **304**, 995 (2004)
4. Huang, C., et al.: *Proc. Natl. Acad. Sci. USA* **106**, 15214 (2009)
5. Ball, P.: *Nature* **452**, 291 (2008)
6. Lee, H.M., Suh, S.B., Lee, J.Y., Tarakeshwar, P., Kim, K.S.: *J. Chem. Phys.* **112**, 9759 (2000)
7. Guyot-Sionnest, P., Hunt, J.H., Shen, Y.R.: *Phys. Rev. Lett.* **59**, 1597 (1987)
8. Ostroverkhov, V., Waychunas, G.A., Shen, Y.R.: *Phys. Rev. Lett.* **94**, 046102 (2005)
9. Stiopkin, I.V., Jayathilake, H.D., Bordenyuk, A.N., Benderskii, A.V.: *J. Am. Chem. Soc.* **130**, 2271 (2008)
10. Sovago, M., Campen, R.K., Wurlpel, G.W., Muller, M., Bakker, H.J., Bonn, M.: *Phys. Rev. Lett.* **100**, 173901 (2008)
11. Raymond, E.A., Tarbuck, T.L., Brown, M.G., Richmond, G.L.: *J. Phys. Chem. B* **107**, 546 (2003)
12. Du, Q., Superfine, R., Freysz, E., Shen, Y.R.: *Phys. Rev. Lett.* **70**, 2313 (1993)
13. Du, Q., Freysz, E., Shen, Y.R.: *Science* **264**, 826 (1994)
14. Gragson, D.E., Richmond, G.L.: *J. Phys. Chem. B* **102**, 3847 (1998)
15. Shultz, M.J., Baldelli, S., Schnitzer, C., Simonelli, D.: *J. Phys. Chem. B* **106**, 5313 (2002)
16. Raymond, E.A., Tarbuck, T.L., Brown, M.G., Richmond, G.L.: *J. Phys. Chem. B* **107**, 546 (2003)
17. Raymond, E.A., Richmond, G.L.: *J. Phys. Chem. B* **108**, 5051 (2004)
18. Buch, V.: *J. Phys. Chem. B* **109**, 17771 (2005)
19. Walker, D.S., Hore, D.K., Richmond, G.L.: *J. Phys. Chem. B* **110**, 20451 (2006)
20. Ji, N., Ostroverkhov, V., Tian, C.S., Shen, Y.R.: *Phys. Rev. Lett.* **100**, 096102 (2008)
21. Tian, C.S., Shen, Y.R.: *J. Am. Chem. Soc.* **131**, 2790 (2009)
22. Sovago, M., Campen, R.K., Bakker, H.J., Bonn, M.: *Chem. Phys. Lett.* **470**, 7 (2009)
23. Morita, A., Hynes, J.T.: *Chem. Phys.* **258**, 371 (2000)
24. Liu, D.F., et al.: *J. Phys. Chem. B* **108**, 2252 (2004)

25. Stipokin, I.V., et al.: *Nature* **474**, 192 (2011)
26. Richardson, S.M., Richardson, J.W.: *Am. Mineral.* **67**, 69 (1982)
27. Boek, E.S., Coveney, P.V., Skipper, N.T.: *J. Am. Chem. Soc.* **117**, 12608 (1995)
28. Karaborni, S., Smit, B., Heidug, W., Urai, J., van Oort, E.: *Science* **271**, 1102 (1996)
29. Edwards, G.R., Evans, L.F., Zipper, A.F.: *Trans. Faraday Soc.* **66**, 220 (1970)
30. Caslavsky, J.L., Vedam, K.: *J. Appl. Phys.* **42**, 516 (1971)
31. Israelachvili, J.N., Adams, G.J.: *Chem. Soc. Faraday Trans 1* **74**, 975 (1978)
32. Israelachvili, J.: *Intermolecular and Surface Forces*. Academic Press, New York (1992)
33. Hu, J., Xiao, X.-D., Ogletree, D.F., Salmeron, M.: *Science* **268**, 267 (1995)
34. Fukuma, T., Ueda, Y., Yoshioka, S., Asakawa, H.: *Phys. Rev. Lett.* **104**, 016101 (2010)
35. Cheng, L., Fenter, P., Nagy, K.L., Schlegel, M.L., Sturchio, N.C.: *Phys. Rev. Lett.* **87**, 156103 (2001)
36. Park, S.-H., Sposito, G.: *Phys. Rev. Lett.* **89**, 085501 (2002)
37. Sposito, G., Prost, R.: *Chem. Rev.* **82**, 553 (1982)
38. Odellius, M., Bernasconi, M., Parrinello, M.: *Phys. Rev. Lett.* **78**, 2855 (1997)
39. Miranda, P.B., Xu, L., Shen, Y.R., Salmeron, M.: *Phys. Rev. Lett.* **81**, 5876 (1998)
40. Bergman, R., Swenson, J.: *Nature* **403**, 283 (2000)
41. Swenson, J., Bergman, R., Howells, W.: *J. Chem. Phys.* **113**, 2873 (2000)
42. Xu, K., Cao, P., Heath, J.R.: *Science* **329**, 1188 (2010)
43. Cantrell, W., Ewing, G.E.: *J. Phys. Chem. B* **105**, 5434 (2001)
44. Beaglehole, D., Radlinska, E.Z., Ninham, B.W., Christenson, H.K.: *Phys. Rev. Lett.* **66**, 2084 (1991)
45. Ball, P.: *Chem. Rev.* **108**, 74 (2008)
46. Cheung, M.S., Garcia, A.E., Onuchic, J.N.: *Proc. Natl. Acad. Sci. USA* **99**, 685 (2002)
47. Pal, S.K., Zewail, A.H.: *Chem. Rev.* **104**, 2099 (2004)
48. Fenimore, P.W., Frauenfelder, H., McMahon, B.H., Parak, F.G.: *Proc. Natl. Acad. Sci. USA* **99**, 16047 (2002)
49. Kropman, M.K., Bakker, H.J.: *Science* **291**, 2118 (2001)
50. Pal, S.K., Peon, J., Bagchi, B., Zewail, A.H.: *J. Phys. Chem. B* **106**, 12376 (2002)
51. Smith, J.D., Saykally, R.J., Geissler, P.L.: *J. Am. Chem. Soc.* **129**, 13847 (2007)
52. Wood, K., et al.: *Proc. Natl. Acad. Sci. USA* **104**, 18049 (2007)
53. Tarek, M., Tobias, D.J.: *Phys. Rev. Lett.* **89**, 275501 (2002)
54. Laage, D., Hynes, J.T.: *Proc. Natl. Am. Soc. U S A* **104**, 1167 (2007)
55. Heugen, U., et al.: *Proc. Natl. Acad. Sci. USA* **103**, 12301 (2006)
56. Zhang, C., Durbin, S.M.: *J. Phys. Chem. B* **110**, 23607 (2006)
57. Ebbinghaus, S., et al.: *Proc. Natl. Acad. Sci. USA* **104**, 20749 (2007)
58. Heydena, M., et al.: *Proc. Natl. Acad. Sci. USA* **107**, 12068 (2010)
59. De, T.K., Maitra, A.: *Adv. Colloid Interface Sci.* **59**, 95 (1995)
60. Zulauf, M., Eicke, H.-F.: *J. Phys. Chem.* **83**, 480 (1979)
61. Moilanen, D.E., Fenn, E.E., Wong, D.B., Fayer, M.D.: *J. Chem. Phys.* **131**, 014704 (2009)
62. Piletic, I.R., Moilanen, D.E., Spry, D.B., Levinger, N.E., Fayer, M.D.: *J. Phys. Chem. A* **110**, 4985 (2006)
63. Piletic, I.R., Moilanen, D.E., Levinger, N.E., Fayer, M.D.: *J. Am. Chem. Soc.* **128**, 10366 (2006)
64. Fenn, E.E., Wong, D.B., Fayer, M.D.: *Proc. Natl. Acad. Sci. USA* **106**, 15243 (2009)
65. Moilanen, D.E., Fenn, E.E., Wong, D.E., Fayer, M.D.: *J. Am. Chem. Soc.* **131**, 8318 (2009)
66. Moilanen, D.E., Fenn, E.E., Wong, D.B., Fayer, M.D.: *J. Phys. Chem. B* **113**, 8560 (2009)
67. Fayer, M.D., Levinger, N.E.: *Annu. Rev. Anal. Chem.* **3**, 89 (2010)
68. Bhushan, B., Israelachvili, J.N., Landman, U.: *Nature (London)* **374**, 607 (1995)
69. Bocquet, L., Charlaix, E., Ciliberto, S., Crassous, J.: *Nature (London)* **396**, 735 (1998)
70. Karaborni, S., Smit, B., Heidug, W., Urai, J., Oort, E.V.: *Science* **271**, 1102 (1996)
71. Piner, R.D., Zhu, J., Xu, F., Hong, S., Mirkin, C.A.: *Science* **283**, 661 (1999)

72. Christenson, H.K., Horn, R.G., Israelachvili, J.N.: *J. Colloid Interface Sci.* **88**, 79 (1982)
73. Klein, J., Kumacheva, E.: *Science* **269**, 817 (1995)
74. Demirel, A.L., Granick, S.: *Phys. Rev. Lett.* **77**, 2261 (1996)
75. Raviv, U., Lurat, P., Klein, J.: *Nature* **413**, 51 (2001)
76. Zhu, Y., Granick, S.: *Phys. Rev. Lett.* **87**, 096104 (2001)
77. Raviv, U., Klein, J.: *Science* **297**, 1540 (2002)
78. Raviv, U., Perkin, S., Lurat, P., Klein, J.: *Langmuir* **20**, 5322 (2004)
79. Antognozzi, M., Humphris, A.D.L., Miles, M.J.: *Appl. Phys. Lett.* **78**, 300 (2001)
80. Jeffery, S., Hoffmann, P.M., Pethica, J.B., Ramanujan, C., Ozer, H.O., Oral, A.: *Phys. Rev. B* **70**, 054114 (2004)
81. Li, T.D., Gao, J.P., Szoszkiewicz, R., Landman, U., Riedo, E.: *Phys. Rev. B* **75**, 115415 (2007)
82. Major, R.C., Houston, J.E., McGrath, M.J., Siepmann, J.I., Zhu, X.Y.: *Phys. Rev. Lett.* **96**, 177803 (2006)
83. Zhu, Y., Granick, S.: *Phys. Rev. Lett.* **87**, 096104 (2001)
84. Khan, S.H., Matei, G., Patil, S., Hoffmann, P.M.: *Phys. Rev. Lett.* **105**, 106101 (2010)
85. Granick, S., Bae, S.C., Kumar, S., Yu, C.: *Physics* **3**, 73 (2010)

Molecular Modeling of Biomolecules

Chaok Seok

1 Introduction

Living organisms are composed of a variety of biomolecules such as proteins, nucleic acids, polysaccharides (polymeric forms of sugar molecules), lipids, metabolites, etc. Those biomolecules interact with each other to sustain life in constantly varying environment. For example, signals from environment (such as existence of food) are transferred to inside the cell as a result of interactions of biomolecules (binding of molecules from food that cause delicious smell to receptor proteins in our nose, and consequent conformational changes of the receptor proteins leading to further changes in other signaling molecules) to generate reactions required for the particular situation (secreting enzymes for digestion). One may therefore say that life phenomenon is governed by complex interaction networks of biomolecules. In that respect, knowing specific structures and interactions of biomolecules with other biomolecules in atomic detail is of tremendous importance. Such knowledge provides understanding of the molecular mechanism underlying biological functions performed by biomolecules, and is also extremely helpful for developing small-molecule drugs that regulate the functions of proteins whose actions lead to diseases. For example, imatinib, a famous drug for chronic myelogenous leukemia, blocks the function of a tyrosine kinase called Bcr-Abl protein that causes the disease. How this drug acts can be clearly seen from the crystal structure [1], as shown in Fig. 1. In the crystal structure, imatinib tightly binds to the active site of Bcr-Abl and thus inhibits the function of the protein. There have appeared drug-resistant proteins since the release of the drug, and new drugs are required for certain patients who have

C. Seok (✉)

Department of Chemistry, Seoul National University, Seoul, Korea
e-mail: chaok@snu.ac.kr

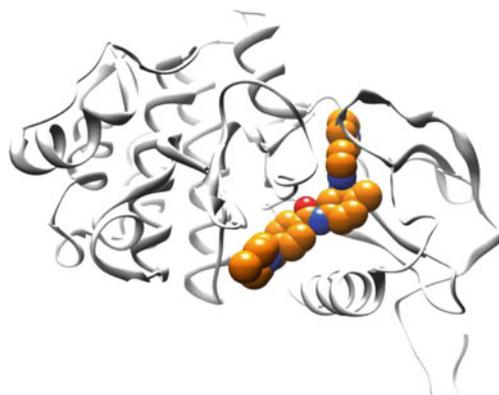


Fig. 1 Three-dimensional molecular structure of imatinib (shown as *orange spheres*) bound to the kinase domain of Bcr-Abl protein (shown as *white ribbon*) that causes chronic myelogenous leukemia [1]. Imatinib tightly binds to the active site of the protein and inhibits its activity in cancer cell. Molecular graphics images of biomolecules in this part were produced using the UCSF Chimera package [2] from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081)

developed drug resistance. Molecular structures of the drug-resistant proteins can be utilized to develop new drugs that can complement imatinib.

Biomolecular modeling, or modeling of three-dimensional structures of biomolecules, has become an indispensable tool in biomedical research. At earlier times, modeling was used just to generate nice-looking pictures. However, these days a variety of commercial and open software programs for molecular modeling has become available, facilitating application to various molecular biology and biochemistry problems. There are now numerous journal papers presenting modeling results published in the literature. Advance of techniques in biomolecular modeling are due in some part to the rapid advance of computer technology and of biotechnology. Faster computers have enabled faster computation of more complicated equations, and advancement in biotechnology lead to large amounts of biological data that can be utilized in modeling. In particular, a large amount of genome sequences and the corresponding protein sequences encoded by the genome became available. Although genome contains all the genetic information for life, how the information is realized in actual life can be understood only when we understand how actual molecules such as proteins play their roles to maintain life. Currently, only a very tiny fraction of proteins with known amino acid sequences have available experimental structures. Typically, a lot of experimental efforts are necessary to study structure and function of proteins because times-taking processes such as expression, purification, and identification of the proteins are necessary for individual proteins. On the other hand, computational modeling studies are very time- and cost-effective because molecular structures and interactions (and therefore functions) can be studied by dealing with virtual molecules without needing to obtain actual molecules.

Biomolecular modeling methods may contribute to THz studies because collective vibrational motions of biomolecules such as proteins or nucleic acids are

in the tera-hertz range. A back of envelop calculation can be done using the relation between mass m , force constant k , and the frequency f of a harmonic oscillator, $f = 1/2\pi\sqrt{k/m}$, as follows to see this. The highest vibration frequency of biomolecules is that of a chemical bond involving the lightest atom, hydrogen. Because the force constant of a chemical bond such as C–H is about $500 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and the effective mass is 1 g mol^{-1} , the frequency of a C–H vibration is roughly $10^{14} \text{ Hz} = 100 \text{ THz}$. A collective vibration mode that consists of about 100 heavy atoms (non-hydrogen atoms such as carbon) has an effective mass on the order of 1,000 times larger than hydrogen. The force constant for a collective motion may be estimated from the Lennard-Jones potential assuming that such a large motion is largely held by non-bonded interactions. A force constant of $\sim 50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ is obtained if the second derivative is calculated at the minimum of a Lennard-Jones potential with the minimum energy $0.1 \text{ kcal mol}^{-1}$ and equilibrium distance of 1 \AA . The force constant of a collective mode is therefore about 1/10 of that of a C–H bond, and the frequency is then approximately $\sqrt{1/10 \cdot 1000}10^{14} \text{ Hz} = 10^{12} \text{ Hz} = 1 \text{ THz}$. Such a collective, low-frequency motion involving a large number of atoms are often related to biological function of the biomolecules. It can be easily expected that such low-frequency modes tend to be more robust and reliable and less affected by thermal fluctuations due to collisions with other molecules in the cell and thus more proper for executing biological roles.

Therefore, precise calculation of the low-frequency modes of biomolecules may help explaining results from THz experiments. Vibrational frequencies may easily be calculated using harmonic approximation if the ‘potential energy function’ and the ‘equilibrium structure’ are known, by applying a process called normal mode analysis. If THz experiments are to be performed to study changes in molecular states, knowledge on ‘conformational changes’ is also required. In this part, methods for biomolecular modeling regarding these aspects are explained. First, types of biomolecules and their structures are first introduced with particular emphasis on proteins, and then the potential energy function that describes the biomolecular interactions in molecular mechanics calculations called ‘force field’ is explained. Next, molecular dynamics simulation methods that can be used to obtain an ensemble of structures that are thermally accessible are explained, and protein structure prediction methods that can be used to predict the native structures when experimental structures are not available are introduced. Finally, normal mode analysis method used for calculating vibration modes in the harmonic approximation is explained.

2 Biomolecules and Protein Structure

A large part of our body consists of high-mass polymers, such as proteins and nucleic acids. There are also non-polymeric, relatively small organic molecules such as glucose, amino acids, or vitamins. Polymers are formed by linking small

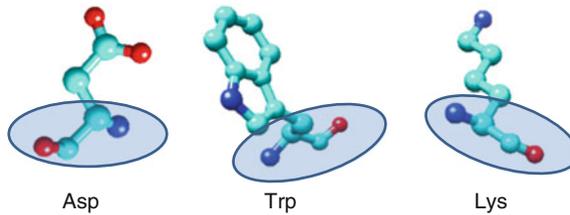
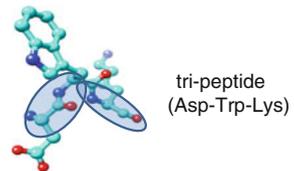


Fig. 2 Three out of twenty amino acids are illustrated here. Carbon atoms are represented by *cyan spheres*, nitrogen in *blue*, and oxygen in *red*. Hydrogen atoms are omitted for clearer view. The backbone areas are indicated with large *blue ovals*. The side-chain of Asp (aspartate) is negatively charged, that of Trp (tryptophan) is neutral, and that of Lys (lysine) is positively charged in the neutral condition. The side-chain of Trp is aromatic (contains benzene ring) and is rather non-polar

Fig. 3 Structure of a tri-peptide of sequence Asp-Trp-Lys



monomer molecules with chemical bonds. For example, proteins are formed by linking amino acid molecules, and nucleic acids by linking nucleotides. Lipid molecules form cellular membrane that covers every cell by assembling into spherical bi-layers without being linked by chemical bonds. Hydrophobic interactions between the tails of lipid molecules and hydrophilic interactions between the heads of lipids and water are the main driving force of such self-assembly.

I will concentrate on protein structures in this part. There are 20 kinds of amino acids of different sizes, shapes, and chemical properties. All amino acids within proteins have the chemical structure of $-\text{NH}-\text{CHR}-\text{CO}-$, where R is called side-chain and can be different kinds of chemical groups. The $-\text{NH}-\text{CH}-\text{CO}-$ part forms the backbone of protein, where H is attached to N, H and R to the central C, and O to the terminal C. Three different amino acids are shown in Fig. 2 to illustrate the chemical diversity. Some amino acids are polar (and some are charged) and hydrophilic (likes water), and others are non-polar and hydrophobic (hates water). Some amino acids are small, and others are big. The 20 amino acids forms the lego block that builds numerous kinds of proteins (tens of thousands in human) whose action leads to the life phenomena.

When the three amino acids in Fig. 2 are linked together, they form a tri-peptide of amino acid sequence Asp-Trp-Lys, as shown in Fig. 3. A larger polymer, called poly-peptide in general, is drawn in Fig. 4. The size of a typical protein is a few hundred amino acids long, but here we show a rather short polypeptide for the purpose of illustration.

The three-dimensional structures of proteins have evolved to make specific interactions with other molecules and to perform specific functions. The structures

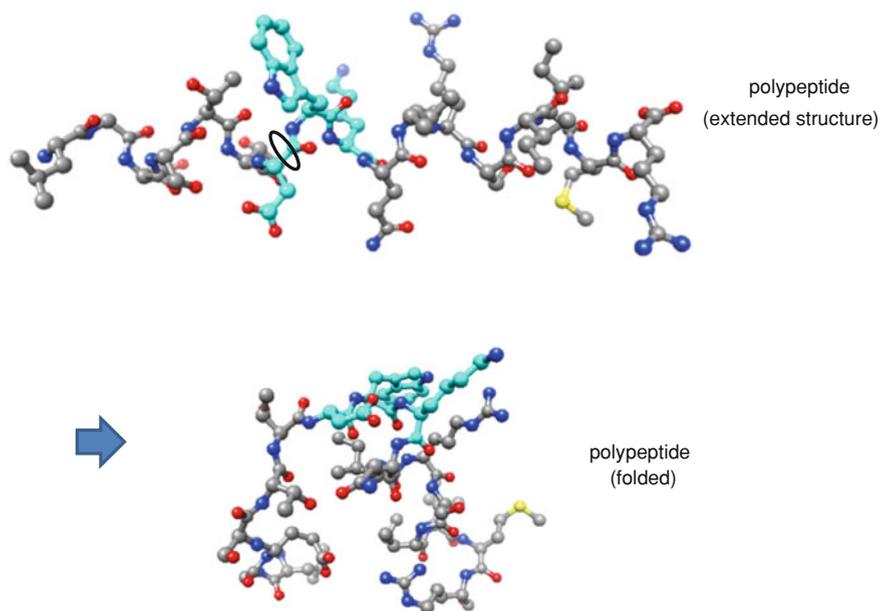


Fig. 4 An extended structure and a folded structure of a polypeptide. The color codes are the same as in Fig. 2, except that carbon atoms are either represented by *cyan* or *grey* and sulfur in *yellow*. The tri-peptide in Fig. 3 is shown in *cyan*. The overall three dimensional structure is determined by the backbone torsion angles, rotation angles about the backbone bonds at fixed bond lengths and bond angles

are largely described by the rotation angles about the backbone bonds (torsion angles). Lengths of chemical bonds and bond angles formed by two consecutive chemical bonds vary very little in general, so the backbone torsion angles are the main degrees of freedom that determine the overall structure. There are two such bonds per amino acids, rotations about N–C and C–C bonds. Rotation about the terminal C of an amino acid and N of the next amino acid is rather fixed because the bond has some double-bond character due to chemical conjugation effect. Therefore, the total number of backbone torsion angles is the number of amino acids times 2. One can imagine that the possible number of structures (conformations) obtained by rotations of all the rotatable backbone bonds would be huge. If three different rotation states are allowed for each amino acid, the number of possible conformations for a protein of 100 amino acids is 3^{100} ! The side-chain parts also have torsion angles, but the effect of such angles to the overall protein structure is rather small. However, detailed interactions between side-chain atoms are important to stabilize particular protein structures and to induce interactions of the protein with other biomolecules. Still, once backbone structures are accurate, it is relatively easy to model side-chain structures accurately.

Protein structures are graphically represented in various ways, and a few examples are shown in Fig. 5. The ball-and-stick representations in the above



Fig. 5 Different representation of protein structure, illustrated for the polypeptide in Fig. 4. Ball-and-stick model for backbone on the *left*, *ribbon model* in the *middle*, and surface representation on the *right* are shown. The tri-peptide of Fig. 3 is colored in *cyan*

figures are already complicated even for small polypeptides, so simpler representations are popular. For example, only backbone may be shown omitting side chains as ball-and-stick models or as ribbon models. When structures are shown in ribbon representation, regular structures such as helices or strands are drawn thicker. Alpha-helices are drawn as helical coils, and beta-strands as arrows. Such regular structures, called secondary structures, are made possible by regular hydrogen bonds between backbone atoms (between NH and CO). Surface representations are also used to show the shapes (and sometimes charges, too) on the molecular surface by using the actual size of each atom. Functional protein structures are typically space-filled, as shown in Fig. 5.

3 Biomolecular Force Field

Modeling of biomolecules requires an energy function that is necessary for measuring stability of various possible structures. The potential energy of a given structure can be calculated most accurately by quantum mechanical methods. However, *ab initio* quantum chemical methods can deal with only small molecules (for example, small peptides that consist of only a few amino acids) with reasonable computational efforts. The computational cost of *ab initio* quantum chemical calculations grows as n^4 with the number of electrons n . Therefore, such methods are not adequate for applications to modeling of biomolecular structures and interactions. Faster quantum chemical methods called semi-empirical methods have been developed to circumvent such difficulties, and computation time of some methods grows almost with n . However, they are still too expensive if a large amount of search in the conformational space is required. For example, more than millions of potential energy evaluations are necessary to obtain convergence in the average thermodynamic quantities even for relatively small systems. Force field is an empirical energy function and is widely used for biomolecular modeling and simulations. Such calculations are called molecular mechanics calculations in contrast to quantum mechanical calculations, and force field is also called molecular mechanics force

field. Energy evaluation is extremely fast for force field because it is enough to calculate the functional value for a given structure. In quantum mechanical calculations, the energy value has to be calculated by complicated iterative methods at each point in the conformational space (for each structure). The force field can be used as the potential energy function to obtain a time trajectory of a biomolecular system by finite difference integration of the Newton's second law. It is called molecular dynamics simulation and is described in more detail in the next section. In this section, I will describe how each energy component of force field is determined, and how they are parameterized.

Typical biomolecular force field has the following functional form:

$$U = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{imp}} + E_{\text{elec}} + E_{\text{vdw}} \quad (1)$$

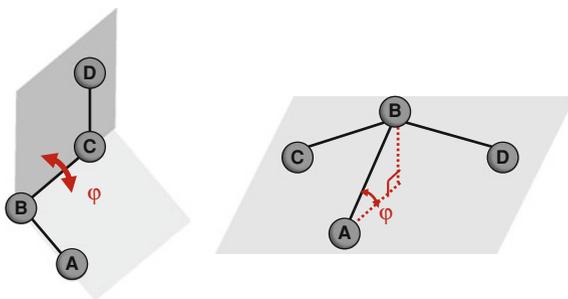
The first two energy components are bond stretching and angle bending energy expressed as sum of harmonic functions as

$$E_{\text{bond}} = \sum_{\text{bonds}} \frac{k_{bi}}{2} (l_i - l_{i,\text{eq}})^2,$$
$$E_{\text{angle}} = \sum_{\text{angles}} \frac{k_{\theta i}}{2} (\theta_i - \theta_{i,\text{eq}})^2$$

where l_i is bond length of the i th bond in the current structure, $l_{i,\text{eq}}$ is the equilibrium bond length of the i th bond, k_{bi} is the bond stretching force constant, θ_i is angle of the i th bond angle, $\theta_{i,\text{eq}}$ is the corresponding equilibrium angle, and $k_{\theta i}$ is the angle bending force constant. Contribution from each bond (or angle) is summed over all bonds (or angles). The parameters $l_{i,\text{eq}}$, k_{bi} , $\theta_{i,\text{eq}}$, and $k_{\theta i}$ are determined for each of pre-determined atom types. Atom types are determined based on hybridization of the involving atoms and the atom types of the directly bonded atoms. For example, a chemical bond between two sp^3 carbons may have the same parameters as other bonds formed by two sp^3 carbons that have different neighboring atoms bonded to the carbons, and the parameters may be different for bonds between sp^2 carbon and sp^3 oxygen. There is no best rule known for determining atom types in energy parameterization. If atom types are set too roughly, chemical differences may not be captured very well. If atom types are too specific and there are too many atom types, the parameters may be over-trained to the training data set, and may not be transferable when applied to other molecules. Usually biomolecular force fields have been developed to have smaller number of atom types than the force fields for small organic molecules. The parameters for bond stretching and angle bending are typically obtained from gas-phase spectroscopy and crystal structures for small organic molecules that can cover the atom types considered in the force field.

The third and the fourth energy terms in Eq. (1) are contributions from rotations of torsion angles and improper torsion angles,

Fig. 6 Definition of proper torsion angle (*left*) and improper torsion angle (*right*)



$$E_{\text{torsion}} = \sum_{\text{torsions}} \sum_n \frac{V_{n,i}}{2} [1 + \cos(n\phi_i - \phi_{i,0})],$$

$$E_{\text{imp}} = \sum_{\text{imptor}} \frac{k_{\phi i}}{2} (\phi_i - \phi_{i,\text{eq}})^2,$$

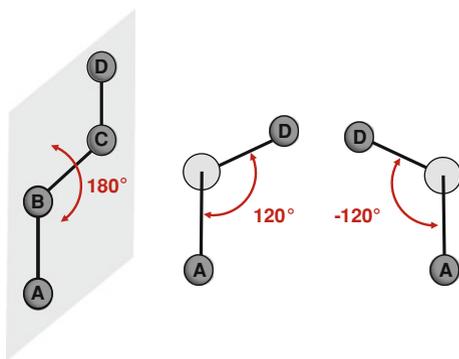
where torsion angle energy is expressed as a Fourier (or cosine) series containing several n terms, and improper torsion angle energy is expressed as harmonic function for convenience. Definition of torsion angle and improper torsion angle in terms of four connected atoms are provided in Fig. 6. Torsion angle (also called proper torsion angles or dihedral angles) formed by four atoms A, B, C, and D is the angle between the two planes formed by ABC and BCD. Improper torsion angle represents the deviation of the central atom B from planarity.

Fourier series is a natural choice for torsion angle energy because of the periodicity of the torsion angle. Bond angles are also periodic, but they do not deviate much from the equilibrium value, so harmonic function is enough. Torsion angles have much lower energy barrier and can cover the full range of rotation angle. For example, rotation about a bond between two sp^3 carbon atoms results in three local minima; 180° (corresponds to the *trans* conformation) and $\pm 120^\circ$ (correspond to the two *gauche* conformations) (see Fig. 7). The highest n value must be at least 3 for this torsion angle. The parameters $V_{n,i}$ and $\phi_{i,0}$ are usually determined by fitting to gas phase quantum chemical calculations for small molecules.

Improper torsion angles were introduced to enforce planarity of four atoms such as those in peptide bonds or in aromatic rings. Although a planar structure can be obtained if the sum of the three bond angles CBA, ABD, CBD in Fig. 6 is 360° , these bond angles can easily deviate from the equilibrium values depending on structures, and even small deviations in the angles can produce rather large deviations from planarity. The improper torsion energy term is included to impose planarity in a simple way. The $\phi_{i,\text{eq}}$ parameter is thus determined by requirement on the geometry, and the force constant $k_{\phi i}$ is determined rather empirically.

The first four energy terms in Eq. (1) are called covalent energy or bonded energy because they originate from chemical bonding. The last two terms in

Fig. 7 Trans (*left*) and gauche structures (*middle and right*). To see the gauche structures clearly, the bond vector from atom B to atom C is positioned perpendicular to the plane of this page



Eq. (1) are called non-covalent energy or non-bonded energy. Non-bonded energy terms are applied between all pairs of atoms i and j that are 3 or more chemical bonds apart. Those atom pairs that are exactly 3 bonds apart are called 1–4 pairs (See Fig. 8), and the interaction energy between them are typically scaled down compared to other non-bonded pairs because their interactions are also considered by the torsion energy term. The two non-bonded interaction terms in Eq. (1) are van der Waals energy described by Lennard-Jones potential and Coulomb interaction between point charges.

$$E_{\text{vdw}} = \sum_{i,j} \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

$$E_{\text{elec}} = \sum_{i,j} \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}}$$

where r_{ij} is distance between atom i and atom j , ε_{ij} and σ_{ij} are the minimum energy and the distance at minimum for Lennard-Jones interaction between atoms i and j , q_i is partial charge assigned to atom i , and ε_0 is the permittivity of vacuum. There are attractive interactions even between non-polar atoms when they are close enough, and such attraction arises from the interactions between dipoles induced due to instantaneous electron correlation. Such interaction, also called dispersion interaction, decays as $\sim 1/r^6$ at long distance, which forms the attractive term of Lennard-Jones potential. Repulsive part of Lennard-Jones potential is set to be proportional to $1/r^{12}$, because of computational efficiency (only a single multiplication is need to compute $1/r^{12}$ from $1/r^6$). There is no theoretical basis for the repulsive part, unlike the attractive part. In fact, repulsion at short distance is expected to exponentially decay with distance from quantum mechanics arguments. The repulsive part of Lennard-Jones potential is too much steep in real practice and sometimes causes numerical problems in molecular mechanics calculations when structures have overlaps between atoms. The Lennard-Jones parameters are obtained from crystal data of small molecules and properties

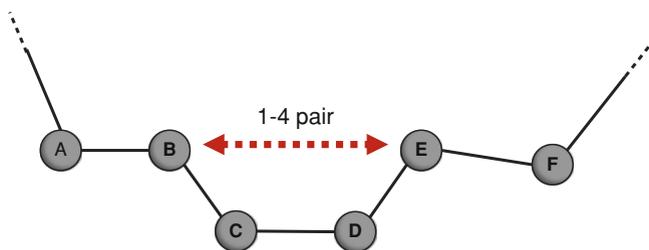


Fig. 8 1–4 interaction pair for non-bonded interactions

obtained from liquid state simulations such as density or heat of vaporization for model systems.

The Coulomb potential between partial charges is used in typical force field. The charges are referred to be ‘partial’ because the magnitudes of the charges are less than unit electronic charge. Such force fields are called ‘fixed-charge force field’ in contrast to ‘polarizable force field’. In a fixed-charge force field, charges are fixed throughout calculations and do not change with structure. In fact, partial charges are not observable physical quantities, so different methods have been developed to assign the partial charge parameters. More sound methods among them consider electronic distributions around different atoms in molecules. The electronic distribution can change depending on local environment, for example, more electronic polarization occurs if highly charged atoms are around. However, polarizable force fields that adopt fluctuating charges or polarizable point dipoles have not yet been widely used in this field because the efficiency and accuracy of the models are not satisfactory yet.

One of the most popular methods for obtaining the partial charges is to fit the electrostatic potential values calculated at grid points around the molecule from the partial charges

$$\phi_{\text{FF}}(\mathbf{r}) = \sum_i \frac{q_i}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}|}$$

to those obtained from the electron probability distribution $\rho(\mathbf{r}')$ which in turn calculated from quantum chemistry

$$\phi_{\text{ESP}}(\mathbf{r}) = \sum_a^{N_{\text{nuc}}} \frac{Z_a e}{4\pi\epsilon_0|\mathbf{R}_a - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')e}{4\pi\epsilon_0|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}',$$

where Z_a is the atomic number, e is the electronic charge, and \mathbf{R}_a is the nuclear coordinate of atom a .

Accurate description of electrostatic interaction of a biomolecular system in the solution phase is not possible without consideration of solvent. Even non-polar solvent can be polarized and show dielectric response to give effective dielectric constant (ratio of solvent permittivity to vacuum permittivity) of around 2. Highly

polar solvent like water has dielectric constant of about 80. In fact, proteins are polarizable also, so dielectric constant of between 2 and 4 instead of 1 are used in certain calculations.

Solvent molecules and salt ions can also be considered explicitly using force field models of solvent. There are a number of different water models available, for example. Explicit water calculations take a lot of computation because the number of interaction points increases dramatically. The number of atoms from water is much larger than that from the macromolecule of interest in a typical simulation system. Moreover, an equilibrium ensemble of water molecule conformations has to be sampled to obtain proper thermodynamic averages. This means a huge amount of additional sampling because of water. For the purpose of fast modeling, various implicit solvation models have been developed such as Poisson-Boltzmann or Generalized Born models. Since these models are based on continuum dielectric approximation, they do not account for molecular nature of water molecules. More detailed integral equation theories have thus been developed. In integral equation theories, solvent is assumed equilibrated at each solute conformation. Such model may produce rather unrealistic dynamics of instant adaptation of solvent molecules to every change of solute conformation in dynamics calculations. When affordable, it is still the best to include solvent molecules explicitly.

4 Molecular Dynamics Simulation

Probably the most popular application of force field is molecular dynamics simulation. Molecular dynamics (MD) simulation is performed by numerically integrating the Newton's second law,

$$-\nabla_i U(\mathbf{r}) = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2},$$

for given initial position $\mathbf{r}_i(t=0)$ and momentum $\mathbf{p}_i(t=0) = m_i \mathbf{v}_i(t=0)$ of each atom i . There are several different numerical methods called integrators, and one of the simplest is the Verlet integrator. The integrator is derived as follows. Taylor expansions of the position at time $t - \delta t$ and $t + \delta t$ are written down first as

$$\begin{aligned} \mathbf{r}(t - \delta t) &= \mathbf{r}(t) - \mathbf{v}(t)\delta t + 1/2\mathbf{a}(t)\delta t^2 + \dots \\ \mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\delta t + 1/2\mathbf{a}(t)\delta t^2 + \dots \end{aligned}$$

and the two equations are added to give

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \mathbf{a}(t)\delta t^2,$$

where higher order terms are truncated, atom index i is omitted, and the acceleration is $\mathbf{a}(t) = -\nabla U(\mathbf{r})/m$. In the above equation, velocity does not appear explicitly. It can be calculated by

$$\mathbf{v}(t) = [\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)]/2\delta t$$

The time step δt is typically taken to be 1 fs (fs = 10^{-15} s) which is approximately one tenth of the shortest period of bond stretching vibration involving the lightest hydrogen atom. A longer, 2 fs time step may be used if bonds involving hydrogen atoms are fixed by methods called SHAKE or RATTLE. There are other integrators called leap frog, velocity Verlet, predictor–corrector, etc. An integrator must be time-reversible and symplectic (does not effectively distort or deform the phase space). Practically, conserved quantities such as total energy must be stable over many integration steps.

Applying the above integrator results in a constant-energy trajectory, but constant-temperature and/or constant-pressure simulations are usually desired because biomolecular systems are under such condition in real situation. Methods called thermostats are used to keep temperature constant, and barostats are used to keep pressure constant. A thermostat must satisfy the relation of the average kinetic energy K and the bath temperature, $T_{\text{bath}} = \left\langle \frac{2}{3Nk_B} K \right\rangle$, where the bracket represents average over simulation trajectory. A simple velocity rescaling method that multiplies velocity by $\sqrt{T_{\text{bath}}/T(t)}$ at each time step, where the instantaneous temperature $T(t)$ is defined by $T(t) = \frac{2}{3Nk_B} K = \frac{2}{3Nk_B} \sum_{i=1}^N \frac{1}{2} m_i v_i^2(t)$, gives correct average $\langle T(t) \rangle = T_{\text{bath}}$. However, this is very artificial because velocity varies abruptly at every step. A weaker coupling to the heat bath can be generated by Berendsen thermostat in which the instant temperature relaxes exponentially by $\frac{dT(t)}{dt} = \frac{1}{\tau} [T_{\text{bath}} - T(t)]$. More desirable thermostats must satisfy the correct fluctuations in the instantaneous temperature $\frac{\sigma_T^2}{\langle T \rangle^2} = \frac{\langle T^2 \rangle - \langle T \rangle^2}{\langle T \rangle^2} = \frac{2}{3N}$, as well as the correct average. Langevin thermostat that solves the following dynamics equation satisfies the correct fluctuations:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -\nabla_i U - m_i \Gamma \frac{d\mathbf{r}_i}{dt} + \mathbf{W}_i(t),$$

where at each time step, all particles receive a random force $\mathbf{W}_i(t)$ and have their velocities lowered due to constant friction Γ . The random force is uncorrelated in time as $\langle \mathbf{W}_i(t) \cdot \mathbf{W}_j(t') \rangle = \delta_{ij} \delta(t - t') 6mk_B T \Gamma$. A good barostat also has to satisfy the correct fluctuations $\frac{\sigma_V^2}{\langle V \rangle^2} = \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V \rangle^2} = k_B T \kappa$ as well as the correct average. There are several different barostats available such as Berendsen, Anderson, etc.

I will further discuss issues related to system preparation that has to be considered for successful application of MD simulations. First, one has to prepare the initial configuration $\{\mathbf{r}_i\}$, and it is usually the best if the initial configuration is closest to the desired state. For example, if an experimental structure of a protein is known, one may start from the experimental structure, relax the structure by energy minimization, run MD to obtain a simulation trajectory, and then calculate

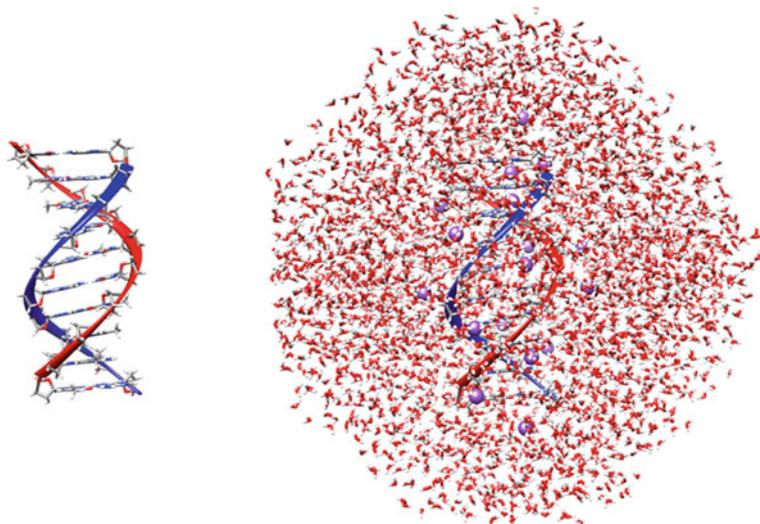


Fig. 9 DNA (shown as a double helix) solvated by water (shown as v-shaped sticks colored in red and white) in a truncated octahedron simulation box. Counter-ions (purple spheres) are added to neutralize the system, and pre-equilibrated water molecules are added to solvate the DNA molecule. Further equilibration is necessary before running production simulation

thermodynamic and kinetic properties from the trajectory. Initial velocities are randomly assigned to give the Maxwell–Boltzmann distribution. Because different random numbers can be generated each time with different ‘seed’ numbers, different trajectories can be obtained even for the same initial configuration by re-assigning initial velocities. If solvent molecules are added, it is important that they are equilibrated well. One may fix the large solute such as protein at the initial stage and energy-minimize and run short MD first before relaxing the whole system. An example of initial configuration is shown in Fig. 9 for a water-solvated DNA. If one wants to see large conformational changes, very long MD simulations must be carried out, or special sampling methods that enable high-energy barrier crossing have to be used. It is also necessary to monitor variables (such as potential energy or conformation change) to confirm that the system reaches equilibrium or at least a stable state. To obtain thermodynamic properties, only a part of trajectory that can be considered to be in equilibrium must be used. The initial part of the simulation that undergoes significant changes is called equilibrium phase, and the later part of stabilized trajectory is called production phase.

Because the simulation box is not of macroscopic size, surface effect can be of a problem. For example, most water molecules in Fig. 9 can be considered to be part of a surface layer facing vacuum. Properties of such a surface are very different from those of bulk water. To avoid the surface effect, one may either simulate a large, but finite system, or simulate an infinite system. When simulating a finite system, the biggest droplet one can afford has to be used. In such cases, restraining

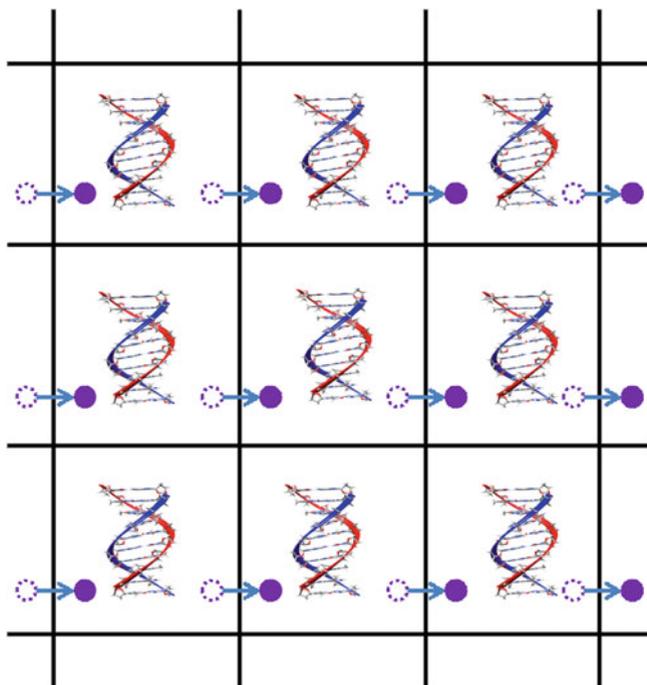


Fig. 10 A schematic picture showing periodic boundary condition for a system of a DNA molecule. A particle that leaves a box (*purple filled circle*) enters the box from the other side (*dotted circle*)

potential is included to keep particles from escaping into vacuum. Simulation of an infinite system is performed in an infinitely replicated lattice of identical systems as illustrated schematically in Fig. 10. Such system is referred to have periodic boundary condition. The geometry of the simulation box may be cubic, rectangular, parallelepiped, etc. A truncated octahedron box is also popular because it is more spherical, so the simulation volume and the number of solvent molecules can be minimized. The periodic boundary condition requires toroidal geometry; when a particle leaves a simulation box, the same particle enters the box from the other side of the box (see Fig. 10). For short-range interactions such as Lennard-Jones potential, the potential energy is not calculated for all atom pairs in the infinite space, but truncated at some distance. In the minimum image convention, each particle in the central cell interacts with only the nearest periodic copy of other particles. Or, potential is cut-off or truncated by incorporating shifting function or switching function that goes smoothly to zero at desired cut-off distance. For longer-range Coulomb interactions, all the interactions of the infinite periodic system are explicitly added up using a method called Ewald summation. An essence of the method is to split the conditionally convergent series into two rapidly convergent series, direct space sum and reciprocal space sum, through the

use of a screening charge. PME (Particle-Mesh Ewald) method is a variant to speed up such calculation.

Periodic boundary condition enables simulation in a bulk-like environment, but it is still very artificial. Periodic boundary conditions may affect some physical properties. For example, phases that are commensurate with the box are stabilized, and phases that are incommensurate such as crystal forms are destabilized. Long-wavelength phenomena such as capillary waves are limited to wavelengths shorter than the longest dimension of the box. One has to be careful to avoid interactions among periodic images and not to affect the system properties by making sure that the effective range of interactions should be shorter than half of the box length. For simulations of an isolated solute, the solute as well as the solvation shell should never be able to interact with itself, and it is ideal to use as large box as possible.

Once a simulation trajectory is obtained, thermodynamic and kinetic properties can be calculated. Simple averages such as energy, pressure are easily calculated by a direct average

$$\langle A \rangle = \frac{1}{M} \sum_{j=1}^M A_j,$$

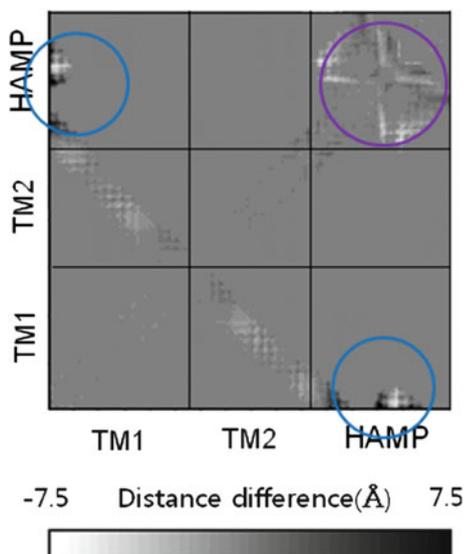
where the summation is over the snapshots of the trajectory, and A_j is the property calculated at the j th snapshot. Typically, not all snapshots are saved. Energy and structure are saved only periodically during simulation because neighboring snapshots are highly correlated. For kinetic properties, characteristic time τ for a process may be obtained by calculating autocorrelation functions or first passage time. One must be careful with thermostat because thermostat can damp activated processes and/or result in nonphysical energy exchange if $\tau \gg \tau_{\text{bath}}$. One may also build up transition matrix models of Markov processes from MD simulations to study longer time scale kinetics.

For biomolecular simulations, structural aspects are of most interest. One may calculate average structure, but average structure is often chemically unrealistic, so one has to be careful about interpreting it. One may also calculate average or fluctuations of distances between particular atoms or torsion angles of particular interest. Comparison to a reference structure is often made to evaluate complex conformational changes by a single measure called RMSD (Root Mean Square Deviation). RMSD is calculated as

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_i^{\text{ref}})^2}$$

after superimposing the structure with the reference structure by the best translation and rotation that minimize the deviation. In addition, distance matrix with ij elements $\Delta r_{ij} - \Delta r_{ij}^{\text{ref}}$, where $\Delta r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, is also frequently used. This measure is convenient to use since no structure fitting is required. It can also show changes

Fig. 11 An example of the use of distance matrix. Difference of distance matrices obtained from simulations of two different states of a protein is shown here. Each axis represents amino acid residue number of the protein. The protein is divided into three subunits called *TM1*, *TM2*, and *HAMP*. It can be easily seen where the most pronounced conformational changes occur, and such regions are indicated with circles

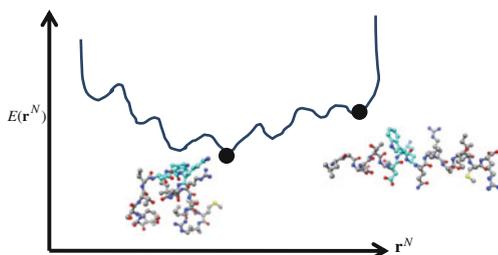


in specific regions in the structure and easy to visualize as a matrix form, as illustrated in Fig. 11.

5 Protein Structure Prediction

Molecular dynamics simulation of biomolecules is the most straightforward when the native structures of the biomolecules known from experiment. Dynamics and energetics around the native state can be investigated reliably. However, in general molecular dynamics simulation fails miserably in predicting the native structure when an initial structure with significant deviation from the native structure is used. It is usually hopeless to start with an arbitrary conformation and to expect to observe phenomena related to the native state. There are two reasons for this; energy and sampling. The energy function, force field here, is not accurate enough, and thus the native structure may not correspond to the free energy minimum obtained from the force field. Also, current molecular dynamics simulations are not efficient and/or long enough, so the free energy minima are hard to be reached. It is hard to prove this clearly though because both energy and sampling are problematic, but many experimental simulations have been indicating that they are both problematic. There still have been reports on successful results for a small number of specific biomolecular systems, but is far from general yet. Therefore, it is the best if the biomolecule of interest has available experimental structure. Even when experimental structure is not available, there may exist experimental structures of related molecules (for example, proteins with similar amino acid

Fig. 12 Ab initio protein structure prediction requires a good free energy function that places a native-like structure at the global minimum and a powerful global optimization method that searches for the global minimum efficiently



sequences), and such structures can be used as templates for structure prediction. Protein modeling method that use such structural templates for structure prediction is called template-based modeling, or more commonly, homology modeling. When there are no structures of related proteins available, protein structures must be predicted ab initio. Such modeling methods are called ab initio modeling or free modeling.

Ab initio protein structure prediction problem is notorious for its difficulty, but it will be of tremendous impact if noticeable improvement in prediction accuracy is possible. In typical ab initio protein structure prediction methods, a free energy model is designed and the conformational space is searched to find the structure corresponding to the lowest free energy state by a technique called global optimization, as illustrated in Fig. 12. As discussed above, current force field models are not satisfactory for such purposes, and many knowledge-based methods that use information from available structure and sequence databases are being currently employed extensively.

Homology modeling typically follows three steps; template selection, sequence alignment, and model building. Template selection requires searching the protein structure database for evolutionally related, homologous proteins that have similar sequences to the target protein. Alignment of target protein sequence to template sequences to find sequence correspondences between related proteins is also a difficult problem, and many bioinformatics techniques are being used. Pairwise sequence alignment generates an alignment between two sequences, and it can be solved exactly and efficiently given a scoring function for each correspondence between amino acids by using a method called dynamic programming. An example is shown in Fig. 13, where the sequence of an adrenergic receptor (ADRB2_1) and that of a dopamine receptor (DRD3_1) are compared. Multiple sequence alignment generates a simultaneous alignment of multiple sequences as in Fig. 13, but finding the exact solution for a given scoring function takes a tremendous amount of computation time (proportional to the product of the lengths of sequences). Moreover, it is also hard to find a good scoring function for multiple sequence alignment. Multiple sequence alignment can be useful for protein structure prediction because different parts (that aligns well with the target) from different template structures can be combined together to build better models than using a single template structure. Many heuristic methods therefore have been developed for multiple sequence alignment.

ADRB2_1	ASIE T LCVIAVD R YFAITSPFKY Q SLLT K NK
DRD3_1	ASIL N LCAISIDRYTAVVMPVHY Q HGTGQSS

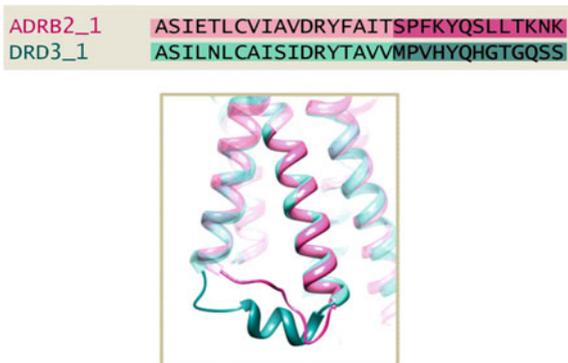
ADRB1	ASIE T LCVIALD R YLAITSPFRY Q SLLT---
ADRB2_1	ASIE T LCVIAVD R YFAITSPFKY Q SLLT---
ADRB3	ASIE T LCALAVD R YLAVTNPLRYGALVT---
DRD1	ASIL N LCVISVD R YWAISSPFRYERKMT---
DRD2	ASIL N LCAISIDRYTAVAMP L Y N TRYS--S
DRD3_1	ASIL N LCAISIDRYTAVVMPVHY Q HGTGQSS
DRD4	ASIF N LCAISVDRFVAVAVPLRY N RQGG---

Fig. 13 Example of sequence alignments. Pairwise sequence alignment above and multiple sequence alignment below. Each letter represents amino acid. Gaps represented as dashes are amino acid deletions (or insertions from the view point of the aligned amino acids). T-COFFEE [3] was used to generate the alignment. Sequence comparison of homologous proteins by such alignment can reveal relation between sequence and function. Because function is intimately related with structure, good sequence alignment is essential for structure prediction

Multiple sequence alignment is also used in many biological studies because it can give a lot of information related to function. Amino acids are mutated over time during protein evolution, and sequence alignment can reveal how protein sequences have evolved to maintain protein functions or to acquire new functions. For example, how important each amino acid is for the function of particular protein can be answered by comparison of sequences of related proteins with similar functions. In the pair alignment in Fig. 13, One may see that A (alanine) in the first column is related to the common aspects of the two receptors as receptors of neurotransmitters. The T (threonine) and N (asparagines) amino acid residues at the fifth site may be thought to be related to the specific functions of the two proteins that binds to different molecules, adrenaline and dopamine. Such relations are seen more clearly when sequences of other adrenergic receptors and dopamine receptors are also aligned together. This can be done even when experimental structures are not available. A at the first column is indeed conserved throughout all receptors, and T at the fifth column are conserved among adrenergic receptors (whose codes starting with ADRB in Fig. 13), and N are conserved among dopamine receptors (whose codes starting with DRD). It is also interesting to see that the colored Q (glutamine) residue that appear later in the sequence in Fig. 13 is common between the two receptors ADRB2_1 and DRD3_1, but they are not well conserved among the two types of receptors in multiple sequence alignment. It is therefore said that “Two homologous sequences whisper, a multiple alignment shouts loudly”, emphasizing that more information can be obtained by multiple alignment of a large number of related proteins.

Since the two sequences are similar, they have similar structures, as shown in Fig. 14. However, in less conserved region colored darker, the structures are a little bit different. This type of less conserved region may be problematic in homology modeling because the model structure constructed using the template information is less reliable. Modeling methods for dealing with such cases are also

Fig. 14 Sequence alignment and corresponding structure for the two proteins with codes ADRB2_1 and DRD3_1. The colors in the structures follow those of the corresponding regions of the sequences



available. For example, protein loop modeling methods can be used to model unreliable connector regions. Loop modeling is a small *ab initio* prediction problem and is much more useful than the full *ab initio* protein structure prediction because of smaller conformational space to search.

In conclusion, homology modeling is accurate enough if good templates are available. Homology model structures can be used for drug design or functional study in such cases. *Ab initio* modeling has found much less applications at the present level of technology so far.

6 Normal Mode Analysis

Normal mode analysis provides an alternative to molecular dynamics simulations for the study of large-magnitude motions of macromolecules. The time scale accessible to theoretical work is extended with normal mode analysis, and this approach has been proven extremely useful for studying collective motions of large biological systems. Low-frequency, slow modes are often related to conformational changes which are directly relevant to biological functions, as mentioned in introduction. For example, in Figs. 15 and 16, low-frequency vibration motion of a HIV protease monomer and a dimer protein are illustrated. Such motions are related to the peptide cleavage function performed by the protein. Such collective motions involving a cooperative movement of a large number of atoms are hard to observe in MD simulations of typical lengths (<100 ns), but have been accessible with longer simulations of $\mu\text{s} \sim \text{ms}$ range.

Normal mode analysis is performed on a system at a local minimum, assuming that the potential is harmonic:

$$U(x_1, x_2, \dots, x_{3N}) = U(x_1^0, x_2^0, \dots, x_{3N}^0) + \frac{1}{2} \sum_i \sum_j \left. \frac{\partial^2 U}{\partial x_i \partial x_j} \right|_{x=x^0} (x_i - x_i^0)(x_j - x_j^0),$$

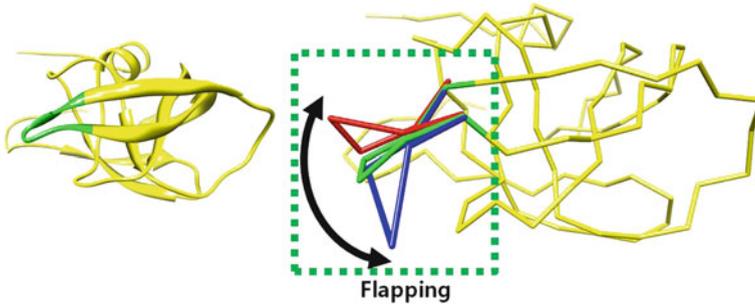
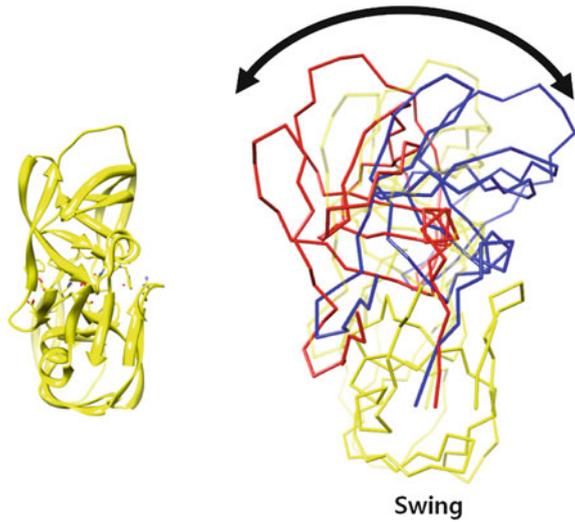


Fig. 15 A low-frequency mode of HIV protease monomer [4] calculated with anisotropic network model [5]. Flapping motion of the beta-turn colored green on the left which covers the active site is captured

Fig. 16 A low-frequency mode of HIV protease dimer [4] calculated with anisotropic network model [5] that shows a swing motion of a monomer relative to the other



where each component of coordinate is written as x explicitly. The Hamiltonian of the system is then

$$H(x) = \frac{1}{2} \sum_i m_i \dot{x}_i^2 + U(x).$$

Working in the mass-weighted coordinates $X_i = \sqrt{m_i}(x_i - x_i^0)$, the Hamiltonian becomes

$$H(X) = \frac{1}{2} \sum_i \dot{X}_i^2 + \frac{1}{2} \sum_i \sum_j \left. \frac{\partial^2 U}{\partial X_i \partial X_j} \right|_{X=X^0} X_i X_j.$$

A transformation from the mass-weighted coordinates to decoupled modes is made by diagonalizing the mass-weighted Hessian (harmonic force constant matrix) to obtain

$$H(q) = \frac{1}{2} \sum_i \dot{q}_i^2 + \frac{1}{2} \sum_i \omega_i^2 q_i^2.$$

The eigenvalue ω_i corresponds to the frequency of each mode, $\nu_i = \omega_i/2\pi$, and the eigenvector corresponding to the eigenvalue gives a set of weights of the mode along the Cartesian coordinates. One can also calculate energy, free energy, entropy of this harmonic system at finite T using statistical mechanics equations.

Force field can be used as the potential energy for normal mode analysis. Because an experimental structure is not at the minimum of the potential, energy minimization has to be carried out first. For normal mode analysis, very thorough energy minimization is necessary with final gradient very close to zero, compared to the relaxation step before MD simulation. However, it is very time consuming to get good convergence in energy minimization for a large, complex structure, and it takes a large amount of computer memory and time to diagonalize a huge Hessian matrix. Using force field would be impractical for large complex systems that have hundreds of thousands of atoms. Alternative methods that consider simple coarse-grained model and assumes a harmonic potential with minimum at the experimental structure and with elastic springs connected between close interaction points in the structure have become very popular. Such models (Gaussian network model, elastic network model, or anisotropic network model) provide excellent description of low frequency modes of complex macromolecules. It is convenient that experimental structure is already at the minimum because no energy minimization is necessary. It is also interesting that the level of coarse graining (how many interaction points are to be considered) do not affect the low-frequency mode a lot. A much smaller size of Hessian matrix can therefore be used compared with all-atom force field. However, unlike molecular mechanics force field, absolute magnitudes of frequencies are not obtained. Only relative values are obtained within a scaling factor, and the scaling factor can be evaluated from experimental data such as crystallographic B-factor. B-factor describes how much thermal fluctuations each atom show in crystal, and can be easily calculated from the eigenvalues and eigenvectors. In Fig. 17, B-factor for each residue in the HIV protease dimer calculated from anisotropic network model is compared with experimental values after fitting the scaling factor for the force constant. Using this scaling factor, the frequency of the mode drawn in Fig. 16 is about 6 THz. The force constant here is calculated to be 2.5 kcal/mol \AA^2 , which is one tenth of that estimated from a Lennard-Jones interaction in introduction, but the mass is about 10 times larger, giving a similar frequency in the THz range.

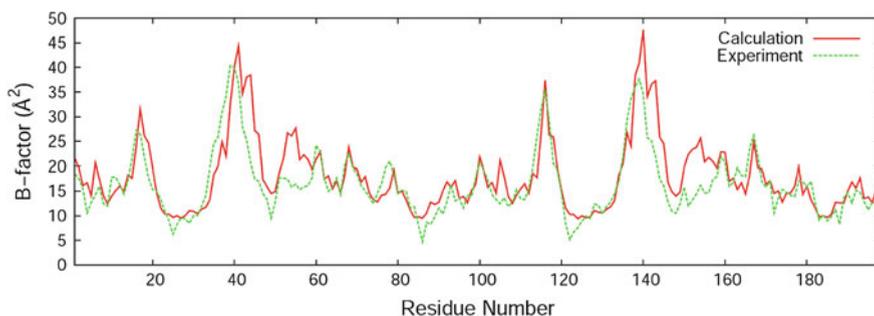


Fig. 17 B-factor calculated from anisotropic network model compared with experimental value for each amino acid residue of HIV protease dimer structure [4]

7 Conclusion

In this part, computational methods for biomolecular modeling such as molecular dynamics simulation, protein structure prediction, and normal mode analysis have been reviewed after introducing protein structure and force field. Biomolecular modeling methods can be useful for studies of their structures, dynamics, and biological functions. There is no unique answer to each modeling problem, and different combinations of different methods may have to be used depending on the nature of the given problem. Therefore, it is very helpful and sometimes necessary to understand pros and cons of different modeling methods clearly. Although molecular modeling has become relatively easy to execute with a bunch of available open and commercial software programs, a naive use of methods often results in meaningless results. THz spectroscopy covers time and length scales that are hard to be accessed with other popular experimental methods such as X-ray crystallography, NMR and, CD spectra. However, it would be challenging to interpret THz spectra of biomolecules satisfactorily. At least the biomolecular modeling methods explained above may play some role in understanding the spectra. Possible difficulties may arise in including the effect of anharmonicity of the potential energy, in obtaining the average over thermodynamic ensembles, and taking into solvation effect in all aspects of calculations, for example, and such problems may be tackled on problem-specific bases.

Acknowledgment I thank Junghyun Choi, Gyu Rie Lee, Dongseon Lee, and Hahnbeom Park for preparing the images in this part.

References

1. Nagar, B., Bornmann, W.G., Pellicena, P., Schindler, T., Veach, D.R., Miller, W.T., Clarkson, B., Kuriyan, J.: Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* **62**(15), 4236–4243 (2002)
2. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E.: UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**(13), 1605–1612 (2004)
3. Notredame, C., Higgins, D., Heringa, J.: T-Coffee: a novel method for multiple sequence alignments. *J. Mol. Biol.* **302**, 205–217 (2000)
4. Kaldor, SW., et al.: Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **40**(24), 3979–3985 (1997)
5. Eyal, E., Yang, L.W., Bahar, I.: Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* **22**, 2619–2627 (2006)

Emotion

Sukwoo Choi

It was widely accepted that emotion such as fear, anger and pleasure could not be studied using a modern scientific tools. During the very early periods of emotion researches, psychologists, but not biologist, dominated in studying emotion and its disorders. Intuitively, one may think that emotion arises from brain first and then bodily responses follow. For example, we are sad first, and then cry. However, groups of psychologists suggested a proposal that our feeling follows bodily responses; that is, we feel sad because we cry! This proposal seems counterintuitive but became a popular hypothesis for emotion. Another example for this hypothesis is as follows. When you accidentally confront a large bear in a mountain, what would be your responses?; you may feel terrified first, and then run, or you may run first, and then feel terrified later on. In fact, the latter explanation is correct! You feel fear after you run (even because you run?). Or, you can imagine that you date with your girl friend who you love so much. Your heart must be beating fast and your body temperature must be elevated! In this situation, if you take a very cold bath, what would you expect? Your hot feeling is usually calmed down after this cold bath; that is, you feel hot because your heart and bodily temperature change. While some evidence supported this hypothesis, others do not. In the case of patients whose cervical vertebrae were severed with an accident, they still retained significant amount of emotion (feelings!) in some cases (but other patients lost most of emotional experience). In addition, one can imagine that there would be a specific set of physical responses for specific emotion if the original hypothesis is correct (e.g. fasten heart beating and reddened face for anger etc.). However, some psychologists failed to find any specific set of physical responses for specific emotion, though others insisted that there existed such specific responses. Based on these controversial observations, another

S. Choi (✉)

Department of Biological Sciences, SNU, Seoul, Korea
e-mail: sukwoo12@snu.ac.kr

hypothesis was proposed; that is, we feel first, and then have physical responses. These two hypotheses seem contradictory to each other, but detailed examination on these hypotheses waited for the development of new analytical tools.

1 Cortical and Subcortical Brain Subregions Involved in Emotion

Roughly speaking, the cortex is in charge of conscious parts of emotion, whereas the subcortical structures take part in unconscious parts of emotion, more like reflex-type functions. The activities of the subcortical structures appear to be directly activated by sensory inputs or imaginary inputs from past experience, and to respond with a minimal delay. These subcortical reactions happen without conscious experience, but they are represented as bodily responses dictated by both the autonomic nervous and hormonal systems. The cortex may detect changes in both the bodily responses and subcortical structures and interpret them along with past emotional experience. So called 'feeling' may result from cortical activation and it may be conscious experience including sensing of emotion-aroused bodily responses. Now we can re-interpret the incident in which someone confronted a big bear in a mountain. When this person saw the bear, his or her amygdala was activated by visual stimuli (bear) before conscious recognition of what's happening. The activated amygdala produced a reflexive reaction, that is, running out of the site. Then this person realized what happened to him or her, and felt fear including sensing bodily responses.

In the following paragraphs, we will review subcortical brain structures involved in emotion. Currently, cortical subregions involved in emotion remain to be elucidated, and hence, description of cortical functions involved in emotion needs more researches in the future.

2 Fear

Among emotion, fear has been extensively studied mainly because animal model, in which fear is measured quantitatively, was available. When the temporal lobe of the cortex including the amygdala, a small almond-like structure, is lesioned in mammals including rats, mice and humans, then the experimental subjects or humans do not feel fear at all or cannot retain fear memory. Astoundingly when the amygdala is lesioned in mice or rats, then these rats or mice play with cats until they got killed by the cats. In addition, whenever the amygdala is electrically stimulated in a patient during brain surgery, then the patient tells a surgeon that he or she feels fear. Therefore, ample evidence strongly suggests that the amygdala is enough and sufficient for fear experience and expression.

Valuable contributions of Drs. James McGaugh (UCLA), Mike Davis (Emory University) and Joseph E. LeDoux (New York University) to the field of fear and its memory has shed light on better understanding of molecular and cellular mechanisms underlying emotion and its memory. Especially, their works have concentrated on fear and its memory. These extraordinary neuroscientists used a simple pavlovian paradigm of fear conditioning. In this simple behavioral testing (called pavlovian fear conditioning), rats or mice are placed in a small chamber, presented with a monotonous tone (usually 1–10 KHz) for 30 s, and then shocked electrically at the end of the tone presentation. These rats or mice learn to associate the tone with the shock, and hence, conditioned rats or mice can retrieve the original fear with the painful shock whenever they are presented with the same tone. This tone-associated fear memory is known to last across the lifetime. Using this particular animal model, there has been tremendous progress in this field.

3 Anger

As shown in the case of fear, anger also has a specific subregion of the brain, called posterior hypothalamus, which is enough and sufficient for anger and its memory. Most of experiment for anger was done using cats in which anger can be easily elicited and measured. One group of experimenters has done clever but bit cruel experiments to characterize brain subregions required for anger. What they did are as follows. They first removed most of the cortex in the cat. This cat seemed to be very volatile (or very easy to make this cat angry); that is, small touch was sufficient to elicit anger. This simple experiment tells us that the cortex somehow inhibits anger in naïve cats. Then this group tried to pinpoint a subregion of the brain, which is required for anger expression, by destroying a potential subregion of the brain. When they destroyed a posterior part of the hypothalamus (hypothalamus is known to involve autonomic nervous reaction and hormonal control), this cortex-removed cat ceased to express anger! Other groups of scientists performed another type of experiment to determine whether the posterior hypothalamus itself is sufficient to elicit anger. They implemented very thin electrode into the hypothalamus and presented very weak electricity. Whenever the electrode-implemented cat was stimulated electrically, the cat became angry! Therefore, the posterior hypothalamus is enough and sufficient for anger just like the case of the amygdala for fear.

4 Pleasure

Unlike the case of fear and anger, it was very difficult to establish an animal model in which pleasure is measured quantitatively. One clever scientist at McGill University improvised a very simple animal model which can be used to study

emotional processes for pleasure (or addiction). This model is simply a self-stimulation paradigm; that is, very thin electrode is inserted into selected subregions of the brain, and then the switch (lever) for this inserted electrode is given to the electrode-inserted rats or mice. These rats or mice first press the lever accidentally, and if the lever press causes pleasure, then these rats or mice keep pressing the lever to get further pleasure. One rat among experimental subjects even did not eat any meals and died because this rat could not cease to press the lever! One of the well-characterized subregion of the brain required for pleasure (even for addiction) is the nucleus accumbens. Just like the case of the amygdala and the posterior hypothalamus, the nucleus accumbens is enough and sufficient for pleasure; that is, destruction of the nucleus accumbens prevents experimental subjects from expressing pleasure responses, whereas stimulation of this subregion elicits pleasure responses.

The nucleus accumbens is also a center for addiction. Narcotic drugs are known to increase monoamine neurotransmitters in this subregion of the brain. Serious problems arise because one needs to inject far greater amount of the narcotic drugs to get similar pleasure as compared with the very first injection of these drugs. Usually these narcotic drugs are severely toxic above a certain dose, and some addicted people may die from injection of overdosed narcotic drugs.

5 Conclusions

Although much more things will have to be elucidated for better understanding of underlying mechanisms for emotion, recent progresses added valuable knowledge to the field of emotion and its memory. First of all, emotion can be understood in terms of a stimulus and response paradigm; that is, subcortical structures involved in emotion receive ongoing stimuli, process them, and respond through our body. These processes occur unconsciously. Our cortex seems to interpret our bodily responses and subcortical activities along with our past experience, and to synthesize so call ‘feelings’! Understanding of emotion and its memory is strongly related with that of our consciousness, one of the fundamental questions in the field of neuroscience.

References

1. Bear, M.F.: *Neurosciences: Exploring the Brain*, 3rd edn. Lippincott Williams & Wilkins (2005)
2. Kandel, E.R.: *Principles of Neural Sciences*, 4th edn. McGraw-Hill Companies, Inc. (2000)

Part III
Terahertz Clinical Imaging

Imaging Biomarker

Hyuck Jae Choi and Joo-Hiuk Son

With recent technical development, new imaging methods are rapidly introduced and creative use of them improves drug development and clinical application. But small numbers of them are integrated in the clinical use. For this reason, understanding the meaning of biomarker and the methodological strategy for validation are important. The purpose of this comment is to explore the basis for the use of imaging biomarker in clinical setting.

1 Definition of Biomarker

The definition of a biomarker is ‘characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention’ [1]. Biomarkers can be grouped into those; ‘proof of mechanism’, ‘proof of principle’, or ‘proof of efficacy’, depending on whether they demonstrate an interaction with and intended target, a modification to a pathway, or effect on the intended disease, respectively. The last type of biomarkers can also be called ‘surrogate endpoints’ [2].

Clinical endpoints means biomarkers that show ‘how patient feels, functions, or survives’ [1] (Fig. 1). Surrogate endpoints are designed to replace clinical endpoints, since clinical endpoints are sometimes difficult to quantify and takes relatively longer time to know the effect of exposure to risk factors. The mortality often takes years to evaluate and the morbidity is often more subjective. For

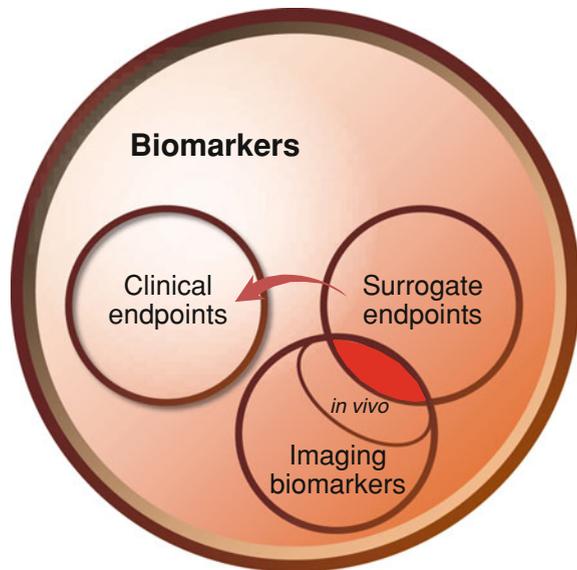
H. J. Choi (✉)

Department of radiology, Asan Medical Center,
University of Ulsan, Seoul, Korea
e-mail: choihj@amc.seoul.kr

J.-H. Son

Department of Physics, University of Seoul, Seoul, Korea
e-mail: joohiuk@uos.ac.kr

Fig. 1 Scheme of imaging biomarker (modified from Schuster [5]) Imaging biomarker, which are adequately validated *in vivo*, can be integrated into surrogate endpoint



example, degree of pain is hard to objectively quantify. For this reason, surrogate biomarkers are valuable if they can show the disease process earlier than clinical endpoint more accurately.

The imaging biomarkers range from simple size measure to functional imaging, such as MR diffusion or perfusion imaging. Imaging biomarker can be defined as ‘anatomic, physiologic, biochemical, or molecular parameters can be detected with imaging methods used to establish the presence or severity of disease’ [3]. By this definition, much of imaging can be a biomarker. Imaging biomarker has several advantages over other clinically used biomarkers, such as biofluid (blood, urine, etc.), tissue (biopsied or excised tissue), physiologic parameters (blood pressure, electroencephalogram, electrocardiogram, etc.) (Table 1). The use of imaging biomarkers in drug development is useful. First, using imaging biomarker, drug company could choose one drug from many candidate drugs. Second, imaging can provide noninvasive validation of results from animal or human. Third, imaging can gradually show more subtle and small changes of disease (Fig. 2).

2 Validation of Biomarker

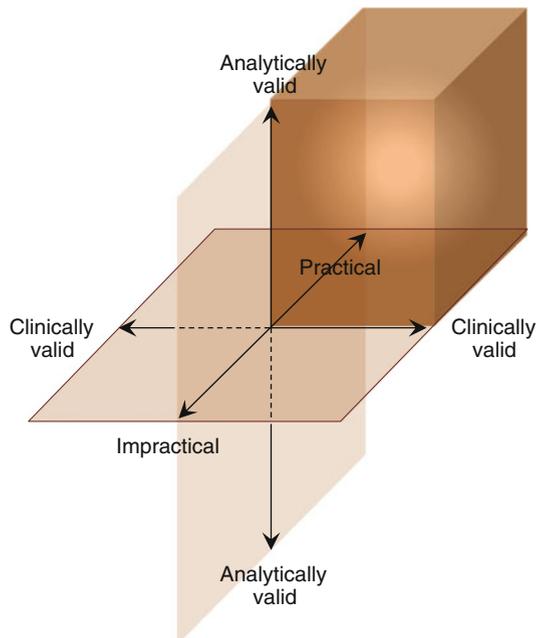
With use of imaging biomarkers we can improve the evaluation of medical therapy. But it is important that imaging biomarkers should be validated before they are used as a preclinical or clinical solution. For the validation, some criteria

Table 1 To and pro in clinically used biomarkers

Biomarkers	Focal disease	Simple Low cost	Non-invasive	Follow up
Biofluids e.g. Blood, urine		✓	✓	✓
Tissue e.g. biopsy, surgery	✓			
Physiologic e.g. blood pressure, EEG, ECG		✓	✓	✓
Imaging e.g. CT, MRI, PET	✓		✓	✓

Imaging biomarker is non-invasive, easy to follow up, and has advantage in localize the disease. But unlike biofluid and physiologic parameters cost is higher. Biopsied or excised tissues are good surrogate biomarker but they have no other clinical advantages over the other clinically used biomarkers

Fig. 2 Validation of biomarker (modified from Schuster [5]) Biomarker should be validated in three aspects (clinical, analytical, and practical aspects). The end results should be in adequately validated area (*thick painted area*)



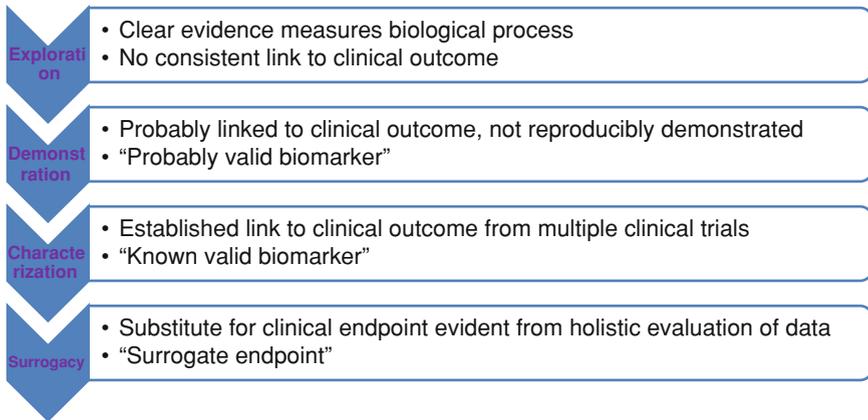


Fig. 3 Life cycle of biomarker: exploration, demonstration, characterization, and surrogacy

should be satisfied. (a) The presence of the imaging biomarker is coupled with the presence of disease or condition (b) the detection or quantitative measurement of imaging biomarker is accurate, reproducible, and feasible (c) the measured changes are closely associated with clinical endpoint [4].

Analytic and clinical validations are important steps in biomarker development [5]. For analytic validation the following issues should be resolved; equipment evaluation, procedure standardization, quality assurance protocols, reproducibility, interrater variability, production and measurement of reference standards, sensitivity to disease state, intersubject variance, image algorithm. Issues in clinical validation are whether there is reliable association between imaging biomarker and clinical endpoint, and whether its use as biomarker has favorable cost-effective impact on disease management.

During development of the new biomarker, it goes through phases. The life cycle of biomarker goes through phases (exploration, demonstration, characterization, and surrogacy) (Fig. 3). In exploration phase, there are clear evidences of measurable biological process but no consistent link to clinical outcome. Usually it is research based and there are multiple platforms. In demonstration phase, there is probably association to clinical outcome but not reproducibly demonstrated. In this phase biomarker is ‘probably valid biomarker’. But the scope is still narrow and continued testing is further needed. In characterization phase, there is established link to clinical outcome from multiple clinical trials. In this phase biomarker is established biomarker and it can enable clinical decisions. In surrogacy phase, biomarker can substitute clinical endpoint.

Smaller numbers of imaging biomarkers are accepted as surrogate biomarkers. This is because of the lack of method standardization, diagnostic performance (sensitivity, specificity, accuracy, etc.), inadequate validation of clinical endpoints. The standardizing imaging methodology is very important issue for potential imaging biomarker [6].

For the understanding of validation process, it is useful to review clinical example of imaging biomarker application. MR scans are often used to evaluate lymph node metastasis in cervical cancer patients [7]. Lymph node size is the imaging biomarker, with size of lymph node at the diagnosis affects the probability of metastasis and the patient mortality. Use of MRI in the detection of lymph node lesions is closely linked to lymph node metastases; MRI measurement of lymph node size is accurate, reproducible and feasible over time. Furthermore, reduction of lymph node size after treatment (such chemotherapy or radiotherapy) is linked to prolonged survival. Since MRI measurement lymph node size is suitable for validation criteria as an imaging biomarker, it can be used for this role. This method may also be used to exclude imaging methods as an imaging biomarker, when there is no strong association with true endpoint. Imaging biomarker which meets the conditions mentioned above has potential to be valuable in daily practice.

3 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) provides information about the resonance of population of nuclei. The resonances are influenced by many environment factors and MRI can detect the variety of physical and biological properties of interest, such as chemical component, diffusion, and perfusion of tissues. MRI is already useful imaging biomarker in daily practice. Aside from detection and size measurement there are three major MR techniques used as imaging biomarkers, MR spectroscopy (MRS), MR diffusion weighted imaging, and MR perfusion imaging.

MR spectroscopic spectra are usually collected from spinning nuclei and are from hydrogen nuclei (proton). The majority of MRS studies have been carried out with proton spectroscopy because water is so abundant in human and readily visible in MRI measurement. But spectra can also be obtained from phosphorus, carbon and fluorine. However, these nuclei are in lower concentration in vivo, so imaging is more difficult. Choline levels are known to be elevated in brain [8], breast [9], prostate [10], colon [11] and cervical cancers [12]. Reduction of normal metabolites as normal tissue is displayed by malignant tissue is also feature. For example, N-acetyl aspartate (NAA), normally abundant amino acid in the brain, is decreased in brain tumor. Although MRS was introduced many decades ago, it is still prebiomarker for many clinicians because the reproducibility, comparability among different centers are not still satisfactory. In addition, it needs additional longer time compared with conventional MRI technique. Additional multicenter study is needed for the validation of this technique.

MR perfusion imaging using dynamic susceptibility contrast (DSC) of brain was developed in 1980s [13]. Recently DSC imaging is more widely available. The contrast leakage in the blood vessel can cause an increase in tumor signal in MRI and this can be quantified. The signal change is typically mathematically

described as permeability or capillary surface area. However, due to the inconsistent results with dynamic contrast enhanced (DCE) images and DSC images they are still in the pre-biomarker stage.

The diffusion or microscopic movement of water can be quantitatively measured with MRI [14]. They are used in the imaging of ischemic stroke. The diffusion coefficient (Apparent diffusion coefficient [ADC]) reported by MRI systems is an average of the water mobility of all directions. Tumors have more water restriction compared with normal tissue and readily detected in MR diffusion image. Although ADC is still pre-biomarker, there is wide agreement how to acquire ADC data, and so this tool is good candidate for becoming imaging biomarker soon.

4 Terahertz Imaging

The electromagnetic (EM) spectrum between microwave and infrared started to be developed about two decades ago. This spectral region is called the terahertz gap, the frequency ranging from 0.1 to 10 THz. The terahertz region has not been taken advantage of due to difficulty of producing and detecting signals. Recently, with the rapid advancement of ultrafast lasers and engineered semiconductors, the application of terahertz became feasible in the late 1980s.

Many groups have studied water in terahertz region to characterize the molecular dynamics of water itself and to determine the effects of water in biological systems [15, 16]. Several groups have performed terahertz interactions with DNAs [17, 18], RNAs [19], proteins [20], and polypeptides [21].

Recently terahertz researchers applied the terahertz techniques to clinical setting, especially to tumors [22–26]. The difference of tumor and normal tissue mainly depend on the degree of hydration in the samples. The sensitive detection of terahertz EM waves to the content of water molecules in the tissue makes it possible for terahertz device to differentiate tumor from normal tissue because cancerous tumors contain more interstitial water compared with adjacent normal tissue. The higher water content, combined with structural changes such as increased cell and protein density, yield larger terahertz absorption and refractive index in tumor compared with normal tissue.

Terahertz EM imaging should be further developed to become surrogate imaging biomarker. The terahertz imaging in medical setting is in developing phase among ‘life cycle of biomarker’. In clinical validation aspect, although there are some studies reporting the difference of tumor and normal tissue in terahertz imaging but there are still no study regarding the association of terahertz imaging with clinical outcome. In analytic validation aspect, there should be considerations about the equipment evaluation, procedure standardization, quality assurance protocols, reproducibility, interrater variability, production and measurement of reference standards, sensitivity to disease state, intersubject variance, and image algorithm. In practical aspect, terahertz should find an adequate position in

medical field. It is skin deep surface imaging due to high attenuation with water, accordingly for terahertz to find a way to contribute in daily practice it should be more portable and has easy access to the body inside, like endoscopy. When terahertz imaging meets all clinical needs mentioned above, it can be a good candidate for surrogate imaging biomarker.

5 Conclusion

New innovative potential biomarkers, including terahertz imaging, are developed recently. Some of them may prove useful for managing cancer patients. However, few biomarkers are accepted and integrated in clinical practice. There are many issues to resolve for potential biomarkers to become surrogate marker. Although it is not easy to resolve these many analytical and clinical validation issues, it is important to know and consider all these necessary issues before and during developing biomarkers. In addition, it takes considerable times for new biomarkers to be accepted to clinicians. But the current trend in medicine, target specific therapy, personalized therapy, may make many chances for the development of new imaging biomarker in the future.

References

1. Lee, J.W., Devanarayan, V., Barrett, Y.C., et al.: Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm. Res.* **23**, 312–328 (2006)
2. Frank, R., Hargreaves, R.: Clinical biomarkers in drug discovery and development. *Nat. Rev. Drug Discov.* **2**, 566–580 (2003)
3. Smith, J.J., Sorensen, A.G., Thrall, J.H.: Biomarkers in imaging: realizing radiology's future. *Radiology* **227**, 633–638 (2003)
4. Annas, G.J.: Faith (healing), hope and charity at the FDA: the politics of AIDS drug trials. *Villanova Law Rev.* **34**, 771–797 (1989)
5. Schuster, D.P.: The opportunities and challenges of developing imaging biomarkers to study lung function and disease. *Am. J. Respir. Crit. Care Med.* **176**, 224–230 (2007)
6. Kelloff, G.J., Hoffman, J.M., Johnson, B., et al.: Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin. Cancer Res.* **11**, 2785–2808 (2005)
7. Choi, H.J., Kim, S.H., Seo, S.S., et al.: MRI for pretreatment lymph node staging in uterine cervical cancer. *AJR Am. J. Roentgenol.* **187**, W538–W543 (2006)
8. Oh, J., Henry, R.G., Pirzkall, A., et al.: Survival analysis in patients with glioblastoma multiforme: predictive value of choline-to-N-acetylaspartate index, apparent diffusion coefficient, and relative cerebral blood volume. *J. Magn. Reson. Imaging* **19**, 546–554 (2004)
9. Katz-Brull, R., Lavin, P.T., Lenkinski, R.E.: Clinical utility of proton magnetic resonance spectroscopy in characterizing breast lesions. *J. Natl. Cancer Inst.* **94**, 1197–1203 (2002)
10. Swindle, P., McCredie, S., Russell, P., et al.: Pathologic characterization of human prostate tissue with proton MR spectroscopy. *Radiology* **228**, 144–151 (2003)
11. Bezabeh, T., Smith, I.C., Krupnik, E., et al.: Diagnostic potential for cancer via 1H magnetic resonance spectroscopy of colon tissue. *Anticancer Res.* **16**, 1553–1558 (1996)

12. Mahon, M.M., Williams, A.D., Soutter, W.P., et al.: ¹H magnetic resonance spectroscopy of invasive cervical cancer: an in vivo study with ex vivo corroboration. *NMR Biomed.* **17**, 1–9 (2004)
13. Villringer, A., Rosen, B.R., Belliveau, J.W., et al.: Dynamic imaging with lanthanide chelates in normal brain: contrast due to magnetic susceptibility effects. *Magn. Reson. Med.* **6**, 164–174 (1988)
14. Le Bihan, D.: Looking into the functional architecture of the brain with diffusion MRI. *Nat. Rev. Neurosci.* **4**, 469–480 (2003)
15. Heugen, U., Schwaab, G., Bründermann, E., et al.: Solute-induced retardation of water dynamics probed directly by terahertz spectroscopy. *Proc. Natl. Acad. Sci.* **103**, 12301–12306 (2006)
16. Vij, J.K., Simpson, D.R.J., Panarina, O.E.: Far infrared spectroscopy of water at different temperatures: GHz to THz dielectric spectroscopy of water. *J. Mol. Liq.* **112**, 125–135 (2004)
17. Zhuang, W., Feng, Y., Prohofsky, E.W.: Self-consistent calculation of localized DNA vibrational properties at a double-helix-single-strand junction with anharmonic potential. *Phys. Rev. A* **41**, 7033–7042 (1990)
18. Young, L., Prabhu, V.V., Prohofsky, E.W., Edwards, G.S.: Prediction of modes with dominant base roll and propeller twist in B-DNA poly(dA)-poly(dT). *Phys. Rev. A* **41**, 7020–7023 (1990)
19. Fischer, B., Hoffmann, M., Helm, H., et al.: Terahertz time-domain spectroscopy and imaging of artificial RNA. *Opt. Express* **13**, 5205–5215 (2005)
20. Markelz, A.G., Knab, J.R., Chen, J.Y., He, Y.: Protein dynamical transition in terahertz dielectric response. *Chem. Phys. Lett.* **442**, 413–417 (2007)
21. Kutteruf, M.R., Brown, C.M., Iwaki, L.K., Campbell, M.B., Korter, T.M., Heilweil, E.J.: Terahertz spectroscopy of short-chain polypeptides. *Chem. Phys. Lett.* **375**, 337–343 (2003)
22. Fitzgerald, A.J., Wallace, V.P., Jimenez-Linan, M., et al.: Terahertz Pulsed Imaging of Human Breast Tumors. *Radiology* **239**, 533–540 (2006)
23. Wallace, V.P., Fitzgerald, A.J., Shankar, S., et al.: Terahertz pulsed imaging of basal cell carcinoma ex vivo and in vivo. *Br. J. Dermatol.* **151**, 424–432 (2004)
24. Woodward, R.M., Wallace, V.P., Pye, R.J., et al.: Terahertz pulse imaging of ex vivo basal cell carcinoma. *J. Investig. Dermatol.* **120**, 72–78 (2003)
25. Park, J.Y., Choi, H.J., Cho, K.-S., Kim, K.-R., Son, J.-H.: Terahertz spectroscopic imaging of a rabbit VX2 hepatoma model. *J. Appl. Phys.* 109:064704-064704-064704
26. Oh, S.J., Kang, J., Maeng, I., et al.: Nanoparticle-enabled terahertz imaging for cancer diagnosis. *Opt. Express* **17**, 3469–3475 (2009)

Terahertz Medical Imaging

Joon Koo Han

1 Introduction

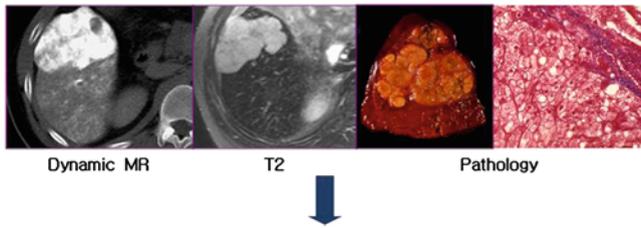
The past 30 years have witnessed the evolution of medical imaging. Imaging technology has drastically improved medical care and, as a result, everyday life. Since the discovery of X-rays by William Roentgen in 1895, doctors have used radiography to diagnose many diseases of the internal organs for several decades. The development of X-rays brought about a revolution in the field of medical diagnostics. The peak of this revolution would begin in the 1970s when computed tomography (CT), magnetic resonance imaging (MRI), and ultrasonography were developed. Enormous progress in the field of computer technology has been boosting the evolution of imaging technology since the 1990s. To date, medical imaging innovations have focused on achieving excellent anatomic resolution and contrast [1].

In the 21st century, the evolution of medical imaging continues to go beyond anatomical and structural imaging (Fig. 1). More emphasis has been placed on functional and molecular imaging techniques that can depict physiological and cellular processes within body tissues [2]. As the demand for imaging at the molecular and cellular levels is increasing, development of the new biophotonics imaging techniques have gained momentum [3]. In the electromagnetic spectrum, γ -rays and X-rays are placed at the top of the frequency spectrum and are already routinely used in clinical practices. Heading downward to low-energy electromagnetic regions, the optical and near-infrared imaging share great success in biophotonics biomedical applications. Most radiologists and imaging researchers

J. K. Han (✉)

Department of Radiology, College of Medicine, Seoul National University,
Daehak-ro, Jongno-gu 103, Seoul 110-799, Republic of South Korea
e-mail: hanjk@snu.ac.kr

20th Century Paradigm : Structural Imaging–Pathology Correlation



21st Century Paradigm: Biochemical / Physiologic/ Functional Imaging

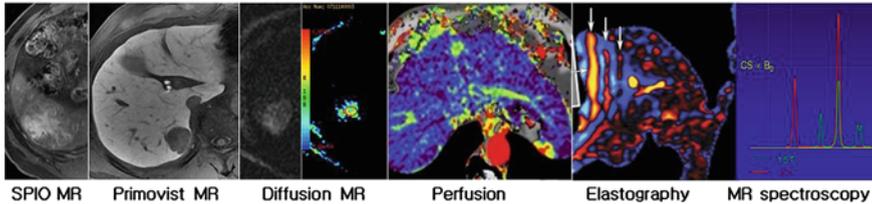


Fig. 1 The changing paradigm of medical imaging (Courtesy of Dr. Choi, Seoul National University Hospital)

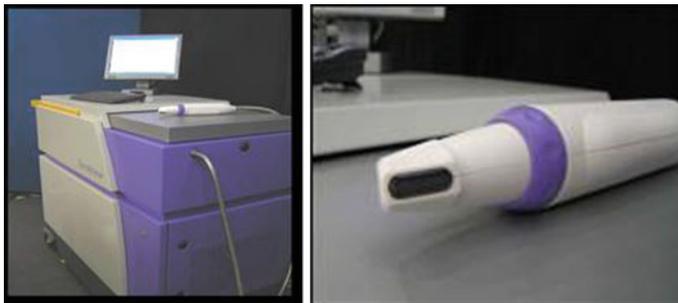


Fig. 2 Commercial terahertz machine and skin cancer detection probe (TPITM imaga 2000; TeraView Ltd, UK). Reprinted with permission from www.teraview.com

are familiar with optical imaging. In particular, optical coherent tomography (OCT) has become an important medical imaging tool for evaluation of the retina. Optical bioluminescence imaging is also a widely used imaging tool in *in vivo* mouse experiments. Heading to terahertz (THz) wavelength region, THz imaging and spectroscopy technologies are rapidly developing. THz imaging has no established clinical application yet. However, THz imaging for skin cancer detection has the potential to be incorporated into routine clinical practice. Several companies (e.g. TeraView Ltd, Cambridge, UK) have developed the probes for skin cancer detection and undergone the clinical trials (Fig. 2).

We believe that THz radiation has unique benefits such as safety due to non-ionizing nature and spectroscopic information over 0.1–3 THz. THz imaging can

survive in the tough battle between the various imaging modalities. In this part, the authors discuss the characteristics of THz radiation and imaging system for biomedical application and provide a short review of recent advances in THz medical imaging application.

2 Terahertz Time-Domain Spectroscopy and Imaging

2.1 THz Imaging System

THz imaging is a novel imaging technique that uses photoconductive generation and THz radiation detection that maps the interaction between THz radiation and biological matter. In this part, we focus on THz imaging using the THz time-domain spectroscopic (TDS) technique, which is widely used in the biomedical applications of THz imaging [4]. This THz-TDS technique uses broadband pulsed THz sources, which are typically photoconductive devices. The THz-TDS technique uses coherent detection to record the THz wave's temporal electric fields, meaning that the THz wave amplitude and phase can be obtained simultaneously. The temporal waveforms can be further Fourier transformed to provide the spectra. Generation and detection of pulsed coherent THz radiation enables spectroscopic analysis and imaging. The THz-TDS system transmission mode is usually used for spectroscopic analysis, while the reflection mode is usually used for biomedical imaging configured by limited penetration of the THz wave into biological tissues.

Figure 3 shows a typical THz-TDS system in reflection mode used for medical imaging [5]. Among various THz-TDS methods, the most established approaches are based on photoconductive antennas [6], which use photoconductive switches that are excited by mode-locked femtosecond pulsed lasers. Femtosecond pulsed laser excitation of photoconductive switches cause rapid transients in the photoconductor carrier density and conductance.

2.2 Advantages of Using THz Imaging for Biomedical Application

THz radiation has a number of properties that make it suitable for biomedical applications. First, THz radiation is a safe, nonionizing radiation with very low photon energy (only 4.14 meV in 1 THz radiation). The frequency of THz radiation (10^{12} Hz) is much lower than that used for X-ray imaging (10^{20} Hz). This frequency is insufficient to cause chemical damage to molecules or knock particles out of atoms. Thus, THz radiation will not cause harmful ionization hazards like those caused by X-rays, which was proven previous researches on keratinocytes and neural cells [7].

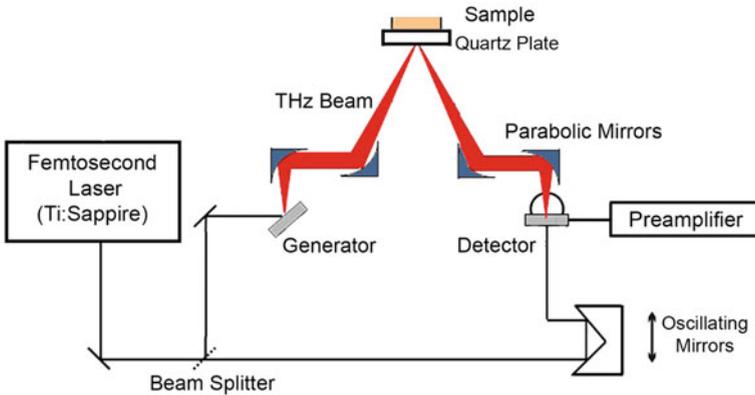


Fig. 3 Terahertz reflection mode imaging setup

Second, THz pulsed imaging explores the THz frequency range of 0.1–3 THz, which corresponds to a wavelength range of 3–100 μm . These wavelengths are significantly larger than the size of scattering structures in tissues; thus, the scattering effects of THz radiation should be considerably reduced compared with those of optical techniques using radiations of shorter wavelengths, e.g. near-infrared or visible radiation [8]. Besides, the wavelength is so short that sub-millimeter resolution is readily achievable. Typical spatial resolution was 250 μm laterally and 20 μm axially [9].

Third, THz pulsed imaging is a coherent technique for recording the THz wave's temporal electric fields. The THz pulse amplitude and phase can be obtained simultaneously, and the temporal waveforms can be further Fourier transformed to give the spectra. This allows precise measurements of the refractive index and absorption coefficient of samples [8]. Broadband detection allows for the acquisition of information over a wide spectral range at THz frequencies, and this provides spectroscopic and potentially diagnostic information not found in other modalities. The absorption is generally due to the chemical constituents of the medium; thus, if specific signatures of disease could be measured, the technique could provide diagnostic information. In addition, THz pulsed imaging is a time-domain technique in which reflections of THz pulses from structures buried in the tissue can be detected; therefore, depth information can also be obtained. Furthermore, the THz pulsed imaging technique has a very high signal-to-noise ratio (SNR) because of efficient elimination of the background noise with coherent time-gated detection [10]. This high SNR allows for small refractive index changes to be detected and imaged. In summary, THz pulsed imaging is a coherent, time-gated, and low-noise technique that has the potential to provide both structural and functional information due to its chemical specificity, which enables the investigation of both morphological and chemical changes [11].

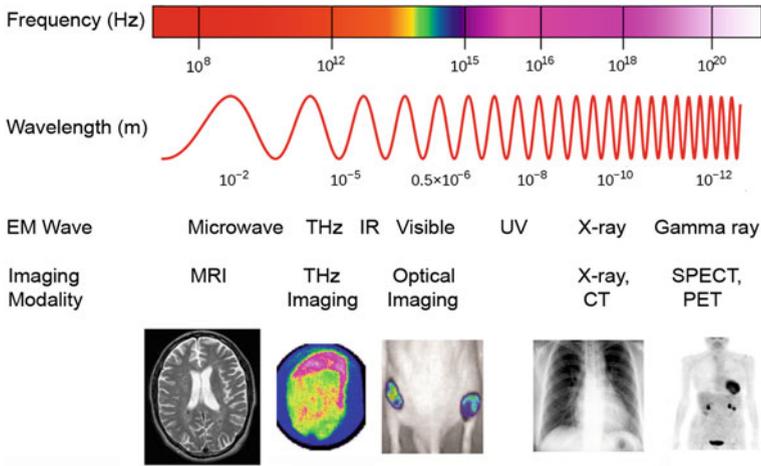


Fig. 4 Biomedical imaging modalities

2.3 Interaction of THz with Biological Materials

A variety of medical imaging modalities make use of photons of all energies (X-rays, γ -rays, optical rays, infrared rays, microwaves, radiofrequency, etc.) or ultrasound waves for imaging (Fig. 4). The interaction with biological materials and the image creation processes differ substantially, but the common principle is that those image modalities probe how biological tissues interact with them. The various ways in which electromagnetic radiations interact with biological materials range from electronic interactions at the very small wavelength X-ray region to the larger molecular motions at the longer wavelength microwave region. THz pulsed imaging explores low-frequency rotational and vibrational motions in molecular systems [12].

THz radiation excites intermolecular interactions such as rotational and vibrational modes in liquid water and probes weak interactions between molecules [13] (Fig. 5). THz radiation’s interaction with biological matter such as water, fat, and protein differ, so this imaging modality can be used to differentiate between biological tissues. Indeed, the THz optical properties of freshly excised healthy human samples of skin, adipose tissue, striated muscle, arteries, veins, and nerves significantly differ in terms of refractive index and attenuation coefficient [14].

Water and protein are the major ingredients of biological tissues, and most diseases are accompanied with changes in the water content and protein structure. Hence, our understanding of the interactions between these molecules and THz radiation is very important. In particular, water shows a very strong absorption spectrum with a broad peak centered at 5.6 THz, which is attributed to resonant

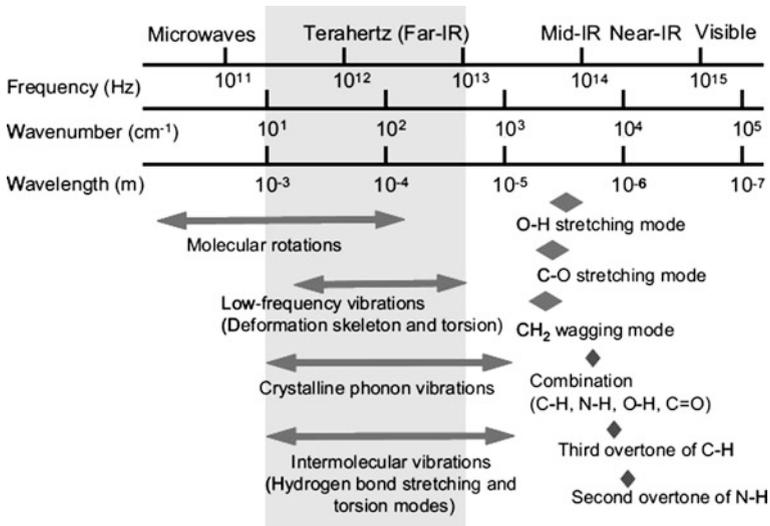


Fig. 5 Characteristic intermolecular interactions on the terahertz wavelength. Reprinted with permission from [15]

stretching of the hydrogen bond between water molecules [15]. Water still has strong absorption in the 0.1–3 THz range used in THz pulsed imaging, making the technique highly sensitive to water content [16].

The sensitivity of THz imaging and spectroscopy to the water content in materials enables its use in the diagnosis of cancers and other diseases. For example, cancerous tumor tissues contain more interstitial water than do healthy tissues [4]. However, according to recent research using paraffin-embedded tissue to exclude the water effect, other possible contrast parameters between cancerous and normal tissues include cell density and protein content [5]. The higher water content combined with structural changes such as increased cell and protein density yields higher THz absorption and refractive index for cancerous tissues.

Since most biological tissues have high water content, the exceptionally high absorption losses at THz frequencies make penetration through biological tissues infeasible. For example, the penetration depth of a commercial THz machine (TPI™ imaga 1000; TeraView) is close to 1 mm in skin and 6 mm in fat [9]. This is an important drawback of THz pulsed imaging for the evaluation of buried lesions in the tissue. However, for investigations of surface features or shallow tissues such as skin, this depth might be acceptable. Besides, the same high absorption coefficient that limits tissue penetration also promotes extreme contrast between substances with lower or higher degrees of water content, and this can help show distinctive contrasts in medical imaging [6].

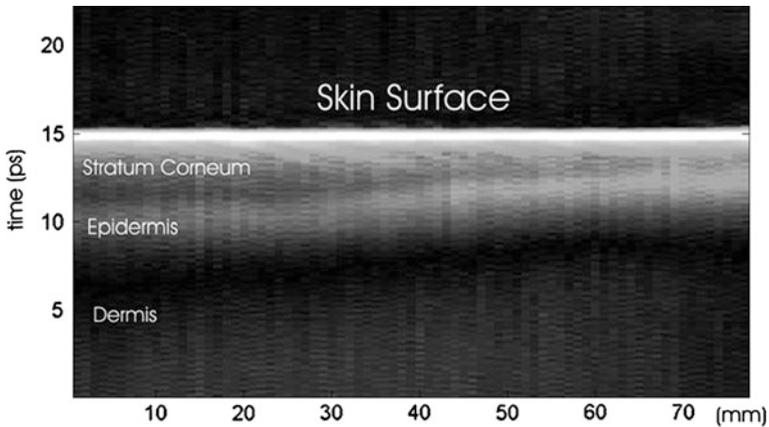


Fig. 6 Terahertz B-scan image of the edge of a volunteer's hand. The *grey-scale* indicates the THz amplitude. It is plotted against the optical delay (vertical axis) and x-position across the scanned area (horizontal axis). The decrease in stratum corneum thickness across the x axis, from the palm side ($x = 30$ mm) to the back side ($x = 70$ mm) of the hand is evident. Reprinted with permission from [9]

3 Applications in Medical Imaging

3.1 Skin Cancer

Skin is the most important target organ of THz pulsed imaging because it is superficial and has thin layers. Normal skin comprises 3 main layers: the stratum corneum (outermost), epidermis, and dermis (innermost). The layered structure of skin is ideal for THz pulsed imaging since the technique uses the reflection from an interface between layers with different refractive indices. In particular, the optical time-delay of the multiple reflections of the different skin layers can be used to estimate depth [17]. The important applications of THz skin imaging are skin cancer detection and invasion depth analysis since deeper invasion indicates worse prognosis. Therefore, the identification of each skin layer in the image is important. The cross-sectional THz B-scan image can be useful for skin imaging since it displays the skin depth profile (Fig. 6).

Considering that the axial resolution of conventional THz pulsed imaging is 20–40 μm , a stratum corneum thinner than the axial resolution cannot be distinguished. For example, on the forearm where the stratum corneum thickness is usually <20 μm , the reflected waveform off the volar forearm was dominated by the reflection off of the epidermis and the optical properties of this layer were determined. However, on the palm of the hand, where the stratum corneum is thicker, the system was able to resolve the stratum corneum [9, 18].

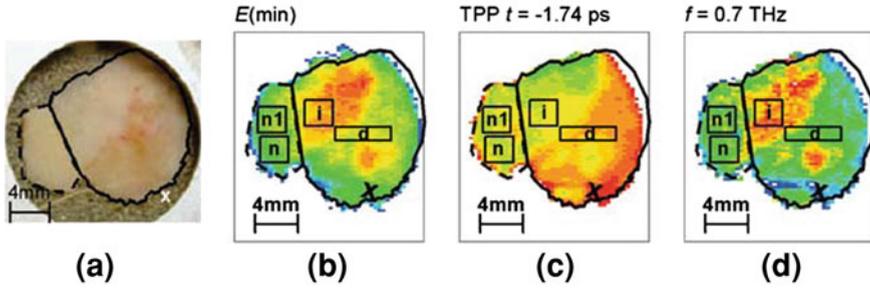


Fig. 7 The visible images and THz images of skin cancer. (Area *d*—diseased; *i*—inflammation; *n*, *n1*—normal). The (a) visible, (b) $E(\min)$ and (c) TPP time-domain images and (d) the frequency-domain image of basal cell carcinoma and normal tissue. The solid and dashed lines mark the boundaries of the diseased and normal tissue respectively. In the visible image, a region of inflammation is present in the top left of the diseased tissue, just visible by the lighter region of tissue, which is clearly visible in the THz images. Reprinted with permission from [28]

Skin cancer is one of the most common cancers worldwide. Non-melanoma skin cancer comprises the majority of skin cancers, with a proportion of approximately 75% basal cell carcinomas and 20% squamous cell carcinomas [19]. Surgical removal is the standard treatment method. However, local recurrence frequently occurs after treatment due to remaining non-visible cancer cells outside the resection margin. In fact, some infiltrative tumors can extend 15 mm beyond the border that is visible to the naked eye at the skin's surface. Surgeons want to remove the cancer with wide safety margins to ensure complete resection; however, patients prefer to preserve as much normal skin as possible for cosmetic reasons. To date, no methods have been established as standard diagnostic imaging tools for skin cancer, even though several imaging modalities such as high-frequency (20 MHz) ultrasonography, optical coherence tomography (OCT), and multi-photon microscopy have been investigated [8]. Based on the data of many ex vivo and in vivo studies presented by many groups in Leeds, Cambridge, and TeraView, THz imaging has the potential to be a standard imaging modality for skin cancer [8, 17, 18, 20–22].

Researchers at TeraView Limited, UK, along with surgeons and histopathologists, have extensively investigated skin cancers [4, 8, 22]. Woodward et al. [22] used the THz pulses to study ex vivo basal cell carcinoma. Their study showed an increase in absorption of the diseased tissue compared to the normal tissue, and this was attributed to either an increase in the interstitial water of the diseased tissue or a change in the vibrational modes of the water molecules with other functional groups. This study showed that THz pulsed imaging is a very sensitive tool. Among a total of 21 samples examined, the 17 samples that showed contrast in THz imaging were confirmed as being diseased tissues, while the other 4 samples, lacking contrast, were not considered as tumors. Wallace et al. [8] also performed in vivo measurements of basal cell carcinoma. In these studies, tumor depth was estimated and compared to histological findings (Fig. 7).

3.2 Skin Burns

Skin burns are very common injuries in daily life. Most burns are superficial and have few sequelae or complications. However, severe burns are morbid and require extensive management including surgical or intensive care unit treatment. Deep burn injuries also result in scarring that may cause joint contractures or extremity loss [23]. Burn injuries are divided into 3 degrees according to depth: superficial partial-thickness burn, deep partial-thickness burn, and full-thickness burn. Distinguishing between superficial and deep partial-thickness burns is important since deep partial-thickness and full-thickness burns require wound excision and skin grafting, whereas more superficial burns can be managed more conservatively [23].

Therefore, burn depth evaluation is essential in the management of burn injuries. However, the most common methods for estimating burn depth are visual and tactile assessment, which are frequently inaccurate. Moreover, these assessments are subjective and dependent on the doctor's experience. Thus, there has been a strong demand for the development of an imaging tool for burn depth assessment. To date, OCT and laser Doppler perfusion imaging have been used for burn depth imaging [24–26]. However, both methods have the limitations of excessive optical scattering and poor depth resolution.

There have been continuous efforts toward using THz techniques for burn assessment. The potential of THz imaging as a tool for burn diagnosis was demonstrated in 1999 using chicken breasts [27]. Recently, a THz reflective imaging system based on direct detection was developed and used to obtain high-resolution images of porcine skin with superficial burns [28]. The burned and unburned regions of porcine skin demonstrated large, clear differences in THz reflectivity (Fig. 8). The burned region contained less water and showed a less reflective signal. Interestingly, the THz images also exhibited a “halo” surrounding the burn areas that may correlate with the burn extent. Their results indicate that the use of THz imaging may be promising for evaluating the severity of skin burns, especially for characterizing the burn areas.

A recent study revealed that *in vivo* THz burn imaging provided valuable information about physiological tissue responses such as edematous changes since it is very sensitive to these fluid shifts [29]. The immediate response of skin tissue following a burn is hyperemia or edema characterized by a sudden rush of fluid to the burned region. After several hours, the response is organized into 3 regions: zones of coagulation, stasis, and hyperemia. The zone of coagulation at the center of the burn is the irreversibly damaged area that contains destroyed cells and denatured proteins. The zone of stasis surrounding the zone of coagulation contains hypoperfused tissues. The zone of hyperemia forms the outer ring and is characterized by edematous, hyperperfused tissues. Concomitant with the appearance of these 3 zones are large local and systemic fluid shifts. Investigations into these fluid shifts have concluded that local water concentration can increase by as much as 80% within 10 min of an injury in a manner that is proportional to

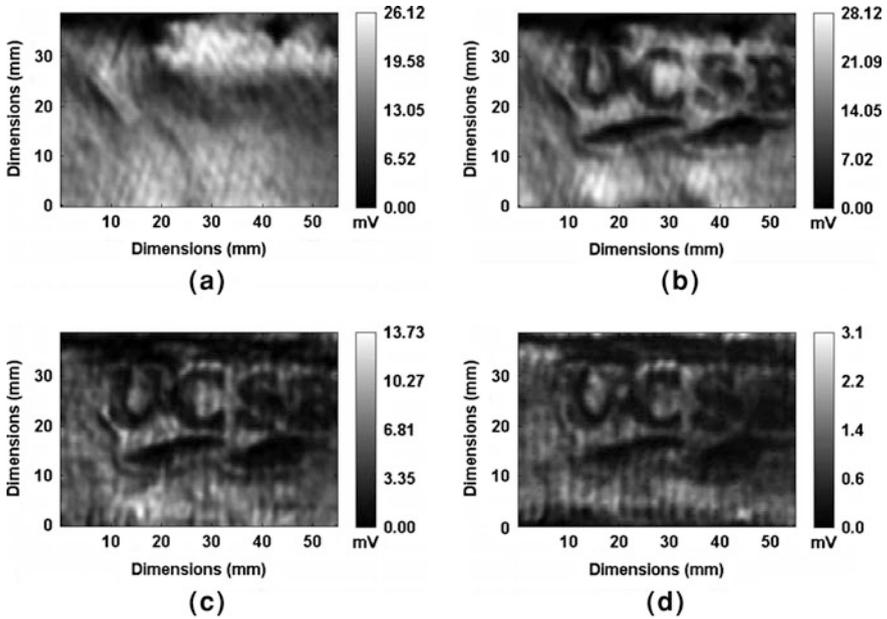


Fig. 8 Terahertz images of a pig skin sample. The lighter areas correspond to areas with more reflective signals. The *gray scale* on each side of the image shows the signal level (mV). **a** Uninjured skin. **b** Burned skin. **c** Burned skin beneath 5 layers of gauze. **d** Burned skin beneath 10 layers of gauze. Reprinted with permission from [29]

the depth of the injury [30]. Figure 9 demonstrates that THz imaging reflects the fluid shift during acute burn injury.

Recently, Huang et al. [31] tried to study the depth information and reflection imaging contrast of THz pulsed imaging of a skin burn. They showed that the time of flight technique is able to reveal the depth profile and thus could be used to evaluate burn depth. They used THz pulsed imaging to measure the thickness of the stratum corneum on the palm beneath a plaster. THz radiation can penetrate nonpolar materials such as gauze or plaster without significant image degradation, and this property is beneficial in monitoring burn wounds without manipulating the wound dressing.

3.3 Breast Cancer

Breast cancer is one of the most common cancers among women, accounting for nearly 1 in 4 cancers diagnosed in US women. Incidence in the US between 2002 and 2006 were of the order of 400 cases in every 100,000 women depending on race and age [32]. Around 90% of patients diagnosed will undergo surgery to treat

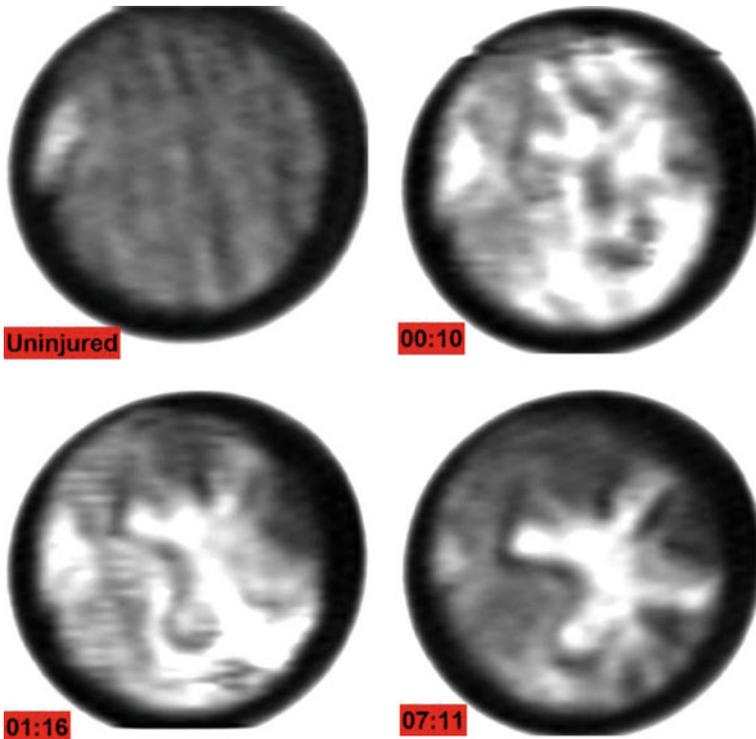


Fig. 9 Terahertz burn images displaying edema formation in an in vivo burn wound. *Top left* a rat belly skin prior to the application of a heated brand. *Top right* a rat belly skin 10 min post-burn showing acute influx of water throughout the region with non-burned areas reflecting as much incident THz radiation as the burned areas. *Bottom left* >1 h post-burn showing a more organized response in which the zone of increased water concentration begins to localize in the region in which the brand contacted the skin. *Bottom right* burn injury after 7 h, by which time the edema localized completely to the brand contact area and the surrounding skin returned to normal, non-burned reflectivity. Additionally a dark, low reflectivity ring of the tissue is seen along the periphery of the hyperhydrated brand contact area. Reprinted with permission from [29]

the disease. Of these, 60% will undergo breast-conserving surgery and the remainder will undergo mastectomy. In breast-conserving surgery, the primary tumor and a margin of normal tissue are removed. At the present time, surgeons use intraoperative clinical palpation of the excised specimen and radiography to determine the tumor margins. These procedures can underestimate the tumor extent. The specimen is then sent for routine histopathology, the gold standard for determining whether the margins are free of malignancy, but is a time- and labor-intensive process. In case of tumorectomy, around 22–59% of patients will require a second operation, as the margins are invariably deemed not free of cancer during histopathology [33]. There is a clinical need to be able to accurately define tumor margins during surgery, conserve normal tissue, and minimize the number of

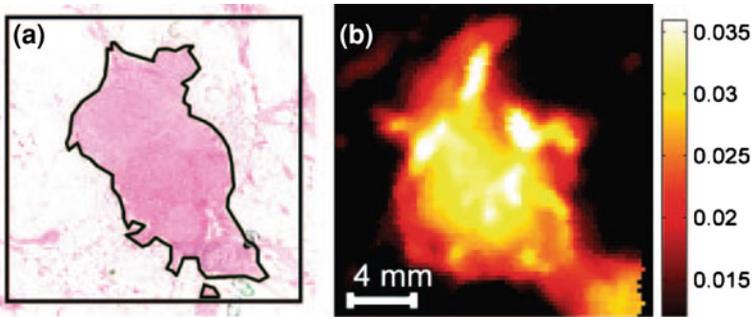


Fig. 10 A photomicrograph (a) and THz image (b) of the breast cancer [black outline in (a)]. Reprinted with permission from [36]

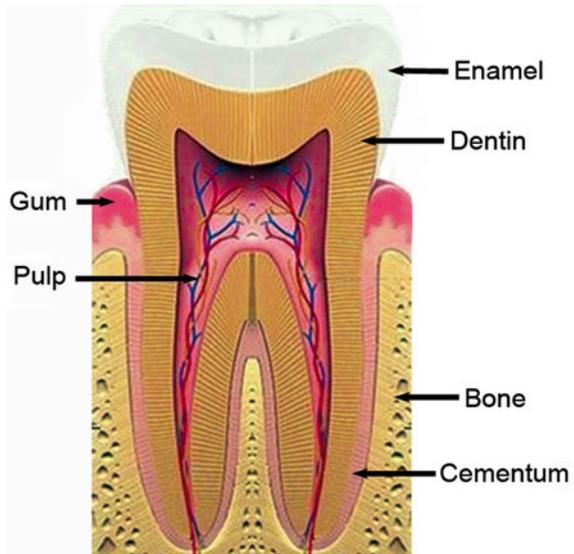
required second surgical procedures. At the present time, no technology is available to accomplish these objectives.

Breast cancer is a good target of THz imaging because of the abundant fat tissue involved. Breasts are composed primarily of fat tissue with low water content and mammary tissue with high water content. Cancer that occurs in the mammary tissue usually infiltrates into the adjacent fat tissue. Therefore, THz imaging, which is very sensitive to tissue water content, can show the extent of breast cancer due to the difference in hydration levels between the cancer and fat tissue. Furthermore, the low water content of fat tissue enables deep penetration of THz radiation. Therefore, THz imaging has the potential to be used as an intraoperative imaging device to detect tumor margins or residual cancer tissues during a surgical procedure. In fact, a hand-held probe was developed as an intraoperative tool for the detection of breast cancer [34, 35].

Another potential advantage of THz imaging for breast cancer is its ability to detect non-calcified breast cancer. In cases of breast cancer with microcalcifications, a surgical procedure can be guided by X-ray imaging. However, non-calcified tumors are often missed during breast-conserving surgery because they are neither visible by X-ray imaging nor palpable [36].

In collaboration with surgeons and histopathologists, Fitzgerald et al. [36] have investigated the feasibility of terahertz imaging for mapping tumor margins in freshly excised human breast tissue. Several samples of excised breast tissue were imaged, and parameters from the time-domain pulse were used to provide contrast. The size and shape of the tumor regions in the THz images were compared with the corresponding histological sections. Figure 10 compares a THz image of excised breast tissue with its histological sample for which a good correlation was found. The correlation coefficient for the tumor area on the THz images compared with those on the photomicrographs of 22 specimens with invasive ductal carcinoma or ductal carcinoma in situ was >0.82 , and the median Spearman rank correlation coefficient, which correlated the shapes on the THz images with those on the photomicrographs, was 0.69. Pickwell and Wallace

Fig. 11 Tooth structure comprising 4 components: the enamel, dentin, cementum, and pulp



also performed a spectroscopic study and observed a higher absorption coefficient and refractive index for cancerous breast tissues compared with those of normal breast tissues [9].

3.4 Dental Imaging

Since the use of THz pulsed imaging in medical applications is currently limited to the evaluation of surface features, another potential application of THz imaging is the diagnosis of dental caries [6]. Interestingly, dental imaging was the first diseased tissue studied using THz imaging [37]. Human teeth have 4 components: the enamel, dentin, cementum, and pulp (Fig. 11) [38]. Enamel is composed primarily of inorganic crystals (96%), with only a small amount of water (3%) and organic matrix (<1%). Dentin constitutes the bulk or body of each tooth and serves as a protective covering for the pulp's connective tissue. It contains approximately 68% inorganic hydroxyapatite complex and is second only to enamel as the most highly mineralized body tissue [39]. Organic components comprise approximately 22% of dentin, and collagen is its major protein. Cementum is the thin, calcified tissue that covers the tooth roots, and it contains fewer minerals than do the bones or dentin. The pulp is the soft-tissue component of the tooth in which blood vessels and nerve branches reside. THz of extracted human teeth is presented in Fig. 12, which shows that each layer of the human tooth structure is characteristically different from the others.

Early caries are a result of mineral loss from enamel, and this causes a change in the refractive index within the enamel [40]. The change in refractive index

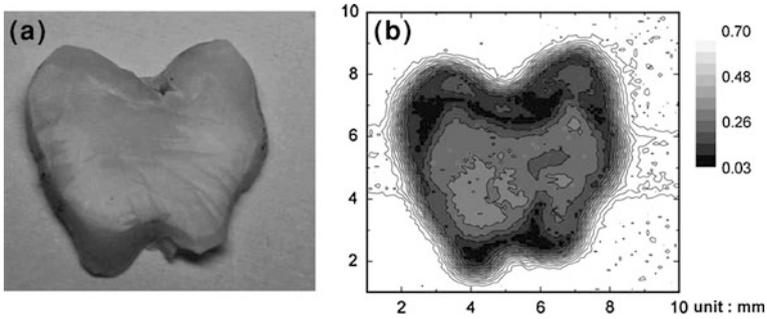


Fig. 12 **(a)** Photograph and **(b)** terahertz transmission amplitude image of an extracted human tooth. Using the differences in terahertz peak amplitude, an image was taken of a sample tooth. Their exposure to THz radiation disclosed that each tissue layer has a characteristic difference from the others. Reprinted with permission from [38]

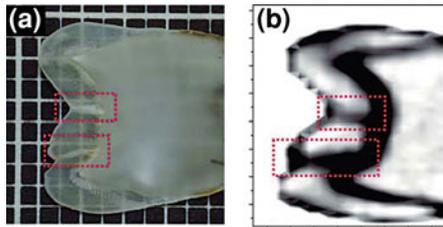


Fig. 13 A 210- μm -thick vertical section of a human molar. The areas highlighted by the *dashed boxes* are enamel caries. **(a)** A light microscopy photograph of the sample. **(b)** An image generated by plotting the terahertz (THz) pulse intensity change as it passes through the sample at different x and y values corresponding to the sample's THz absorption map. Note that caries are detected in this way and dentin and enamel are differentiated. Reprinted with permission from [53]

means that small surface or subsurface caries that are not detected by the human eye can be detected using THz imaging (Fig. 13). Early caries detected at the initial stages of demineralization are reversible and can be treated without drilling [6].

3.5 Corneal Hydration Measurement

Corneal hydration measurement is a good potential application for THz imaging. The cornea is the transparent front part of the eye that covers the iris, pupil, and anterior chamber. The cornea refracts light to contribute to the eye's focusing power. Some kinds of corneal dystrophies and dry eyes are related with the disturbed cornea's water-regulating process [41]. Keratoconus exhibits nonuniform changes in the water concentration of the cornea [42]. Corneal hydration is also perturbed by LASIK surgery or excimer laser ablation, both of which are photo-refractive correction procedures [43]. However, there is no available clinical tool

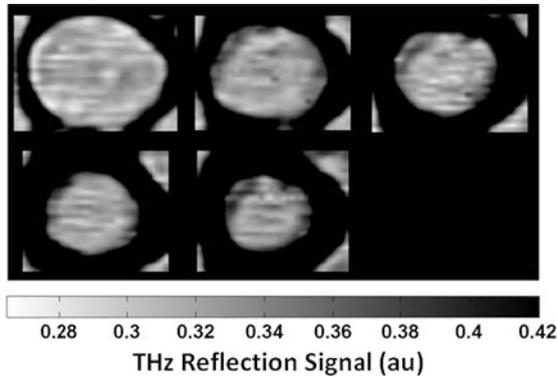


Fig. 14 THz image of an drying porcine cornea at (*upper-left to lower-right*) 84.74%, 78.64%, 75.27%, 70.25%, and 66.06% water content by mass. Reprinted with permission from [44]

to measure corneal hydration. Theoretically, THz imaging may be a suitable tool for measuring corneal hydration *in vivo* because the cornea surface is smooth and the cornea contains a homogenous water content [29].

A recent study with extracted *ex vivo* porcine corneas showed that THz imaging can measure the changes in corneal hydration while drying the cornea [44]. In this experiment, the extracted cornea was mounted onto a flat geometry by pressing against a quartz window. THz imaging showed clear preferential drying from the outer edges, as the border closes further in towards the center with each subsequent frame (Fig. 14). Comparison of 9 corneas hydrated from 79.1 to 91.5% demonstrated an approximately linear relationship between THz reflectivity and water concentration. These results suggest that THz imaging has the potential to be used for corneal hydration measurement.

3.6 Other Potential Applications

A THz endoscopy system may extend the medical application of THz imaging to the internal organs of the body such as the gastrointestinal (GI) tract or joints. Endoscopy is widely used for the diagnosis of diseases in the GI tract, including the stomach and colon. Arthroscopy is also an essential diagnostic and surgical tool in orthopedic surgery. Additionally, the use of laparoscopic surgery has been rapidly increasing. There have been efforts to develop a fiber-coupled THz-TDS system [45, 46]. Recently, a miniaturized fiber-coupled THz endoscope system was developed [47] that utilizes a photoconductive generator and detector driven by a mode-locked Ti:sapphire laser through dispersion-compensated optical fibers. With the generator and detector attached in a parallel manner, the endoscope head is only 8×6 mm in size, small enough to be inserted into a human body (Fig. 15). In the near future, a THz endoscope system will be integrated into a clinical endoscope system.

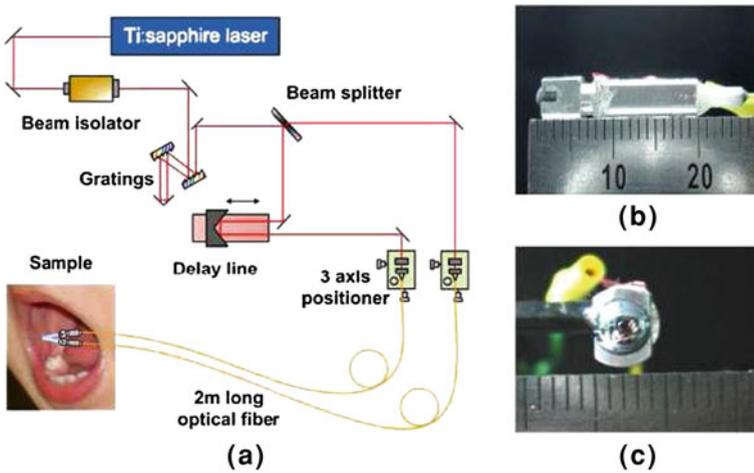


Fig. 15 **a** Experimental set-up for the terahertz endoscope. **b** Side view of the module. **c** Front view of the module. Reprinted with permission from [47]

There has been much research regarding the potential applications of a THz endoscopy system such as THz imaging of the GI tracts and articular cartilage. In a recent thesis paper, THz imaging was performed for ex vivo samples of both normal and diseased human colonic tissues [12]. The results of the colon cancer and adjacent normal tissue study, which used binary logistic regression analysis of time-domain reflected signals, revealed a sensitivity of 89.2% and a specificity of 71.4% between regions identified as healthy and those deemed dysplastic in the tissue samples. A recent study for THz imaging of ex vivo knee cartilage showed that THz imaging has the potential to estimate cartilage thickness using optical delay between the reflections of different layers within the cartilage [48].

4 Technical Consideration for Medical Imaging

Several technical issues must be considered for implementation of the THz imaging systems for in vivo medical imaging. Recently, Pickwell-MacPherson et al. [34] reported an excellent review of the practical considerations for in vivo THz imaging. In this part, they briefly discuss the technical considerations from the point of view of clinical doctors.

The penetration depth is the most crucial issue in in vivo imaging. The role of imaging is not only to diagnose the disease but also to assess its extent. For example, the standard treatment of skin cancers other than melanoma is surgical skin excision, Mohs micrographic surgery, etc. [19]. Preoperative information about tumor invasion depth is essential for surgical planning. Even during the

surgical procedure, a continuous effort to identify the tumor margin is necessary to achieve complete excision. As incomplete tumor removal is the most common cause of recurrence, identification of tumor extent is directly related to prognosis. It is similar in the diagnosis and treatment of breast cancer.

As water has strong absorption across the THz range, the attenuation of THz radiation of various tissues differs according to their water content. The penetration depths range from a few hundred microns in high water-containing tissues to several centimeters in high fat-containing tissues [9, 49].

The penetration depth of THz radiation also depends on the SNR of THz imaging technology. If a THz imaging system uses high-power THz radiation for greater penetration or if the weak signal reflected from the deep layer can be sensitively detected, information about the deep tissue layer can be collected. THz radiation can penetrate much deeper in fat tissue than in water or skin for a given SNR [34].

Current THz imaging systems have a relatively low SNR, limiting their penetration depths. Indeed, systems such as quantum cascade lasers [50] and higher power systems [51] were introduced in order to deliver high power and increase the SNR [12]. The SNR may also be improved through the development of more sensitive detectors [52].

Spatial resolution is another important issue in *in vivo* imaging. The resolution of conventional THz imaging systems is limited by the wavelength [4]. Higher the frequency of THz radiation, better is its spatial resolution. For example, the lateral and axial resolutions obtained with a THz system at 3 THz are 150 and 20 μm , respectively, while those obtained at 1.5 THz are approximately 300 and 40 μm , respectively [9]. However, high-frequency THz radiation cannot penetrate as far as low-frequency THz radiation. Besides, the THz signal power is also lower at higher frequencies. Therefore, the penetration and beam power limits of high-frequency THz radiation reduce its spatial resolution [34].

The conventional resolution is not enough for cellular or molecular imaging trends in medical imaging technology. Doctors expect the imaging at least at the cellular level to enable cell tracking. The current THz imaging is used in a far field with its beam focused using off-axis parabolic mirrors or low-loss dielectric lenses. However, various near-field imaging technologies are under development to overcome the current resolution limitation.

Other critical issues include scan time and a user-friendly interface. Scan time has not gained much interest in the research laboratory. Physicists in the laboratory can easily understand the basic structure of the THz imaging system. However, the clinical setting is totally different from the laboratory setting, and clinical doctors are rather customer-minded. Clinical doctors always consider the advantages and disadvantages when choosing a diagnostic imaging modality. The most important considerations are safety and diagnostic information. The next important considerations are a user-friendly interface, scan time, and cost effectiveness. For example, in the 1970s, when the CT scan was in its early stages of development, it was scientifically regarded as a fascinating novel imaging tool. However, the clinical use was greatly hampered by radiation exposure, poor image quality, long

scan time, large-sized equipment, and cost. CT scanners become an essential diagnostic tool only after many commercial vendors improved them to meet doctors' demands such as short scanning time, image reconstruction for easy understanding of the patient's condition, and affordability.

Currently, a commercial THz imaging probe (TeraView) features a line scan mode that scans the length of the quartz window to acquire a 2D image. In this way, the probe can acquire up to 100 pulses per second [34]. As such, it takes 200 s to obtain a 2D data of a 3×4 cm image with 250- μm spatial resolution that is composed of approximately 20,000 pixels. This acquisition rate has been achieved over the past decade by extensive research for the engineering of THz imaging systems. We hope and expect to achieve a faster acquisition rate, nearly real-time imaging, in order to seamlessly use THz imaging during surgical procedures.

The system geometry, acquisition rate, image resolution, and penetration depth are important aspects that must be carefully considered during the development of a THz system for *in vivo* imaging. It is crucial to realize that both the quality of the THz source and the fundamental optical properties of the sample of interest affect the resulting image resolution and penetration depth. Specifically, the bandwidth and SNR of the THz source affect the axial resolution and penetration depth, respectively. The axial resolution and penetration depth also depend on the refractive index and absorption coefficient: a given bandwidth and SNR samples with a higher refractive index can achieve a lower axial resolution while samples with a lower absorption coefficient have a deeper penetration depth. Thus, it is important to be aware of source and sample parameters to deduce the theoretical capabilities of the system.

5 Conclusion

Still, THz imaging is in the early stages of development. However, it has great potential to be a valuable imaging technique in the future. Over the past decade, great advancements in THz generation, THz detection, and THz imaging have been made. Biomedical applications of THz imaging technology also have drawn immense interest. However, other research-based imaging technologies such as optical coherence tomography, high-frequency ultrasonography, near-infrared spectroscopy, confocal Raman microspectroscopy, multiphoton microscopy are also rapidly developing and currently offer greater penetration, better resolution, and specifically targeted contrast mechanism [6]. Competition between these imaging modalities is fierce and THz imaging should find the niche. It might be high sensitivity of THz signal to water content or imaging capability of sub-surface structure. Indeed, the first few hundred micrometers from the tissue surface are hard to image with other imaging technologies.

In order to move THz imaging from the laboratory to the clinic, acceptance by the medical community and further inclusion of clinicians in THz imaging research field are very important. In addition, continuing improvements of imaging

technology in terms of compact size, low cost, robust and user-friendly systems should be also achieved as well as improvements in fundamental technologies. We hope that THz imaging becomes a household word within a few decades, as have radiography, CT, MRI, and ultrasonography.

References

1. Bodley, R.: A century of medical imaging. *Paraplegia* **33**(12), 685–686 (1995)
2. Padhani, A.R., Miles, K.A.: Multiparametric imaging of tumor response to therapy. *Radiology* **256**(2):348–364
3. Wolbarst, A.B., Hendee, W.R.: Evolving and experimental technologies in medical imaging. *Radiology* **238**(1), 16–39 (2006)
4. Son, J.H.: Terahertz electromagnetic interactions with biological matter and their applications. *J. Appl. Phys.* **105**(10), 102033-1–102033-10 (2009)
5. Park, J.Y., Choi, H.J., Cho, K.-S., Kim, K.-R., Son, J.-H.: Terahertz spectroscopic imaging of a rabbit VX2 hepatoma model. *J. Appl. Phys.* **109**(6), 064704 (2011)
6. Sun, Y., Sy, M.Y., Wang, Y.X., Ahuja, A.T., Zhang, Y.T., Pickwell-Macpherson, E.: A promising diagnostic method: terahertz pulsed imaging and spectroscopy. *World J. Radiol.* **3**(3), 55–65 (2011)
7. Bourne, N., Clothier, R.H., D'Arienzo, M., Harrison, P.: The effects of terahertz radiation on human keratinocyte primary cultures and neural cell cultures. *Altern. Lab. Anim.* **36**(6), 667–684 (2008)
8. Wallace, V.P., Fitzgerald, A.J., Shankar, S., et al.: Terahertz pulsed imaging of basal cell carcinoma ex vivo and in vivo. *Br. J. Dermatol.* **151**(2), 424–432 (2004)
9. Pickwell, E., Wallace, V.P.: Biomedical applications of terahertz technology. *J. Phys. D Appl. Phys.* **39**(17), R301–R310 (2006)
10. Mittleman, D.M., Jacobsen, R.H., Nuss, M.C.: T-ray imaging. *IEEE J. Sel. Top. Quantum Electr.* **2**(3), 679–692 (1996)
11. Wallace, V.P., Macpherson, E., Zeitler, J.A., Reid, C.: Three-dimensional imaging of optically opaque materials using nonionizing terahertz radiation. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **25**(12), 3120–3133 (2008)
12. Reid, C.: Spectroscopic methods for medical diagnosis at terahertz wavelengths. In: Doctoral Thesis, University College London (2009)
13. Zelsmann, H.R.: Temperature dependence of the optical constants for liquid H₂O and D₂O in the far IR region. *J. Mol. Struct.* **350**(2), 95–114 (1995)
14. Fitzgerald, A.J., Berry, E., Zinov'ev, N.N., et al.: Catalogue of human tissue optical properties at terahertz frequencies. *J. Biol. Phys.* **29**(2), 123–128 (2003)
15. Ueno, Y., Ajito, K.: Analytical terahertz spectroscopy. *Anal. Sci.* **24**(2), 185–192 (2008)
16. Tewari, P., Taylor, Z.D., Bennett, D., et al.: Terahertz imaging of biological tissues. *Stud. Health Technol. Inform.* **163**, 653–657 (2011)
17. Woodward, R.M., Cole, B.E., Wallace, V.P., et al.: Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue. *Phys. Med. Biol.* **47**(21), 3853–3863 (2002)
18. Pickwell, E., Cole, B.E., Fitzgerald, A.J., Pepper, M., Wallace, V.P.: In vivo study of human skin using pulsed terahertz radiation. *Phys. Med. Biol.* **49**(9), 1595–1607 (2004)
19. Neville, J.A., Welch, E., Leffell, D.J.: Management of nonmelanoma skin cancer in 2007. *Nat. Clin. Pract. Oncol.* **4**(8), 462–469 (2007)
20. Pickwell, E., Fitzgerald, A.J., Cole, B.E., et al.: Simulating the response of terahertz radiation to basal cell carcinoma using ex vivo spectroscopy measurements. *J. Biomed. Opt.* **10**(6), 064021 (2005)

21. Wallace, V.P., Fitzgerald, A.J., Pickwell, E., et al.: Terahertz pulsed spectroscopy of human Basal cell carcinoma. *Appl. Spectrosc.* **60**(10), 1127–1133 (2006)
22. Woodward, R.M., Wallace, V.P., Arnone, D.D., Linfield, E.H., Pepper, M.: Terahertz pulse imaging of skin cancer in the time and frequency domain. *J. Biol. Phys.* **29**(2–3), 257–261 (2003)
23. Jaskille, A.D., Shupp, J.W., Jordan, M.H., Jeng, J.C.: Critical review of burn depth assessment techniques: part I: historical review. *J. Burn Care Res.* **30**(6), 937–947 (2009)
24. Zhang, H.F., Maslov, K., Stoica, G., Wang, L.V.: Imaging acute thermal burns by photoacoustic microscopy. *J. Biomed. Opt.* **11**(5), 054033 (2006)
25. Atiyeh, B.S., Gunn, S.W., Hayek, S.N.: State of the art in burn treatment. *World J. Surg.* **29**(2), 131–148 (2005)
26. Ruminski, J., Kaczmarek, M., Renkielska, A., Nowakowski, A.: Thermal parametric imaging in the evaluation of skin burn depth. *IEEE Trans. Biomed. Eng.* **54**(2), 303–312 (2007)
27. Mittleman, D.M., Gupta, M., Neelamani, R., Baraniuk, R.G., Rudd, J.V., Koch, M.: Recent advances in terahertz imaging. *Appl. Phys. B Laser. Optic.* **68**(6), 1085–1094 (1999)
28. Taylor, Z.D., Singh, R.S., Culjat, M.O., et al.: Reflective terahertz imaging of porcine skin burns. *Opt. Lett.* **33**(11), 1258–1260 (2008)
29. Taylor, Z.D., Singh, R.S., Bennett, D.B., Tewari, P., Kealey, C.P., Bajwa, N., Culjat, M.O., Stojadinovic, A., Hubschman, J.-P., Brown, E.R., Grundfest, W.S.: THz medical imaging: in vivo hydration sensing. *IEEE Trans. Terahertz Sci. Technol.* **1**(1), 0034(0031–0020) (2011)
30. Morehouse, J.D., Finkelstein, J.L., Marano, M.A., Madden, M.R., Goodwin, C.W.: Resuscitation of the thermally injured patient. *Crit. Care Clin.* **8**(2), 355–365 (1992)
31. Huang, S.Y., Macpherson, E., Zhang, Y.T.: A feasibility study of burn wound depth assessment using terahertz pulsed imaging. In: 4th IEEE/EMBS International Summer School and Symposium on Medical Devices and Biosensors, ISSS-MDBS 2007, pp. 132–135 (2007)
32. Breast Cancer Facts and Figures, Atlanta: American Cancer Society, Inc. (2009–2010) www.cancer.org/acs/groups/content/@nho/documents/document/f861009final90809pdf.pdf
33. von Smitten, K.: Margin status after breast-conserving treatment of breast cancer: how much free margin is enough? *J. Surg. Oncol.* **98**(8), 585–587 (2008)
34. Pickwell-Macpherson, E.: Practical considerations for in vivo THz imaging. *Terahertz Sci. Technol.* **3**(4), 163–171 (2010)
35. Ashworth, P.C., Pickwell-MacPherson, E., Provenzano, E., et al.: Terahertz pulsed spectroscopy of freshly excised human breast cancer. *Opt. Express* **17**(15), 12444–12454 (2009)
36. Fitzgerald, A.J., Wallace, V.P., Jimenez-Linan, M., et al.: Terahertz pulsed imaging of human breast tumors I. *Radiology* **239**(2), 533–540 (2006)
37. Ciesla, C.M., Arnone, D.D., Corchia, A., et al.: Biomedical applications of terahertz pulse imaging. In: Proceedings of SPIE, vol. 3934, pp. 73–81 (2000)
38. Sim, Y.C., Maeng, I., Son, J.-H.: Frequency-dependent characteristics of terahertz radiation on the enamel and dentin of human tooth. *Curr. Appl. Phys.* **9**(5), 946–949 (2009)
39. Goldberg, M., Kulkarni, A.B., Young, M., Boskey, A.: Dentin: structure, composition and mineralization. *Front. Biosci. (Elite Ed)* **3**, 711–735 (2011)
40. Pickwell, E., Wallace, V.P., Cole, B.E., et al.: Using terahertz pulsed imaging to measure enamel demineralisation in teeth. In: Joint 31st International Conference on Infrared Millimeter Waves and 14th International Conference on Terahertz Electronics, IRMMW-THz 2006, pp. 578–578 (2006)
41. Klintworth, G.K.: Corneal dystrophies. *Orphanet J. Rare Dis.* **4**, 7 (2009)
42. Bron, A.J.: Keratoconus. *Cornea* **7**(3), 163–169 (1988)
43. Dougherty, P.J., Wellish, K.L., Maloney, R.K.: Excimer laser ablation rate and corneal hydration. *Am. J. Ophthalmol.* **118**(2), 169–176 (1994)
44. Bennett, D.B., Taylor, Z.D., Tewari, P., et al.: Terahertz sensing in corneal tissues. *J. Biomed. Opt.* **16**(5), 057003 (2011)
45. Van Rudd, J., Mittleman, D.M.: Influence of substrate-lens design in terahertz time-domain spectroscopy. *J. Opt. Soc. Am. B* **19**(2), 319–329 (2002)

46. Jordens, C., Krumbholz, N., Hasek, T., et al.: Fibre-coupled terahertz transceiver head. *Electr. Lett.* **44**(25), 1473–1475 (2008)
47. Ji, Y.B., Lee, E.S., Kim, S.H., Son, J.H., Jeon, T.I.: A miniaturized fiber-coupled terahertz endoscope system. *Opt. Express* **17**(19), 17082–17087 (2009)
48. Kan, W.C., Lee, W.S., Cheung, W.H., Wallace, V.P., Pickwell-Macpherson, E.: Terahertz pulsed imaging of knee cartilage. *Biomed. Opt. Express* **1**(3), 967–974 (2010)
49. Arnone, D.D., Ciesla, C.M., Corchia, A., et al.: Applications of terahertz (THz) technology to medical imaging. In: *Proceedings of SPIE*, vol. 3828(1), pp. 209–219 (1999)
50. Williams, B.S.: Terahertz quantum-cascade lasers. *Nat. Photon.* **1**(9), 517–525 (2007)
51. Tachiki, M.: Emission of terahertz electromagnetic waves driven by an external current in intrinsic Josephson junctions. *Phys. C Supercond.* **468**(7–10), 631–633 (2008)
52. Stillman, W.J., Shur, M.S.: Closing the gap: plasma wave electronic terahertz detectors. *J. Nanoelectr. Optoelectr.* **2**, 209–221 (2007)
53. Crawley, D.A., Longbottom, C., Cole, B.E., et al.: Terahertz pulse imaging: a pilot study of potential applications in dentistry. *Caries Res.* **37**(5), 352–359 (2003)

Principles of Tomography Imaging: From X-ray CT, MRI to THz Tomography

Chang-Beom Ahn and Hochong Park

1 Introduction to Digital Image

A digital image is a digital representation of a 2D (dimensional) image which is expressed mathematically by a 2D matrix of pixel (picture element) values where each element position corresponds to the pixel location in 2D space. To specify a digital image, some parameters related to the image format need to be defined. The image width, w , and the image height, h , in the number of pixels are necessary. Then, the image consists of $w \times h$ pixels and is called a $w \times h$ pixel image, and the dimension of corresponding matrix is also $w \times h$. In a digital image, the pixel values are expressed in a bit format. Hence, the number of bits assigned to each pixel value needs to be defined. If the value is represented with b bits, each value is an integer between 0 and $2^b - 1$, and can represent 2^b different colors. In this way, the overall number of bits required to represent the given digital image becomes $b \times (w \times h)$.

For a grayscale image, the pixel value defines the darkness at each pixel position, and 0 represents black and the maximum value, $2^b - 1$, represents white. For a color image, in general, a specific color is defined by a superposition of three fundamental colors—red, green and blue, and each pixel value is expressed by a 3D vector, $[R \ G \ B]^T$, where element defines the superposition weight in color mixture. Therefore, if b bits are assigned to each pixel of color image, $b/3$ bits are usually assigned to each fundamental color. In this case, $[R \ G \ B] = [0 \ 0 \ 0]$ represents black and $[R \ G \ B] = [2^{b/3} - 1 \ 2^{b/3} - 1 \ 2^{b/3} - 1]$ represents white.

C.-B. Ahn (✉)

Department of Electrical Engineering, Kwangwoon University,
Seoul, 139-701, Korea
e-mail: cbahn@kw.ac.kr

H. Park

Department of Electronics Engineering, Kwangwoon University,
Seoul, 139-701, Korea

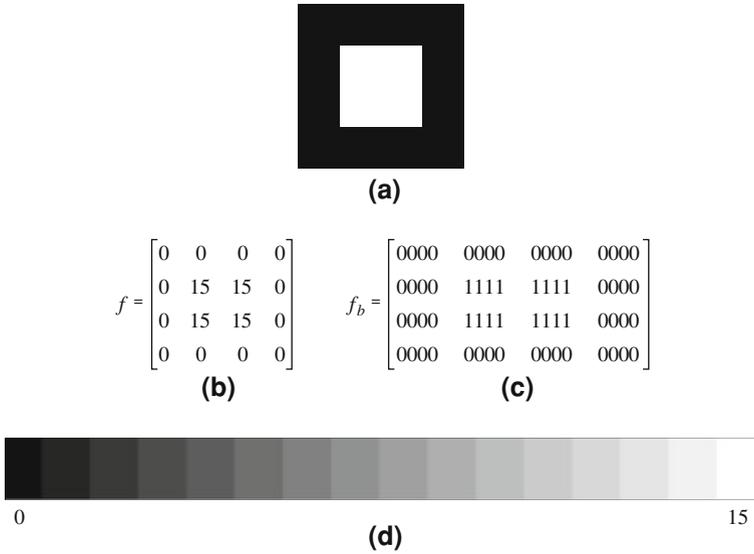


Fig. 1 **a** A 4×4 pixel grayscale image. **b** A matrix representing the image **(a)** with $b = 4$. **c** A matrix representing the image **(a)** with $b = 4$ in a bit format. **d** 16 different grayscale when $b = 4$

For example, a simple 4×4 pixel grayscale image in Fig. 1a is defined mathematically by a 2D matrix in Fig. 1b if $b = 4$. Figure 1c shows the same matrix in a bit format, and Fig. 1d shows the 16 different colors for 4-bit grayscale image.

There exist many factors that determine the quality of image. When a physical size of image is given, the number of pixels defines the spatial resolution. Figure 2a–d shows images with the different number of pixels; with the small number of pixels, the fine detail of object cannot be expressed and a very coarse image in space results. The number of bits per pixel determines the color depth of image; as shown in Fig. 2e and f, the small number of bits results in a very coarse image in depth. Therefore, when the total number of bits, q , for an image is given, it is important to design the appropriate number of pixels, $w \times h$, and the number of bits per pixel, b , such that $q = b \times (w \times h)$.

2 Imaging System

The simplest method to obtain a digital image of a given object is the serial acquisition of the pixel values in a 2D grid. As depicted in Fig. 3, a transmitter outputs a certain pulse to the object such that it is focused at a point of object surface, and a detector receives the reflected waveform. The shape of received waveform depends on the surface characteristic of the object at a given position, and its energy is stored in a digital format with the given number of bits. The object moves sequentially in a 2D grid, and the values at all positions are acquired,

Fig. 2 Examples of image with the different number of pixels and the different number of bits per pixel. **a** 256×256 pixels, 8 bits/pixel. **b** 128×128 pixels, 8 bits/pixel. **c** 64×64 pixels, 8 bits/pixel. **d** 32×32 pixels, 8 bits/pixel. **e** 256×256 pixels, 4 bits/pixel. **f** 256×256 pixels, 1 bit/pixel

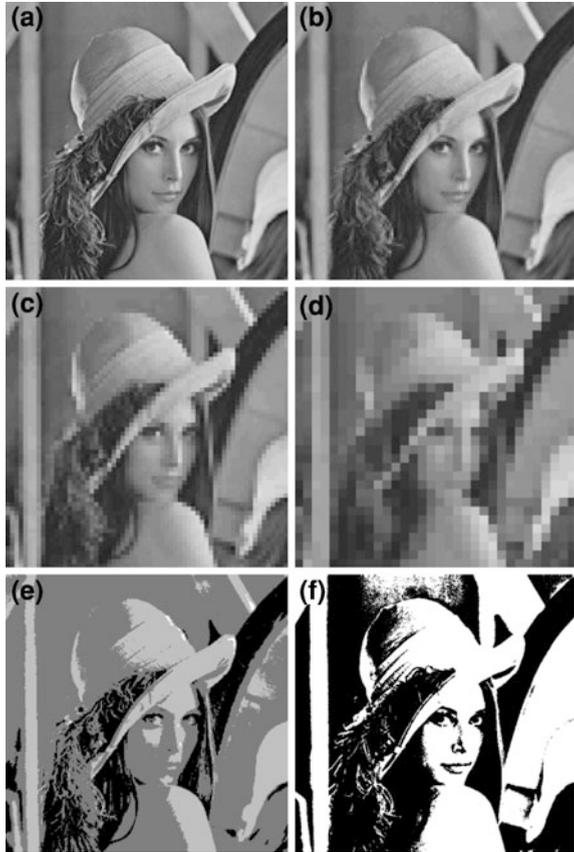
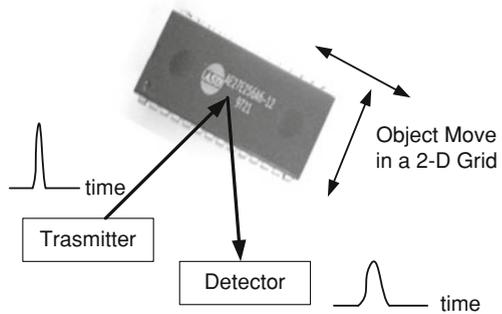


Fig. 3 Imaging set-up with a serial acquisition of pixel value with a single transmitter and a single detector



resulting in the image of the object surface. Of course, the specific image data varies with the kind of pulse used because the reflectivity depends on the physical characteristics of the pulse.

Fig. 4 **a** The pulse reflection at each discontinuity. **b** The received waveform consisting of 3 reflection layers

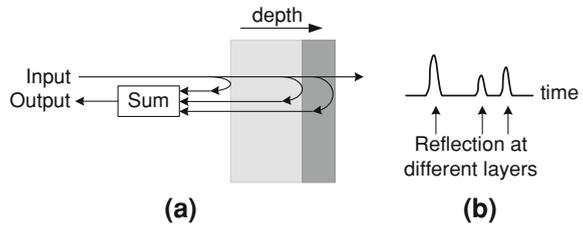
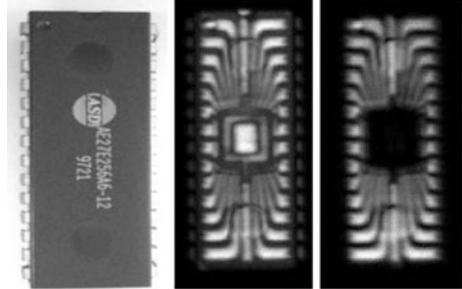


Fig. 5 Example of tomographic images of IC chip obtained using a THz pulse



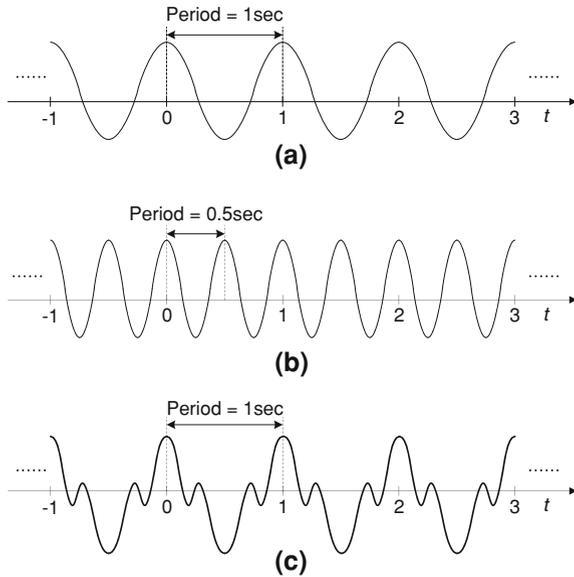
If the pulse penetrates into the object, the same scheme can yield the tomographic image of the object. A portion of pulse is reflected when it meets the discontinuity in material as depicted in Fig. 4a. Hence, the waveform consisting of all reflected pulses at different depth, as shown in Fig. 4b, can be acquired. In this waveform, the gap between the pulses corresponds to the depth, and the energy of each pulse corresponds to the characteristics of each discontinuity. Therefore, if an image is generated using the pulse energy at each time location independently, the tomographic image at a different depth can result. For example, Fig. 5 shows the tomographic images of a surface and two different layers of integrated circuit (IC) chip using a THz pulse.

3 Frequency and Spectrum

3.1 Introduction

To analyze and modify the property of image and to reconstruct the image from the insufficient or degraded data, the concept of frequency is required. The frequency is a well-known concept in physics; a frequency is the number of occurrences of a repeating event per unit time, and 1 Hz (hertz) means one occurrence per 1 s. For example, Fig. 6a shows a cosine waveform as a function of time t , expressed by an equation $f(t) = \cos 2\pi t$. Its frequency is 1 Hz since the same shape repeats in every 1 s. Of course, it is also true to say that its frequency is 0.5 Hz if one event is defined as the waveform of 2 s; one repetition in 2 s yields the frequency of 0.5 Hz. In fact, the waveform in Fig. 6a has a frequency of $1/n$ Hz

Fig. 6 **a** A cosine waveform with a fundamental frequency of 1 Hz. **b** A cosine waveform with a fundamental frequency of 2 Hz. **c** A waveform with a fundamental frequency of 1 Hz and with other higher frequency components



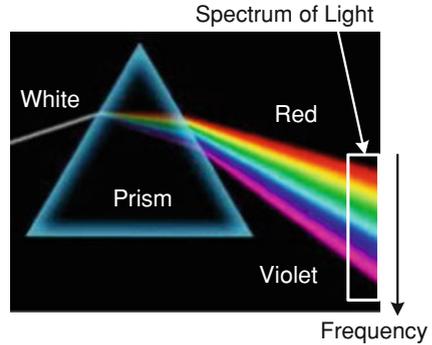
with any positive integer n . To eliminate this ambiguity, a *fundamental frequency* is defined as the largest possible frequencies. Then, all integer fractions of fundamental frequency are also the valid frequencies. A *period* is defined as a reciprocal of frequency, and corresponds to the duration of repeated event in time.

A waveform in Fig. 6b has a fundamental frequency of 2 Hz, and it changes faster than Fig. 6a because of a higher fundamental frequency. Now, let us compare the waveform in Fig. 6a with that in Fig. 6c. Both have the same fundamental frequency of 1 Hz. However, it is obvious that Fig. 6c changes faster than Fig. 6a, and so the fundamental frequency alone cannot identify the waveform characteristics related to the rate of change. The reason is that the fundamental frequency only defines the rate of a repeated event. If the event itself changes very fast, which is the case of Fig. 6c, the overall waveform changes very fast even with a low fundamental frequency; a part of Fig. 6c from 0 to 1 s is changing faster than that in Fig. 6a. This means that Fig. 6c contains the frequency components other than its fundamental frequency of 1 Hz, and those might be higher than 1 Hz, while Fig. 6a contains only a 1 Hz component. Then, an important question on frequency analysis arises: what frequency components are contained in Fig. 6c and what are their magnitudes?

If a waveform $f(t)$ contains many frequency components, it can be represented by Eq. (1), where N different components are assumed.

$$f(t) = \sum_{k=1}^N A_k \cos(2\pi f_k t + \phi_k) \tag{1}$$

Fig. 7 The prism disperses the light according to the frequencies, resulting in the spectrum



Therefore, the problem of frequency analysis comes down to the computation of A_k , f_k , and ϕ_k in Eq. (1) when a waveform $f(t)$ is given.

We already have experienced this kind of frequency analysis in physics. The electro-magnetic wave has different characteristics as a function of frequencies, and for visible light, the color is related to the frequency. If the white light meets a prism, it is refracted at the prism surface and enters the prism at a different angle as a function of frequencies, as shown in Fig. 7. Then, the dispersed light with different colors results and is called the *spectrum* of light. Therefore, the role of prism is a frequency analysis of light, and shows each frequency component as a function of frequencies.

As a prism analyzes the frequency of light, there exists a mathematical tool which analyzes the frequency of waveform and yields its spectrum. Let the time index be n and the frequency index be u . Then, a spectrum $F(u)$, which is a function of u , is another way to represent the waveform $f(n)$. Two signals $f(n)$ and $F(u)$ have the identical information, and only differ in the representation domain which is called the *time domain* and the *frequency domain*, respectively.

3.2 Fourier Transform

When $f(n)$ is given, its spectrum $F(u)$ can be computed by the *Fourier transform*, which is defined by Eq. (2), where $j = \sqrt{-1}$.

$$F(u) = \sum_{n=-\infty}^{\infty} f(n) \exp(-j2\pi un) \quad (2)$$

Also, $F(u)$ is converted back to $f(n)$ by the *inverse Fourier transform* defined by Eq. (3).

$$f(n) = \int_{-0.5}^{0.5} F(u) \exp(j2\pi un) du \quad (3)$$

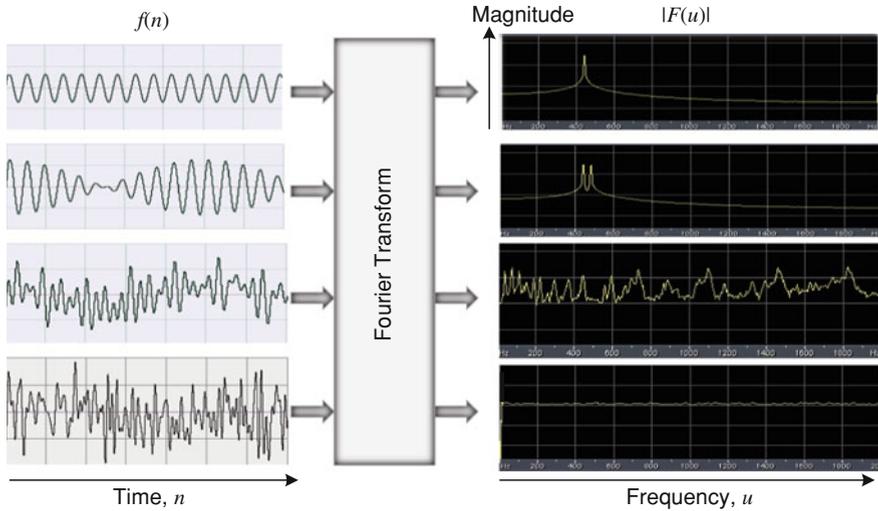


Fig. 8 Examples of waveforms and their spectra

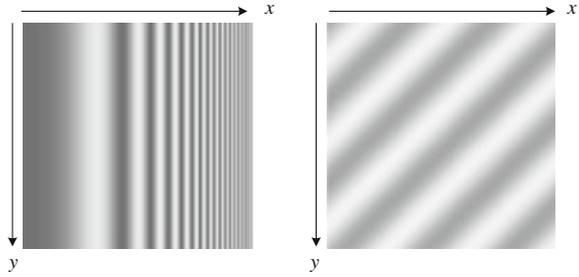
These formulae contain a complex number, j , and in general, the spectrum $F(u)$ has complex numbers and is represented by $F(u) = |F(u)|\angle F(u)$. Therefore, when plotting the spectrum, only its magnitude $|F(u)|$ is generally used. When $f(n)$ is a real-numbered waveform, $F(u)$ has a property of symmetry defined as Eq. (4). That is, $F(u)$ of a negative u is just a complex conjugate of $F(u)$ of a positive u .

$$F(u) = F^*(-u), |F(u)| = |F(-u)| \tag{4}$$

Figure 8 shows some examples of the waveform $f(n)$ and the magnitude of spectrum $|F(u)|$. Figure 8a has only one frequency component, and Fig. 8b has two frequency components with the same magnitude. Figure 8c has a very complex spectrum with many frequency components with varying magnitudes, and Fig. 8d has almost a flat spectrum where all frequency have uniform magnitude. This waveform is called a *white* signal, similar to a white light with all colors in equal strength. Computing the spectrum in the frequency domain is called a *frequency analysis*, and returning to the time domain is called a *frequency synthesis* because all frequency components defined in $F(u)$ are integrated to result in $f(n)$.

So far, the frequency and the spectrum of 1D time-domain signal, $f(n)$, are explained. Now, the concept of frequency is expanded to the 2D spatial domain, since the frequency analysis of 2D image is required. In the 2D spatial domain, a frequency is the number of occurrences of a repeating event per unit length in the 2D space. Let two directions in space are indexed by x and y . For an image in Fig. 9a, a vertical line in y direction at a fixed x position has no change in pixel value. Hence, the image has zero frequency in the y direction at all x positions. At a fixed y position, the pixel values of horizontal line in the x direction experience the variation, which shows that the image has non-zero frequency in the

Fig. 9 **a** An image with zero frequency in the y direction and non-zero frequency in the x direction. **b** An image with the same frequency property in the x and y directions



x direction. Furthermore, from left to right, the rate of change increases steadily, and the image has a low frequency in the x direction in small x and a high frequency in large x . For an image in Fig. 9b, the change of pixel values in the x and y directions have the same property and the frequency in both directions is the same. In this way, the *spatial frequency* in 2D spatial domain (x, y) is defined, and the spectrum of image $f(x, y)$ is expressed as $F(u, v)$, where u and v are the frequency indices in x and y directions, respectively.

The *2-D Fourier transform* and the *2-D inverse Fourier transform* are defined by Eq. (5), which are the modified equations of Eqs. (3) and (4) from 1D to 2D. It is assumed that $f(x, y)$ is an $N \times N$ pixel image. The symmetry property, $F(u, v) = F^*(u, v)$, also holds.

$$F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \exp[-j2\pi(ux + vy)] \quad (5)$$

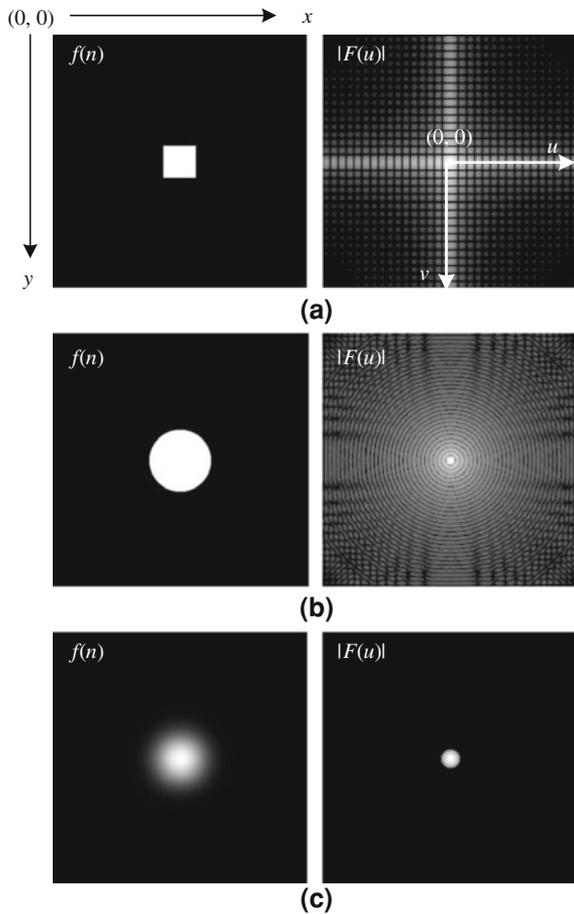
$$f(x, y) = \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} F(u, v) \exp[j2\pi(ux + vy)] dudv$$

Figure 10 shows some examples of images and their spectra, where the origin of spectrum, $(u, v) = (0, 0)$, is located at the center of spectrum. In all spectra, the zero frequency has the largest magnitude since most of the image is a flat region of constant pixel values. In Fig. 10a and b, the abrupt change in pixel values between black and white at the edge of shape yields the high frequency components. In Fig. 10c, the change of pixel values at the circle edge is very smooth, and no high frequency components exist.

3.3 Filtering

There are many image operations that modify the characteristics of image according to a certain target. For example, an amplification by a factor of 2 can be realized simply by $g(u, v) = 2f(u, v)$. When a modification of spectrum is necessary, a more sophisticated operation must be implemented, and the spectral shaping of image is called an image *filtering*.

Fig. 10 Examples of images and their spectra.
a Rectangular. **b** Circle.
c Smoothed circle



The filtering of signal is first explained in the 1D domain for a better representation and understanding. A 1D signal $f(x)$ is given, and a new signal $g(x)$ is to be generated by removing the high frequency components in $f(x)$. Since no information in frequency of $f(x)$ is revealed in the x domain, $f(x)$ is first transformed to the frequency domain and $F(u)$ is obtained. Assuming that $F(u)$ is given by Fig. 11a, the required spectrum $G(u)$ is generated as in Fig. 11b after the high frequency components with $u > u_0$ are removed with $G(u) = 0, u > u_0$ and $G(u) = F(u), u < u_0$. Then, $G(u)$ corresponds to the spectrum of the target signal $g(x)$, and $G(u)$ is finally transformed into the x domain, resulting in the required $g(x)$.

Obtaining $G(u)$ from $F(u)$ is simply expressed by the spectral multiplication, $G(u) = H(u)F(u)$, where $H(u)$ is the rectangular spectrum shown in Fig. 11c. Hence, a multiplication by $H(u)$ performs the specific spectral shaping in the frequency domain, and $H(u)$ is called a *frequency response* of filter. If $H(u)$ is designed arbitrarily, any kind of spectral shaping can be realized, and so the filter

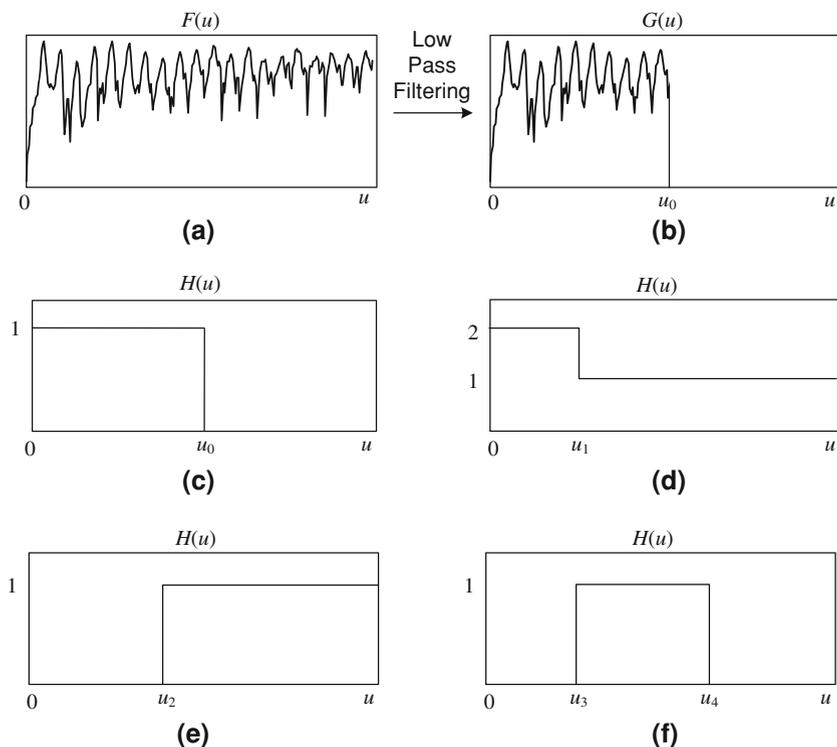


Fig. 11 Spectral shaping. **a** Spectrum of input image. **b** Spectrum of output image. **c** Frequency response of low-pass filter. **d** Frequency response of filter for low frequency boosting. **e** Frequency response of high-pass filter. **f** Frequency response of band-pass filter

design corresponds to the design of $H(u)$. For example, if $H(u)$ in Fig. 11d is applied to $F(u)$, the function of filter becomes the boosting of low frequency by amplifying the low frequency components below u_1 by a factor of 2.

In general, the removal of high frequency component is called a low-pass filtering. Similarly, a high-pass filtering is the removal of the low-frequency components, and a band-pass filtering is the removal of low- and high-frequency components, keeping the frequency band. Figures 11e and f shows the examples of $H(u)$ for a high-pass filtering and a band-pass filtering.

An image filtering from $f(x, y)$ to $g(x, y)$ consists of three steps—Fourier transform of $f(x, y)$, multiplication by $H(u, v)$, and inverse Fourier transform of $G(u, v)$, as depicted in Fig. 12. Then, is it possible to compute $g(x, y)$ directly from $f(x, y)$ without computing the spectra?

By a mathematical derivation, the direct relation between $g(x, y)$ and $f(x, y)$ is given by Eq. (6), where FT and IFT stand for the Fourier and the inverse Fourier transforms, respectively.

Fig. 12 Filtering procedure through the spectrum computation

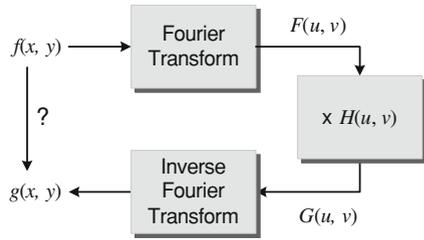


Fig. 13 Image filtering in the spatial and the frequency domains



$$\begin{aligned}
 g(x, y) &= \text{IFT}\{H(u, v)F(u, v)\} = \text{IFT}\{\text{FT}[h(x, y)] \text{FT}[f(x, y)]\} \\
 &= \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} h(x - p, y - q)f(p, q)
 \end{aligned}
 \tag{6}$$

$h(x, y)$, which is the inverse Fourier transformed signal of $H(u, v)$, is called a *impulse response* of filter. This equation contains a complicated operation between $h(x, y)$ and $f(x, y)$, and is simply defined as the *convolution* between $h(x, y)$ and $f(x, y)$, and expressed using a symbol $*$ as in Eq. (7).

$$g(x, y) = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} h(x - p, y - q)f(p, q) \equiv h(x, y) * f(x, y)
 \tag{7}$$

Hence, the filtering is defined in the spatial domain by the convolution between the input image and the impulse response of filter. Finally, a filtering can be described both in the spatial and the frequency domains, summarized in Fig. 13; a convolution in the spatial domain or a multiplication in the frequency domain.

4 Reconstruction Algorithm of the Computed Tomography

Image reconstruction from projection data is a well-known mathematical problem (also known as inverse Radon transform) whose solution is used in many scientific and medical areas such as seismic data processing, computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI).

From X-ray physics, X-ray is attenuated when it passes through an object. The detected signal is mathematically represented as a function of line integral of the attenuation coefficients along the path of the X-ray given by

$$I(x) = I_o \cdot \exp\left[-\int f(x, y)dy\right],
 \tag{8}$$

where $I(x)$ is the intensity of the detected ray, I_o is the intensity of the incident ray, and $f(x, y)$ is the attenuation coefficient at (x, y) .

Thus one-dimensional line-integral data or projection data is obtained if a linear array of X-ray detector is used.

From two dimensional Fourier transform,

$$F(u, v) = \iint f(x, y) e^{-j2\pi(ux+vy)} dy dx \quad (9)$$

Thus

$$\begin{aligned} F(u, 0) &= \int \left[\int f(x, y) dy \right] e^{-j2\pi(ux)} dx \\ &= \int p(x) e^{-j2\pi ux} dx \\ &= P_0(u), \end{aligned} \quad (10)$$

where $P_0(u)$ is the spatial frequency data along the line passing through the origin with a rotational angle of 0° from the u axis and $p(x)$ denotes the projection data at x projected along y axis.

Since Eqs. (9) and (10) are hold for an arbitrary angle of rotation of the coordinate, Eq. (10) may be generally written as

$$\begin{aligned} P_\phi(w) &= \int p_\phi(r) e^{-j2\pi wr} dr \\ &= \mathfrak{F}_1 \{ p_\phi(r) \}, \end{aligned} \quad (11)$$

where $P_\phi(w)$ denotes the spatial frequency data on the line passing through the origin with a rotational angle of ϕ from the u axis, and $p_\phi(r)$ is the projection data projected normal to the line passing through the origin with a rotational angle of ϕ from the x axis. The X-ray detected at $r = r_l$ shown in Fig. 14 passes through the points given by

$$x \cos \phi + y \sin \phi = r_l \quad (12)$$

Note that ϕ is the same angle in both the spatial and the spatial frequency domains. Thus the spatial frequency data is obtained by the one-dimensional Fourier transform of the projection data along the line normal to the projection path in the rotated coordinate as shown in Fig. 14. This relation is also known as projection theorem or central slice theorem.

If the projection data is measured regularly with a constant incremental angle, two dimensional spatial frequency data is obtained in a polar coordinate by one-dimensional Fourier transform of the projection data. Thus the reconstruction is to find $f(x, y)$ by a two dimensional inverse Fourier transform of the spatial frequency data in the polar coordinate. From two dimensional Fourier transform in the rectangular coordinate,

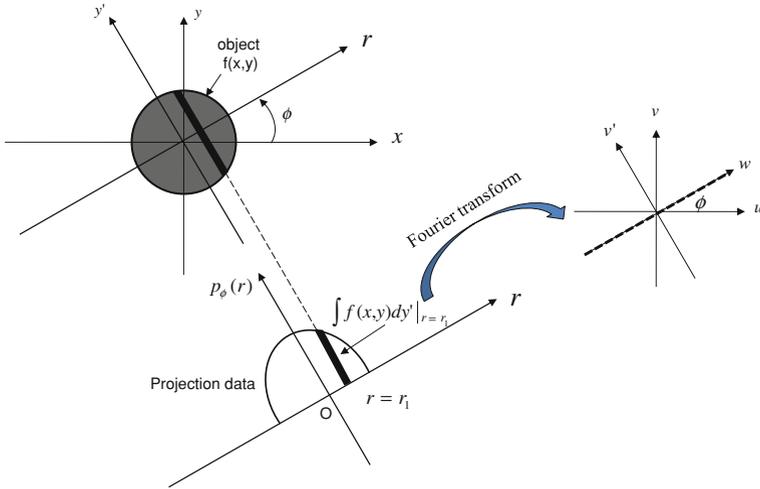


Fig. 14 A line of spatial frequency data with a rotational angle ϕ from u axis in the spatial frequency domain is obtained from one-dimensional Fourier transform of the projection data with the same rotational angle ϕ from x axis in the spatial domain

$$\begin{aligned} \hat{f}(x, y) &= \mathfrak{F}_2^{-1}[F(u, v)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) \exp[i2\pi(xu + yv)] \, dudv \end{aligned} \tag{13}$$

By the coordinate transformation from the rectangular coordinate to the polar coordinate in the spatial frequency domain, i.e., $u = w \cos \phi$ and $v = w \sin \phi$, then Eq. (13) becomes

$$\hat{f}(x, y) = \int_0^\pi \int_{-\infty}^{\infty} P_\phi(w) \exp[i2\pi w(x \cos \phi + y \sin \phi)] |J| \, dw d\phi, \tag{14}$$

where J is Jacobian given by

$$|J| = \begin{vmatrix} \frac{\partial u}{\partial w} & \frac{\partial v}{\partial w} \\ \frac{\partial u}{\partial \phi} & \frac{\partial v}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos \phi & \sin \phi \\ -w \sin \phi & w \cos \phi \end{vmatrix} = |w| \tag{15}$$

By substituting Eq. (12) for general r and Eq. (13) into Eq. (14), Eq. (14) becomes

$$\begin{aligned} \hat{f}(x, y) &= \int_0^\pi \int_{-\infty}^{\infty} |w| P_\phi(w) \exp[i2\pi wr] \, dw d\phi \\ &= \int_0^\pi \check{p}_\phi(r) \, d\phi, \end{aligned} \tag{16}$$

where

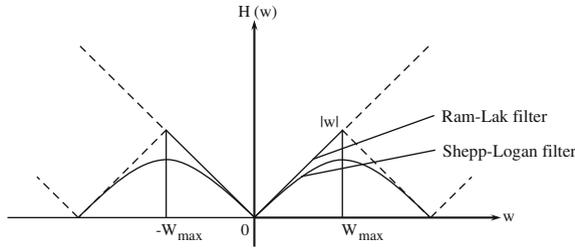


Fig. 15 Two filter kernels for the filtered backprojection are shown: Theoretical filter, $|w|$ known as Ram-Lak filter and Shepp-Logan filter, $\frac{2W_{max}}{\pi} \left| \sin\left(\frac{\pi}{2} \cdot \frac{w}{W_{max}}\right) \right|$ whose derivatives are continuous at the boundaries $+W_{max}, -W_{max}$ to reduce Gibb's phenomena

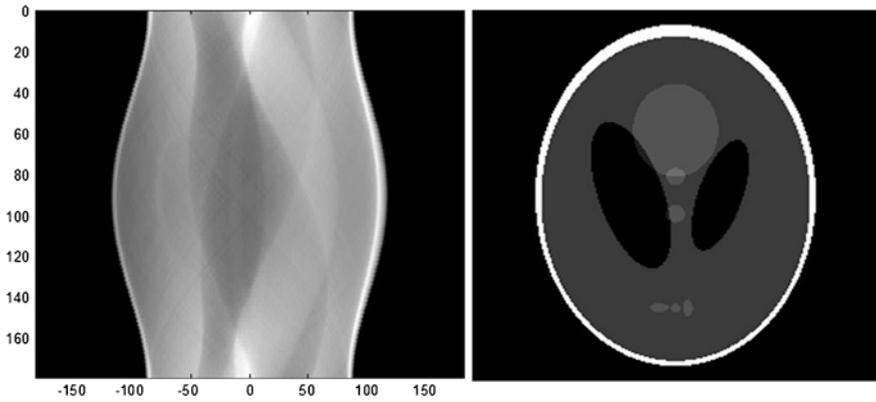


Fig. 16 An example of projection data (known as sinogram) and corresponding reconstructed image by the filtered backprojection is shown

$$\begin{aligned}
 \tilde{p}_\phi(r) &= \int_{-\infty}^{\infty} |w| P_\phi(w) \exp(i2\pi wr) dw \\
 &= \mathfrak{F}_1^{-1} \{ |w| P_\phi(w) \} \\
 &= \mathfrak{F}_1^{-1} \{ |w| \} * p_\phi(r)
 \end{aligned}
 \tag{17}$$

In Eq. (17) $\mathfrak{F}_1^{-1} \{ |w| \}$ is the inverse Fourier transform of the Jacobian which becomes a filter kernel pre-calculated. In deriving Eq. (17), convolution property of the Fourier transform is used. Some variation of the filter kernel is shown in Fig. 15. The reconstruction process given by Eqs. (16) and (17) is also known as filtered backprojection.

An example of projection data and corresponding reconstructed image is shown in Fig. 16 using a mathematical phantom. The phantom is composed of a set of ellipses whose projection data is mathematically known. The projection data displayed in (r, ϕ) space is called sinogram.

5 Principles of Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a tomographic imaging technique based on the nuclear magnetic resonance (NMR) phenomena. It produces high resolution and high contrast images with various imaging parameters, such as density, T_1 and T_2 relaxation times, chemical shift, flow, diffusion, perfusion, etc. The spins in a magnetic field have a precession frequency proportional to the magnetic field given by Larmor equation as

$$f = \gamma/2\pi \cdot B = \gamma \cdot B, \quad (18)$$

where γ is gyromagnetic ratio divided by 2π given by 42.57 MHz/T. A super-conductive magnet with bore diameter of 0.8–1 m is usually used to make a main magnetic field of 0.5–7 T. Main field strength lower than 0.5 T is obtained by a permanent magnet with a gap of 0.4–0.5 m. The signal-to-noise ratio of the magnetic resonance imaging is generally proportional to the field strength. Radio frequency (RF) excitation produces a free induction decay (FID) signal which is used for NMR spectroscopy and MRI. The RF excitation and reception of the FID signal are achieved by RF coils. Various RF coils are designed to fit specific organ imaging. The FID signal decays exponentially with a time constant of T_2 , and returns to equilibrium state with a time constant of T_1 . Since the hydrogen spins in normal and malignant tissues have different T_1 and T_2 values in general, the T_1 and T_2 relaxation times play important role in the mechanism of contrast in MRI.

For MRI, gradient fields are added to the main magnetic field. The gradient field is a linearly increasing magnetic field as a function of space. If we assume main magnetic field and gradient fields are in z direction without loss of generality, the magnetic field at location (x, y, z) is represented as

$$B(x, y, z; t) = [B_o + G_x(t) \cdot x + G_y(t) \cdot y + G_z(t) \cdot z] \vec{z}, \quad (19)$$

where B_o denotes main magnetic field, G_x , G_y , and G_z are slopes of x , y , and z directional gradient fields, respectively. Note that all the three gradient fields are in z direction with variations along x , y , and z directions, respectively.

Spatial position of the spins is encoded by the gradient field (Eq. (19)), which is then modulated by the precession frequency of the spins by Larmor equation (Eq. (18)). The positional information is restored by the Fourier transform of the frequency encoded signal as schematically shown in Fig. 17. All the hydrogen spins experience the same magnetic field on the left side of Fig. 17 without gradient field. Thus the FID signal is close to a sine wave and Fourier transform of the FID signal shows a single peak. On the other hand, the hydrogen spins on the right side of Fig. 17 experience magnetic field proportional to the spatial position due to the linearly increasing gradient field. The FID signal is a summation of these frequencies, and a projection data is obtained after Fourier transform of the FID signal.

A timing diagram (pulse sequence) of the application of RF field, three gradient fields, and acquisition of the signal is shown in Fig. 18. A 90° selective RF pulse is

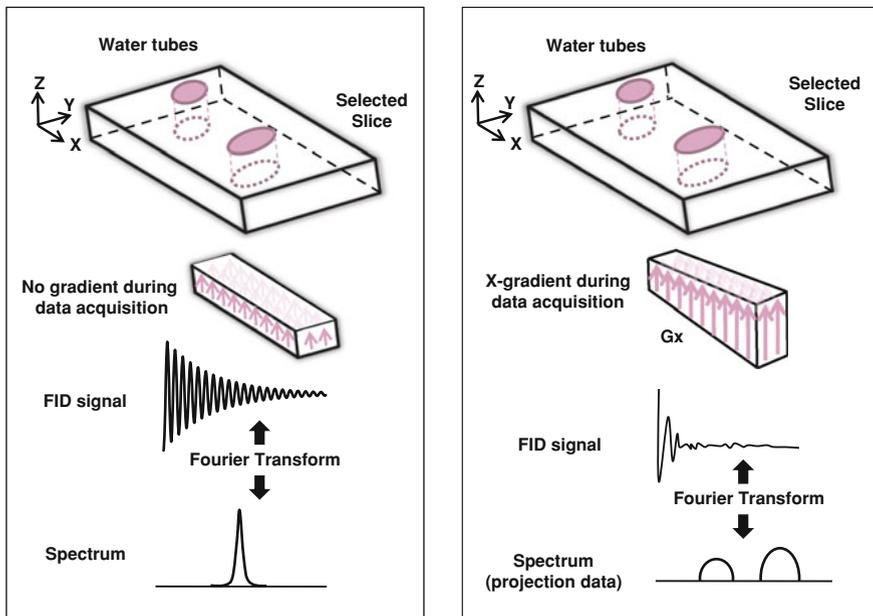


Fig. 17 FID signals with and without gradient field during the data acquisition period are schematically shown. Single peak is detected after Fourier transform of the FID signal when the gradient field is off (*left*), while the projection data is obtained after Fourier transform of the FID signal when the gradient field is on (*right*)

applied with z directional gradient field to make a slice selection ($z = z_0$). A 180° selective RF pulse is applied to refocus an echo signal (spin echo). The acquired signal is expressed as

$$s(g, t) = \iint \rho(x, y; z_0) \exp\{i \gamma (g_x T + G_y t)\} dx dy \tag{20}$$

By change of variables,

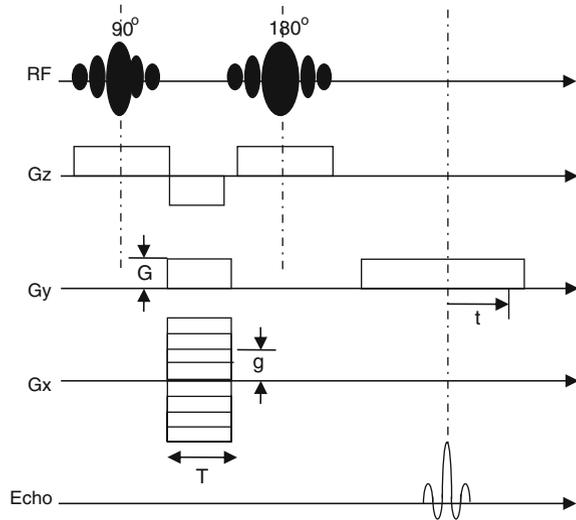
$$\begin{aligned} s\left(\frac{k_x}{\gamma T}, \frac{k_y}{\gamma G}\right) &= \iint \rho(x, y; z_0) \exp\{i 2\pi (k_x x + k_y y)\} dx dy \\ &= \mathfrak{S}_2^{-1}\{\rho(x, y; z_0)\}, \end{aligned} \tag{21}$$

where

$$2\pi k_x = \gamma g T, \quad 2\pi k_y = \gamma G t \tag{22}$$

Reconstruction can be made by the two dimensional Fourier transform of Eq. (21) given by

Fig. 18 Pulse sequence for the spin-echo based two dimensional Fourier imaging. Selection gradient is applied along z direction, readout (or frequency encoding) gradient is applied along y direction, and phase encoding gradient is applied along x direction, respectively



$$\begin{aligned} \rho(x, y; z_o) &= \mathfrak{I}_2 \left\{ s \left(\frac{k_x}{\gamma T}, \frac{k_y}{\gamma G} \right) \right\} \\ &= \iint s \left(\frac{k_x}{\gamma T}, \frac{k_y}{\gamma G} \right) \exp \{ -i 2\pi (k_x x + k_y y) \} dx dy \end{aligned} \quad (23)$$

With more general gradient waveforms, Eq. (22) may be written as

$$k_x(t) = \gamma \int_0^t G_x(t) dt \quad \text{and} \quad k_y(t) = \gamma \int_0^t G_y(t) dt \quad (24)$$

Using Eq. (24) more general pulse sequences can be designed, such as echo planar imaging (EPI) or spiral scan imaging (SSI) for ultra fast MRI. An example of the acquired spin echo signal at 1.0 T whole body MRI system and corresponding reconstructed image is shown in Fig. 19.

6 Principle of Terahertz Tomography

Terahertz imaging and spectroscopy systems with terahertz electromagnetic wave (T-ray) have been successfully applied to various fields including security, inspection, and biomedical imaging. T-ray biomedical imaging has great potential due to the interaction of T-ray with water molecules in the biological tissue. Terahertz tomography is an emerging area for investigating small biological samples nondestructively.

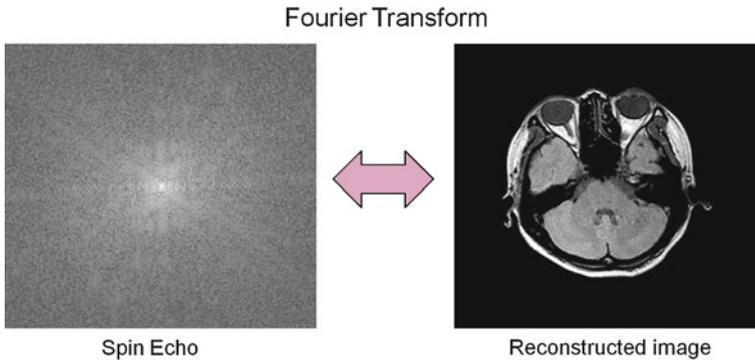


Fig. 19 Acquired signal by the 2D spin-echo based Fourier imaging sequence and the corresponding reconstructed image are shown

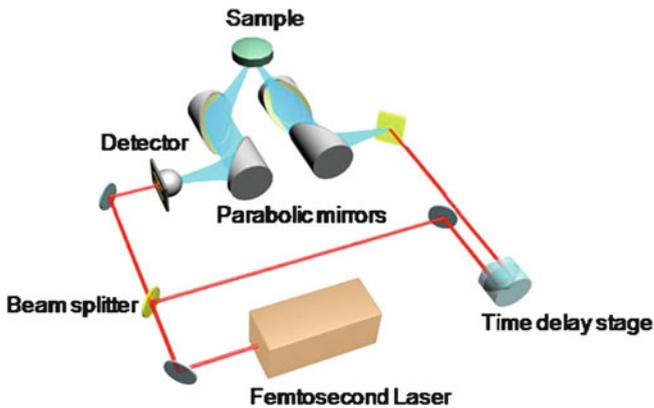


Fig. 20 Terahertz reflection-type time-domain spectroscopy system is used for the terahertz tomography

A typical reflection-type terahertz time-domain spectroscopy (TDS) system shown in Fig. 20 can be used for tomography. The T-ray is emitted from the terahertz transmitter (femtosecond laser) to the object; the reflected time domain signal of the T-ray is acquired by a detector. Translational and rotational scanning mechanisms are equipped with the system. The time resolution of the T-ray is set by the time delay stage. The transmission distance of the T-ray by the unit time delay becomes the depth resolution. The spatial resolution in the transverse plane is set by the scanning mechanism. Since the depth information is directly obtained from the time delay of the T-ray, tomographic image is obtained from two dimensional raster scanned images for a given range of time delays.

The principle of T-Hz tomography is examined with a stair-like phantom in Fig. 21 (top left). The phantom is made by stacking plastics of height of 0.3 mm

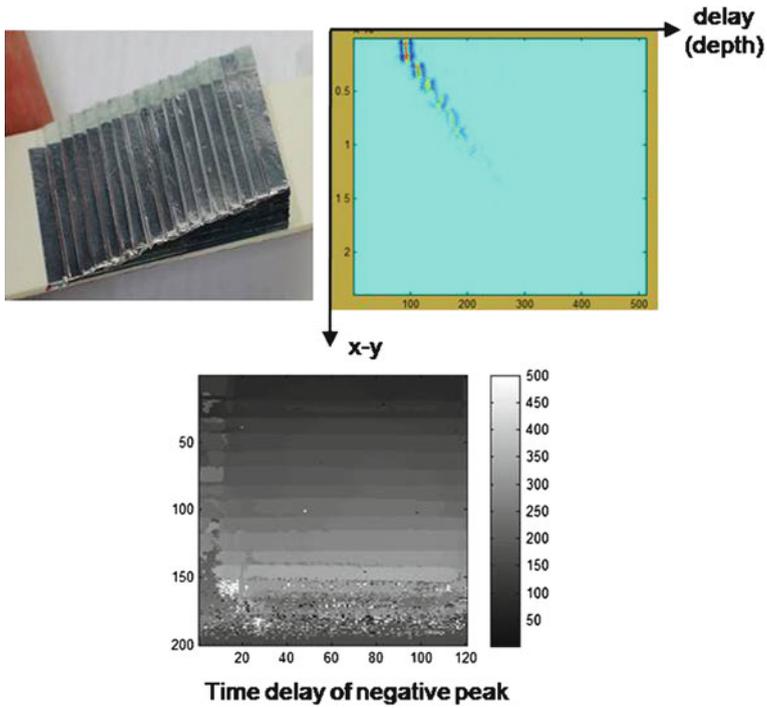


Fig. 21 Principle of terahertz tomography by the time delay of the T-ray is shown with a stair-like phantom (*top left*). Detected signal is shown in *top right*. The time delay is displayed at the *bottom* which is equivalent to the depth of the phantom. The spatial resolution in the transverse plane is 250 μm and the depth resolution is 21 μm

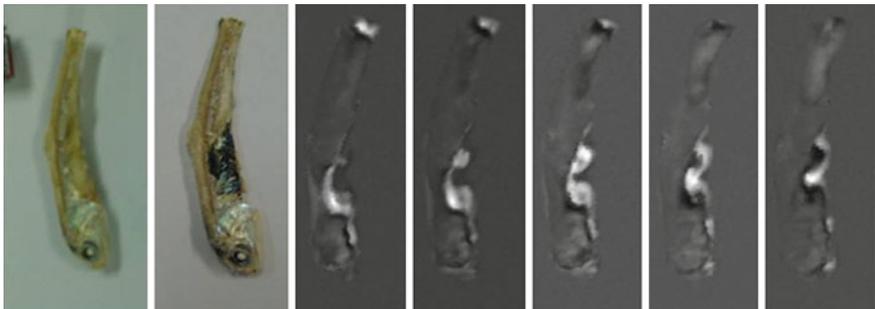


Fig. 22 Terahertz tomographic images of the parched anchovy. Photograph of the sample, dissection image, and terahertz tomographic images from *top to bottom* are shown sequentially from *left to right*. The transverse resolution is 250 μm and the tomographic slice thickness is 70 μm

with aluminum tape on the surface. The size of the phantom in the transverse plane is 50 mm \times 30 mm. The detected T-ray is shown in top right. Note a stair-like pattern due to the different time delays from the stairs of the phantom. The image of the time delay equivalent to the depth of the phantom is shown at the bottom. The spatial resolution in the transverse plane is 250 μ m and the depth resolution is 21 μ m.

Terahertz tomography of a patched anchovy is shown in Fig. 22. The sample size is 10 mm \times 38 mm in the transverse plane. A photograph of the sample, a dissection image, and five terahertz sectional images from top to bottom are shown sequentially from left to right. The transverse resolution is 250 μ m and the slice thickness is 70 μ m. Maximum intensity is projected among 5 time delays (70 μ m). Some internal organs are observed in the terahertz images similar to those in the dissection image.

Introduction to Quantum Sensors in Cryogenic Particle Detection

Yong-Hamb Kim and Sun Kee Kim

Abstract Cryogenic detectors have been important tools in many aspects of science because their sensitivities can provide more than extreme limits of conventional semiconductor based detectors. The sensor developments in cryogenic particle detection are based on the precise measurement of noble properties of condensed matter in low temperatures. The major measurement technologies originate from quantum measurements, phase transitions and superconducting electronics. Although the early developments of cryogenic detectors were initiated by applications to elementary particle physics, they have been adopted in biology, forensics, and security as well as astronomy and nuclear science. Various types of cryogenic detectors cover a wide energy range from THz radiations to hundreds MeV particles. We review the recent development of sensor technologies in cryogenic particle detection. The measurement principles are covered together with applications to elementary particle physics and THz measurement.

1 Introduction

Scientific endeavors searching for the answers to the fundamental questions of nature are often limited by the signal detection technology. For example, the search for neutrinoless double beta decays to study the nature of neutrino, one of

Y.-H. Kim (✉)
Korea Research Institute of Standards and Science,
Daejeon 305-340, Korea
e-mail: yhkim@kriss.re.kr

S. K. Kim
National University, Seoul 151-742, Korea
e-mail: skkim@hep1.snu.ac.kr

the fundamental particles of nature, requires high energy resolution to separate the irreducible background from two neutrino double beta decay from the same isotope. The direct detection of dark matter particle, the major portion of the matter in the universe with yet unknown identity, requires the measurement of energies as low as few keV and additional capability of separation of nuclear recoils from electron recoil background. The anisotropy measurement in microwave background of the universe requires extreme resolution for radiation power of cosmic microwave background with frequencies up to tenth of THz.

These kinds of experiments involves measurement of energy spectrum of certain type of radiations. For double beta decay search, energy of two electrons from the decay needs to be measured precisely.

The energy deposited in the matter by an interaction of radiation with atoms in the matter can be converted into measurable quantities such as ionized charged particles, scintillation lights and phonons. The ionization can be measured by collecting the charges with the applied electric field. Scintillation light are easily measurable by photon sensors such as photomultiplier tubes or photo-diodes. However, the major portion of energy deposition are converted into the phonons. It infers that phonon measurement can be more effective than ionization and scintillation measurements.

Accurate measurement of phonons created by a particle or radiation absorption is not tribal particularity at room temperature. The phonons exist naturally in any condensed matter. Their statistical frequency/energy distribution determines the by the temperature of the matter. When an absorber is thermally attached to a heat reservoir the thermal energy of the absorber fluctuates resulting in changing phonon distribution in time. This thermal fluctuation can easily overwhelm over the phonons created by particle detection. At low temperatures, however, the available thermal energy in other words heat capacity is reduced so as the fluctuations are reduced. Moreover, because the heat capacity of the absorber is typically lower at low temperatures, the resulting temperature increase caused by an energy input is increased.

The intrinsic energy resolution of the detector is determined by the fluctuation of produced quanta in the energy loss procedure. The energy needed to produce an electron-hole pair or a scintillation photon is more than a few to 10 eV. However, in phonon measurement case, the average energy in a temperature (T) is about $k_B T$ where k_B is the Boltzmann constant. At 100 mK, $k_B T$ is close to 10^{-5} eV.

In cases of temperature measurements that include electron temperature in metal absorber, the statistical fluctuation in thermal energy limits the intrinsic resolution of thermal detectors. When the total thermal energy of a detector with a heat capacity C is approximated as CT , the average number of energy quanta can be expressed by $N \approx CT/k_B T$. Eventually, the statistical thermal noise or the limit of energy resolution due to thermal energy fluctuation is

$$(\delta E)_{\text{rms}} \approx k_B T (N)^{1/2} \approx (k_B T^2 C)^{1/2}. \quad (1)$$

The fluctuation limit the resolution of a detector with 0.2 pJ/K of heat capacity at 100 mK to 1 eV in the full width at half maximum (FWHM). In an application to X-ray spectrometers 1.8 keV FWHM was measured for 6 keV X-rays at NASA GSFC [1]. In contrast, in an ideal Si-based charge collection detector, the resolution is suffered by the number statistics of electron-hole pairs and limited to 120 eV FWHM in 6 keV measurement even though the theoretical Fano factor 0.115 of Si is used [2].

The scope of this chapter is to introduce superconducting quantum sensors to new scientists who are not experts on the cryogenics and particle detection. The detectors that have been developed recently for high resolution and are capable of THz measurements are mainly discussed with their working principles. Ones who research for more details on the fabrications and specific concerns on some applications are recommended to look up the references listed in this chapter.

In the following section, the measurement principle of cryogenic detectors is discussed. Various quantum sensors widely used in the cryogenic detector is introduced with basic properties and measurement circuits in Sect. 3. Applications to particle physics experiments such as dark matter and double beta decay search using the cryogenic detectors are reviewed in Sect. 4. Section 5 is especially focused on the recent development is THz radiation detection techniques.

2 Principle of the Cryogenic Detectors

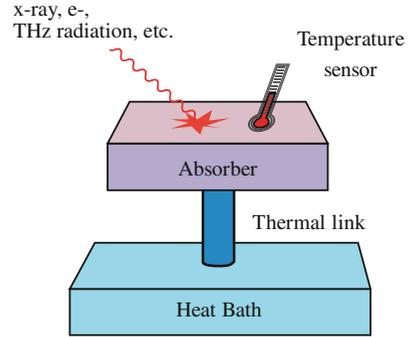
Research in cryogenic detectors is primarily motivated by particle physics and astrophysics. This is because low temperature detectors can provide energy sensitivities and thresholds beyond the limit of conventional semiconductor based detectors. Remarkable improvements for the sensitivities and thresholds have been achieved. The principles, motivations, and applications of the low temperature detectors have been reviewed in Ref. [3].

The low temperature detectors are found to be useful not only in particle and astrophysics, but also in condensed matter and biophysics. Currently, these cryogenic detectors are used in many different fields of physics, such as cosmological dark matter searches, X-ray astrophysics, double beta decay experiments and direct measurements of neutrino mass, as well as THz applications for astronomy and security.

2.1 Interaction of Radiation with Matter

A charged particle in the material goes through atomic interactions and continuously loses its energy. The process is well described by the Bethe–Block formula. X-rays and gamma rays interact with matter by means of photoelectric effect, Compton scattering. For the gamma rays above few MeV, the major energy loss is

Fig. 1 Typical structure of cryogenic detectors. An energy transfer into the absorber is converted into thermal energy. The temperature sensor is meant to measure the temperature increase. The detector setup is connected to a heat bath at low temperatures



by pair creation. Then, the electrons lose energies by the similar process as other charged particles. The interaction of a charged particle with an atom causes the excitation and the ionization of the atom. The average energy required for an electron-hole pair production is usually more than twice higher than the band-gap of the crystals. This means that more than 60% of energy is converted to phonons. In the case of scintillating crystals, one of the brightest scintillators produces about 60,000 photons per MeV energy deposition. Therefore less than a few percent of energy is converted to scintillation photons. The most of energy goes to the lattice vibration and eventually thermalized. For radiations with higher stopping power, this fraction is even smaller, due to the quenching effect. Recoiled nuclei also encounter the interaction with the nucleus in the detector of which some portion of energy cannot be used for the atomic excitation or ionization. Therefore, further reduction of ionization or scintillation yield are observed. This feature allows the separation of nuclear recoil signals from electron recoil signals, by the combination of phonons and lights or phonons and ionization signals.

2.2 Principle of Thermal Detection

When a particle interacts with an absorber, an energy transfer occurs to an electron in the solid or a nucleus in the absorber. Most of the energy transferred into the absorber converts into thermal energy of the solid eventually. With a thermometer attached to the absorber the temperature change caused by the initial energy transfer can be measured.

Typical cryogenic calorimeters consist of two parts as seen in Fig. 1. One is an absorber to make initial interactions with the particles. The other is a temperature sensor to measure the temperature change of the absorber. Typically either the absorber or the temperature sensor is connected to a thermal reservoir, heat bath, with a weak thermal link while they are connected each other tightly. The bath is assumed to have a very large heat capacity. Moreover, its temperature does not change in time, typically below 1 K.

Fig. 2 a A simplified thermal model of typical cryogenic detectors.
b Temperature response of the simplified model. C_{tot} represents the sum of the heat capacities in the detector system

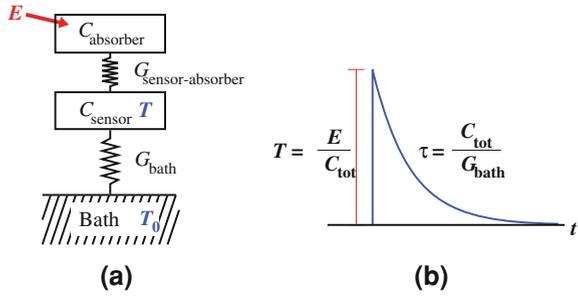


Figure 2 shows a thermal diagram and a typical temperature signal of thermal detectors in general. The absorber and the temperature sensor are connected to each other with a thermal conductance $G_{\text{absorber-sensor}}$. The thermal energy is eventually released to the heat bath. The thermal link of the absorber to the bath is G_{bath} . The detection principle of a calorimetric particle detector is straightforward. Assuming that the heat transfer from the absorber to the sensor is much more effective than that from the heat link to the bath (i.e. $G_{\text{absorber-sensor}} \gg G_{\text{bath}}$). The temperature change, $T - T_0$ in the sensor due to the original energy deposition, E , from a particle is

$$T - T_0 = \frac{E}{C_a + C_s} e^{-t/\tau} \quad (2)$$

with

$$\tau = \frac{C_a + C_s}{G_{\text{bath}}} \quad (3)$$

where T_0 is the bath temperature, and C_a and C_s are heat capacities of the absorber and of the temperature sensor respectively. As shown in Fig. 2b, a typical response of the temperature sensor, a measurement of the initial temperature increase can provide the amount of energy transferred into the absorber. The decay time constant, τ , has the information of the event rate of the detector system without significant pile-ups of signals.

The signals in the calorimetric particle detection are closely related to the heat capacities of the absorber and the sensor as well as the thermal conductance to the thermal reservoir as Eq. 2 implies. At low temperatures the heat capacity of many materials becomes very small, with the result that a temperature rise may be sufficiently large to measure accurately. With a pure dielectric material as an absorber, a small heat capacity can be obtained even with a large volume or large mass. This is because the specific heat of dielectric decreases strongly as the temperature decreases. Generally, the choice of an absorber for a low temperature particle detector depends upon the application. Depending upon the interactions of the absorber with the particles, the mass, compositions and geometries as well as the heat capacities should be considered when choosing an

absorber. The other crucial component of low temperature detectors is the temperature sensor. The temperature sensors have been developed using superconducting electronics to optimize the signal to noise ratio. The *quantum* phenomena at low temperatures provides the essential tool of the temperature sensors, which are often called *quantum sensors*. Temperature quantum sensors can be designed without an absorber, or with one that has a heat capacity comparable to the sensor itself. When these temperature sensors are attached to an absorber with a large heat capacity, the resolution and the threshold are significantly adversely affected. In the following section the currently available quantum sensors are introduced with the detection principles.

3 Quantum Sensors

3.1 Introduction to Temperature Sensors

Most thermometers used in everyday life, industrial or scientific purposes utilize a change of a physical quantity depending upon the temperature. For instance, mercury thermometers work because the volume of the liquid changes on different temperatures. In thermocouple gauges, the temperature dependence of contact potential of two dissimilar metals results in voltage signals of the temperature. Platinum thermometers have the resistance proportional to the temperature in their working temperature range.

In cryogenic particle detection an energy input (E) to the absorber causes a temperature change that is $\Delta T = E/C$ as discussed in the previous section. If the temperature sensor is a resistance thermometer which has its electrical resistance is a function of temperature, the resistance change is resulted from the initial energy input as

$$\Delta R = \left(\frac{\partial R}{\partial T} \right) \frac{E}{C}. \quad (4)$$

In magnetic thermometers, the magnetization of the sensor material changes depending upon the temperature such that

$$\Delta M = \left(\frac{\partial M}{\partial T} \right) \frac{E}{C}. \quad (5)$$

In this section, recently developed temperature sensors are categorized with their detection principles as resistance thermometers and magnetically coupled thermometers. Quasi-particle detectors do not measure the temperature but quasi-particles in a superconductor created by a radiation input. These detectors are discussed in this part because they use similar measurement electronics and operate at similar temperatures to other temperature sensors. These sensor

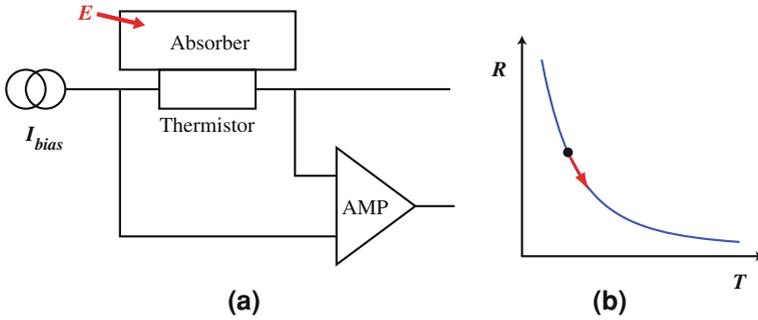


Fig. 3 **a** Simplified circuit diagram of a thermistor setup with an absorber. **b** Typical R–T characteristic of a thermistor. An energy input to an absorber results in a change of resistance of the thermistor attached to the absorber

technologies have reached their sensitivity to quantum limits of the measurement methods, and they are still under development in many world-wide institutions.

3.2 Resistive Thermometers

3.2.1 Semiconductor Thermistors

One of the commonly used temperature sensors is a thermistor, which is, here, a critically doped semiconductor near but below the metal-insulator transition [4]. Ion implantation doped Si and neutron transmutation doped (NTD) Ge thermistor are two kinds of semiconductor thermistors used for low temperature detectors. The electric conductivity of the thermistors strongly depends upon the temperature at the temperature range of interest. In such thermistors, the resistance can be found as

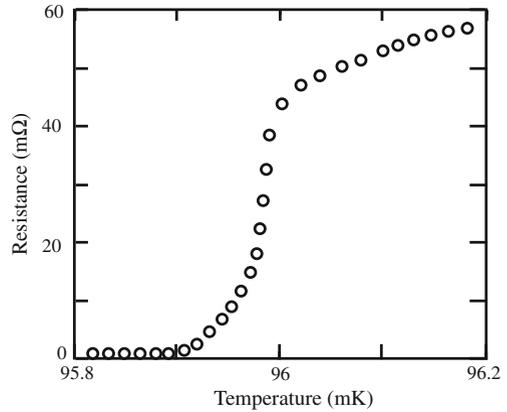
$$R(T) = R_r \exp\left(\frac{T_r}{T}\right)^{1/2} \quad (6)$$

with two characteristic constants, R_r and T_r of the thermistor [5]. The resistance of them is typically 1–100 M Ω near 50 mK.

These thermistors are fairly easy to use because they can be operated with conventional electronics like Field Effect Transistors (FETs), and do not require sophisticated superconducting electronics. They are typically current-biased and measures the voltage across the thermistor amplified with a FET located at 100 K. A simplified circuit diagram and detection concept of semiconductor thermistors for particle detection are illustrated in Fig. 3.

NTD Ge sensors are widely used for various thermal detectors with large crystals because of their reproducibility and uniformity in doping density.

Fig. 4 Resistance of a Mo/Cu TES near the superconducting transition measured at NIST [9]



Mass production is also possible for the NTD sensors. CUORE and EDELWEISS projects are those that use NTD Ge thermistors as their temperature sensors [6, 7]. Doped Si devices have been a popular choice for X-ray spectrometers. They also have an advantage of integration possibility to fabricate many pixel detectors [5]. A high-resolution X-ray spectrometer with 36 pixels of the doped Si devices is scheduled to launch to the space for astronomical X-rays [8].

The response time of thermistors are, however, slow because of poor coupling between the conduction electrons and the lattice of the thermistor. The dependence of the resistance of the thermistor on the applied electric field can affect the performance of the temperature sensor. Joule heating from the bias current is another disadvantage of those sensors for many pixel detectors.

3.2.2 Transition Edge Sensor

Transition Edge Sensors (TESs) are one of the most developed quantum sensors in these kinds. A TES is a superconducting strip operating at the temperature of its superconducting-normal transition. The superconducting strip is often made from a thin pure metal film or a bilayer of superconductor and noble metal. Mo/Au, Mo/Cu and Ti/Au are popular choice of the bilayer. The transition width of the films is typically a few mK or less. The resistance in the normal state is usually a few tens of mΩ. Figure 4 shows the resistance of a Mo/Cu bilayer near the superconducting-normal phase transition edge [9]. Here, the transition width is about 0.1 mK. The temperature dependence of the resistance ($\partial R/\partial T$) is very large at the transition. It infers that it is a very sensitive thermometer in a narrow operating temperature range.

A cross-sectional view of a TES structure is shown in Fig. 5. The bilayer is fabricated on a SiN_x membrane which services as a weak thermal link to the silicone substrate, heat bath here. Recent fabrication developments have realized a overhanging structure of an absorber onto each of TES pixel in order to have larger

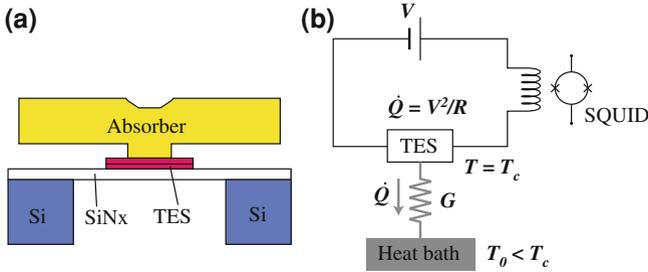


Fig. 5 **a** Simplified cross-sectional view of a TES pixel with an absorber (not to scale). **b** Electric circuit of the TES measurement system and thermal diagram of the detector system

active area for an array of pixels [1, 10]. It has been made not just for a single TES, but also for arrays of many TESs and techniques of multiplexing them [9].

TESs are, in general, operated in a voltage bias mode, so that the resistance change causes the current change of the circuit. This current change is measured with a Superconducting QUantum Interference Device (SQUID). A simplified TES measurement circuit is illustrated in Fig. 5b. The state of the art development of SQUID technologies guarantees an accurate and fast measurement of the current changes with low noise.

In addition to the electric measurement circuit, a thermal circuit can be employed as shown in Fig. 5b. The substrate is attached to a sample holder whose temperature is regulated at a temperature below the transition temperature of the bilayer, T_c . The temperature of the bilayer film can be set at its transition temperature with its own heat generation by the bias voltage. In this case, the amount of heat, \dot{Q} , generated by Joule heating flows to the heat bath via the membrane, weak thermal link with a thermal conductance. Once the circuit is in this mode, the resistance increase reduces the power dissipation since $\dot{Q} = V^2/R$ where V is the bias voltage and R . It leads to the temperature of the film lower. The temperature of the film maintains due to the Electro-Thermal Feedback (ETF) while the current of the circuit changes, and is measured with the SQUID [11]. This self-biasing ETF mode makes several advantages for particle detection with TESs. It helps to find an operation temperature by itself. The linearity of the detection energy is improved. The Dynamic range of the measurement is also increased.

The additional heat converted from the energy input into the absorber has to flow via the thermal conductance. With the ETF of a TES measurement system, the current of the circuit is reduced for the amount of the heat. In other word, the total change of power dissipation in time should be equal to the energy input as

$$E = V \int_0^{\infty} I_0 - I(t) dt. \quad (7)$$

where I_0 is the current on the TES to keep its temperature at the transition temperature with the Joule heating.

One advantage of using a TES for particle detection is that it can be directly evaporated on the surface of an absorber. The direct contact provides an efficient heat transfer between the absorber to the TES, which results in much faster response time of TESs than that of NTD sensors. It is suitable for detecting athermal phonons which directly deposit their energy to the sensor.

TESs have been used as the temperature sensor for dark matter search projects like CDMS and CRESST although they have different crystals as their absorbers [12, 13]. 1.8 keV FWHM for 6 keV was obtained using a Mo/Au TES by NASA GSFC group [1]. It is the best energy resolution in the energy region published so far. In nuclear spectrometry, Quantum sensor group at NIST demonstrated 25 keV FWHM for 100 keV gamma rays and 1 keV FWHM for 5 MeV alpha particles [14, 15]. Their energy resolutions obtained by Mo/Cu TESs and superconducting Sn absorbers are orders of magnitudes improved from commercial semiconductor counterparts, High Purity Ge detectors for gamma rays and Passive Implanted Planar Silicon detectors for alpha particles.

3.3 *Magnetic Thermometers*

Magnetic thermometers utilize a magnetic material whose magnetization is a function of temperature. Unlike resistance thermometers, the pick-up loop of their circuit is superconducting in magnetization measurement. Metallic Magnetic Calorimeters (MMCs) have been developed for high resolution X-ray spectrometers and various particle detections [16]. MMCs employ a paramagnetic material in a magnetic field for the sensor material. Magnetic Penetration Thermometers (MPTs) has recently been realized for a high resolution X-ray measurement [17]. A paramagnetic material of an MMC is replaced with a superconducting film in an MPT setup. Basic concepts of the two noble methods are discussed in following sections.

3.3.1 *Metallic Magnetic Calorimeter*

MMCs uses a small concentration of erbium in a gold host as a sensor material so called Au:Er [18, 19]. This diluted magnetic ions in a metallic host have paramagnetic properties which can be approximated as a spin 1/2 system with a g value of 6.8.

The magnetization is inversely proportional to the temperature a simple paramagnetic system known as Curie's law. It indicates a measurement of magnetization can be used as the temperature information of a paramagnetic material, which makes a magnetic thermometer in other words a magnetic calorimeter to be used in particle detection. Au:Er keeps its paramagnetic properties at tens of mK temperatures. A simplified setup of an MMC is illustrated in Fig. 6.

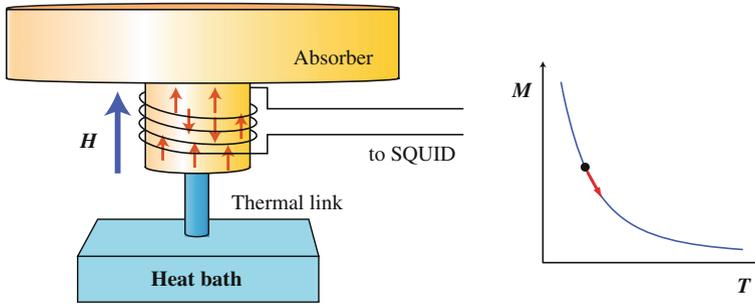


Fig. 6 **a** Simplified MMC setup with an absorber. **b** Typical M - T characteristics of an Au:Er sensor in a magnetic field. An energy input to an absorber results in a change of magnetization of the MMC sensor attached to the absorber

The gold host is used for a fast thermalization time that takes to reach a thermal equilibrium in sensor material between sub-thermal systems of electrons and magnetic spins. The exchange interactions between the magnetic spins that cause negative effect on the temperature dependence of magnetization and heat capacity are well understood using a mean field theory [19].

The thermalization time of MMCs is very fast, within a microsecond, so that the rise time of signals is mostly limited by the SQUID electronics that are used to measure the change of magnetization of a magnetic sample in the SQUID pick-up coil. The resolution of MMCs is slightly better than those of NTDs and TESs. The fact that the heat capacity of an MMC itself can be much larger than those of other sensors makes MMC sensors suitable to be attached to an absorber with a large heat capacity. In addition, the working temperature of MMC sensors is very wide. MMC sensors have various parameters that can be fine tuned to optimize their sensitivity and the threshold, even with an absorber with a large heat capacity [20] or with a large crystal [21].

The development of MMCs in particle detection was made in a setup of a Au:Er sensor placed inside the loop of a SQUID. In this early design of an MMC, the SQUID loop itself was used as a pickup coil [16]. The state of the art development of SQUID technologies guaranteed an accurate and fast measurement of the magnetization change with low noise upon any temperature change caused by an energy absorption. The SQUID converts the change of magnetic flux into a measurable voltage signal in the basis of a quantum interference measurement operating at low temperatures.

The detection principle of an MMC can be characterized as $E \rightarrow \Delta T \rightarrow \Delta M \rightarrow \Delta \Phi \rightarrow \Delta V$ where E , T , M , Φ and V indicate the physical quantities of energy, temperature, magnetization, magnetic flux and voltage, respectively. An MMC setup with a two-stage SQUID system obtained an energy resolution of 2.7 eV FWHM for the X-rays of ^{55}Fe in Heidelberg [22]. Moreover this detection method has recently been adopted in KRISS for the measurements of MeV region in alpha

spectrometry [23], decay energy spectrometry [24] and in double beta decay experiments [25].

3.3.2 Magnetic Penetration Thermometer

A superconducting metal below the transition temperature has not just the electrical resistance, but its magnetization properties follow as a perfect diamagnetic material. As TESs use a sharp transition in electrical resistance, MPTs use a sharp transition in magnetization of a superconductor.

The measurement circuits in MPT have been adapted from MMC meander setups [17]. The Au:Er layer is just replaced by a superconducting film layer. Hf and Mo/Au layers has demonstrated high resolutions for 6 keV X-rays so far. For an uniform and sharp transition the superconducting film are made using a state-of-art technologies developed for TES sensors.

One of the important advantages is that MPT devices do not have any heat dissipation in the sensor circuit like TES and Thermistor cases. It is correct for the other magnetic thermometer case. The heat dissipation from 1000 pixels or more of resistive sensors is not negligible in typical refrigerators used in cryogenic detection. MPTs or MMCs may provide a future spectrometer technology for many pixel sensor for high energy and spatial resolutions.

3.4 *Quasi-Particle Detectors*

In superconducting metals some of electrons are paired to form Cooper pairs at temperatures below their T_c . These Cooper paired electrons occupy in a lower energy state than the energy levels for unpaired electrons separated by an energy gap E_g . They move without any electrical resistivity and do not carry heat. Heat conduction in a superconductor is made by the lattice vibrations (phonons) and the unpaired electrons only. The unpaired electrons in a superconductor are called quasi-particles. The number of quasi-particles decreases exponentially with respect to E_g/kT below the transition temperature where k is the Boltzmann constant.

An absorption of a photon or a charged particle in a superconductor can break the paired electrons to occupy the levels above the energy gap. These quasi-particles may result a tunneling current through a narrow potential barrier or change the inductance of a microwave circuit made with the superconductor. The tunneling current is measured in Superconducting Tunneling Junction (STJ) devices. The inductance change can be a measurable quantity in particle detection with Microwave Kinetic Inductance Devices (MKIDs).

These quasi-particle detectors are not an equilibrium detector. It means that they do not measure the temperature although the major part of the energy transferred into the superconductor converts into the lattice vibrations and eventually phononic temperature. For this reason their energy resolutions are not as good as other

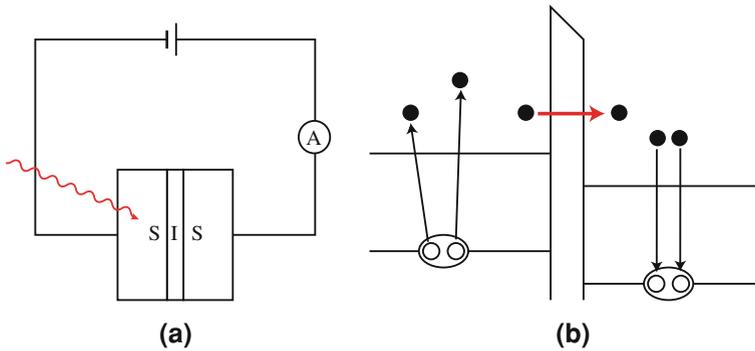


Fig. 7 Basic concept of STJ devices for particle detection **(a)** SIS junction. **(b)** Schematic procedures of particle detection in a SIS junction

thermal detectors like TESs and MMCs. However, the quasi-particle detectors are extremely useful because of their response time and multiplexing possibilities.

3.4.1 Superconducting Tunneling Junction Devices

STJs are realized in a Superconductor-Insulator-Superconductor (SIS) junction. An input radiation with the energy is larger than twice of the gap energy, $E > E_g$, the Cooper pairs may be broken apart creating two quasi-particles above the gap as shown in Fig. 7. With a small electric potential across the junction the quasi-particle may have a chance to tunnel through the barrier. The electric potential is supplied by a dc voltage just below the gap voltage ($V = 2E_g/e$) where e is the charge of an electron. The operation temperatures of STJ devices are typically much lower than the critical temperature such that quasi-particle population is negligible by thermal excitation.

For the detection of X-rays about 10 eV FWHM at 6 keV were achieved with STJ detectors [26]. The energy resolution for X-ray detection is limited by non-equilibrium effects. The non-equilibrium effects, however, provide high speed operation for a given working temperature.

STJ detectors are currently used in various fields like submillimeter (THz), optical/UV and X-ray spectroscopy, and time-of-flight mass spectrometry [26–28].

3.4.2 Microwave Kinetic Inductance Devices

Superconductors have zero resistance for a dc current. However, for an ac current, they have non-zero resistance, but a complex surface impedance. When radiation absorption occurs in a superconductor, quasi-particles are created as discussed with STJ devices. In MKIDs, quasi-particles in a strip of a superconductor cause the change in the surface impedance of a microwave resonance circuit. In other

words, the radiation absorption produces quasi-particles leading to the change in kinetic inductance of a resonance circuit. Eventually, resonance frequency and phase are shifted with a radiation absorption [29].

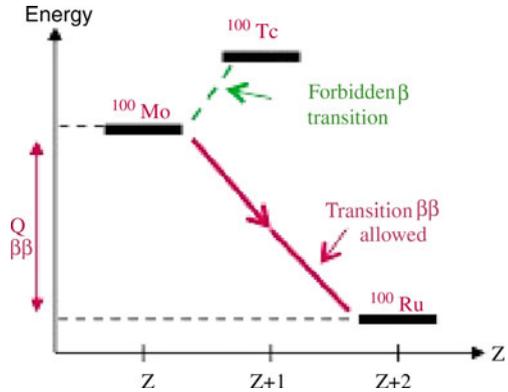
Because MKIDs uses resonance circuits with high quality factors, multiplexing of them are fairly easy. It is a great advantage of using MKID technology for particle detection which requires many pixels of sensors and position information.

4 Application to Elementary Particle Physics

4.1 Dark Matter Searches

The existence of dark matter in our universe is suggested by Zwicky through the observation of velocities of galaxies in a galaxy cluster in 1930s. Since then, astronomical observations such as Rotational velocity curves of many galaxies and gravitational lensing by cluster of galaxies added evidence of dark matter. Recent precision measurement of cosmic microwave background map also suggested that the major portion of our universe is composed of unknown contents, dark energy and dark matter. However, the identification of dark matter is yet unknown. There have been several experiments searching for direct interaction of dark matter with the detectors at deep underground laboratories. An elastic interaction of dark matter particle causes the recoil of the nucleus in the detector, which deposits the energy in the detector. The recoil energy is typically below 100 keV. The deposited energy is transferred to scintillation lights, ionizations, and lattice excitations-phonons. Detection of phonons with or without the readout of additional information—ionization or scintillation lights—has been developed. The CDMS experiment developed Transition Edge Sensor (TES) on silicon and germanium absorber [12]. CDMS also reads the ionization signal by applying the electric field. Since the relative ratio of ionization yield to the phonon excitation of nuclear recoil event is different from that of electron recoil event, the read out of two independent signals provide an efficient mean of separation of gamma background. The CRESST group also developed the TES sensor based detector with CaWO_4 crystal absorber [13]. CRESST employed an additional read out of scintillation signals as CaWO_4 crystal is a good scintillator as well. This also provides a method of background rejection from the nuclear signals. EDELWEISS developed the detector with germanium absorber but with the Neutron Transmutation Doped (NTD) Germanium thermometer [7]. Recently the AMoRE collaboration showed that Metallic Magnetic Calorimeter (MMC) based CaMoO_4 crystal developed for the double beta decay search can be used as a good dark matter search detector as well [25]. One of the outstanding issues in dark matter search detector development is the way to discriminate the surface background events. Since the ionization collection efficiency is lower for the energy loss near the surface, the electron recoil background events leak into the nuclear recoil signal region and potentially very dangerous background.

Fig. 8 Energy level scheme of beta decays of isotopes of $A = 100$



4.2 Neutrino Mass Measurement with Beta Decay

Neutrino oscillation phenomena have been observed in solar neutrino and atmospheric neutrinos and in artificial neutrinos from accelerators and nuclear reactors. These observations indicate that the neutrinos have small but non-zero masses and the mass differences and mixing angles are measured. However, absolute mass scale of neutrinos needs to be measured yet. The precise measurement of the end point of beta decays is to directly measure the neutrino mass. The most successful achievement so far is done by the KATRIN experiment [30], which is based on electrostatic spectrometry of ^3H beta decays. Recently the high energy resolution achieved with cryogenic detectors enable the measurement with calorimetric technique. Microcalorimeter Array for Rhenium Experiment (MARE) is an experiment trying to measure endpoint of beta decay of ^{187}Re , whose end point energy is 2.47 keV [31]. To compete with the sensitivity of KATRIN, $m < 0.2 \text{ eV}/c^2$, MARE is planning to use 10000 detectors followed by the first stage of using 300 detectors. As a large amount of readout channels of MMC or TES sensors involved, the multiplexing is an important issue for MARE.

4.3 Neutrinoless Double Beta Decay Searches

Nuclides of atomic number Z with an excess of neutrons tends to beta decay reduce the number of neutrons to become the nuclide with $Z + 1$. But, for some nuclides, this is not possible because the daughter nucleus is in higher mass state. In that case decay to next to the neighbor can occur by having two neutrons decay as illustrated for ^{100}Mo in Fig. 8.

Double beta decays with two neutrinos are allowed processes in nuclear physics and have been observed in various isotopes. Double beta decay without the emission of neutrinos could occurs if the neutrinos are Majorana particles with

non-zero masses. Since the two electrons are emitted without any other particles the signature is very clean with mono-energetic peak in total measured energy. However, the double beta decay with two neutrinos with shorter half life occurs and the total energy spectrum extends to the end point. Therefore, neutrinoless double beta decay signal can be easily swamped by the irreducible background from two neutrino double beta decay if the energy resolution is poor. Therefore, the high energy resolution is one of the most critical requirement of the detectors for neutrinoless double beta decay. Detectors based on tracking such as NEMO [32] can positively identify two electron tracks and therefore remove the most of the background. However the efficiency of the tracking method is not so high because the source is usually separated from the detector. On the other hand calorimetric method can use the source in the detector, and often the source itself can be used as the detector resulting in very high efficiency. However, due to lack of positive identification of two electrons, the high energy resolution is important to separate the background. The cryogenic detector can be used for sensitive experiments for the neutrinoless double beta decay search if the absorber contains double beta decay elements. CUORE [6] is the one of the first double beta decay search experiment using cryogenic detector technique. Their absorber is TeO_2 crystal and the NTD Ge thermometer is used for the temperature sensor. ^{130}Te is the double beta decay isotope with the natural abundance of 33.8%. The energy resolution achieved by CUORICINO, the pilot experiment of CUORE, is 7–9% at 2,530 keV [33]. The target mass of CUORE is 760 kg without enrichment. Recently, the AMoRE collaboration developed MMC based detector with the CaMoO_4 crystal absorber to search for the neutrinoless double beta decay of ^{100}Mo . With the crystal of the 5 g mass, they achieved energy resolution of 1.7 keV for the 60 keV gammas and 11 keV for the 5.5 MeV alphas [25]. Development of the detector based on 200 g crystal is in progress.

5 Application to THz Wave Measurements

Superconducting detectors have been played a major role in for radiation detection and measurement of THz signals. It is mainly because their noise equivalent power (NEP) are exceptionally better than other uncooled detectors or those cooled as low as liquid nitrogen temperatures. TESs, NTD sensors, STJ devices and MKIDs are those which are being developed for THz detection.

In THz measurement the detectors measure the radiation power of the radiation. The detectors consists of a sensor, an absorber and a thermal link as shown in Fig. 1. The temperature increase can be expected as

$$T - T_0 = \frac{W}{G_{\text{bath}}} \quad (8)$$

with a constant radiation power input, W . It is a similar expression to Eq. 2, but with W and G instead of E and C when the energy is measured. Cryogenic detectors to measure the radiation power are generally called *bolometers* while those to measure the energy are called *calorimeters*. It is noted that the temperature increase is not influenced by the heat capacity, C , in bolometric measurement case as Eq. 8. C and G_{bath} determine the time constant, τ , to reach to the new temperature. The measurement of the temperature increase should be made after reaching a new equilibrium, at least several times longer than the time constant.

TESs are one of the most popular choices for THz detectors which require extremely low NEP [34]. The TES bolometers have already demonstrated a low NEP with a time division multiplexing for 1,000 pixel operations in astronomical millimeter and sub-millimeter radiations [35].

THz detectors with TES bolometers made possible to make a complete passive a scanner for human black body radiation [36]. A weapon hidden inside clothing can be found because its temperature is somewhat lower than that of human body. It is different from infrared (IR) imaging sensors because IR radiations do not penetrate the clothing. A frequency band between 335 and 375 GHz is chosen for this passive weapon search project. The radiation in this range does not have significant the absorption in clothing materials as well as in the air particularly the moisture. With an active mirror system the black body radiation in the band is focused to TES bolometers to measure the radiation power.

This passive scanning THz monitor requires I to distinguish 310 K (37°C) signals from 293 K (20°C) backgrounds. The sensors have to have a good temperature resolution about at least a few degrees as well as spatial and time resolution about 1 cm. TES bolometer used in Ref. [36] can provide enough resolution to find a test weapon hidden inside a shirt. However, this work still need to improve the fast scanning methods and for moving objects. A THz TES camera combined with optical and IR cameras would provide images of hidden objects in weapon safety control and airport security applications.

Superconducting detectors have demonstrated their NEP as good as $10^{-19} \text{ W}/\sqrt{\text{Hz}}$ with STJ and TES bolometers [34]. This value is orders of magnitude better than room temperature detectors. These detectors have already been applied in THz astronomy. Although the NEP of MKID has been measured about $10^{-17} \text{ W}/\sqrt{\text{Hz}}$ so far, their multiplexing capability will accelerate this method to be used in wider applications.

6 Conclusion

Particle detectors with higher energy resolution do not just make fine corrections on existing measurements, but can make it possible for new discoveries in fundamental research of science. Cryogenic particle detectors using quantum sensor technologies have provide a number of successful and on-going experiments in

elementary particle physics. Recent developments of THz bolometers have demonstrated outstanding performance on noise equivalent power for THz radiations. In near future, these quantum sensors are expected to play a major role in astronomy and security and other applications. Applications to X-ray spectroscopy, measurement of radioactivity for radiation standard are not discussed in the current scope, but they are also very active fields of application of cryogenic particle detection.

Acknowledgments The work at KRISS was supported by the National Research Foundation of Korea Grant funded by the Korean Government(NRF-2011-220-C00006) and Radiation Technology Development Program funded by MEST.

References

1. Bandler, S.R., Brekosky, R.P., Brown, A.D., Chervenak, J.A., Figueroa-Feliciano, E., Finkbeiner, F.M., Iyomoto, N., Kelley, R.L., Kilbourne, C.A., Porter, F.S., et al.: *J. Low Temp. Phys.* **151**(1), 400 (2008)
2. Alig, R.C., Bloom, S., Struck, C.W.: *Phys. Rev. B* **22**(12), 5565 (1980)
3. Enss, C. (ed.): *Cryogenic Particle Detection*. Springer, Berlin (2005)
4. Mott, N.F., Twose, W.D.: *Adv. Phys.* **10**, 107 (1961)
5. McCammon, D.: *Top. Appl. Phys.* **99**, 35 (2005)
6. Arnaboldi, C., Avignone Iii, F., Beeman, J., Barucci, M., Balata, M., Brofferio, C., Bucci, C., Cebrian, S., Creswick, R., Capelli, S., et al.: *Nucl. Instr. Meth. A* **518**(3), 775 (2004)
7. Armengaud, E., Augier, C., Benoît, A., Bergé, L., Besida, O., Blümer, J., Broniatowski, A., Chantelauze, A., Chapellier, M., Chardin, G., et al.: *Phys. Lett. B* **687**(4-5), 294 (2010)
8. Takahashi T., Mitsuda K., Kelley R., Aharonian F., Akimoto F., Allen S., Anabuki N., Angelini L., Arnaud K., Awaki H., et al.: *Proc. SPIE* **7732**, 77320Z-77320Z-18 Arxiv preprint arXiv:1010.4972 (2010)
9. Irwin, K., Hilton, G.: *Top. Appl. Phys.* **99**, 63 (2005)
10. Lee, Y.H., Bae, J.B., Kim, I.H., Kim, Y.H., Lee, H.Y., Lee, K.B., Lee, M.K., Lee, S.J.: *IEEE T. Appl. Supercond.* **19**(3), 469 (2009)
11. Irwin, K.: *Appl. Phys. Lett.* **66**, 1998 (1995)
12. Ahmed, Z., Akerib, D., Arrenberg, S., Bailey, C., Balakishiyeva, D., Baudis, L., Bauer, D., Brink, P., Bruch, T., Bunker, R., et al.: *Science* **327**, 1619 (2010)
13. Angloher, G., Bauer, M., Bavykina, I., Bento, A., Brown, A., Bucci, C., Ciemiński, C., Coppi, C., Deuter, G., Von Feilitzsch, F., et al.: *Astropart. Phys.* **31**(4), 270 (2009)
14. Dorise, W.B., Ullom, J.N., Beall, J.A., Duncan, W.D., Ferreira, L., Hilton, G.C., Horansky, R.D., Irwin, K.D., Mates, J.A.B., Reintsema, C.D., Vale, L.R., Xu, Y., Zink, B.L., Rabin, M.W., Hoover, A.S., Rudy, C.R., Vo, D.T.: *Appl. Phys. Lett.* **90**, 193508 (2007)
15. Horansky, R.D., Stiehl, G.M., Beall, J.A., Irwin, K.D., Plionis, A.A., Rabin, M.W., Ullom, J.N.: *J. Appl. Phys.* **107**, 044512 (2010)
16. Fleischmann, A., Enss, C., Seidel, G.M.: *Top. Appl. Phys.* **99**, 151 (2005)
17. Bandler S.R., Irwin K.D., Kelly D., Nagler P.N., Porst J.P., Rotzinger H., Sadleir J.E., Seidel G., Smith S., Stevenson T.: *Magnetically coupled microcalorimeters*. In: 14th International Workshop on Low Temperature Detectors, Heidelberg, Germany (2011)
18. Fleischmann, A., Schönefeld, J., Sollner, J., Enss, C., Adams, J.S., Bandler, S.R., Kim, Y.H., Seidel, G.M.: *J. Low Temp. Phys.* **118**, 7 (2000)
19. Enss, C., Fleischmann, A., Horst, K., Schönefeld, J., Sollner, J., Adams, J.S., Huang, Y.H., Kim, Y.H., Seidel, G.M.: *J. Low Temp. Phys.* **121**(3), 137 (2000)

20. Fleischmann, A., Enss, C., Schönefeld, J., Sollner, J., Horst, K., Adams, J.S., Kim, Y.H., Seidel, G.M., Bandler, S.R.: Nucl. Instr. Meth. A **444**(1–2), 100 (2000)
21. Kim, Y.H., Eguchi, H., Enss, C., Huang, Y.H., Lanou, R.E., Maris, H.J., Mocharnuk-Macchia, A.N., Seidel, G.M., Sethumadhavan, B., Yao, W.: Nucl. Instr. Meth. A **520**(1–3), 208 (2004)
22. Fleischmann, L., Linck, M., Burck, A., Domesle, C., Kempf, S., Pabinger, A., Pies, C., Porst, J.P., Rotzinger, H., Schäfer, S., Weldle, R., Fleischmann, A., Enss, C., Seidel, G.: IEEE T. Appl. Supercond. **19**, 63 (2009)
23. Ranitzsch, P.C., Kempf, S., Pabinger, A., Pies, C., Porst, J.P., Schafer, S., Fleischmann, A., Gastaldo, L., Enss, C., Jang, Y.S., et al.: Nucl. Instr. Meth. A **652**, 299 (2011)
24. Lee, S.J., Lee, M.K., Jang, Y.S., Kim, I.H., Kim, S.K., Lee, J.S., Lee, K.B., Lee, Y.H., Kim, Y.H.: J. Phys. G Nucl. Part. Phys. **37**, 055103 (2010)
25. Lee, S.J., Choi, J.H., Danevich, F.A., Jang, Y.S., Kang, W.G., Khanbekov, N., Kim, H.J., Kim, I.H., Kim, S.C., Kim, S.K., et al.: Astropart. Phys. **34**, 732 (2011)
26. Lerch, P., Zehnder, A.: Top. Appl. Phys. **99**, 217 (2005)
27. Schoelkopf, R., Moseley, S., Stahle, C., Wahlgren, P., Delsing, P.: IEEE T. Appl. Supercond. **9**, 2935 (1999)
28. Ohkubo, M., Shiki, S., Ukibe, M., Tomita, S., Hayakawa, S.: Int. J. Mass Spectrom. **299**, 94 (2011)
29. Day, P.K., LeDuc, H.G., Mazin, B.A., Vayonakis, A., Zmuidzinas, J.: Nature **425**(6960), 817 (2003)
30. Osipowicz, A., et al.: arXiv:hep-ex/0109033 (2001), also available at <http://arxiv.org/abs/hep-ex/0109033>
31. Sangiorgio, S., et al.: Prog. Part. Nucl. Phys. **57**, 68 (2006)
32. Argyriades, J., et al.: Nucl. Phys. A **847**, 168 (2010)
33. Arnaboldi, C., Artusa, D., Avignone, F. III, Balata, M., Bandac, I., Barucci, M., Beeman, J., Brofferio, C., Bucci, C., Capelli, S., et al.: Phys. Rev. C **78**, 035502 (2008)
34. Sizov, F., Rogalski, A.: Prog. Quantum Electr. **34**(5), 278 (2010)
35. Battistelli, E.S., Amiri, M., Burger, B., Halpern, M., Knotek, S., Ellis, M., Gao, X., Kelly, D., Macintosh, M., Irwin, K., et al.: J. Low Temp. Phys. **151**(3), 908 (2008)
36. May, T., Zieger, G., Anders, S., Zakosarenko, V., Meyer, H., Schubert, M., Starkloff, M., Rößler, M., Thorwirth, G., Krause, U.: Proc. SPIE **7309**, p. 73090E (2009)

Intravital Microscopy for THz-Bio Analysis

Pilhan Kim

1 Introduction

Intravital microscopy is a high-resolution imaging technique to observe biological phenomena in living organisms. It often also stated as *in vivo* microscopy. Literal meaning of *in vivo* is “within the living” and there is another term, *ex vivo* of which literal meaning is “out of the living”. Both terms are commonly used to describe the status of sample at the moment of biological manipulations or investigations are done. *In vivo* study is a form of research using whole living organism in experiment to investigate a certain biological phenomenon in its natural environment, whereas *ex vivo* study uses non-living subjects such as tissues or organs dissected from dead animal. *In vitro* of which literal meaning is “within the glass” is another commonly used term. *In vitro* study is a form of research using small living subject such as cell in a controlled environment such as petri dish or test tube. Cell culture, the process of growing cells in a petri dish, is the most common form of *in vitro* study. Figure 1 summarizes the status of samples for biological study categorized by *in vivo*, *in vitro* and *ex vivo*.

1.1 *In Vivo and In Vitro Research*

Living organism is extremely complex system comprised of multiple biological subjects such as gene, protein and cell. For example, adult human contains approximately 70 trillion cells. And a single cell has 23,000 protein-coding genes

P. Kim (✉)

Graduate School of Nanoscience and Technology, Korea Advanced Institute of Science and Technology, Daejeon, Yuseong-gu, Korea
e-mail: pilhan.kim@kaist.ac.kr

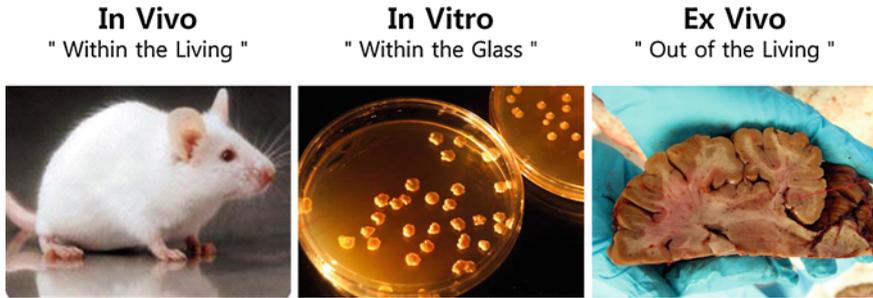


Fig. 1 The status of sampled for biological study. In vivo sample is a commonly used albino laboratory mouse. In vitro sample is colonies of fungus grown in petri dish. Ex vivo sample is a sliced section of brain dissected from rabbit

those are strictly controlled in a specific way depending on the cell type to function properly. For the survival of organism, all of individual cells form complex network and interact with each other extensively to maintain homeostasis of whole body against environment. This extreme complexity is the biggest barrier for biologists to identify individual components and analyze their specific function. The most important advantage of in vitro study is that it provides a highly controlled artificial environment to simplify the system, which allows biologists to focus on the function of individual component under study. Amplifying the outcome of manipulating individual biological subject such as gene, protein or even specific cell type in vitro has been a key enabler for biologists to identify them in the first place and analyze their mechanism in detail. In many biomedical fields including genetics, cellular and molecular biology, in vitro experiments is extensively performed. However, this simplicity of in vitro study has potential limitation that there are always chances of over-interpretation of its result, because fundamentally it fails to reconstitute the complex environment in specific tissue under physiological condition that can only be found in in vivo model. The failure in follow-up in vivo study using living organism to reproduce the results of in vitro study is so common that the in vivo experiment for a confirmation is indispensable.

1.2 Model Organism for In Vivo Research

A model organism is non-human species that can be used to investigate specific biological phenomena through various experimental approaches, which can provide deeper understanding about the similar phenomena in other species, potentially human. There have been many model organisms used in biomedical science for in vivo experiments. Figure 2 summarizes the popular model organism commonly used in wide field of biology. The bacterium, *Escherichia coli* (*E. coli*) and the yeast, *Saccharomyces cerevisiae* has been popular model organisms for genetics and cell biology. More developed multicellular organisms such as the



Fig. 2 Common model organism for biological study. *E. coli* is widely used in microbiology and modern bio-engineering. Yeast, *S. cerevisiae*, is a eukaryotic model organism extensively used in molecular and cell biology. Its genome is the first that was completely sequenced as a eukaryote. *C. elegans* is a multi-cellular eukaryotic organism widely used in developmental biology, especially neural development. Also *C. elegans* is the first multi-cellular organism that its genome is completely sequenced. Fruit fly, *D. melanogaster*, is the most popular organism in genetics due to the well-established genetic transformation technique. Zebrafish is getting more attention as a vertebrate model organism in vertebrate development and gene function. Finally, mouse is a mammalian organism that distinguishes it from former model organism. Since it shares almost all of physiological system with human, the mouse has been a most widely used animal in basic and translational biomedical study in preclinical stage

roundworm (*Caenorhabditis elegans*, abbreviated to *C. elegans*), fruit fly (*Drosophila melanogaster*) and zebrafish (*Danio rerio*) have been extensively used to investigate systemic biological processes such as development, physiological or neurological functions.

To perform biomedical studies aiming to provide potential clinical implications for human disease, the mouse is the most widely used model organism. Since the

mouse is a mammal like human, it has molecular functional mechanism similar to human in very high level. And it holds complex physiological systems resembling human such as immune, endocrine, nervous, cardiovascular, respiratory, gastrointestinal, reproductive, skeletal and muscular system. Other advantages includes a short generation time—10 weeks from being born to giving birth, a highly prolific breeding—5 to 10 pups per each birth and an immediate recovery of fertility after birth and an relatively easy handling and maintenance in laboratory environment. In addition, there are well-established molecular and genetic techniques applicable only to the mouse, which enables an analysis of the complex underlying mechanisms in great detail. And the easiness in the manipulation of mouse embryos for over- or under-expression of certain gene allows the creation of genetically modified transgenic mouse, which can greatly help in vivo experiment by revealing the specific gene function more clearly. Finally, a more than 100 years of research employing the mouse has resulted in abundant documentation about phenotypes of mouse strains and mutant. All these factors make the mouse an ideal platform for basic and translation biomedical research toward bench-to-bedside translation for clinical application that THz-Bio researches ultimately pursue.

1.3 In Vivo and Ex Vivo Analysis of Mouse Model

Every experiment needs a way to measure the outcome of manipulating subjects under study. The uniqueness of in vivo study using living organism as an experimental test bed pose substantial challenges in performing detailed analysis of biological process happening inside the living model organism—need to minimize the alteration of natural in vivo environment while extract accurate measurement data with high sensitivity out of the highly complex in vivo situation. Due to these technical challenges, ex vivo analysis using a specimen dissected from the body of model organism has been the most prevalent form of measurement done to analyze various biological processes rather than in vivo analysis. In other words, although the experiment is performed with living animal in vivo, still the detailed analysis is mainly performed ex vivo. For a biomedical study using the mouse as model organism, there are two most extensively used ex vivo analysis methods, histology and flow cytometry, to obtain cellular level information.

Histological analysis is a microscopic scale observation of processed tissue dissected from organism by light microscope or electron microscope. Preparation of tissue for histological analysis is performed in three major steps—fixation, sectioning and staining. Fixation is a step to preserve tissue from degradation and maintain its microscopic structure. The most common way is dipping the tissue in a chemical fixative such as formaldehyde. After fixation, tissues are progressively dehydrated and embed into a liquid material, then hardened in a mold. For light microscopy observation, the tissue blocks are then sliced into sections as thin as a single cell layer, typically 5–10 μm . Since most biological tissues have very little inherent optical contrast when it is thinly sliced, staining to provide proper level of

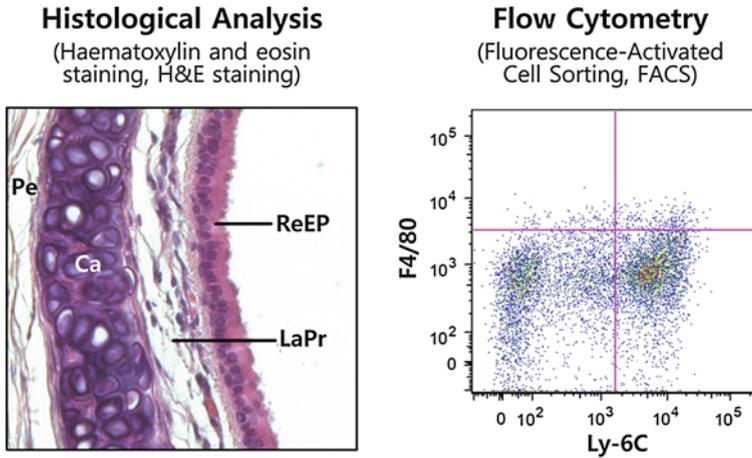


Fig. 3 Examples of ex vivo analysis. Histological analysis is a microscopic observation of ex vivo tissue, typically processed for preservation and then stained to provide contrast in cellular level. Hematoxylin and eosin (H&E) staining that typically stains the nucleus in *blue* and the remaining tissue in various shade of *red* is the most common method for histological analysis. The presented example is an H&E stained trachea of mouse. Microscopic structures such as perichondrium (*PE*), cartilage (*CA*), lamina propria (*LaPr*) and respiratory epithelium (*ReEp*) and their sub cellular features are clearly distinguishable. Flow cytometry is a technique to count the number of cells extracted from ex vivo tissue and then suspended in aqueous medium. Similar to the histological analysis, it is a microscopic analysis that provides quantified information about cellular composition of tissue instead of producing an image. Fluorescence-activated cell sorter (*FACS*) is a most widely used modern flow cytometer that relies on the fluorescence signal to distinguish specific cell labeled by fluorophore, which is explained in detail later in this chapter. The presented example is a FACS analysis of the dissected heart of mouse showing the population of subset of cells expressing high level of Ly-6C, lymphocyte antigen six complex, and low level of F4/80, a macrophage specific transmembrane protein. This result provides the information about the population of Ly-6C^{high} monocytes in the heart while excluding macrophage

contrast and highlight the specific feature of interest is essential. Hematoxylin and eosin (H&E) that stain the nucleus in blue and cytoplasm, muscle and extracellular matrix in various shade of red, pink and orange is the most commonly used staining method. In modern medicine, histopathology, the microscopic observation of dissected tissue with histological staining is the most important tool to provide accurate diagnosis of various diseases such as cancer and inflammatory disease by providing cellular level information. Naturally, for the preclinical biomedical studies using a mouse model of human disease, histological analysis is extensively used (Fig. 3).

Flow cytometry is a technique for counting the number of cells suspended in an aqueous medium. Similar to the histological analysis, it is a microscopic cellular-level analysis that provides quantified information about cellular composition of tissue instead of producing an image. A hydrodynamically-focus stream of fluid containing cells is implemented in microscopic channel so that the cells are carried

in an aligned manner to pass a single spot one by one. A beam of light is focus on the spot where individual cell pass. Multiple detectors are used to sense various optical signal generated by the flowing cells. Typically, one detector is aligned in line with the propagating light to detect scattering signal in forward direction and several detectors are placed in perpendicular to detect scattering signal from side. In addition, several detectors to detect fluorescence signal from the cells labeled with one or more fluorophores are commonly used. For biological cellular analysis, the fluorescence labeling technique is very widely used due to its ability to provide high specificity by employing antibody-based labeling technique, which will be explained in detail later in this part. Fluorescence-Activated Cell Sorter (FACS) is modern flow cytometer based on this fluorescence labeling, which can analyze several thousands of cells every second in real-time and separate or isolate cells with specific fluorescence properties. In modern medicine, FACS is widely used to quantify the population of specific cell type in either blood or dissected tissue, especially for immune cell, in the unit of transplantation and oncology combined with chemotherapy or radiation. In addition, it has been extensively used in basic biomedical studies in the field of immunology, pathology and general molecular biology.

The fundamental limitation of these *ex vivo* analysis is that it is terminal measurement providing a one time observation of cellular process in a living organism that is highly dynamic in nature. As described previously, a living organism is an extremely complex system in which numerous biological processes is dynamically initiated, regulated and interacted with each other in genetic, molecular and cellular level. Basically, *ex vivo* analysis provides us a snapshot of this orchestra of life, failing to track the dynamics maintaining it. Furthermore, the procedure to perform these *ex vivo* analysis, for example the fixation and slicing of tissue for histological analysis, the dissociation of tissue to cell level for flow cytometry, modify the natural microenvironment of cells under *in vivo* situation. To overcome these limitations, intravital microscopy that can directly visualize dynamic cellular process in living organism *in vivo* has actively developed.

2 Intravital (In Vivo) Microscopy

During the last decade, intravital microscopy has become an indispensable technique for wide area of biomedical science such as molecular and cell biology, immunology, neuroscience, developmental and tumor biology. Real-time direct visualization of dynamic cellular processes in action under its natural environment *in vivo* by intravital microscopy has provided unprecedented insights those were almost impossible to obtain by static observation of thinly-sliced chemically-fixed tissues *ex vivo* by conventional light microscopy (i.e. histological analysis). In addition, although *in vitro* model studies have provided many significant findings with its advanced flexibility and manipulability over *in vivo* model studies, it fails to reconstitute the exact microenvironment in tissue under physiological condition

that can only be achieved in in vivo model. It makes the intravital microscopy based analysis become invaluable for many biomedical investigations.

2.1 Fluorescence Detection

Currently, intravital microscopy is mostly performed by observing *fluorescence* emitted from a *fluorophore* as a contrast to create image, which is categorized as *fluorescence microscopy* [1, 2]. It can reveal objects of interests exclusively labeled by fluorophore from an otherwise black background, providing high contrast and sensitivity. There are multiple advantages in fluorescence microscopy over other types of microscopy. First, over the past decades, tens of thousands fluorescent probes were synthesized by organic chemists to provide a toolbox to label virtually any components in biological systems. In addition, numerous protocols for labeling the object of interest in any status of samples, in vivo, ex vivo and in vitro have been well established. Second, the development of fluorescent proteins, most notably green fluorescent protein (GFP), has allowed the intrinsic endogenous labeling of numerous components in living organisms by genetic engineering, which eliminates the difficult challenge of devising in vivo labeling scheme using exogenous fluorescent probes. Its profound impact on biomedical science is recognized by Nobel Prize in Chemistry awarded at 2008 for the discovery and development of the GFP. Lastly, rapid technical advances in light sources and detectors optimized for fluorescence detection aids fluorescence microscopy to be an ideal technique for the observation of all kinds of biological specimens by allowing the detection of molecule at very small concentration with superior signal-to-background noise ratio.

Fluorescence is the emission of light from a fluorophore through after its absorption of light typically within nanoseconds. Figure 4a shows typical electronic states of a fluorophore in Jablonski diagram. Each electronic states have a multiple vibrational state. The energy of absorbed photon makes an electron to be excited, elevating to a higher energy level, which is the process called excitation. Subsequently, the electron returns to the ground state and emit a lower energy photon, which is the process called as emission. Due to the energy loss involved in the vibrational state transition in excited state, the wavelength of emitted light is longer than that of absorbed light. This difference in the wavelength between the absorbed photon and emitted photon is called Stokes shift. This is the fundamental property that enables the fluorescence imaging have high contrast and high sensitivity by completely filtering out the unwanted background light spectrally without affecting the signal emission light. Thus, the proper selection of spectral filters is critical for the separation of excitation and emission wavelengths in fluorescence imaging. On the other hand, since the energy transition by absorption and emission can occur between multiple vibration states in each state, a fluorophore typically has a broadened spectrum of absorption and emission as shown in Fig. 4b. These spectral properties of a fluorophore are determined by its electronic

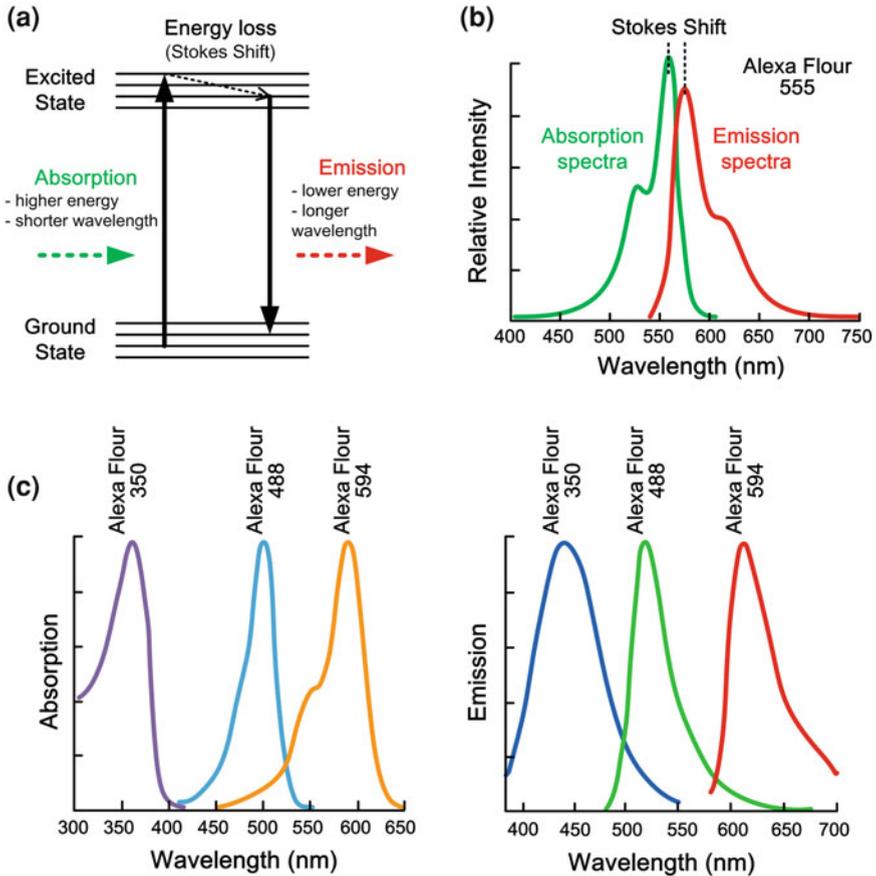


Fig. 4 Fundamentals of fluorescence and fluorophore. **a** Jablonski diagram depicting the electronic states of a fluorophore and energy transition involved in fluorescence. **b** Typical spectral properties of fluorophore. Absorption and emission spectra of Alexa Flour 555 from invitrogen™ is illustrated. **c** Example of fluorophores for multi-color fluorescence imaging, Alexa Flour 350, 488 and 594 from invitrogen™ is illustrated. Emission spectra of these fluorophores are largely non-overlapping, which minimize the crosstalk between them

energy structure that can be tuned to have many different emission wavelengths, i.e. colors. During past decades, variety of colorful fluorophore have developed in ultraviolet (UV), visible and near infrared (NIR) spectrum from 300 nm to 800 nm in wavelength. Fortunately, the emission spectrum of several fluorophore is spectrally narrow and separated enough that it is possible to use multiple fluorophores with different color at the same time as illustrated in Fig. 4c. This enables simultaneous visualization of multiple components labeled by fluorophores of difference colors, respectively. It is extremely useful to analyze complex processes involving interaction between components, for example, interaction between different subsets

Table 1 Common fluorophores

Color	Blue	Green	Orange	Red	NIR
Emission wavelength (nm)	450–480	500–520	570–600	660–700	760–850
Fluorophore	BFP, CFP DAPI Hoechst	GFP, YFP FITC, Cy2, Alexa Fluor 488	DsRed, RFP, TAMRA, Cy3 Alexa Fluor 555	APC Cy5 Cy5.5 Qdot® 705 Alexa Fluor 647	Cy7 Qdot® 800

of immune cells. Common fluorophores with different colors for multi-color imaging are summarized in Table 1. As long as instrument allows, these fluorophores can be simultaneously visualized after spectral separation.

Fluorescein Isothiocyanate (FITC) and Tetramethyl Rhodamine (TAMRA) are traditional fluorescent dyes those have been successfully utilized for many years. The quantum yield of fluorescein is very high and its physical and chemical characteristics have been well known. But, fluorescein is very sensitive to environmental factors such as pH, changing its emission intensity. 4',6-diamidino-2-phenylindole (DAPI) and Hoechst dyes are very popular to stain DNA and chromatin by its binding property to the double helix. And their blue fluorescence makes them to be easily combined with traditional dyes with green and red fluorescence. Alexa Fluor dyes, marketed by invitrogen™, have several improved features such as higher quantum yield than spectrally similar traditional dyes, enhanced photostability under continuous excitation and pH insensitivity. In addition, the absorption spectra of derivatives are well matched to common laser lines and water solubility is high. These advantages makes the Alexa Fluor dyes to be one of the most popular dyes for many fluorescence imaging application, but its relatively high price is the drawback. Cyanine dyes, Cy2, Cy3, Cy5, Cy5.5 and Cy7 are more environmentally stable and thus less sensitive to pH and organic media mounting samples. The cyanine dyes generally have broader absorption spectra than Alexa Fluor family, making them somewhat more versatile in the choice of laser excitation. As described before, due to the unique advantage of endogenous labeling, fluorescent proteins in various colors in blue (BFP), cyan (CFP), green (GFP) and red (RFP, DsRed) are widely used. Qdot®, marketed by invitrogen™, is newly emerging semiconductor quantum dot with diameter of tens of nanometers. Unlike typical organic fluorophores those have highly specific spectral absorption profiles, quantum dot have a unique absorption property that absorption efficiency continuously increases with shorter wavelength, independent of emission wavelength. Thus, multiple kinds of quantum dots of different color can be excited by single UV light source. Another unique feature is that the emission wavelength of quantum dot is determined by the dot size, which enables the manufacturing of quantum dot with emission at NIR range. And the photostability of quantum dot is several orders of magnitude higher than organic dyes. These exceptional properties and the reduced scattering/absorption by most

Table 2 Common light source and output wavelength peak

Light source type	UV	Violet	Blue	Green	Yellow	Red
	Output wavelength (nm)					
Argon-ion laser	351, 364	–	457, 488	514	–	–
Krypton-ion laser	–	413	–	531	569	647
Helium–Neon laser	–	–	–	543	594	633
DPSS laser	365	405	488	532	561	641
Mercury-arc lamp	365	405, 436	–	546	579	–
Xenon-arc lamp	–	–	467	–	–	–
Metal–Halide lamp	365	435	495	520, 545	575	625

biological tissue in NIR range makes the NIR quantum dot to have great potential for in vivo observation of living animal, even for clinical application such as fluorescence-guided surgery of human patient [3]. To perform a fluorescence imaging with a certain fluorophore, it is crucial to select a light source that output wavelength is closely matched to the peak wavelength of absorption spectra of fluorophore in the viewpoint of relatively low emission light intensity compared to excitation light intensities. Common light sources type and output wavelengths are summarized in Table 2.

2.2 Sectioning

Traditional fluorescence microscopy with simple configuration of Köhler illumination and wide field detection can provide excellent images of thin transparent specimen, mostly sliced tissue sample after fixation or living cell cultured in vitro [1, 2]. However, for much thicker sample such as live mouse, it fails to reveal microscopic-scale details at the focus plane due to the reduced contrast by the fluorescence arise from the out-of-focus plane, because the fluorophore distributed over the entire pass of excitation light in sample is indiscriminately excited. For in vivo observation of living organism, in a given two-dimension (2D) image acquired by wide field detection, the fluorescence originated from the out-of-focus plane can easily occupy 90% of total fluorescence collected by objective lens, which completely obscure the in-focus microscopic detail. In addition, the strong scattering of live tissue gives more adverse effect on the contrast of acquired image. Thus, intravital microscopy must equip with a capability that can exclusively acquire fluorescence signal only from the focus by eliminating the out-of-focus fluorescence signal, which is termed as *sectioning* and this type of microscopy is categorized as *sectioning microscopy*. The removal of unnecessary fluorescence provides tomographic snapshot of thick living sample with greatly enhanced contrast, which enables three-dimensional (3D) reconstructions by integrating the image data at multiple focal planes computationally.

2.3 Confocal Microscopy

As a sectioning microscopy, there exist two major types of microscopies, *confocal microscopy* [4, 5] and *two-photon microscopy* [6, 7]. In both cases, the illumination of sample with the light source to excite a fluorophore is performed by 2D raster-pattern scanning of tightly focused beam, which is termed as *scanning*. To achieve high-resolution with *scanning*, it is crucial to implement light illumination that is focused to the smallest spot possible, the diffraction-limited spot, in the plane of focus. This tight focus is normally created by sending collimated beam to pass the high power, high numerical aperture (NA) objective lens that transform the beam to converge. The higher the NA of objective lens and the shorter the wavelength of light for excitation, the smaller the spot becomes to the diffraction-limit that is calculated to $0.61 \lambda/\text{NA}$ in radius, where λ is the wavelength of the light. Although this diffraction-limited spot at the focal plane ensures the highly localized fluorescence signal generation for high-resolution information, still the excitation of fluorophore happens non-selectively all along the converging and diverging beam path out of focal plane. Thus, it is necessary to implement a rejection mechanism to block the fluorescence from our-of-focus plane. In confocal microscopy, this is achieved by placing a pinhole in front of the detector at a plane optically conjugated to the focal plane such that pinhole aperture is aligned with the focus. This arrangement makes the only fluorescence arise from the focus exclusively pass the pinhole to reach detector.

The operation principle of confocal microscopy for fluorescence imaging with sectioning is illustrated in Fig. 5. The arrangement of major components including pinhole, objective lens, sample, dichroic beam splitter and detector as well as the path of excitation light are illustrated in Fig. 5a. Dichroic beam splitter is a specially engineered optical component that reflects shorter wavelength excitation light while transmit longer wavelength fluorescence. Excitation light under 2D scanning is directed to the sample by dichroic beam splitter and focused by objective lens to create tightly focused spot inside the sample. Figure 5b illustrates the light path of fluorescence that is emitted from the focal plane. This in-focus fluorescence signal is back-collected by the objective lens and then separated by the dichroic filter to transmit toward the detector. Then the fluorescence signal passes the pinhole of which aperture is precisely aligned to the optical conjugation point with the focus of excitation light inside sample and reach the detector as illustrated in Fig. 5b. In contrast, the fluorescence emitted from the out-focus plane is focused onto the non-transparent part of the pinhole and thus fails to reach the detector as illustrated in Fig. 5c. The effectiveness of pinhole in terms of sectioning is greatly related to the size of opening in the pinhole. Normally, for the observation of a thin specimen, it is recommended to match aperture size to the calculated spot size at focus, $0.61 \lambda/\text{NA}$ in radius (often referred to Airy Disk size). But, for in vivo observation of thick living specimen, increasing the aperture size up to 2–4 Airy Disk size often provides significantly brighter image with small penalty in resolution due to excessive light scattering.

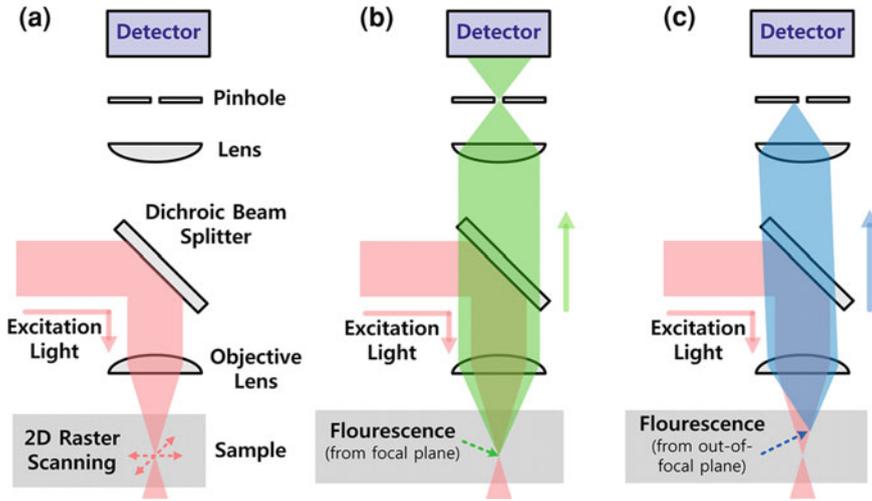


Fig. 5 Principle of confocal microscopy. **a** Schematic layout of confocal microscopy for fluorescence imaging with sectioning. **b** Light path of fluorescence emitted from the focus. **c** Light path of fluorescence emitted from the out-of-focus

2.4 Two-Photon Microscopy

Two-photon microscopy achieves *sectioning* based on nonlinear optical process between excitation light and fluorophore, two-photon absorption. It is a phenomenon occurs when two individual photons arrive at the same fluorophore simultaneously (within ~ 0.5 fs interval) and are absorbed together to combine their energy to excite the fluorophore. Since the energy of photon is inversely proportional to the wavelength, two absorbed photons substituting one photon of twice energy should have twice longer wavelength. For example, a fluorophore normally absorb UV light at 350 nm can be excited by two-photon absorption with NIR light at 700 nm. Figure 6a illustrates the two-photon absorption along with one-photon absorption. As depicted, once fluorophore is excited in either absorption process, it returns to the ground state in exact same manner emitting a fluorescence photon. Due to the requirement that two photon hit the same fluorophore within extremely short interval, compared to one-photon absorption, about a million times higher intensity of light is required to generate the same level of fluorescence with two-photon absorption. In practice, to achieve this high intensity, the excitation light facilitating two-photon absorption has to be extensively concentrated not only in space by high NA objective lens but also in time. Mode-locked Titanium: sapphire laser generating ultrashort pulse light of 50–200 fs pulsewidth at about 80 MHz repetition rate is the most common light source for two-photon microscopy. Even with this level of concentration in time domain, the high photon density sufficient for two-photon absorption is achieved only at the

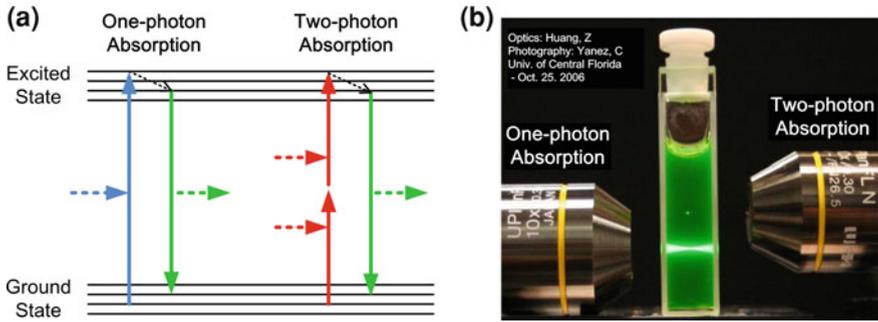


Fig. 6 Principle of two-photon Microscopy. **a** Jablonski diagram, illustrating one- and two-photon absorption. **b** Photograph showing the spatial confinement of fluorescence generation with two-photon absorption, compared to one-photon absorption

very small volume at the focus. In other words, the fluorescence excited by two-photon absorption is generated only from the focal plane as shown in Fig. 6b, which facilitates intrinsic *sectioning* without additional trick like pinhole in front of detector like confocal microscopy. In Fig. 6b, a quartz cuvette with the solution of fluorophore, Fluorescein having green fluorescence, is placed between two identical objective lenses. From the right-side objective lens ultrashort pulse light with 200 fs pulsewidth and 780 nm wavelength is delivered into the cuvette, demonstrating the highly confined generation of fluorescence signal by two-photon absorption. In contrast, when illumination light of 380 nm wavelength tuned for one-photon absorption is delivered through the left-side objective lens, the fluorescence is generate indiscriminately in whole illuminated volume. Table 3 summarizes the major features of confocal and two-photon microscopy. De-scanning is the back-propagation of the fluorescence through the beam scanner, which is necessary only for confocal microscopy to implement stationary focused spot on the aperture of pinhole that is conjugated with 2D scanning spot [4].

2.5 Beam Scanning

Most common method to implement a *scanning* for commercial confocal microscope is using two galvanometer-based oscillating mirrors to deflect the angle of reflected light. Typically, one mirror scans the excitation light along the ‘fast axis’ (mostly horizontal direction) and the other mirror scans the ‘slow axis’ (mostly vertical direction) in much slower speed to sequentially cover 2D plane. Naturally, the speed of image acquisition is limited by the maximum scanning speed of fast-axis scanner. Typical scanning speed of galvanometer based oscillating mirror is from 500 to 1 kHz, which corresponds to 1–2 frames per second with 512×512 pixels. However, for the in vivo observation such as intravital microscopy, it is often very useful to have a capability of video-rate image acquisition at 24–30

Table 3 Comparison between confocal and two-photon microscopy

	Confocal microscopy	Two-photon microscopy
Sectioning principle	Pinhole rejecting out-of-focus fluorescence	Nonlinear fluorescence generation exclusively in focus
Light source type	Continuous wave laser	Femtosecond pulsed laser
Light source wavelength	360–640 nm	700–1,050 nm
Excitation process	Single-photon absorption	Two-photon absorption
Pinhole before detector	Yes	No
De-scanning	Yes	No

frames per second. Fast dynamics such as blood circulation, muscle reflex and neural signal are several examples. One common method to achieve video-rate scanning is using resonant oscillating mirror that can scan at very high speed from 7 to 14 kHz by sacrificing scanning linearity. One commercial confocal microscope employing this resonant scanning is available in market, Nikon's A1R-A1 confocal microscope. Another method is using polygonal mirror rotating at very high speed from 400 to 900 revolution per second, which corresponds to 14.4–32.4 kHz with 36 facet polygonal mirror. In practice, for *in vivo* observation, the image acquisition speed often needs to be compromised to obtain high quality image depending on the brightness of the fluorescence signal, since the achievable signal to noise ratio is proportional to the square root mean of exposure time per pixel with given signal intensity and frame rate.

3 Applications of Intravital Microscopy

Unique capability of intravital microscopy to monitor the dynamic cellular process in natural microenvironment inside living animal *in vivo* has enabled numerous new biomedical studies in unprecedented way. Combined with rapidly developing fluorescent probes, it has opened up new avenue to investigate the complex pathophysiology of human disease with animal model in much greater details at cellular and molecular level. Intravital microscopy has been utilized to monitor gene expression, protein activity, drug delivery, cell trafficking, cell interaction, physiological response under external stimuli in live animal *in vivo*, which provides new insights unobtainable by conventional *ex vivo* and *in vitro* study. Here, several examples of intravital microscopy based analysis in biomedical studies will be introduced.

3.1 *In Vivo* Monitoring of Nanoparticle for Gene Delivery

In the field of nanotechnology, there have been tremendous amount of efforts to develop new functionalized nanoparticles for biomedical applications such as gene

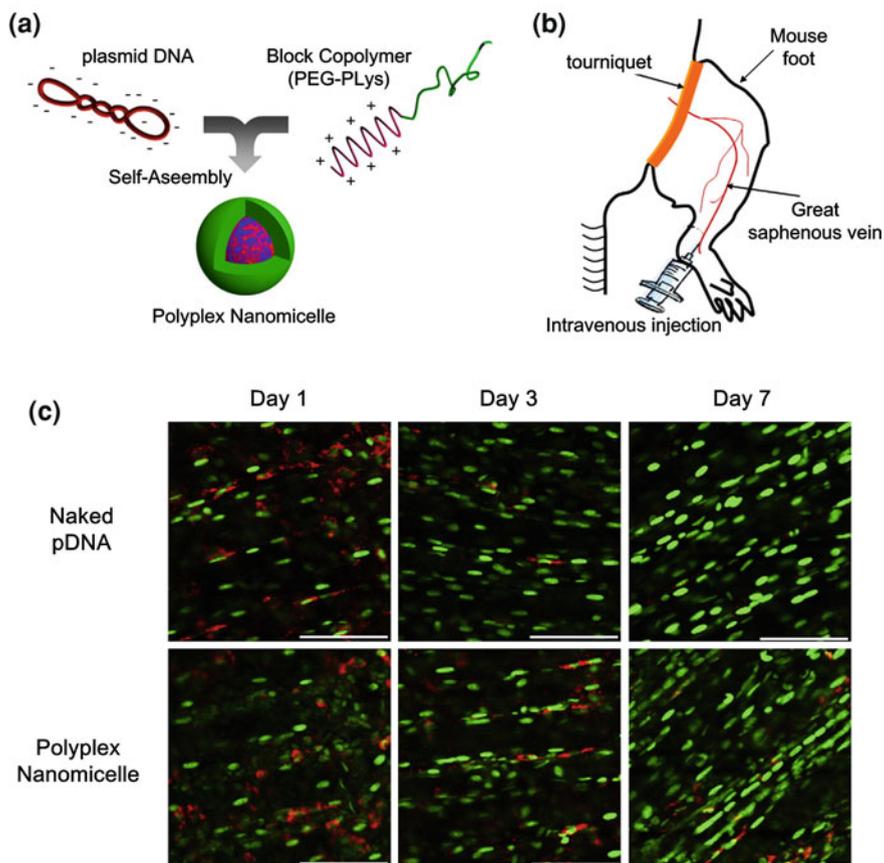


Fig. 7 In vivo tracking of polyplex nanomicelle hydrodynamically administered for gene delivery to thigh muscle over a week in live mice. **a** Polyplex nanomicelle composed of plasmid DNA and block copolymer. **b** Hydrodynamic gene delivery to the thigh muscle by intravenous injection and tourniquet. **c** Longitudinal in vivo tracking of Cy5-labelled plasmid DNA (red) distributed in the muscle fibers of Histone-GFP mice, in which every nucleus of cell express GFP (green). Scale bars are 100 μm. Modified from Ref. [8]

or drug delivery, diagnosis and therapeutics. In the process of the translation from the chemical bench to the bed in clinic, extensive analysis using experimental animal, especially mouse, to assess the efficacy and safety in vivo is commonly required. The intravital microscopy can visualize a single animal over period of hours to weeks, which enables direct monitoring of nanoparticle introduced into the animal in chronological view, like a movie.

In this example, polyplex nanomicelle administered for gene delivery to skeletal muscle was monitored of over a week [8]. The polyplex nanomicelle is synthesized by self-assembly process in the mixture of plasmid DNA (pDNA) and block copolymer as illustrated in Fig. 7a. To image the pDNA in vivo, it is

fluorescently labeled by red fluorophore, Cy5. As an experimental animal model to monitor the delivery of Cy5-labeled pDNA *in vivo*, a transgenic Histone-GFP fusion mouse (B6.Cg-Tg (HIST1H2BB/EGFP)1 Pa/J, Jackson Lab) expressing GFP in the nucleus of every cell is used. To increase the efficacy of gene delivery, the vascular circulation of right foot is isolated by placing tourniquet and then the polyplex nanomicelle containing Cy5-labeled pDNA solution is injected through the Great saphenous vein as illustrated in Fig. 7b. The gene delivery efficacy of naked pDNA and polyplex nanomicelle in terms of distribution in the muscle and retention over time is monitored by intravital microscopy analysis using custom-built video-rate confocal microscopy system [9, 10]. To evaluate the time-dependent change, the injected mouse is imaged at multiple time point of day 1, 3 and 7. GFP and Cy5 are excited by 491 nm and 635 nm continuous-wave DPSS laser, respectively. Figure 7c shows the representative images at each time point. On day 1, for both of naked pDNA and nanomicelle, the Cy5-labeled pDNA (red) is widely distributed, revealing no difference. With image analysis to identify the co-localization of GFP and Cy5 fluorescence, the ratio of nuclei with delivered pDNA can be assessed. For naked pDNA, the Cy5-labeled pDNA is significantly decreased on day 3 and completely disappeared on day 7. In contrast, for the nanomicelle, much larger amount of pDNA is detectable until day 7. This intravital microscopy analysis clearly demonstrates the greatly improved gene delivery efficacy of nanomicelle over naked pDNA.

3.2 In Vivo Monitoring of T Cell Migration in Lymph Node

For immunological studied, intravital microscopy has been one of most extensively utilized tool due to the significant importance of analyzing the dynamic immune cell behavior in natural microenvironment *in vivo* [11]. Intravital microscopy has provided many important implications about how, where and when difference subsets of immune cells interact and migrate to initiate and regulate immune response in systemic way. Among many subset of immune cells, T cell has been considered as a major player to deliver effector function in adaptive immunity and its dysfunction cause various kinds of complications in human disease. Excessive immune reaction such as autoimmune disease is often mediated by T cell migrated to the lesion. Therefore, the migratory behavior of T cell has been actively targeted for therapeutic purpose with novel drug.

In this example, the underlying mechanism of immunosuppressive drug, FTY720, recently approved for treating multiple sclerosis, is investigated by directly observing the T cell migratory behavior in lymph node with intravital microscopy [12]. With *ex vivo* histological analysis, FTY720 treatment induces the rapid accumulation of T cell in the lymph node by blocking its egress. And it has been suggested that G protein family, especially Gzi2 has implication in the intracellular signaling of T cell affected by FTY720. To visualize the migratory behavior of T cell in presence or absence of Gzi2, same number of normal

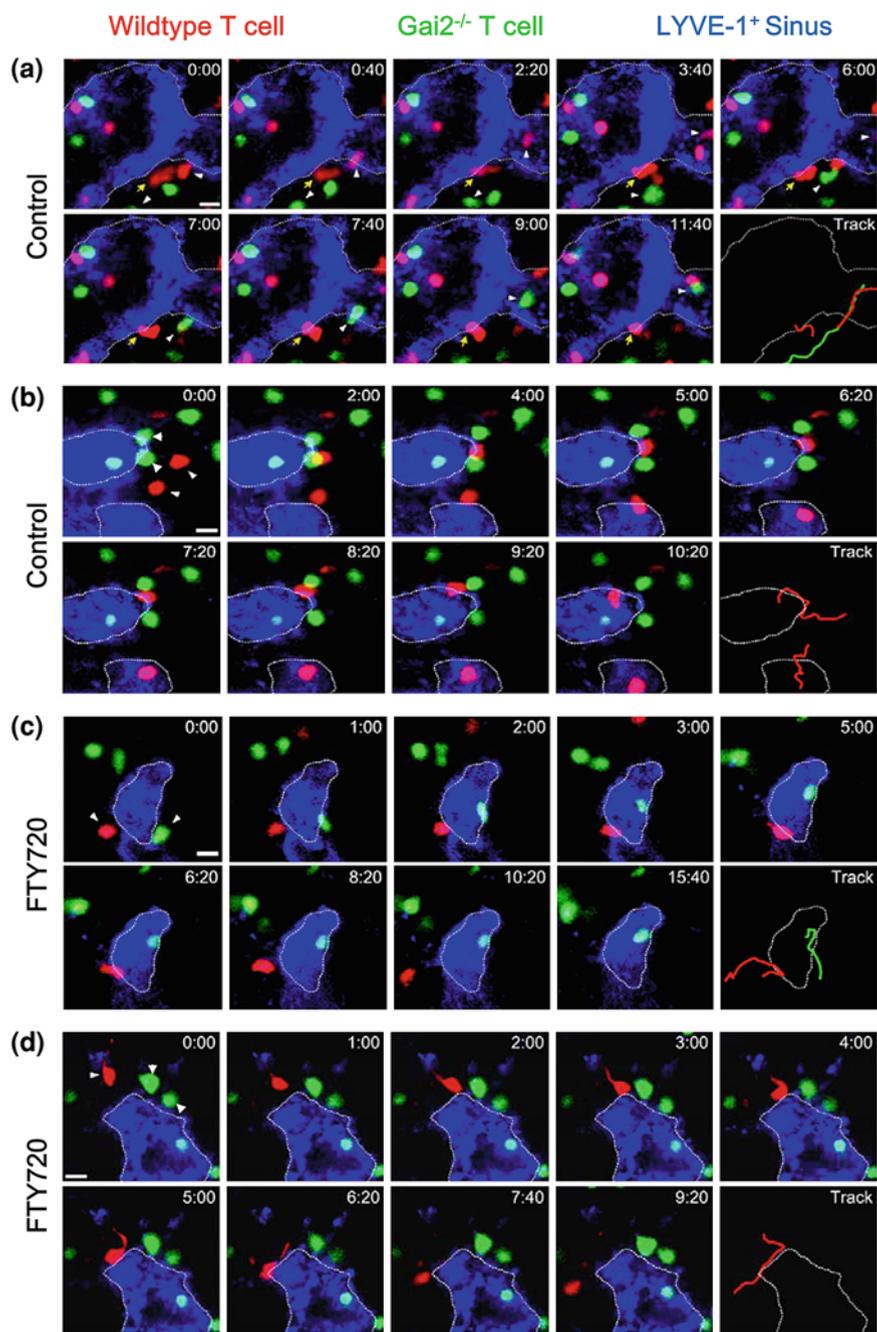
wildtype T cells and *Gzi2* deficient T cells (*Gzi2*^{-/-} T cell) are intravenously injected 24 h prior to the imaging. Wildtype and *Gzi2*^{-/-} T cells are stained by CFSE and CMTMR, respectively. CFSE and CMTMR are cell membrane permeable fluorophore derived from fluorescein and rhodamine, respectively. 4 h after the injection, the mouse was treated either FTY720 or saline as control vehicle. Popliteal lymph node located behind the knee of mouse foot was selected to be monitored by intravital microscopy analysis using custom-built video-rate confocal microscopy system [9, 10, 12]. To visualize the cortical sinus structure in popliteal lymph node, an anti-LYVE-1 antibody conjugated with Alexa Fluor 647 was injected subcutaneously in the foot pad, which then was drained to the popliteal lymph node to stain the sinus. CFSE, CMTMR and Alexa Fluor 647 are excited by 491, 532 and 635 nm continuous-wave DPSS laser, respectively.

Figure 8 shows the time-lapse image sequence tracking the migratory behavior of wildtype T cell (red) and *Gzi2* deficient T cell (green) around the sinus (blue). Without FTY720 treatment, both of wildtype and *Gzi2*^{-/-} T cell freely migrate through the sinus as shown in Fig. 8a in which one wildtype T cell (red, arrowhead), followed by a *Gzi2*^{-/-} T cell (green, arrowhead), enter the sinus through the same location. In another site, two wildtype T cells (red, small arrowhead) approached and engaged different sinus for 3–7 min, then crossed the wall of sinus as shown in Fig. 8b. In contrast, *Gzi2*^{-/-} T cell (green, large arrowhead) shows extended adhesion to the sinus followed by arresting-like behavior near the sinus, which is now observed in the migratory behavior of wildtype T cell. With FTY720 treatment, in sharp contrast, wildtype T cells approached, engaged and sometimes even partially protruded itself into the sinus, but eventually retracted and went away as shown in Fig. 8c–d. Interestingly, low number of *Gzi2*^{-/-} T cell still managed to enter the sinus though it requires more extended time for engagement with the sinus as shown in Fig. 8c. And, unlike the wildtype T cell that does not stay near the sinus, many *Gzi2*^{-/-} T cell stayed near the sinus as shown in Fig. 8d. These observations clearly demonstrate the change in the migratory behavior of T cell by FTY720 and its dependence on the G-protein subunit, *Gzi2*.

3.3 In Vivo Monitoring of Colorectal Tumor Development and Angiogenesis

Cancer is a leading cause of death worldwide and colorectal cancer is the third in both incidence and death. With extensive investments of resource, our understanding of the genetic and molecular origin of cancer has been greatly improved. In last decade, intravital microscopy has provided novel insights into gene expression and function in tumors, revealed the crucial role of microenvironments surrounding the tumors, and their response to therapy, which enables new approaches to improve cancer treatment [13].

In this example, the development of colorectal tumor and angiogenesis, a formation of new vessel to supply oxygen and nutrition to the growing tumor, are



◀ **Fig. 8** In vivo tracking of distinct behaviors of individual T cells in lymph node. **a–d** Time-lapse image sequence shows the migration of T cell around the sinus. Wildtype T cell with *Gzi2* (red), engineered T cell absence of *Gzi2* (green) and LYVE-1 expressing sinus (blue) is visualized in the mice treated with control vehicle **a, b** and FTY720 **c, d**. Border of the sinus is marked by dotted white line. Cumulative cell trajectories are marked by solid lines of red (wildtype T cell) or green (*Gzi2*^{-/-} T cell). Scale bars are 10 μm . Modified from Ref [12]

monitored in vivo in the colon of living mouse over several month by intravital microscopy modified in endoscopic form [10]. To visualize the inner mucosa of colon non-invasively through the anus of living mouse, a miniaturized side-view endoscopic probe with narrow diameter is fabricated by using gradient refractive index (GRIN) lenses. GRIN lens is a rod shaped optical element in which the refractive index is gradually varied along the radial direction so that light can be focused like a sinusoidal wave as illustrated in Fig. 9a. To relay the beam focused by objective lens, a GRIN triplet with 1 mm diameter was constructed by attaching coupling GRIN lens (pitch = 0.25, NA = 0.5) and imaging GRIN lens (pitch = 0.16, NA = 0.5) to each surfaces of a longer relay lens (pitch = 1, NA = 0.1). Then a micro-prism with aluminum coating at slanted surface (base length = 0.7 mm) was attached to the surface of the imaging lens to steer the light to side. The light path in the side-view endoscope is illustrated in Fig. 9a. The assembled side-view endoscope is packaged by metal sheath with diameter of 1.25 mm to provide protection from environment and mechanical strength for in vivo live animal imaging as shown in Fig. 9b. Then, the side-view endoscope is integrated to a custom-built XYZ translation/rotation mount of custom-built video-rate confocal microscopy system [9, 10]. The side-view endoscope can be inserted into the descending colon of anesthetized mouse through anus. Angular direction of laser beam emitting from the endoscope can be rotated inside the colon as shown in Fig. 9c. Axial translation of imaging site on the colonic wall can be accomplished by translating the mouse as shown in Fig. 9d.

As an animal model for colorectal tumor, we used *Apc*-cKO-GFP mouse originally developed by Dr. Raju Kucheralapati at the Harvard Medical School [14]. The transduction of adenoviral *Cre* into the colon epithelia can inactivate the *Apc* gene, which mimics somatic *Apc* mutation in human patients with sporadic colorectal cancer. To monitor the development of tumor, GFP is subsequently activated in *Apc* mutated cell. This modification enables the long-term tracking and quantifying the growth of *Apc*-inactivated cells in colonic mucosa by visualizing the GFP expressing cells. A large polyp expressing GFP at 11 weeks after *Apc* inactivation was observed in vivo as shown in Fig. 10a. Blood vasculature is simultaneously visualized by intravenously injected TAMRA-dextran conjugates. Dilated vessel with irregular shaped lumen suggesting overgrowth of endothelial cells is clearly distinguishable to surround the polyp, which is typical characteristic feature in tumor-associated vessel. In another mouse, a longitudinal observation of the same site in the descending colon was performed. Figure 10b shows images

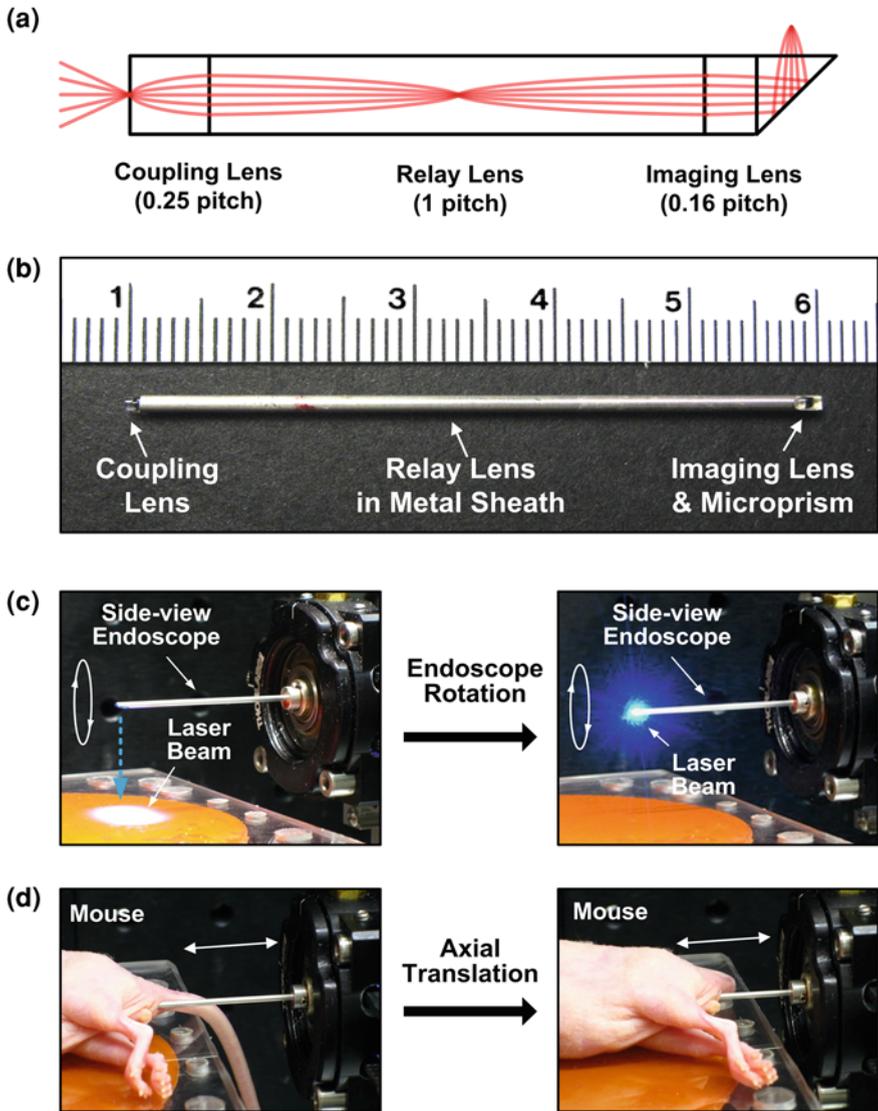


Fig. 9 In vivo side-view endomicroscopy system for the colon imaging. **a** Illustration of light path inside the side-view endoscope comprised by GRIN lenses and right-angle micro-prism. **b** Photo of a side-view endoscope packaged by metal sheath to provide mechanical strength. Unit of number in ruler is centimeter and the length of side-view endoscope is 5 cm. **c** Photo of the side-view endoscope integrated to the custom rotation holder of intravital microscopy system and the relayed laser beam. **d** Photo of translation along axial axis with the side-view endoscope non-invasively inserted to the colon through the anus of living mouse

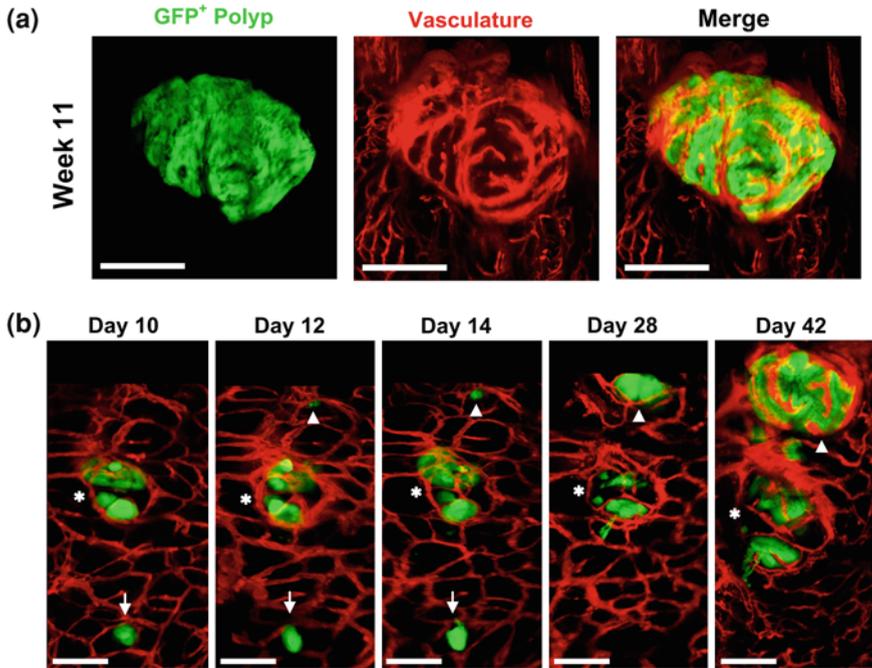


Fig. 10 In vivo longitudinal imaging of colorectal tumor growth and angiogenesis. **a** Images of a large polyp with GFP expressing *Apc* inactivated cells (green) and vasculature (red) obtained at 11 weeks after *Apc* inactivation. Blood vessels were contrast-enhanced by intravenously injected TAMRA-dextran. Dilated blood vessels are surrounding the polyp. Scale bars are 500 μm . **b** Longitudinal repeated imaging of *Apc* inactivated GFP⁺ cells at the same site on day 10, 12, 14, 28 and 42 after *Apc* inactivation. The images show a GFP⁺ lesion that appears to grow (arrowhead), remains unchanged (asterisks) or disappeared (arrow) as well as associated changes in blood vessels visualized by TAMRA-dextran. Scale bars are 200 μm . Figure 10b is modified from Ref. [10]

obtained at day 10, 12, 14, 28 and 42. Multiple groups of GFP⁺ cells were observed at day 10. These micro-nodules showed different behaviors over time in the following imaging. One nodule (arrowhead) rapidly grows, whereas others either remained almost unchanged (asterisks) or vanished (arrow). These results demonstrate the capability of endoscopic intravital microscopy to monitor the fate of these cells in vivo from the moment of genetic mutation to the formation of large adenomas. In addition to genetic predisposition, the stromal microenvironment is important factor in the initiation and development of tumors. Side-view endomicroscopy can be utilized to visualize a variety of events in tumorigenesis such as vascular changes, matrix modulation and circulating cell infiltration, and thus is expected to be a powerful research tool in tumor biology.

4 Future Perspective of Intravital Microscopy for THz-Bio Analysis

Recently, as an unexplored spectrum of electromagnetic wave, generation, transmission and detection of Terahertz wave has been heavily studied through various technological developments. In addition, new application of Terahertz wave, especially in biomedical field, has been actively searched. However, as a new electromagnetic wave, its effect on living organism *in vivo* has remained as almost completely unknown [15]. Intravital microscopy can be an ideal tool to analyze the unexplored effect of Terahertz wave on biological sample in molecular and cellular level. Quantitative and qualitative analysis of interactions between biological samples and Terahertz wave in various conditions by intravital microscopy can provide basic information for future biomedical application, such as *in vivo* damage threshold to various cellular compartment or any detrimental effects in physiological level. For this purpose, *in situ* monitoring of biological sample under Terahertz waves radiation is highly desirable, which can be achieved by modified intravital microscopy. Detailed analysis about the molecular specificity of Terahertz wave in interactions with *in vivo* tissue and related cellular physiological effects would be the basic cornerstone to establish the Terahertz wave as a novel tool to control biological process. And, based on the analysis results, new biomedical therapy can be expected to be developed by implementing a setup for highly localized radiation of Terahertz wave in high temporal and spatial resolution by utilizing intrinsic high-absorption of Terahertz wave in biological sample,

In conclusion, as illustrated in examples, intravital microscopy observing dynamic process happening in living animal with sub-micron resolution can provide valuable new insights those are unobtainable by other analysis. For the field of Terahertz science aiming biomedical applications, intravital microscopy analysis can be highly useful tool to extend our understanding about interactions of Terahertz wave and biological tissue in living animal *in vivo*.

References

1. Diaspro, A.: *Optical Fluorescence Microscopy*. Springer, New York (2010)
2. Lichtman, J.W., Conchello, J.-A.: Fluorescence microscopy. *Nat. Methods* **2**, 910–919 (2005)
3. Kim, S., Lim, Y.T., Soltesz, E.G., et al.: Near-infrared fluorescent type II quantum dots for sentinel lymph node mapping. *Nat. Biotechnol.* **22**, 93–97 (2004)
4. Pawley, J.: *Handbook of Biological Confocal Microscopy*. Springer, New York (2006)
5. Conchello, J.-A., Lichtman, J.W.: Optical sectioning microscopy. *Nat. Methods* **2**, 920–931 (2005)
6. Masters, B.R., So, P. T. C.: *Handbook of Biomedical Nonlinear Optical Microscopy*. Oxford University Press, New York (2008)
7. Helmchen, F., Denk, W.: Deep tissue two-photon microscopy. *Nat. Methods* **2**, 932–940 (2005)

8. Itaka, K., Osada, K., Morii, K., Kim, P., Yun, S.-H., Kataoka, K.: Polyplex nanomicelle promotes hydrodynamic gene introduction to skeletal muscle. *J. Control. Release* **143**, 112–129 (2010)
9. Kim, P., Puoris'haag, M., Cote, D., Lin, C.P., Yun, S.-H.: In vivo confocal and multiphoton microendoscopy. *J. Biomed. Opt.* **13**, 010501 (2008)
10. Kim, P., Chung, E., Yamashita, H., Hung, K.E., Mizoguchi, A., Kucherlapati, R., Fukumura, D., Jain, R.K., Yun, S.-H.: In vivo wide-area cellular imaging by side-view Endomicroscopy. *Nat. Methods* **7**, 303–305 (2010)
11. Germain, R.N., Miller, M.J., Dustin, M.L., et al.: Dynamic imaging of the immune system: Progress, pitfalls and promise. *Nat. Rev. Immunol.* **6**, 497–507 (2006)
12. Zhi, L., Kim, P., Thompson, B.D., et al.: FTY720 blocks egress of T cells in part by abrogation of their adhesion on the lymph node sinus. *J. Immunol.* **187**, 2244–2251 (2011)
13. Jain, R.K., Munn, L.L., Fukumura, D.: Dissecting tumour pathophysiology using intravital microscopy. *Nat. Rev. Cancer* **2**, 266–276 (2002)
14. Hung, K., Maricevich, M., Richard, L., et al.: Development of a mouse model for sporadic and metastatic colon tumors and its use in assessing drug treatment. *Proc. Natl. Acad. Sci. USA* **107**, 1565–1570 (2010)
15. Wilmink G. Grundt J.: Invited review article: current state of research on biological effects of terahertz radiation. *J. Infrared Mili. Terahz Waves.* doi:[10.1007/s10762-011-9794-5](https://doi.org/10.1007/s10762-011-9794-5) (2011)