KAIST Research Series

Chong-Min Kyung *Editor*

Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting

KAIST

KAIST Research Series

Series editors

Chan Beum Park, Daejeon, Korea, Republic of (South Korea)
Bumki Min, Daejeon, Korea, Republic of (South Korea)
Jae Woo Lee, Daejeon, Korea, Republic of (South Korea)
Jae Seung Jeong, Daejeon, Korea, Republic of (South Korea)
Sang Ouk Kim, Daejeon, Korea, Republic of (South Korea)
Insung S. Choi, Daejeon, Korea, Republic of (South Korea)

More information about this series at http://www.springer.com/series/11753

Chong-Min Kyung
Editor

Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting

Editor
Chong-Min Kyung
Department of Electrical Engineering,
Center for Integrated Smart Sensors
KAIST
Daejoen
Korea, Republic of (South Korea)

ISSN 2214-2541 ISSN 2214-255X (electronic)
KAIST Research Series
ISBN 978-94-017-9989-8 ISBN 978-94-017-9990-4 (eBook)
DOI 10.1007/978-94-017-9990-4

Library of Congress Control Number: 2015942486

Springer Dordrecht Heidelberg New York London
© Springer Science+Business Media Dordrecht 2016
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.
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.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

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

Contents

Part I Devices for Low Energy

Tunneling Field-Effect Transistors for Ultra-Low-Power Application . . . 3
Byung-Gook Park

Bulk FinFETs: Design at 14 nm Node and Key Characteristics 33
Jong-Ho Lee

Micro and Nanoelectromechanical Contact Switches for Logic, Memory, and Power Applications 65
Yong-Ha Song and Jun-Bo Yoon

Part II Systems and Circuits for Low Energy

Ultralow Power Processor Design with 3D IC Operating at Sub/Near-Threshold Voltages 121
Sandeep Samal and Sung Kyu Lim

Energy Harvesting from the Human Body and Powering up Implant Devices 147
Ross Kerley, Xiucheng Huang and Dong Sam Ha

Reconfigurable Photovoltaic Array Systems for Adaptive and Fault-Tolerant Energy Harvesting 181
Naehyuck Chang, Massoud Pedram, Hyung Gyu Lee, Yanzhi Wang and Younghyun Kim

Low-Power Circuit Techniques for Efficient Energy Harvesting 211
Jaeha Kim, Myeong-Jae Park, Joonseok Yang and Wootaek Lim

Part III Materials for Low Energy

Preparation of Porous Graphene-Based Nanomaterials for Electrochemical Energy Storage Devices 229
Yuanzhe Piao

Graphene and Two-Dimensional Transition Metal Dichalcogenide Materials for Energy-Related Applications 253
Gyeong Sook Bang and Sung-Yool Choi

Part I
Devices for Low Energy

Tunneling Field-Effect Transistors for Ultra-Low-Power Application

Byung-Gook Park

Abstract One of the major roadblocks to further scaling of complementary metal-oxide semiconductor (CMOS) devices is power consumption. Reduction of power consumption requires low operation voltage, which requires low threshold voltage. In order to decrease the threshold voltage without excessive increase of OFF current, reduction of the subthreshold swing is essential. To reduce the subthreshold swing, various carrier injection mechanisms other than thermal carrier injection have been proposed. Currently, interband tunneling is the most promising mechanism and the device that utilizes such a mechanism is a tunneling field-effect transistor (TFET). After the introduction to the fundamentals of TFETs, various approaches to increase the drain current of Si TFETs by device structure engineering are described. The last section focuses on bandgap engineering to enhance the drain current of TFETs.

Keywords Tunneling field-effect transistor (TFET) · Ultra-low power · Subthreshold swing · Nanowire · Doping engineering · Tunneling area · Bandgap engineering

1 Introduction

Complementary metal-oxide semiconductor (CMOS) device scaling has been going on for more than 50 years. The average design rule has shrunk to 70 % of the previous generation. Such a rate of shrinkage brings roughly an order of magnitude reduction every 20 years. Starting with a few tens of μm design rules in the 1960s, devices with a physical gate length of less than 20 nm are in production now. In a little more than 50 years, some devices have been shrunk down by more than three orders of magnitude in size.

B.-G. Park (✉)
Seoul National University, Seoul, Korea
e-mail: bgpark@snu.ac.kr

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_1

In the early 1990s, the first semiconductor technology roadmap (National Technology Roadmap for Semiconductors, Semiconductor Industry Association, 1992) was developed in the United States of America to guide the semiconductor industry and set up milestones for technology development. In 1999, the roadmap was internationalized with a new name, the International Technology Roadmap for Semiconductors (ITRS). According to the recent ITRS (2013 version) [1], we may be able to see logic chips made of CMOS devices with a 5 nm feature size in 2028 (Fig. 1).

One of the major roadblocks to further scaling of CMOS devices is power consumption, since the power density of a semiconductor chip increases with scaling and there is a definite limit of heat dissipation in the conventional cooling system. The increase in power density originates from the fact that the power consumption of a single device does not scale in proportion to the device area. In the carrier velocity saturation limit, the current of a device scales with the channel width, while the voltage applied to the device is difficult to scale with the channel length. Thus, the power consumption given by the product of the current and the voltage does not scale with the area.

Reduction of power consumption requires low operation voltage, which, in turn, requires low threshold voltage. Reduction of threshold voltage, however, is accompanied by an exponential increase in the OFF current. In order to decrease the threshold voltage while maintaining the OFF current low, reduction of the subthreshold swing, which is defined as the voltage change required for an order of magnitude change in current, is essential.

To reduce the subthreshold swing below 60 mV/decade at room temperature, various carrier injection mechanisms other than thermal carrier injection have been proposed [2–11]. Interband tunneling can be used for injection of carriers and

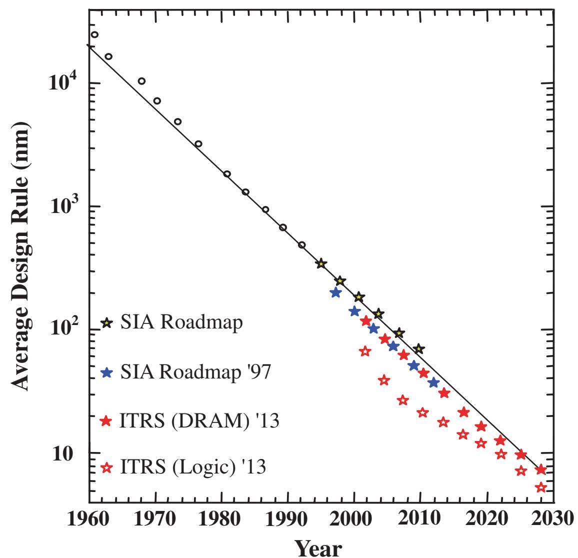


Fig. 1 Average design rule as a function of year. The predictions of the SIA roadmap and the ITRS are also plotted

the device that utilizes such a mechanism is a **tunneling field-effect transistor** (TFET) [12–16]. To take advantage of the infrastructure of the currently dominant CMOS technology, silicon (Si) is a preferred material for TFETs. Si TFETs, however, suffer from low ON current due to the relatively large bandgap of Si.

In this chapter, we survey various approaches to improve the performance of TFETs. In Sect. 2, we calculate the power consumption in a CMOS inverter and establish the importance of subthreshold swing in low power operation. Section 3 deals with the fundamentals of TFET. In Sect. 4, various approaches to increase the drain current of Si TFETs by device structure engineering are described. On the other hand, Sect. 5 focuses on the bandgap engineering to enhance the drain current of TFETs.

2 Low Power Operation and Subthreshold Swing

2.1 Switching and Leakage Power Consumption

Let us consider the power consumption in a CMOS inverter composed of an n-channel MOSFET and a p-channel MOSFET (Fig. 2). Assuming that the input of the CMOS inverter is a pulse train (usually called a clock) with a frequency *f*, we first evaluate the energy consumed during one complete clock cycle. Multiplying it by the clock frequency, we can obtain the power consumption for the switching operation. Let us first consider a high–low transition (discharging process) at the output node. During this transition, the capacitance, *C*, that represents the total parasitic capacitance connected to the output node, will be discharged from V_{DD} to 0. The instantaneous power, *p*, that the n-channel MOSFET consumes is

$$p = v_O i_D = -v_O C \frac{\mathrm{d}v_O}{\mathrm{d}t}. \tag{1}$$

The energy lost during this high-low transition can be obtained by integrating *p* with respect to time for the complete transition.

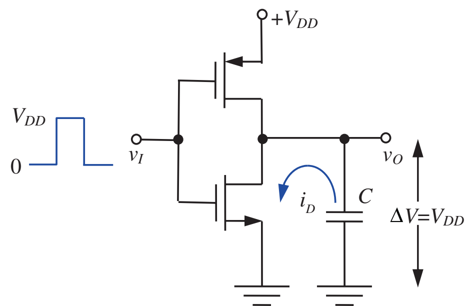


Fig. 2 CMOS inverter circuit for the calculation of power consumption. v_I is the input voltage, v_O is the output voltage, V_{DD} is the power supply voltage, and i_D is the drain current of a MOSFET

$$W_{\mathrm{HL}} = \frac{1}{2}CV_{\mathrm{DD}}^2 \tag{2}$$

The instantaneous power that the p-channel MOSFET consumes during the low–high transition (charging process) is given as

$$p = (V_{\mathrm{DD}} - v_O)i_D = (V_{\mathrm{DD}} - v_O)C\frac{\mathrm{d}v_O}{\mathrm{d}t}. \tag{3}$$

The energy loss during the low–high transition is

$$W_{\mathrm{LH}} = \frac{1}{2}CV_{\mathrm{DD}}^2. \tag{4}$$

One clock cycle includes both the high–low and low–high transition, so that we need to add up Eqs. (2) and (4) in order to obtain the total energy loss per clock. Now, if we multiply it by the clock frequency, f, we obtain the switching power consumption.

$$P_{\mathrm{switching}} = fCV_{\mathrm{DD}}^2 \tag{5}$$

In this equation, we can see that the power supply voltage (V_{DD}) plays an important role in power consumption. For low power operation, V_{DD} reduction is essential. But if we want to reduce V_{DD}, the threshold voltage of MOSFETs should also be reduced. Otherwise, the performance of the circuit will be degraded significantly. The reduced threshold voltage, however, brings in the exponential increase in the OFF current as shown in Fig. 3. The large OFF current is directly related to the high leakage power consumption, since the leakage power is given as the product of the supply voltage and the OFF current, I_{OFF}.

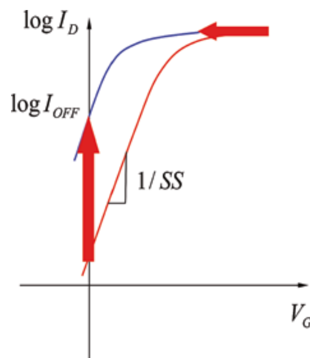


Fig. 3 Threshold voltage reduction and exponential increase in the OFF current (I_{OFF}) in MOSFETs. Such a drastic increase in the OFF current is originated from the exponential dependence of the subthreshold characteristic on the gate bias voltage (V_G). Since the subthreshold swing is finite and similar in most MOSFETs, I_{OFF} increases exponentially as the threshold voltage decreases

$$P_{\text{leakage}} = V_{\text{DD}} I_{\text{OFF}} \tag{6}$$

This leakage power is consumed not only during the period of switching but also during the standby time. For applications where the equipment stays mostly in the standby mode, large leakage power consumption can be fatal.

2.2 Subthreshold Swing and Its Limit in MOSFETs

Figure 3 shows the tight linkage between the threshold voltage and the OFF current in field-effect transistors (FETs). Since log I_D is linearly dependent on the gate bias voltage below the threshold voltage (that is, in the subthreshold region), the OFF current increases exponentially as the threshold voltage decreases linearly. Thus, reduction in the threshold voltage entails a drastic increase in the OFF current.

In order to deal with this phenomenon quantitatively, the concept of **subthreshold swing** (*SS*) is convenient and useful. The subthreshold swing is defined as the inverse of the slope of the log I_D versus V_G curve in the subthreshold region. The meaning of the subthreshold swing is the amount of gate voltage required for one-decade change of the drain current. At room temperature, the typical value of subthreshold swing is 70–90 mV/decade in MOSFETs and 60 mV/decade in bipolar transistors. For the ON/OFF current ratio ($I_{\text{ON}}/I_{\text{OFF}}$) of 10^5, the threshold voltage of a MOSFET should be at least 0.35 V. This is the reason why it is difficult to reduce the power supply voltage of MOSFETs below 1 V.

In order to reduce the switching power of CMOS circuits, the subthreshold swing of MOSFETs should be reduced. Unfortunately, there has been a strict limit in reducing the subthreshold swing of conventional FETs. The lowest value of the subthreshold swing that a conventional FET can achieve at room temperature is 60 mV/decade.

Let us consider the origin of the 60 mV/decade subthreshold swing limit in conventional FETs. The origin resides in the thermal carrier injection mechanism from the source to the channel as shown in Fig. 4. The gate bias controls the energy barrier height between the source and the channel, and carriers in the source have energy distributed by the Boltzmann law. Since the carriers can be injected into the channel only when they have a kinetic energy higher than the barrier height, the current injected into the channel is proportional to the Boltzmann factor, $e^{-q(V_0-V)/k_BT}$, where q is the electron charge, V_0 is the built-in potential, k_B is the Boltzmann constant, and T is temperature. The current through a p–n junction is

$$I = I_0 e^{-q(V_0-V)/k_BT}, \tag{7}$$

where I_0 is a constant called the reverse saturation current.

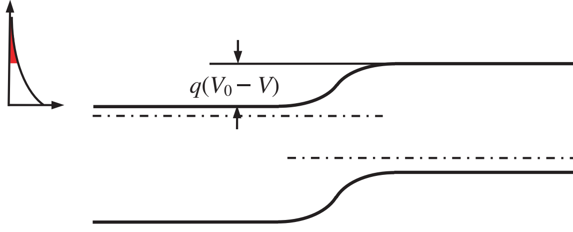


Fig. 4 Origin of the 60 mV/decade subthreshold swing limit in a p–n junction. The bias voltage across the junction controls the energy barrier height, and carriers in the injection source have energy distributed by the Boltzmann law, resulting in exponential dependence of the current on the bias voltage and inverse temperature

If we take the logarithm of Eq. (7),

$$\log I = \log I_0 - \frac{q(V_0 - V)}{k_B T} \log e. \tag{8}$$

By differentiating Eq. (8) and taking its inverse, we obtain

$$\frac{dV}{d(\log I)} = \frac{k_B T}{q \log e}. \tag{9}$$

At room temperature (300 K), this value is about 60 mV/decade. Due to voltage drop in the gate oxide, only a portion of the gate voltage can be used to lower the barrier between the source and the channel. Considering the voltage division between the channel and the gate oxide in a field-effect transistor, we obtain

$$SS = \frac{\mathrm{d}V_G}{\mathrm{d}(\log I_D)} = \frac{k_B T}{q \log e}\left(1 + \frac{C_{\mathrm{ch}}}{C_{\mathrm{ox}}}\right). \tag{10}$$

where C_{ch} is the channel capacitance per unit area and C_{ox} is the oxide capacitance per unit area. Thus, the subthreshold swing of conventional FETs must be larger than 60 mV/decade at room temperature. The only way of reducing the subthreshold swing below 60 mV/decade in conventional FETs is to reduce the temperature, but the power required for cooling exceeds the power saving from the lower supply voltage.

To overcome the 60 mV/decade limit in the subthreshold swing, a few mechanisms that do not depend on thermal carrier injection have been proposed. One of such mechanisms is injection by impact ionization, which was proposed by Gopalakrishnan [2]. **Impact-ionization MOS (IMOS)** devices based on this mechanism use modulation of avalanche breakdown voltage of a gated p-i-n diode [3–6]. Another mechanism is injection by interband tunneling [12–16] and the device that utilizes such a mechanism is called a TFET. The third mechanism is injection by mechanical contact [7–11], which is implemented by a **mechanical switch**. In this chapter, we will discuss TFETs only, since they appear to be the most promising and CMOS-compatible devices.

3 Fundamentals of Tunneling Field-Effect Transistor

3.1 Interband Tunneling Mechanism

Figure 5 shows the comparison between two carrier injection mechanisms: (a) thermal carrier injection and (b) tunneling injection. In thermal carrier injection, only the carriers with kinetic energy higher than the barrier height can be injected into the channel. Since the number of injected carriers is proportional to the Boltzmann factor, $e^{-q(V_0-V)/k_BT}$, the minimum subthreshold swing is 60 mV/decade at room temperature. Tunneling carrier injection in a reverse-biased p^+-n^+ junction, however, utilizes quite different carrier distribution in the source region. The energy gap in the p^+ region eliminates a large portion of the Femi–Dirac tail distribution of electrons and the number of electrons in the conduction band is almost completely negligible. Thus, we may consider only the tunneling of valence band electrons and such tunneling results in majority carrier injection into the n^+ region. Most of the tunneling electrons see the same triangular energy barrier whose height and width are fixed at a given bias voltage. The valence band electrons with energy higher than the conduction band edge of the channel can tunnel into the channel.

In order to calculate the tunneling probability in interband tunneling, we usually assume that the electric field, E, in the depletion region of p^+-n^+ junction with the bandgap, E_g, is constant, as shown in Fig. 6. Using Wentzel–Kramers–Brillouin (WKB) approximation, the wave function, $\psi(x)$ can be calculated as follows.

$$\psi(x) \cong \psi(0)\exp\left\{-\frac{1}{\hbar}\int_0^x \sqrt{2m^*[U(x)-E]}\mathrm{d}x\right\}, \tag{11}$$

where ħ is the reduced Planck constant, m_e is the effective electron mass, x is the position, $U(x)$ is the potential energy, and E is the total energy of the tunneling electron. Since $U(x) = E_g - q\mathrm{E}x$, we obtain

$$\psi(x) \cong \psi(0)\exp\left\{-\frac{1}{\hbar}\int_0^x \sqrt{2m^*[E_g - qEx]}\mathrm{d}x\right\}. \tag{12}$$

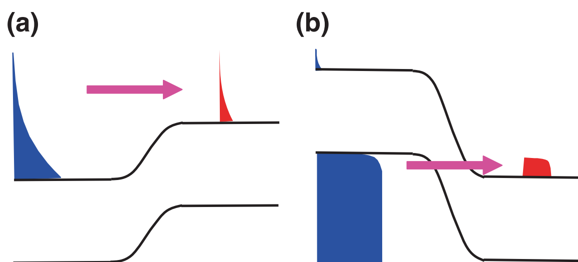


Fig. 5 Comparison of carrier injection mechanism: **a** thermal carrier injection and **b** tunneling carrier injection

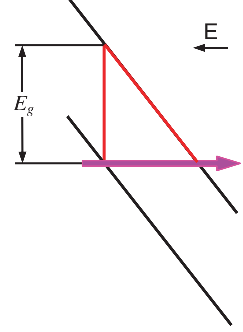


Fig. 6 Approximate band structure for tunneling probability calculation in interband tunneling. E_g is the bandgap and E is the electric field

Tunneling probability can be calculated as

$$T = \left| \frac{\psi\left(E_g/E\right)}{\psi(0)} \right|^2 \cong \exp\left[-\frac{4\sqrt{2m^*}E_g^{3/2}}{3qE\hbar} \right]. \tag{13}$$

In this equation, we can see that the tunneling probability is exponentially dependent on the bandgap and electric field. Especially the dependency on the bandgap is stronger, since the energy gap determines both the height and width of tunneling barrier. This property makes the bandgap engineering an effective method to control the current density and subthreshold swing.

In Fig. 6, we notice that the tunneling probability should be the same for all electrons independent of their energy, since the tunneling barrier is identical for all electrons. The interband tunneling current density, J_{IB}, can be obtained by multiplying the tunneling probability with a factor determined by the density of states and distribution function. The following current–voltage relationship is obtained by such a method [17] and includes the tunneling probability in Eq. (13) as an exponential factor.

$$J_{\mathrm{IB}} = \frac{\sqrt{2m^*}q^3\mathrm{E}V}{4\pi^3\hbar^2E_g^{1/2}} \exp\left(-\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\mathrm{E}\hbar} \right), \tag{14}$$

where

$$E = \sqrt{\frac{2qN_{\mathrm{ch}}(V+V_0)}{\varepsilon_{\mathrm{si}}}}. \tag{15}$$

Since the valence band electrons do not follow the Boltzmann distribution and the injection mechanism is tunneling through a barrier whose width is changing as a function of the bias voltage, interband tunneling is a nonthermal process. The exponential factor does not have a dependence on temperature, confirming that the process is nonthermal. Furthermore, the overall equation does not show any explicit dependence on temperature.

3.2 *Tunneling Field-Effect Transistor*

If we substitute the n^+-source with a p^+-source in an n-channel MOSFET as in Fig. 7, we obtain a tunneling field-effect transistor (TFET). Usually, TFETs are made as an ultrathin body (UTB) structure (Fig. 7a) or a double-gate (DG) structure (Fig. 7b), in order to isolate the channel from the substrate electrically. In these devices, carrier injection from the source occurs through an induced tunnel junction between the source and the channel. Unlike normal MOSFETs, TFETs have an asymmetric structure.

Figure 8 shows the band diagrams of a TFET for its OFF and ON states. When the gate bias voltage is below the threshold voltage, electron tunneling from the valence band of the source to the conduction band of the channel is blocked. Holes can be injected into the channel because of the lowered barrier, but they are blocked at the large potential barrier formed at the reverse-biased channel-drain

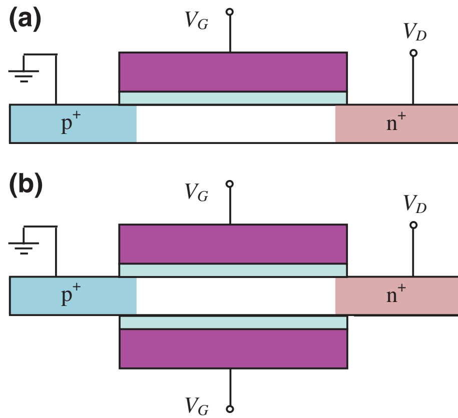


Fig. 7 Tunneling field-effect transistors (TFETs): **a** ultrathin body (UTB) TFET and **b** double-gate (DG) TFET. In both cases, a p^+-source is used instead of the n^+-source in order to form an induced tunnel junction between the source and the channel

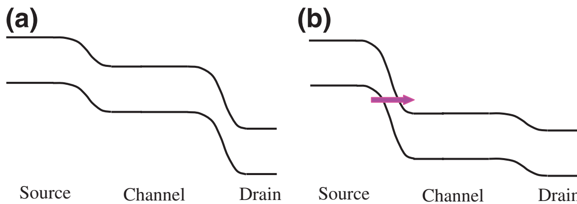


Fig. 8 Band diagram of a TFET. **a** OFF state: Electron tunneling from the valence band of the source to the conduction band of the channel is blocked. **b** ON state: Electrons tunnel from the valence band of the source to the conduction band of the channel

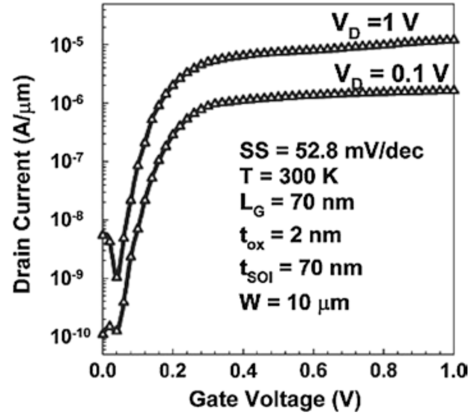


Fig. 9 Subthreshold characteristics of a TFET. This device shows a subthreshold swing value less than 60 mV/decade at room temperature. Reproduced with permission from Ref. [16] © 2007 IEEE

junction. If the gate bias voltage is above the threshold voltage, electrons tunnel from the valence band of the source to the conduction band of the channel. Once injected into the channel, the electrons can be transported as in a conventional MOSFET.

The advantage of a TFET lies in its carrier injection mechanism. Since tunneling rate is independent of temperature and the carrier injection mechanism is nonthermal, the subthreshold swing can be smaller than the thermal injection limit (60 mV/decade at room temperature). Figure 9 shows such an example in a fabricated TFET [16]. The TFET is fabricated on a silicon-on-insulator (SOI) structure, and both the gate oxide and the body are very thin (UTB structure). Such a structure enhances the gate control over the channel, so that the intensity of the electric field in the tunneling region is maintained high. In addition, the buried oxide in the SOI structure provides a blockage to the possible leakage path from the source to the substrate. Such a blockage is automatically provided in most of the three-dimensional device structures with a floating channel.

3.3 *Point Subthreshold Swing*

In a MOSFET, the subthreshold characteristic can be approximated by a simple exponential function over the entire range of the subthreshold region (Fig. 10a), since the subthreshold current is determined mainly by the thermal injection from the source to the channel. Accordingly, the subthreshold swing in a MOSFET is almost constant throughout the subthreshold region, as shown in Fig. 10c. Although we have defined the subthreshold swing on the basis of differentiation operation at the specific gate bias voltage (Eq. (10)), the point-by-point evaluation of subthreshold swing is not necessary. In fact, it is a general practice to extract the subthreshold swing from two points that are separated by at least 4–5 decades of current level in the subthreshold characteristic (Fig. 10a).

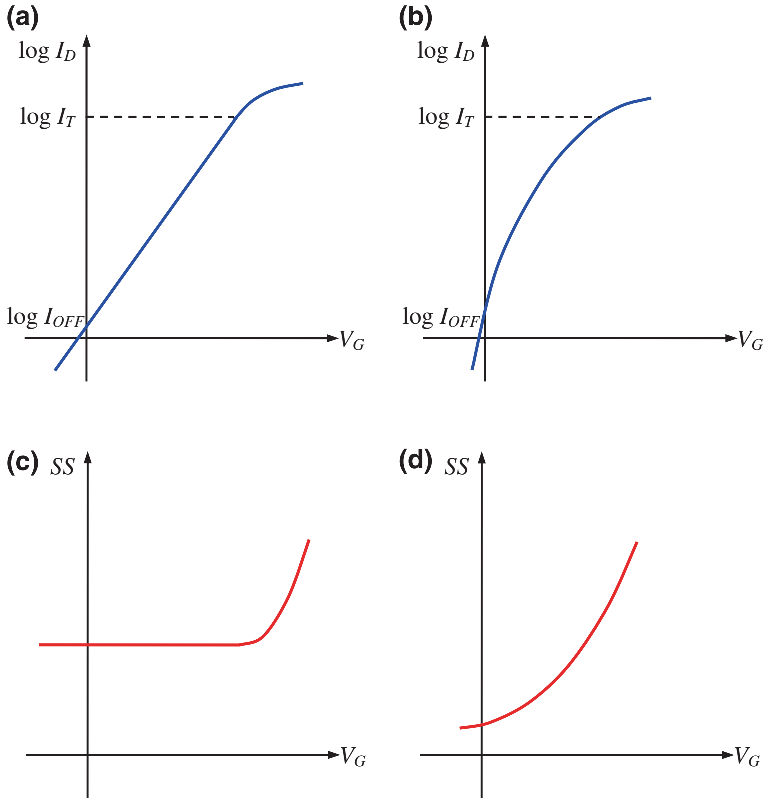


Fig. 10 Subthreshold characteristics of **a** MOSFET and **b** TFET. In MOSFETs, the log I_D versus V_G curve is almost a straight line, while, in TFETs, it is a curve with continuously decreasing slopes. Subthreshold swings of **c** MOSFET and **d** TFET are also plotted

TFET subthreshold swing, on the contrary, varies significantly as the gate bias voltage changes. In the case of TFETs, we need to use **point subthreshold swing** based on the differentiation operation.

$$\mathrm{SS} = \frac{\mathrm{d}V_G}{\mathrm{d}(\log I_D)} \neq \mathrm{const.} \tag{16}$$

If we assume that the interband tunneling at the source–channel junction determines the drain current, log I_D can be written as

$$\log(I_D) = -\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\mathrm{E}\hbar}\log e + \log\frac{\sqrt{2m^*}q^3\mathrm{E}VA_t}{4\pi^3\hbar^2 E_g^{1/2}}. \tag{17}$$

Noticing that the first term is varying more rapidly than the second term as a function of the electric field, we can estimate that the subthreshold swing would be

proportional to $(V_G + V_0')^{3/2}$. Figure 10b, d show the subthreshold characteristic and swing of TFET as a function of the gate bias voltage. At low gate bias voltage, the subthreshold swing of TFET is lower than that of MOSFET. The TFET subthreshold swing, however, increases rapidly and exceeds that of MOSFET below the threshold voltage. As expected, the point subthreshold swing is an essential and important concept in TFETs.

3.4 Ambipolar Characteristics

One problem associated with TFETs is their **ambipolar characteristic**. The ambipolar behavior is originated from tunneling at the drain junction and is basically the same phenomenon as the gate-induced drain leakage (GIDL) in MOSFETs. Figure 11a, b show the band diagrams of an n-channel TFET in its ON and OFF states, respectively. Under the condition of high gate voltage, the normal ON current flows due to the large tunneling current at the source junction. Under the condition of low (or negative) gate bias voltage, the bandgap in the drain edge of the channel may also form a relatively thin tunnel barrier, so that tunneling at the drain junction would not be negligible. Once this tunneling occurs, holes are generated near the drain junction and the hole current in the channel and source maintains the current continuity. This current will increase exponentially, as the gate bias voltage decreases (or becomes more negative).

Figure 12 shows the typical ambipolar characteristic of a TFET. If the doping profile of the drain is similar to that of the source, the OFF characteristic should be roughly the mirror image of the ON characteristic. In real TFETs, however, the OFF current is much smaller than the ON current, since the doping profile of the drain is usually gradual, while that of the source is maintained quite steep. Lightly doped drain (LDD) or underlapped drain structure can suppress the ambipolar behavior further.

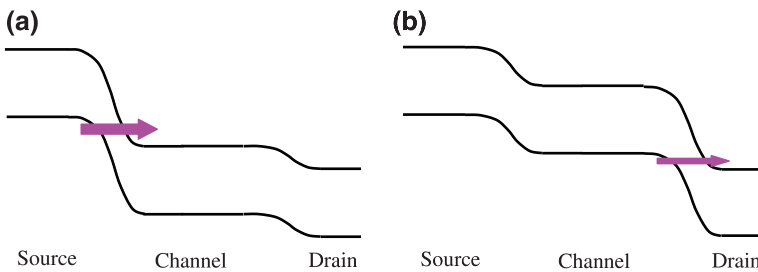


Fig. 11 Band diagrams of a TFET: **a** under high gate bias voltage (ON state), and **b** under low gate bias voltage (OFF state)

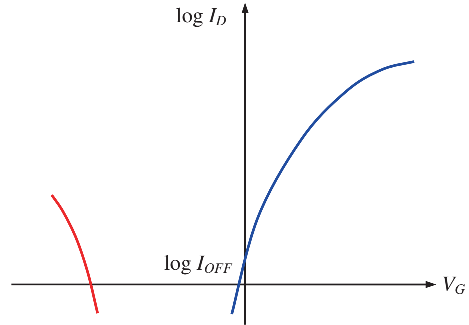


Fig. 12 Ambipolar behavior of a TFET. If the doping profile of the drain is similar to that of the source, the OFF characteristic (*red line*) should be roughly the mirror image of the ON characteristic (*blue line*)

4 Structure Engineering for Tunneling Field-Effect Transistors

4.1 Silicon Tunneling Field-Effect Transistors and Related Issues

One major issue with silicon (Si) TFETs is the relatively low current density compared with that of MOSFETs. As can be seen in Fig. 9, the typical current density of a TFET is more than an order of magnitude smaller than that of a typical MOSFET. The reason behind such a low current density is the relatively large bandgap of silicon. The bandgap of silicon is about 1.1 eV and this value is significantly larger than that of germanium (0.67 eV). This is the reason why germanium has been used for tunnel diodes, instead of silicon.

One of the methods to increase the current density of Si TFETs is to enhance the electric field, so that we can compensate the effect of large bandgap. In Eq. (11), it is clear that the ratio of $E_g^{3/2}$ and the electric field determines the exponential factor and we can increase this factor by increasing the electric field. In order to increase the electric field at the source–channel junction, the gate should maintain a tighter control over the channel. Cylindrical nanowires are the best structure in enhancing the gate control over the channel, since the gate surrounds the channel and the electric field is concentrated around and inside the channel. Another method of enhancing the electric field is to insert a thin counter-doping layer between the source and the channel in a TFET. The layer can enhance the electric field by space-charge effect.

In addition to the stronger electric field, a larger injection area can also increase the tunneling current. Since the inversion layer is confined to the surface of the channel in a planar TFET, the effective injection area is usually very small. If we can increase the effective injection area by structure engineering, the tunneling current can be enhanced significantly because the tunneling current is proportional to the effective injection area.

The ultimate remedy for the consequence of relatively large bandgap is the use of smaller bandgap. This solution, however, requires a smaller bandgap material such as germanium. Since material change is quite a different topic, a separate section (Sect. 5) will be devoted for that purpose. In this section, we will focus on the structure engineering of Si TFETs.

4.2 Cylindrical Nanowire Channel

Figure 13 shows the schematic diagram of a cylindrical **nanowire** TFET and its cross section at the channel region. The cross-sectional shape of the channel is chosen to be circular, not rectangular, because the channel with a rectangular cross-section suffers from **corner effect**. In rectangular channels, the field will be concentrated near the corners only, while, in circular channels, the field will be concentrated evenly around the circle. Thus, circular channels do not suffer from the corner effect.

The cross section of a circular channel is shown in Fig. 13b. The gate oxide thickness is t_{ox}, and the radius of the channel is R_{ch}. The capacitance per unit area of the channel surface is given as

$$C_{\mathrm{ox}} = \frac{\varepsilon_{\mathrm{ox}}}{R \ln\left(1 + \frac{t_{\mathrm{ox}}}{R}\right)}. \tag{16}$$

In this equation, the equivalent oxide thickness is $R \ln\left(1 + \frac{t_{\mathrm{ox}}}{R}\right)$. Let us divide it by t_{ox} to normalize it and define $x = R/t_{\mathrm{ox}}$. We can easily show that the normalized equivalent oxide thickness is smaller than 1 for all $x > 0$. Thus, the equivalent oxide thickness is always smaller than t_{ox}. Such a smaller equivalent oxide thickness reflects the effects of enhanced field in the circular channel and results in increased capacitance. Using a high dielectric constant material, we can further decrease the equivalent oxide thickness and increase the gate capacitance.

The increased capacitance tightens the gate control over the channel, and the tighter gate control enhances the electric field at the junction between the source and the channel. In addition to the increased gate capacitance, the cylindrical channel structure further enhances the electric field at the junction between the source and the channel because of the three-dimensional field enhancement effect

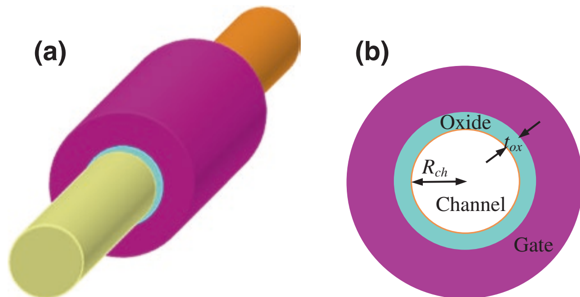


Fig. 13 Nanowire tunneling field-effect transistor: **a** schematic diagram and **b** cross section at the channel region. The gate oxide thickness is t_{ox}, and the radius of the channel is R_{ch}

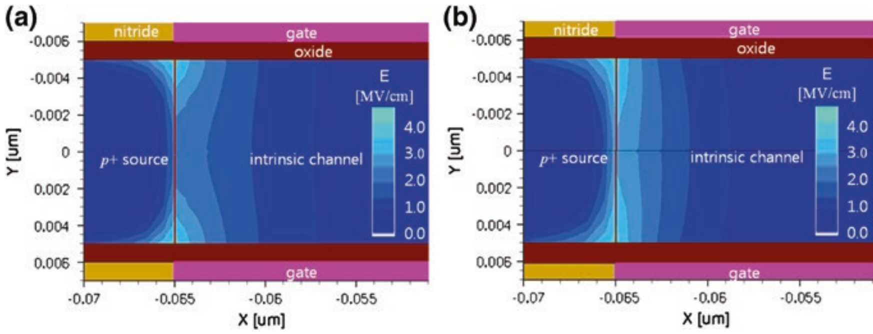


Fig. 14 Comparison of electric field in the channel: **a** double-gate structure and **b** cylindrical channel structure. To eliminate the effect of the smaller equivalent oxide thickness in the cylindrical channel device, a thicker oxide with the same equivalent oxide thickness (1 nm) is used. In addition, the channel diameter (10 nm) of the cylindrical channel structure is the same as the channel thickness of the double-gate structure

within the channel. Figure 14 shows the comparison of a double-gate structure and a cylindrical channel structure in terms of electric field in the channel. To eliminate the effect of the smaller equivalent oxide thickness in the cylindrical channel device, a thicker oxide with the same equivalent oxide thickness is used. As can be seen in this figure, the high field region in the cylindrical channel is extended toward the center ($y = 0$), further than that in the double-gate structure.

Although the difference in the electric field of double-gate and cylindrical channel structures is not dramatic, the carrier generation rate by tunneling can be very different due to its exponential dependence on the electric field. The comparison of carrier generation rate by tunneling in these two structures is shown in Fig. 15. As expected, the region of high carrier generation rate is confined to the small area near the oxide–silicon interface in the case of the double-gate TFET. The region of high carrier generation rate in the cylindrical channel TFET is much

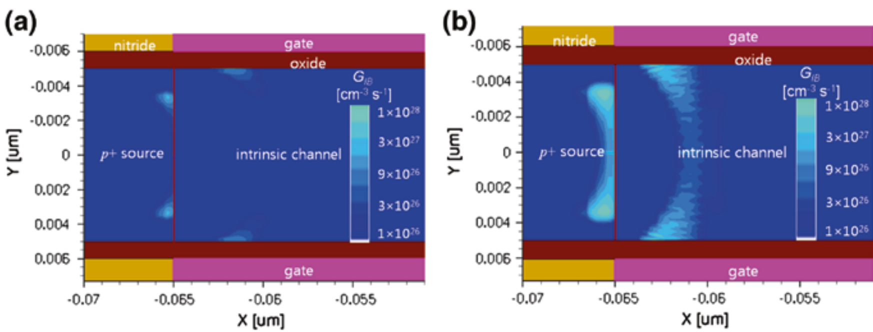


Fig. 15 Comparison of carrier generation rate (G_{IB}) by tunneling in the channel: **a** double-gate structure and **b** cylindrical channel structure. All the dimensions of the devices are maintained the same as those in Fig. 14

larger and extended toward the center ($y = 0$) of the cylinder. These results clearly demonstrate that not only the thinner effective thickness of oxide but also the three-dimensional effect within the channel enhances the field in the cylindrical channel. Thus, the cylindrical channel is almost ideal structure for TFETs.

Experimentally, both vertical and horizontal channel nanowire TFETs have been implemented. Vertical channel TFETs are relatively easy to fabricate. Vertical nanowires with a cylindrical shape can be grown with a metal nanoparticle as a catalyst [18], or etched (and oxidized) using a patterned mask [19–21]. The vertical nanowires can be turned into nanowire TFETs by forming the drain, gate electrode, and source.

Figure 16 shows the schematic cross section of a vertical nanowire TFET structure. The gate is formed around the nanowire, and the drain, the channel, and the source are formed in sequence vertically. In order to implement a steep doping profile at the source junction, dopant segregated silicidation process that causes dopants to pile up at the silicide edge is used.

Horizontal channel formation is usually more difficult than vertical channel formation. To form a horizontal channel, a nanowire should be patterned by lithography and etch [22]. For the isolation of channel from the substrate, a silicon-on-insulator (SOI) or silicon-on-nothing (SON) structure is often used. Figure 17 shows a typical shape of a horizontal channel nanowire TFET. The main advantage of horizontal channel devices over vertical channel devices is the

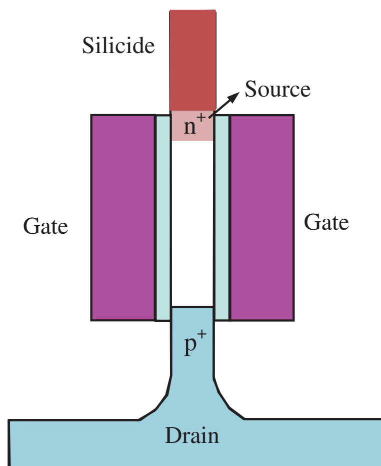


Fig. 16 Schematic cross section of a vertical nanowire TFET structure

Fig. 17 Schematic diagram of a horizontal nanowire TFET structure

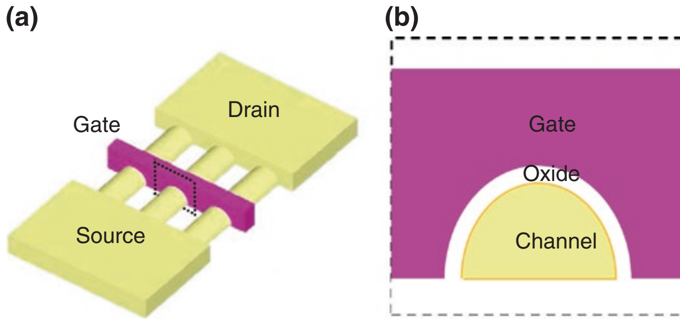


Fig. 18 Hemicylindrical nanowire TFET: **a** schematic diagram and **b** cross section of the channel and the gate. This is the close-up of the dotted area in (**a**)

ease of gate and contact formation. As can be seen in Fig. 17, the arrangement of source, channel, and drain in the horizontal nanowire device is the same as that of a planar device, facilitating the gate and contact formation.

As a variation of cylindrical nanowire TFETs, hemicylindrical nanowire TFETs have been proposed (Fig. 18) [23]. The reason for adopting a hemicylindrical channel instead of a cylindrical channel is that the hemicylindrical channel is easier to fabricate with planar MOSFETs, while the field concentration effect of cylindrical channel is preserved. Since we can consider the hemicylindrical channel as the cylindrical channel cut into half, symmetry ensures that the field pattern of the latter would be maintained in the former. Integration of TFETs and MOSFETs on the same substrate is important since there are some functions of MOSFETs that TFETs cannot implement easily.

If we form another gate (back gate) under the channel of each device, we can control the threshold voltage of the hemicylindrical nanowire TFET. The threshold voltage control of individual devices can be a very useful feature in a few perspectives: (1) accurate adjustment of threshold voltage is possible in the case it is required, (2) multiple threshold voltage circuit can be easily implemented, and (3) variable threshold voltage devices can be employed in low power circuits.

4.3 Doping Engineering

Another method of enhancing the electric field is to insert a thin n-(p-) layer between the p^+-(n^+-) source and the channel in an n-(p-) channel TFET. Such a structure is depicted in Fig. 19. Figure 19a shows source, channel, and drain doping for an n-channel conventional TFET. If we incorporate a thin n-doped layer between p^+- and i- region, we obtain a channel with an n-doped layer near the source (Fig. 19b). This n-doped layer enhances the electric field by the space-charge effect in the depletion region.

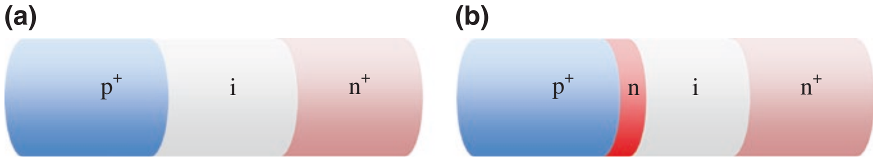


Fig. 19 Insertion of n-doped layer between the p^+-source and the channel in an n-channel TFET: **a** without an n-doped layer and **b** with an n-doped layer. The purpose of the n-doped layer is to enhance the electric field at the junction between the source and the channel

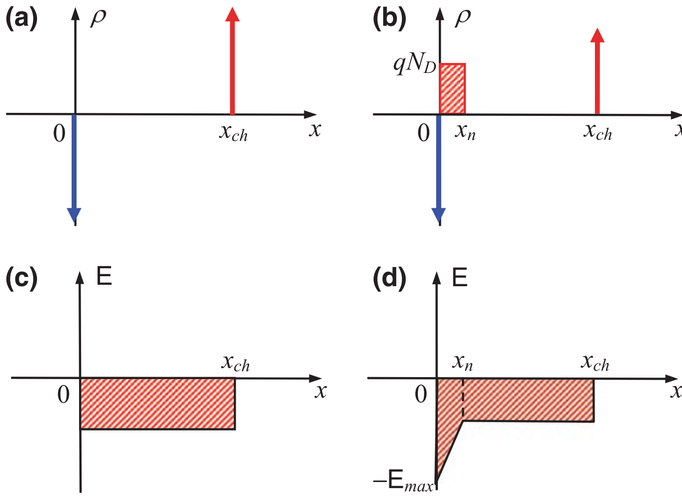


Fig. 20 Space charge and electric fields in the n-channel TFET structure shown in Fig. 19: **a** charge density in the device without n-doped layer, **b** charge density in the device with n-doped layer, **c** electric field the device without n-doped layer, and **d** electric field in the device with n-doped layer

We can understand such a field enhancement effect by considering a simplified model. To illustrate the basic principle of field enhancement, we just consider the one-dimensional effect in this model. We further assume that the source/drain doping is high enough to be treated as infinite concentration. In addition, the n-doped layer is assumed to be completely depleted. The same bias voltage is applied to both structures. Such a model is depicted in Fig. 20. If we consider the charge densities in the whole structure, they should appear as shown in Fig. 20a, b. Due to the infinite doping concentration, there should be delta function charge densities at the source/drain edge of the channel. In the n-channel TFET without an n-doped layer, there should be no charge density. In the n-channel TFET with an n-doped layer, there should be nonzero charge density in the n-doped layer. The electric fields for the n-channel TFETs without and with the n-doped layer are shown in Figs. 20c, d.

Since the integration of electric field gives the potential difference across the channel, the area marked by slanting lines in the electric field plots should be the same. This condition gives us the following relationship between the applied reverse bias, V_R, and the maximum electric field, E_{max}.

$$V_R + V_{bi} = \frac{qN_D x_n}{2\varepsilon_{si}} + \left(E\mathrm{max} - \frac{\rho_n}{\varepsilon_{si}}\right) x_{ch} \tag{17}$$

From this equation, we can obtain the maximum electric field as follows.

$$E\mathrm{max} = \frac{V_R + V_{bi}}{x_{ch}} + \frac{qN_D}{\varepsilon_{si}}\left(1 - \frac{x_n}{2x_{ch}}\right) \tag{18}$$

Since $x_n < x_{ch}$ and the electric field in the structure without the n-doped layer is $\frac{V_R+V_{bi}}{x_{ch}}$, the maximum electric field with the n-doped layer is higher than that without the n-doped layer.

In order to confirm the effect of the n-doped layer on the drain current in a cylindrical channel TFET, device simulation has been carried out and the results are shown in Fig. 21. In this device, the channel diameter (d_{ch}) is 10 nm, the oxide thickness is 1.1 nm, and the channel length is 22 nm. The doping concentration of the n-doped layer (N_D) is 3×10^{19} cm^{-3}, and its thickness (x_n) is 4 nm. The device with the n-doped layer shows a dramatic improvement in the drain current characteristics. The average subthreshold swing decreases significantly, and the ON current is improved at least by an order of magnitude. It is also confirmed that these characteristics are improved further as the channel diameter decreases.

Doping concentration of the source region is also an important parameter that should be optimized to improve the subthreshold characteristics of TFETs. If we consider only the electric field intensity, higher doping concentration is preferable to the lower one. When the subthreshold characteristics are considered, however, a careful analysis is required in terms of point subthreshold swing that can be a function of doping concentration in the source of a TFET. Equations 14 and 15 do not show any dependence of tunneling current on the source doping concentration,

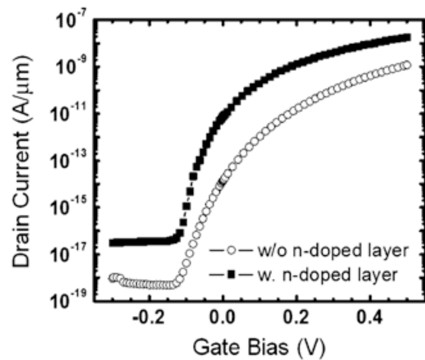


Fig. 21 Effect of a thin n-doped layer on the drain current in a cylindrical channel TFET. The channel diameter (d_{ch}) is 10 nm, the oxide thickness is 1.1 nm, and the channel length is 22 nm. The doping concentration of the n-doped layer (N_D) is $3\text{x}10^{19}$ cm^{-3}, and its thickness (x_n) is 4 nm

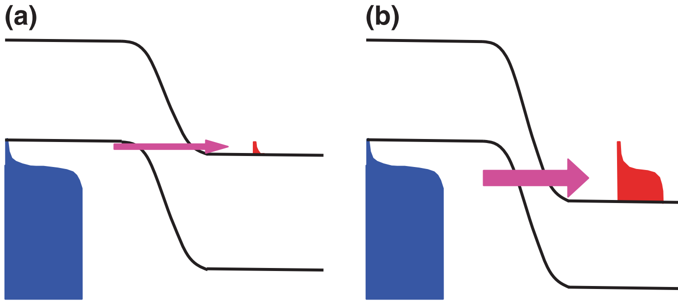


Fig. 22 Band diagram and the Fermi–Dirac distribution function at the source: **a** with a low gate bias voltage, and **b** with a high gate bias voltage. The Fermi energy is located at least a few times the thermal energy ($k_B T$) below the valence band edge

since the electron occupation probability (Fermi–Dirac distribution function) in the source is assumed to be one. In TFETs with a degenerately doped source, however, the Fermi energy level resides within the valence band and the electron occupation probability can be much smaller than one, especially at the energy near the valence band edge.

Figure 22 shows the band diagram and the Fermi–Dirac distribution function at the source, when the Fermi energy is located at least a few $k_B T$ below the valence band edge. If the gate bias voltage is low, only the electrons with a energy near the valence band edge can tunnel through the barrier as shown in Fig. 22a. Since these electrons follow the Boltzmann distribution, the tunneling current can show a characteristic similar to that of thermal carrier injection. That is, the Boltzmann factor will appear in the current–voltage characteristics. Such modification can increase the subthreshold swing at low gate bias voltages and degrade the performance of TFETs.

If the gate bias voltage is high, however, the influence of the Boltzmann factor will disappear and the assumption of full occupation of states becomes valid, even though the energy range where that assumption is valid is somewhat shrunk as shown in Fig. 22b. Without the Boltzmann factor, the original tunneling current equation (Eq. (14)) will be restored and the point subthreshold swing of typical TFET will also be recovered.

In summary, the source doping concentration should not be too high. If it is too high, it might increase the subthreshold swing at low gate bias voltages, degrading the performance of the TFET. If it is too low, the electric field can decrease and the ON current may degrade. The appropriate source doping concentration can be found by adjusting the doping level to the value at which the Fermi level is close to the valence band edge.

4.4 Tunneling Area

In a planar TFET structure, the tunneling region is confined to a tiny area of the source–channel junction near the surface of the channel, as shown in Fig. 23. This is because tunneling is extremely sensitive to the electric field and high electric field is generated only in the region where the inversion layer and the source meet. Since the inversion layer is very thin, the high electric field region cannot be extended toward the body along the source junction, resulting in a small tunneling area. Such a characteristic is also important in MOSFETs, but TFETs usually suffer more from it, since the current is almost completely determined by tunneling at the source–channel junction.

If we increase the tunneling area by structure engineering, the tunneling current can be increased significantly. One of such approaches is to use the extended source structure shown in Fig. 24. In this structure, the source region is extended into the channel, while leaving a narrow undoped region between the channel surface and the source region. The interband tunneling occurs mainly in the direction perpendicular to the channel and the tunneling area is increased drastically.

In addition to the increase in tunneling area, the change in the tunneling direction helps the enhancement of electric field. Since the tunneling occurs toward the gate, the direction of tunneling current coincides with that of gate electric field.

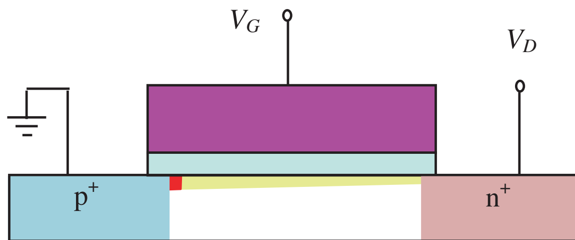


Fig. 23 Tunneling region in a planar TFET. It is confined to a tiny spot of the source–channel junction near the surface of the channel. The tunneling region is marked in red

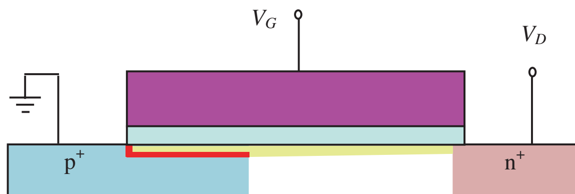


Fig. 24 Planar TFET structure with an extended source. The source region is extended into the channel, while leaving a narrow undoped region between the channel surface and the source region. The interband tunneling occurs mainly in the direction perpendicular to the channel and the tunneling area is increased drastically

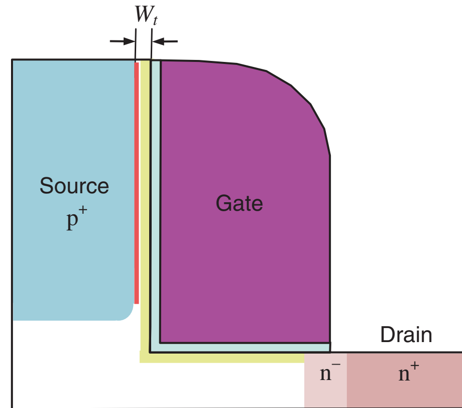


Fig. 25 L-shaped TFET structure. The extended channel region is rotated by 90° and the direction of tunneling is horizontal. W_t is the width of the narrow undoped channel

This coincidence enhances the effectiveness of the gate bias voltage in inducing high electric field at the tunneling area. In the case of conventional TFET (Fig. 23), the direction of tunneling current is perpendicular to the gate electric field. Because of this change in direction, the influence of the gate bias voltage should be indirect and weak in the conventional TFET.

One major drawback of the planar TFET with an extended source is its large footprint. That is, the active area increases significantly due to the extended source region. In order to solve the problem of large footprint, an **L-shaped TFET** structure (Fig. 25) is proposed [24]. It is named as L-shaped TFET, because its channel resembles the alphabet L. In L-shaped TFETs, the extended source region is rotated by 90°, so that the extended source and the narrow undoped channel "stand up" like the vertical line of the letter L. The direction of tunneling is now horizontal, but still perpendicular to the vertical portion of the channel. Since the tunneling area is independent of the footprint in L-shaped TFETs, it can be increased as much as the fabrication capability tolerates.

In this structure, two additional features are introduced. One is the rounding of the lower right corner of the source region. If we leave a sharp edge at this corner, tunneling current will be crowded near it. Since this type of current crowding is not desirable, the corner is rounded to avoid it. Another feature is the lightly doped drain (LDD) located at the drain edge of the channel. The purpose of LDD is to suppress the ambipolar characteristic, which can increase the OFF current significantly. Since the cause of the ambipolar characteristic is unwanted tunneling at the drain junction, LDD can effectively suppress the OFF current by reducing the electric field at the drain junction.

One key parameter that needs to be optimized in this structure is the width, W_t, of the narrow undoped channel (Fig. 25). W_t is roughly the width of the tunneling barrier between the source and the channel, so that the tunneling probability would

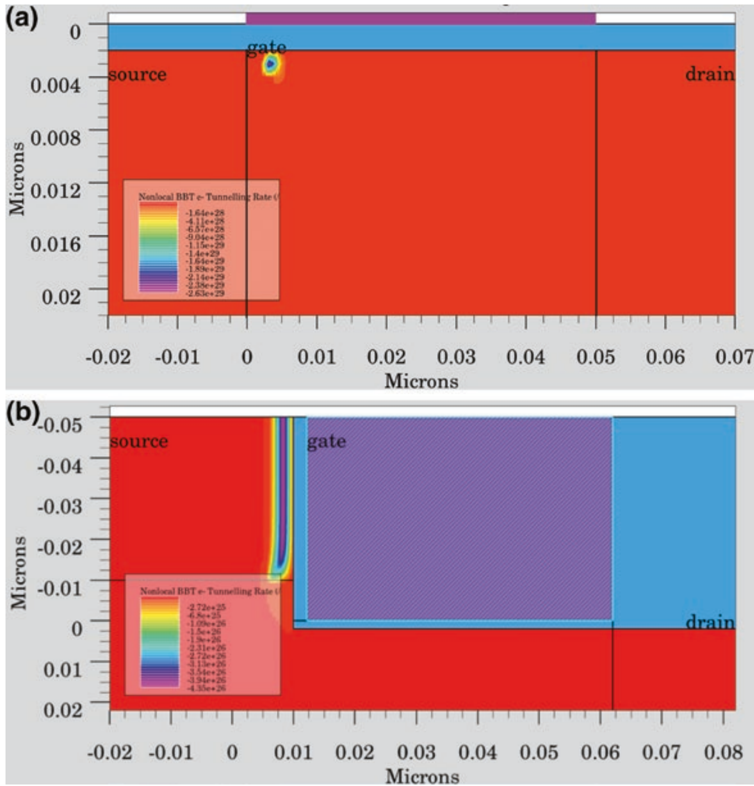


Fig. 26 Two-dimensional (2D) contour plot of electron tunneling rates for **a** a conventional planar TFET, and **b** an L-shaped TFET

be strongly dependent on it. If we consider only the subthreshold swing, W_t should be as small as possible. If W_t is too small, however, the threshold voltage becomes too large, which is not desirable for low power operation. Thus, $W_t \approx 4$ nm is found to be optimum. In order to form an ultrathin undoped layer on top of a heavily doped region, low-temperature epitaxial growth is required.

Figure 26 shows two-dimensional (2D) contour plot of electron tunneling rates for a conventional planar TFET and an L-shaped TFET, when the device is fully turned on. In the planar TFET, the region with high tunneling rate is confined to a tiny spot. On the other hand, the L-shaped TFET shows that the region with high tunneling rate is extended along the source junction that faces the gate. It is notable that the high tunneling rate is maintained throughout the entire length of the narrow undoped channel.

5 Bandgap Engineering for Tunneling Field-Effect Transistors

5.1 Impact of Bandgap on Tunneling Probability

We can boost the current density in a silicon-based TFET by changing the channel material with a narrow bandgap material such as germanium. The advantage of using a narrow bandgap material can be easily seen by examining the interband tunneling current density given in Eq. (14). Since the dominant factor in this equation is the exponential function whose argument is proportional to the 3/2 power of the bandgap, the current density should be a sensitive function of the bandgap. Due to the negative sign in the argument of the exponential function, the current density should increase exponentially as the bandgap decreases. Thus, using a narrow bandgap material as the channel material is an efficient method of boosting the current density.

Figure 27 shows the impact of bandgap on tunneling probability. The tunneling probability is calculated as a function of electric field. To analyze only the effect of bandgap, we assume that the effective mass of electron, m^*, is 0.25 m_0, where m_0 is the electron mass in vacuum. As expected, the tunneling probability increases at least by an order of magnitude when the bandgap decreases by 0.3–0.4 V. The difference in the tunneling probability increases rapidly as the electric field decreases. For example, the differences become several orders of magnitude when the electric field is 1.5 MV/cm. From this observation, we can deduce that the average subthreshold swing will decrease significantly as the bandgap decreases.

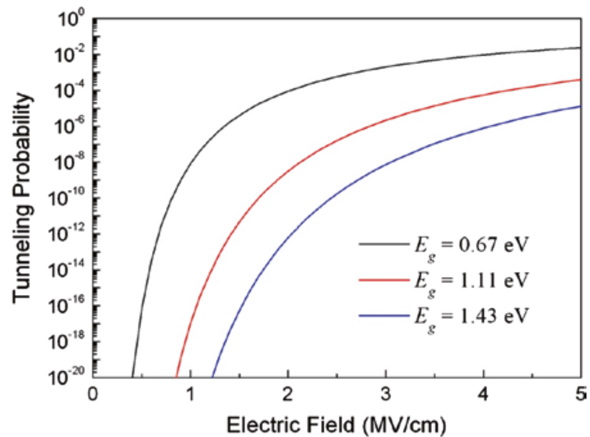


Fig. 27 Tunneling probability as a function of electric field for three bandgaps: ($E_g = 0.67$ eV, $E_g = 1.11$ eV, and $E_g = 1.43$ eV). The three bandgaps correspond to germanium (Ge), silicon (Si), and gallium arsenide (GaAs), respectively

5.2 Ambipolar Characteristics in Devices with a Narrow Bandgap Channel

One problem associated with the narrow bandgap channel material is its worse ambipolar characteristic compared with that of the wide bandgap material. When the bandgap of the channel material is large, the tunneling probability is small at the low gate bias voltage (OFF state) as shown in Fig. 28a, since the width and the height of the tunnel barrier are still relatively large. The narrow bandgap material, however, can cause trouble in the supposedly OFF current under the condition of low gate and high drain bias. As shown in Fig. 28b, the narrow bandgap in the drain edge of the channel can form a low and thin tunnel barrier, so that the tunneling current at the drain junction would be much higher than the case of the wide bandgap material.

Figure 29 compares the ambipolar characteristics of TFETs with a narrow and wide bandgap channel material. In the TFET with a wide bandgap channel, the OFF characteristic stays far away from the ON characteristic, since it takes quite a large negative shift in the gate bias voltage to create the tunneling condition at the drain junction (Fig. 28a). On the contrary, the TFET with a narrow bandgap channel requires only a small or no shift in the gate bias voltage for a similar tunneling condition (Fig. 28b). Thus, the ambipolar characteristic is quite strong in devices with a narrow bandgap channel and generates a huge current in the OFF characteristic. If the OFF characteristic overlaps the ON characteristic as in Fig. 29, the most valuable portion of the ON characteristic, which scores the smallest subthreshold swing, is buried under the OFF characteristic. Then, the major advantage of the TFET, which is a very small OFF current, disappears.

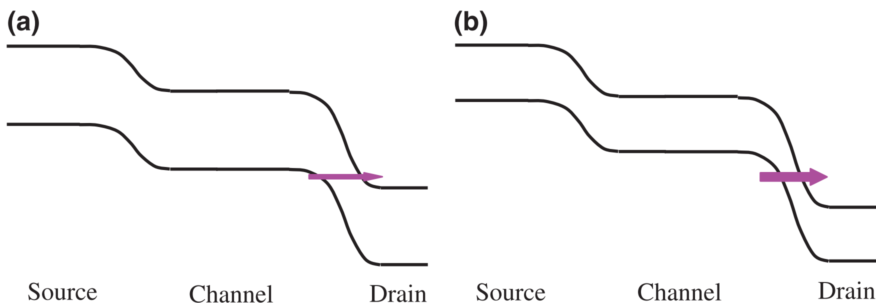


Fig. 28 Band diagrams of n-channel TFETs: **a** with a wide bandgap channel material, and **b** with a narrow bandgap channel material. The gate bias voltage is low in both cases (OFF state)

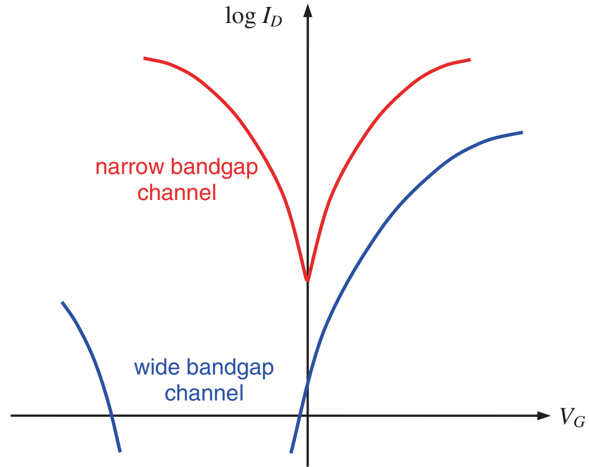


Fig. 29 Ambipolar behavior of TFETs with a narrow and wide bandgap channel material

5.3 Bandgap Engineering

One method of suppressing the ambipolar behavior in devices with a narrow bandgap is to use a narrow bandgap material only at the source region, while maintaining a large bandgap material near the drain region. Figure 30 shows a bandgap-engineered TFET structure which is designed to have a large ON current while the ambipolar behavior is suppressed. Under the high gate and high drain bias voltage (Fig. 30a), the source junction is reverse-biased and the current density increases due to the narrow bandgap of the source material. Under the low gate and high drain bias voltage (Fig. 30b), the drain junction is reverse-biased, but not much current can flow due to the thick tunnel barrier formed by the wide bandgap material.

Incorporation of a narrow bandgap material into the source region is relatively easy in a vertical channel nanowire TFETs, since the lattice mismatch between

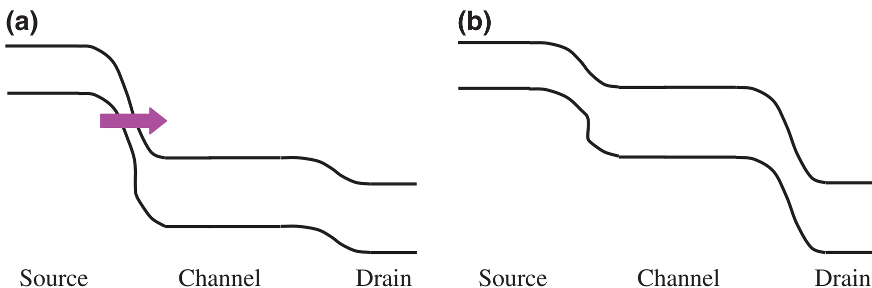


Fig. 30 Bandgap-engineered n-channel TFET with a narrow bandgap channel material at the source region and a wide bandgap material at the channel and the drain region: **a** under high gate and high drain bias voltage, and **b** under low gate and high drain bias voltage

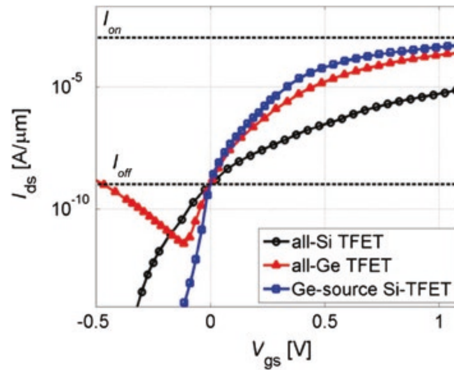


Fig. 31 Comparison between the homostructure and heterostructure TFETs by simulation. The dashed lines represent the ON current (10^{-3} A/μm) and OFF current (10^{-9} A/μm) of the ITRS 2003 device specifications. The subthreshold characteristics are shifted horizontally such that the required OFF current is achieved at a gate bias voltage, $V_{gs} = 0$ V. Reproduced with permission from Ref. [25] © 2008 AIP

two materials does not matter much in the vertical nanowire structure due to the small cross-sectional area. Nanowire TFETs with a germanium source have been one of the most sought-after structures up to now [25, 26]. Germanium (Ge) is a group IV material just like silicon (Si), so that the material compatibility with silicon is good except for the lattice mismatch. In a vertical nanowire structure grown with a metal nanoparticle as a catalyst, the formation of germanium source on top of silicon can be easily done by switching the source gas.

Simulated subthreshold characteristics for all-Si, all-Ge, and Ge-source Si-TFETs are shown in Fig. 31 [25]. The all-Si TFET shows about two orders of magnitude lower current than the all-Ge and Ge-source Si TFET. In the all-Ge TFET, however, ambipolar behavior is clearly observed. The Ge-source Si TFET exhibits the highest ON current for the same OFF current.

What is important in determining the current density is not only the bandgap but also the band offset of the **heterojunction**. In Fig. 30, the band offset occurs only in the valence band, and the conduction band has no offset at the heterointerface. This type of band alignment is not favorable to the p-channel TFET. As shown in Fig. 32a, there would be an additional potential barrier caused by the band offset. The effect of the narrow bandgap in the source is compensated by the additional potential barrier. In order to utilize the advantage of the narrow bandgap in the source, the band offset should occur only in the conduction band as shown in Fig. 32b. Note that the band diagram in Fig. 32b is the mirror image of that in Fig. 30a. Thus by considering hole injection from the source to the channel, we can easily explain how the current density is increased and the ambipolar behavior is suppressed in this bandgap-engineered device.

To implement the band diagram in Fig. 32b, indium arsenide (InAs) is often used as a source material [26, 27]. InAs has a bandgap of 0.36 eV and an electron

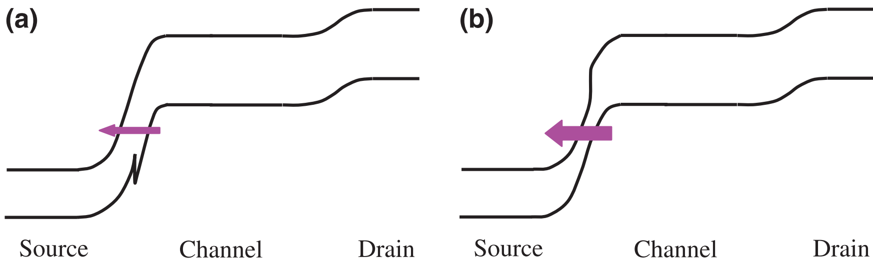


Fig. 32 Effect of band alignment on the tunneling current of p-channel TFETs: **a** when the band offset occurs in the valence band, and **b** when the band offset occurs in the conduction band

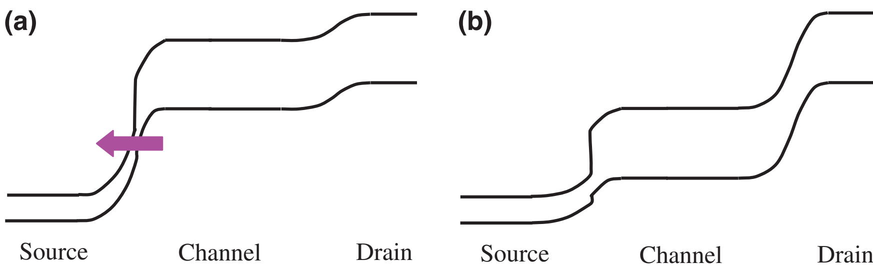


Fig. 33 Bandgap-engineered p-channel TFET with a narrow bandgap channel material (InAs) at the source region and a wide bandgap material (Si) at the channel and the drain region: **a** under high gate and high drain bias voltage, and **b** under low gate and high drain bias voltage

affinity of 4.9 eV. Thus, InAs forms a staggered gap heterojunction with Si as shown in Fig. 33. The conduction band offset is 0.85 eV and the valence band offset is −0.11 eV. The staggered gap heterojunction can provide a very low and thin tunnel barrier near the heterojunction. Such a characteristic is very useful in boosting current. In addition to the small bandgap of InAs, the favorable band offset in the InAs-Si heterojunction can enable a large ON current in the p-channel TFET.

Acknowledgments This work is supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT and Future Planning as the Global Frontier Project.

References

1. ITRS Committee (2013) International technology roadmap for semiconductors. http://public.itrs.net/
2. Gopalakrishnan K, Griffin PB, Plummer JD (2002) I-MOS: A novel semiconductor device with a subthreshold slope lower than kT/q. In: IEDM Technical Digest, p 289
3. Gopalakrishnan K, Griffin PB, Plummer JD (2005) Impact ionization MOS (I-MOS)-part I: device and circuit simulations. IEEE Trans. Electron Dev 52:69
4. Gopalakrishnan K, Woo R, Jungemann C, Griffin PB, Plummer JD (2005) Impact ionization MOS (I-MOS)-part II: experimental results. IEEE Trans. Electron Dev 52:77

5. Choi WY, Song JY, Lee JD, Park YJ, Park B-G (2005) A novel biasing scheme for I-MOS (impact-ionization MOS) devices. IEEE Trans Nanotechnol 4:322
6. Choi WY, Song JY, Lee JD, Park YJ, Park BG (2005) 70-nm impact-ionization metal-oxide-semiconductor (I-MOS) devices integrated with tunneling field-effect transistors (TFETs). In: IEDM Technical Digest, p 975
7. Kinaret J, Nord T, Viefers S (2003) A carbon-nanotube-based nanorelay. Appl Phys Lett 82:1287
8. Kam H, Lee DT, Howe RT, King T-J (2005) A new nano-electro-mechanical field effect transistor (NEMFET) design for low-power electronics. In: IEDM Technical Digest, p 463
9. Abele N, Fritschi N, Boucart K, Casset F, Ancey P, Ionescu AM (2005) Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor. In: IEDM Technical Digest, p 1075
10. Yousif M, Lundgren P, Ghavanini F, Enoksson P, Bengtsson S (2008) CMOS considerations in nanoelectromechanical carbon nanotube-based switches. Nanotechnology 19:285204
11. Loh OY, Espinosa HD (2012) Nanoelectromechanical contact switches. Nature Nanotechnol 7:283
12. Appenzeller J, Lin Y-M, Knoch J, Avouris P (2004) Band-to-band tunneling in carbon nanotube field-effect transistors. Phys Rev Lett 93:196805
13. Bhuwalka KK, Schulze J, Eisele I (2004) Performance enhancement of vertical tunnel field-effect transistor with SiGe in the δp^+ layer. Jpn J Appl Phys 43:4073
14. Zhang Q, Shao W, Seabaugh A (2006) Low-subthreshold-swing tunnel transistors. IEEE Electron Device Lett 27:297
15. Nirschl T, Weis M, Fulde M, Schmitt-Landsiedel D (2007) Correction to revision of tunneling field-effect transistor in standard CMOS technologies. IEEE Electron Device Lett 28:315
16. Choi WY, Park B-G, Lee JD, King Liu T-J (2007) Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec. IEEE Electron Device Lett 28(8):743
17. Fair RB, Wivell HW (1976) Zener and avalanche breakdown in As-implanted low voltage Si n-p junctions. IEEE Trans Electron Devices 23:512
18. Vallett AL, Minassian S, Kaszuba P, Datta S, Redwing JM, Mayer HS (2010) Fabrication and characterization of axially doped silicon nanowire tunnel field-effect transistors. Nano Lett 10:4813
19. Chen ZX, Yu HY, Singh N, Shen NS, Sayanthan RD, Lo GQ, Kwong D-L (2009) Demonstration of tunneling FETs based on highly scalable vertical silicon nanowires. IEEE Electron Device Lett 30(7):754
20. Yang B, Buddharaju KD, Teo SHG, Singh N, Lo GQ, Kwong DL (2008) Vertical silicon-nanowire formation and gate-all-around MOSFET. IEEE Electron Device Lett 29(7):791
21. Kwong D-L, Li X, Sun Y, Ramanathan G, Chen ZX, Wong SM, Li Y, Shen NS, Buddharaju K, Yu YH, Lee SJ, Singh N, Lo GQ (2012) Vertical silicon nanowire platform for low power electronics and clean energy applications. J Nanotechnol 2012:492121
22. Lee M, Jeon Y, Jung J-C, Koo S-M, Kim S (2012) Multiple silicon nanowire complementary tunnel transistors for ultralow-power flexible logic applications. Appl Phys Lett 100:253506
23. Park BG, Sun MC, Kim SW (2014) Silicon tunneling field effect transistors with a hemi-cylindrical nanowire channel for ultra-low power application. In: Conference Proceedings, 012021
24. Kim SW, Choi WY, Sun M-C, Kim HW, Park B-G (2012) Design guideline of Si-based L-shaped tunneling field-effect transistors. Japan J Appl Phys 51(06FE09)
25. A.S. Verhulst, William G. Vandenberghe, K. Maex, and G. Groeseneken, "Boosting the on-current of a n-channel nanowire tunnel field-effect transistor by source material optimization," Journal of Applied Physics, 104, 064514 (2008)
26. Ionescu AM, Riel H (2011) Tunnel field-effect transistors as energy-efficient electronic switches. Nature 479:329
27. Tomioka K, Fukui T (2011) Tunnel field-effect transistor using InAs nanowire/Si heterojunction. Appl Phys Lett 98:083114

Bulk FinFETs: Design at 14 nm Node and Key Characteristics

Jong-Ho Lee

Abstract In contrast to conventional 2-D MOSFETs, FinFETs are able to be scaled down to 20 nm and beyond, and have superior performance. There are two types of FinFETs:SOI FinFETs and bulk FinFETs. Bulk FinFETs are built on bulk-Si wafers, which have less defect density and are cheaper than SOI wafers, while also having better heat transfer rate to the substrate compared to SOI FinFETs. In 2011, Intel announced the world's first 3-D transistors in the mass production of a 22 nm microprocessor (code-named Ivy Bridge). The 3-D transistors adopted by Intel are actually bulk FinFETs. In this chapter, we provide the design guidelines for bulk FinFETs at the 14 nm node, and compare bulk and SOI FinFETs in terms of scalability, parasitic capacitance, and heat dissipation. Decrease of the drain current by parasitic resistance in the source (S) and drain (D) regions is also addressed. Drain current fluctuation by single charge trap is studied in terms of the trap depth, trap position, and percolation path. In the design of 14 nm bulk FinFETs, a punch-through stopper at a position just under the S/D junction depth is required to suppress unwanted cross-talk between S and D. The peak concentration of the stopper needs to be 2–3 $\times$ 10^{18} cm^{-3}. The S/D junction depth should be equal or slightly smaller than the height of fin body, defined from the surface of the isolation oxide region to the top of the fin body. Considering the short channel effect and drain current drivability, the reasonable doping concentration of uniformly doped fin body is 2–3 $\times$ 10^{17} cm^{-3}. To keep the drain-induced barrier below 100 mV/V when the length between the S and D junctions is the same as the gate length (14 nm), the width of the fin body should be ~9 nm. Under the same doping concentration and geometry, both 14 nm SOI and bulk FinFETs have nearly the same *I–V* characteristics, which mean nearly the same scalability. Since thin fin bodies protruding from the substrate are easily depleted, the junction capacitance of the S/D to fin body can be reduced to similar or even lower values than that of SOI FinFETs. To achieve a similar heat transfer rate to the substrate as

J.-H. Lee (✉)
School of ECE, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea
e-mail: jhl@snu.ac.kr

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_2

bulk FinFETs, the buried oxide in SOI FinFETs should be thinned down to 20 nm or beyond, which could cause unwanted increase in the parasitic capacitance. The contact area between the metal electrode and the S/D region should be as wide as possible to reduce the S/D parasitic resistance.

Keywords Bulk FinFET · Punch-through stopper · Body width · Body doping · Channel length · Fin height · Contact resistance · Capacitance · Temperature

1 Introduction

The basic concept of the field effect transistor (FET) was patented by Lilienfeld [1] in 1930, but was not implemented at that time because of the unavailability of practical means. The first oxides grown in high pressure steam were used by Kahang and Atalla [2] to fabricate the first MOSFET structures in 1960. In 1963, Grove, Deal, and Snow developed the first commercially available process for the construction of MOSFETs with stable characteristics [3]. Since around 1960, MOSFETs with planar (or two-dimensional) channel structure have been applied to integrated circuits (ICs) for the cheap and efficient performance of functions such as digital and mixed signal processing, signal amplification, computing, and information storage. Thus, silicon MOSFETs have emerged since the 1980s as the predominant technology of the microelectronics industry. MOSFETs have been scaled down consistently over many generations of technology manufacturing, resulting in steady improvements in circuit performance (speed), integration density, and cost per function (or bit). The gate lengths of MOSFETs were reduced from the μm range to several tens of nm. Conventional nanoscale bulk-Si MOSFETs suffer from increase of the off-current (I_{OFF}) as one of the short channel effects (SCEs). One method to suppress such short channel effects is to reduce the thickness of the gate oxide, but use of a thin gate oxide increases the gate leakage current. Nanoscale devices require low-resistance, ultra-shallow source/drain junction depth, ultrathin equivalent gate oxide thickness using high-κ dielectrics, and low-resistance metal (or silicide) gate electrodes to meet the specifications of the ITRS [4]. Because of these requirements, it is well known that the scaling-down of bulk MOSFETs with planar channel structure beyond 20 nm is nearly impossible. One of the technology boosters for future technology generations is device structure. Advanced device structures include ultrathin body silicon on insulator (SOI) single-gate transistors [5, 6] and double/triple-gate (or multiple gate) transistors [7–9], both of which have better device scalability than classic bulk-Si transistors [10]. Thus, better device architecture will be necessary to continue achieving the benefits obtained from device scaling in the past. Double-gate (DG) MOSFETs that have top and bottom gates or left and right gates have long been recognized for their potential to better control the SCEs. Triple-gate (TG) MOSFETs display similar behaviors to the DG devices, but have three gate electrodes on the three surfaces of the rectangular-shaped body. Several different

structures for DG and TG MOSFETs have been proposed or demonstrated experimentally [11, 12]. For decades, various kinds of three-dimensional (3-D) transistors have been reported [7–48]. Among them, FinFETs have become the predominant technology because the fabrication process of the device structure is relatively easy and compatible to that of conventional planar channel MOSFETs. Intel announced the mass production of CPUs based on 22 nm 3-D transistors in 2011 [44]. The 3-D transistor was called tri-gate or bulk FinFET. The bulk FinFET was developed at the 14 nm technology node, and is ready to be applied in mass production. Research on the development of FinFET is ongoing at 10 nm [40, 46, 47] and even 7 nm [36, 38] technology node. However, no systematic design guideline on the design of channel and source/drain contact has been presented. Therefore, in this chapter, we focus on the analysis of intrinsic device performance and source/drain contact resistance of the bulk FinFET at the 14 nm node and provide a design guideline of the device.

2 SOI and Bulk FinFETs

First of all, the key properties of SOI and bulk FinFETs are explained briefly in this section. FinFETs [7, 8, 15–48] seem to be very promising for the future CMOS technology including logic and memory applications. As mentioned in the introduction, FinFETs basically have a double-gate structure which yields excellent scaling-down characteristics and high performance. FinFETs have been in development for various applications, such as high speed digital integrated circuits (ICs) [17–21], analog ICs [22, 23], SRAMs [24–26], flash memories [27–31], and DRAMs [32–35].

FinFETs were known to most of us as those fabricated on SOI wafers in the late 1990s and early 2000s, which were basically SOI MOSFETs. When double-gate transistors were first reported, they were mainly demonstrated on SOI substrates to overcome the problems associated with short channel effect (SCE) [7, 9, 11, 15–20]. The FinFETs built on SOI wafers are referred to as SOI FinFETs. These devices are known to have the advantages of easy fabrication and excellent scalability, because no shallow trench isolation (STI) process is required, and they have no leakage path near the junction depth of the source/drain regions. These SOI devices are normally applied to high speed circuits due to their low parasitic capacitances. In general, floating body SOI devices, including single- and double-gate devices, may have floating body problems [49] depending on their doping, Si film thickness, and bias conditions. SOI wafers have higher wafer cost and higher defect density than bulk-Si wafers. Note that the heat generated in the channel of SOI FETs cannot be easily dissipated to the substrate due to the thick buried oxide with a very poor heat transfer rate. Switching from conventional planar MOSFETs to the SOI FinFET means not only active splitting [16] for narrow and depleted channels, but also changing from conventional four-terminal devices to three-terminal (body floating) ones. Consequently, this narrows the circuit operation

windows because of the three-terminal characteristics. Therefore, it would be more appropriate to consider four-terminal FinFETs achieved by connecting the fin body directly to the Si substrate. These four-terminal FinFETs are built on bulk-Si wafers and called body-tied FinFETs, or preferably, bulk FinFETs [26]. When body-tied FinFETs were first reported, they were referred to as Omega (Ω) MOSFETs because the cross-section of the body resembles the Greek letter Ω. F.-L. Yang et al. called their MOSFET as omega FET when it was reported in 2002 IEDM, because the gate structure looks like Ω [16]. To differentiate the body-tied FinFET from the SOI FinFET and the omega FET, we therefore called the body-tied FinFET a bulk FinFET. A brief explanation of FinFET classification is provided in Fig. 1, which shows cross-sectional views of the SOI and bulk FinFETs with the cut fin body. The fin bodies of the SOI and bulk FinFETs are floated and tied to the substrate, respectively. “G” stands for gate electrode.

Bulk FinFETs are more familiar to IC design engineers compared to three-terminal FETs, and the fabrication steps of the devices are compatible with those of the conventional planar (or 2-D) channel CMOS devices fabricated on bulk-Si wafers. A schematic 3-D view of the bulk FinFET is shown in Fig. 2. Here, W_{fin}

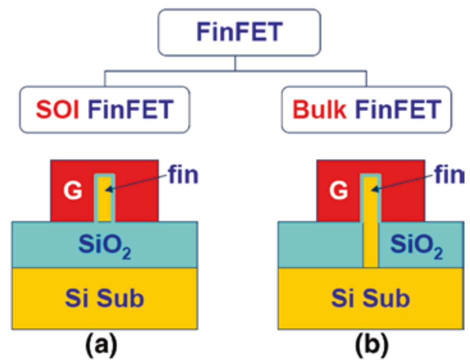


Fig. 1 FinFET classification, and cross-sectional views of SOI (**a**) and bulk (**b**) FinFETs

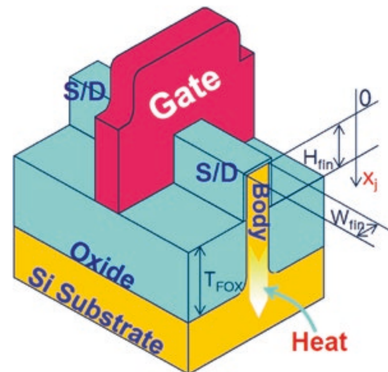


Fig. 2 Three-dimensional schematic of bulk FinFET. H_{fin} and W_{fin} represent fin height and width, respectively. T_{FOX} and x_{j} stand for the field oxide thickness and the source/drain junction depth, respectively

and H_{fin} represent the fin body width and fin height, respectively. The fin height is defined as the height from the surface of the oxide isolation region to the top of the fin body. T_{FOX} represents the thickness of the field oxide for device isolation. The source/drain junction depth is represented by x_{j}. The heat generated in the channel can be transferred to the substrate through the fin body, which is connected to the substrate.

The design of bulk FinFETs eliminates the problems associated with SOI FinFETS, such as expensive wafer cost, high defect density, floating body effect, and poor heat dissipation. In addition, they keep nearly the same scalability as SOI FinFETs while having better heat dissipation characteristics [50]. The key properties of the bulk FinFET will be explained in detail in the chapter on device design at the 14 nm node, and later compared with those of SOI FinFETs.

3 Design of 14 nm Bulk FinFET

In this section, the key parameters are examined, including local doping to suppress punch-through, fin body doping, junction-to-junction length, and fin body width at the technology node of 14 nm. In the 22 nm technology node, the fin body shape looks trapezoidal (or tapered) [51, 52]. Such a body profile seems to easily form isolation oxide between the fins, though a rectangular (or vertical) profile gives better electrical performance [52]. However, at 14 nm and beyond, it becomes difficult to keep the same fin body profile in terms of device scalability and control of the fin body profile. Although bulk FinFET with a tapered fin body at 14 nm has been reported [36], it is reasonable to utilize a vertical fin body profile in the channel region and then a trapezoidal body which increases in width approaching the substrate [38, 39]. In this chapter, we adopt a vertical body profile in the channel region and provide key properties and design guidelines. For simplicity, the fin body protruding from the substrate has vertical side surfaces (rectangular body shape).

Figure 3a shows a 3-D view of a 14 nm bulk FinFET wherein the source and drain regions for the metal contact pad have an epitaxial layer. The top view of the FinFET is shown in Fig. 3b. The gate length (L_{g}) and fin width (W_{fin}) are 14 and 10 nm, respectively. The source and drain regions for metal the contact pad are 40×40 nm^2, and are located 15 nm away from the gate electrode. The size of the metal contact formed on the source and drain regions is 30×30 nm^2, and ideal ohmic contact (interfacial contact resistivity, $\rho_{\text{c}} = 0\ \Omega\ \text{cm}^2$) is assumed. The gate oxide thickness is 1 nm. The source and drains are uniformly doped at the concentration of 5×10^{20} cm^{-3}. In Fig. 3c, a cross-sectional view cut along the source or drain region across the channel length is shown. The fin height (H_{fin}) and field oxide thickness (T_{FOX}) are 100 and 300 nm, respectively. The metal contact for the substrate is formed on the bottom of the 200 nm thick substrate.

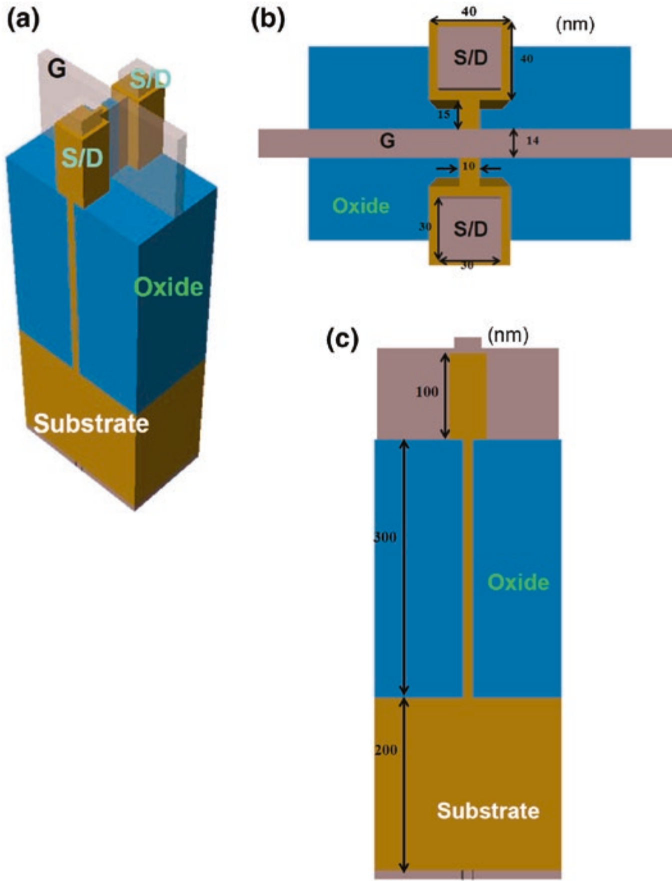


Fig. 3 Three-dimensional view of a 14 nm bulk FinFET for device simulation (**a**). *Top view* of the FinFET (**b**). Gate length (L_g) and Fin width (W_{fin}) are 14 and 10 nm, respectively. Source and drain regions for the metal contact pad are 40 × 40 nm^2. Cross-sectional view cut along the source or drain across the channel length (**c**). Fin height (H_{fin}) and field oxide thickness (T_{FOX}) are 100 and 300 nm, respectively. The gate oxide thickness is 1 nm. The source and drains are doped uniformly at the concentration of 5×10^{20} cm^{-3}. The side surfaces of the fin body profile inside the field oxide are vertical for simplicity of simulation

3.1 Effect of Local Doping (LD) Profile in the Fin Body

The local doping (LD) of bulk FinFETs is required to suppress unwanted bulk punch-through near the junction depth (x_j) of the source and drain regions. In SOI FinFETs, this type of doping is not needed because the buried oxide (BOX) effectively suppresses the punch-through. Figure 4a shows a cross-sectional view cut along the channel length of the bulk FinFET. Local doping is located near the bottom of the source and drain regions, and is called a punch-through stopper (PTS)

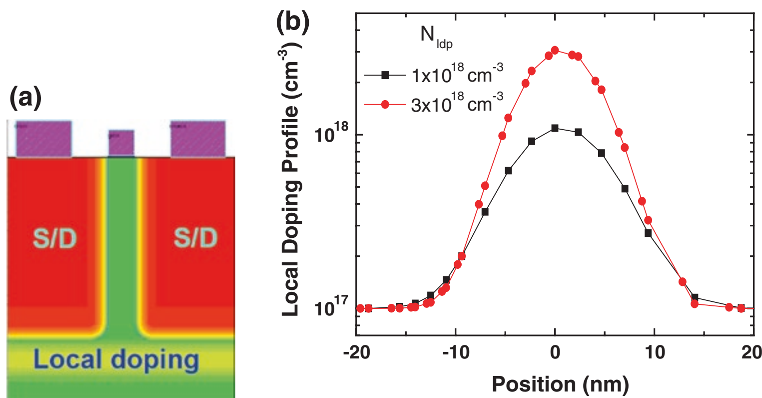


Fig. 4 **a** Cross-sectional view of the bulk FinFET cut along the channel length. The local doping runs laterally near the *bottom* of the source and drain regions. **b** An example of the local doping profile for the punch-through stopper. The peak position was set to 0 nm. N_{ldp} stands for peak doping concentration. Here, the fin body doping concentration was fixed at 1×10^{17} cm^{-3}

[21, 46]. The peak doping concentration and standard deviation of the local doping represent important parameters. Figure 4b shows an example of a local doping profile, wherein N_{ldp} stands for the peak doping concentration of the PTS.

In this chapter, the standard deviation of the doping profile is in the range of ~8 nm, because there is a processing method for vertical localization of the doping profile. A new PTS formation technique utilizing lateral doping by straggling ions from the isolation region (SiO_2) was proposed [21]. In this technology, ions for the PTS are implanted into the isolation region at a tilt angle of 0°. The implanted ions are scattered in the isolation region, and some of the ions scattered laterally penetrate into the fin region, resulting in localized doping. In [21], ions were implanted into the isolation region formed on both sides of the vertical fin body and the top hard mask (Si_3N_4) formed on the fin body. Figure 5a shows a schematic cross-sectional view explaining the implantation into the isolation oxide at a tilt angle of 0° and the fin body profile, which consists of two parts: the vertical fin body where the channel is formed and the tapered fin body wherein the body width is increased as it gets closer to the substrate. In Fig. 5b, the implantation angle is slightly tilted and a hard mask (a layer of nitride, Si_3N_4 in this case) is formed on both sides of the fin body protruding from the isolation oxide. Since the nitride layer is more resistant to the penetration of implanted ions than the SiO_2 layer, the local doping region can be formed effectively using the method shown in Fig. 5b. If doping for the punch-through stopper is distributed vertically in the fin body where the channel is formed, the threshold voltage will be increased, causing clear degradation of the current drivability due to increased impurity scattering. In that case, there would be variation in the threshold voltage due to random doping fluctuation. Thus, it is important that the doping profile be localized vertically to suppress the punch-through.

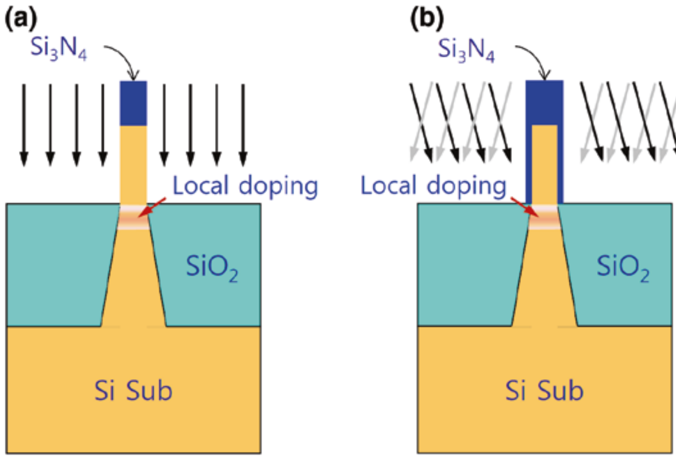


Fig. 5 Cross-sectional views of the bulk FinFET cut across the channel length. Ion implantations in **a** and **b** were performed with a tilt angle of 0° and a slight tilt, respectively

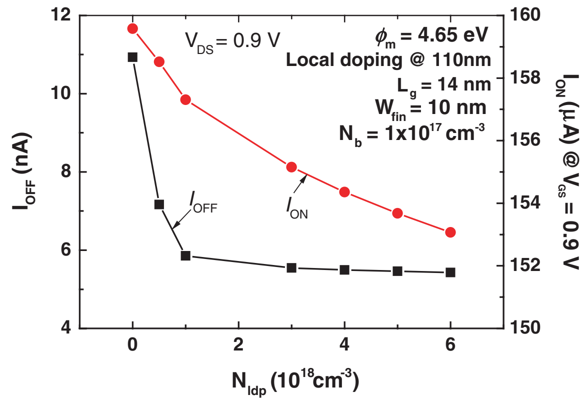


Fig. 6 On-current (I_{ON}) and off-current (I_{OFF}) of 14 nm bulk FinFET with the peak doping concentration of the local doping (N_{ldp})

The peak doping concentration (N_{ldp}) of the PTS is investigated in terms of the drain current characteristics. Figure 6 shows the I_{ON} and I_{OFF} behavior with the N_{ldp} in the 14 nm FinFET at a fixed fin height (H_{fin}) of 100 nm. The gate oxide thickness is 1 nm in the simulated structures. The position of the peak doping concentration is 110 nm, and the source (or drain) junction depth is 100 nm. As can be seen in the figure, when the N_{ldp} increases from 0 to 6 × 10^{18} cm^{-3}, the I_{ON} decreases slowly by 4 %. However, the I_{OFF} demonstrates a significant decrease from 11 to 6 nA with increase of the N_{ldp} from 0 to 6 × 10^{18} cm^{-3}, after which only slight decrease occurs on further increase of the N_{ldp}. From the figure, N_{ldp} values above 1 × 10^{18} cm^{-3} could be considered acceptable. However, if a process margin and the junction capacitance between the source (or drain) to the substrate are considered, the N_{ldp} of 2–3 × 10^{18} cm^{-3} is reasonable.

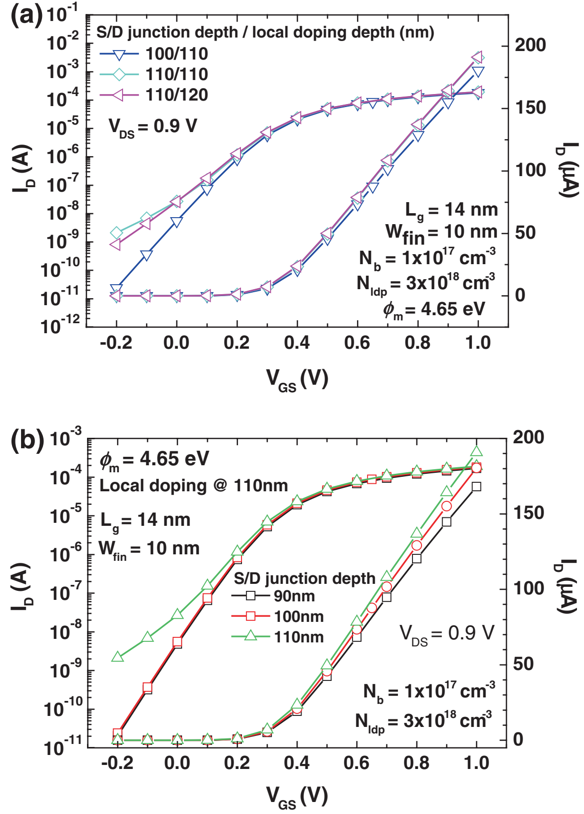


Fig. 7 I_D–V_{GS} curves of 14 nm bulk FinFETsat a fixed V_{DS} of 0.9 V. **a** Examination of position of the peak local doping and x_j as parameters. **b** Change of the x_j at a fixed position of peak local doping of 110 nm

The position effect of the peak doping concentration by changing the source and drain junction depth (x_j) is now discussed. Figure 7a shows the I_D–V_{GS} characteristics of 14 nm FinFETs according to the source (or drain) junction depth and the position of the peak local doping concentration (equivalent local doping depth). The drain bias is set to 0.9 V. The H_{fin}, W_{fin}, and N_{ldp} are 100 nm, 10 nm, and 3×10^{18} cm^{-3}, respectively. In this figure, uniform body doping (N_b), which will be examined in the next section, is assumed to be 1×10^{17} cm^{-3}. The workfunction of the gate electrode (ϕ_m) is 4.65 eV. Three different cases show nearly the same I_{ON}, but with differences of about 4 times in the I_{OFF} values. Inverse triangle symbols represent a reasonable I_D–V_{GS} curve of 14 nm FinFETs when the x_j and position of the peak concentration are 100 and 110 nm, respectively, at a fixed H_{fin} of 100 nm. If the x_j is increased to 110 nm at a fixed H_{fin} of 100 nm, a significant increase in the I_{OFF} occurs, regardless of the position of the peak local doping concentration, as depicted by the diamond and triangle symbols. The results give the important message that the x_j should be less than the H_{fin}, since unwanted cross-talk between the source and drain can be suppressed effectively by the electric field from the gate electrode under such conditions.

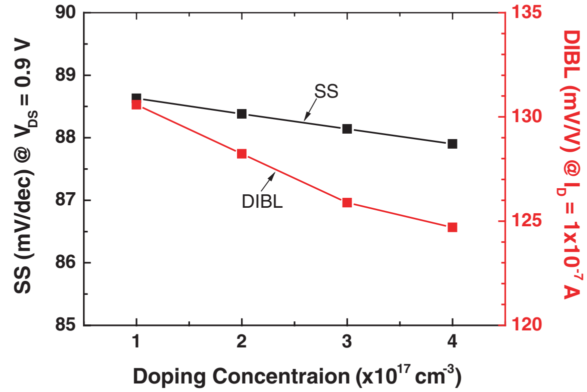


Fig. 8 Subthreshold swing (SS) and drain-induced barrier lowering (DIBL) of 14 nm FinFETs as a parameter of fin body doping concentration. Here, V_{DS} was 0.9 V. The fin body doping was assumed to be uniform

Figure 7b shows similar results to those obtained in Fig. 7a as a function of x_j at a fixed H_{fin} of 100 nm. The I_D–V_{GS} curves of the 14 nm FinFETs with x_j of less than 100 nm are reasonable since the punch-through is suppressed by the electric field from the gate bias. However, if the x_j is increased to 110 nm, a significant increase in the I_{OFF} occurs, as depicted by the triangle symbols. Note that decreasing x_j causes a slight decrease in the I_{ON} because the effective channel width is decreased.

3.2 *Effect of Fin Body Doping*

The doping concentration in the fin body is a key design parameter affecting the subthreshold slope (*SS*), drain-induced barrier lowering (*DIBL*), and carrier mobility. In terms of carrier mobility, the body doping should be kept as low as possible. However, decrease in the body doping causes increase in the *DIBL* and *SS,* as shown in Fig. 8. In the figure, the body doping is assumed to be uniform. As the body doping is increased from 1×10^{17} to 4×10^{17} cm^{-3}, the I_{ON} decreases by about 5 %, while the I_{OFF} decreases more significantly (~40 %). Note that the V_{th} keeps nearly the same values for different concentrations of body doping because it is mainly determined by the work-function of the gate electrode (4.65 eV). When the body doping is increased from 1×10^{17} to 2×10^{17} cm^{-3}, a significant decrease of the I_{OFF} occurs. In Fig. 8, as the body doping increases from 1×10^{17} to 4×10^{17} cm^{-3}, the *DIBL* at an I_D of 10^{-7} A changes within 2.5 %, while the SS at a V_{DS} of 0.9 V changes by about 2 %. Considering all factors mentioned above, the body doping of 2–3 $\times$ 10^{17} cm^{-3} is reasonable.

3.3 *Effect of Junction-to-Junction Length*

Now, the effect of the length between the source and drain junctions (junction-to-junction length) at a fixed channel length of 14 nm is examined.

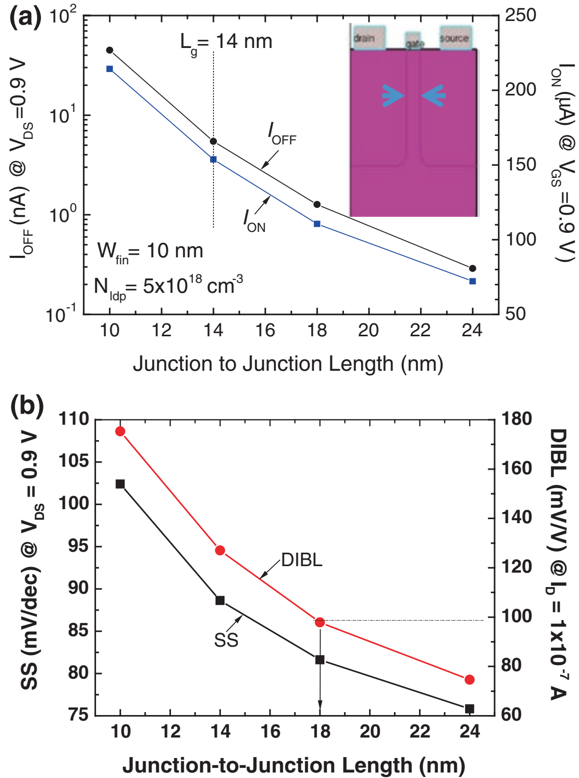


Fig. 9 I_{ON} and I_{OFF} behaviors according to the length between the source and drain (**a**). *SS* and *DIBL* behaviors according to the length (**b**). The *inset* in figure (**a**) represents a cross-sectional view, cut along the channel length. The *arrows* depicts the length (14 nm)

The inset of Fig. 9a illustrates the cross-sectional view cut along the channel length of a bulk FinFET. The arrows were prepared to show the length between the source and drain junctions (junction-to-junction length), which is 14 nm of the example provided. In this figure, the body doping (N_b) is 1 × 10^{17} cm^{-3}. Looking at Fig. 9a, it can be seen that the I_{ON} and I_{OFF} decrease significantly with increase in length. Especially high changes in the I_{OFF} can be observed, by ~100 times with change in the length from 10 to 24 nm. If the length is larger than 14 nm, it can be said that the source and drain are underlapped. When the length is 14 nm, the I_{ON} at the V_{GS} and V_{DS} of 0.9 V is ~728 μA/μm, while the I_{OFF} at the V_{GS} of 0 V and V_{DS} of 0.9 V is 26 nA/μm. Note that only a drift and diffusion model was applied in this simulation. Figure 9b shows the *SS* and *DIBL* characteristics according to the length. As the junction-to-junction length increases, the *SS* and *DIBL* are greatly improved. However, appreciable degradation of the I_{ON} and I_{OFF} occurs with increasing length. If the *DIBL* is kept under 100 mV/V, the length needs to be larger than 18 nm when the body doping is 1 × 10^{17} cm^{-3}. If the reasonable body doping of 2–3 × 10^{17} cm^{-3} is considered, the length can be reduced to ~14 nm. If there is no mention to the junction-to-junction length in the following sections, the length is 14 nm.

3.4 Effect of Fin Body Width

The control of the fin body width (W_{fin}) is very important at a given technology node because the width greatly affects the short channel effect. The fin body width is defined by two arrows in the inset of Fig. 10a. By decreasing the fin body width from 12 to 8 nm, the I_{ON} decreases by 18 %, while the I_{OFF} decreases by ~100 times, as shown in Fig. 10a. Thus, the I_{OFF} is very sensitive to the fin body width, which should be controlled accurately to maintain a narrow I_{OFF} distribution. Note the total effective channel width ($2 \times H_{\text{fin}} + W_{\text{fin}}$) decreases from 212 to 208 nm as the W_{fin} decreases from 12 to 8 nm. The decrease of the total width is just less than 2 % of the W_{fin}. The reason for the decrease of I_{ON} can mainly be attributed to the decrease of V_{th} with decreasing W_{fin}. Conversely, the decrease of I_{OFF} with decreasing W_{fin} can be explained by two reasons: the decrease in V_{th} and increase in *SS* with decreasing W_{fin}. The V_{th} defined at an I_D of 1 μA/μm is decreased by 0.14 V by decreasing the W_{fin} from 12 to 8 nm. Figure 10b shows change of the *SS* and *DIBL* with the W_{fin}. As the W_{fin} decreases from 12 to 8 nm, the *SS* decreases from 106 to 77 mV/dec, while the *DIBL* decreases from 183 to 80 mV/V. Therefore, W_{fin} also has a significant effect on the *SS* and

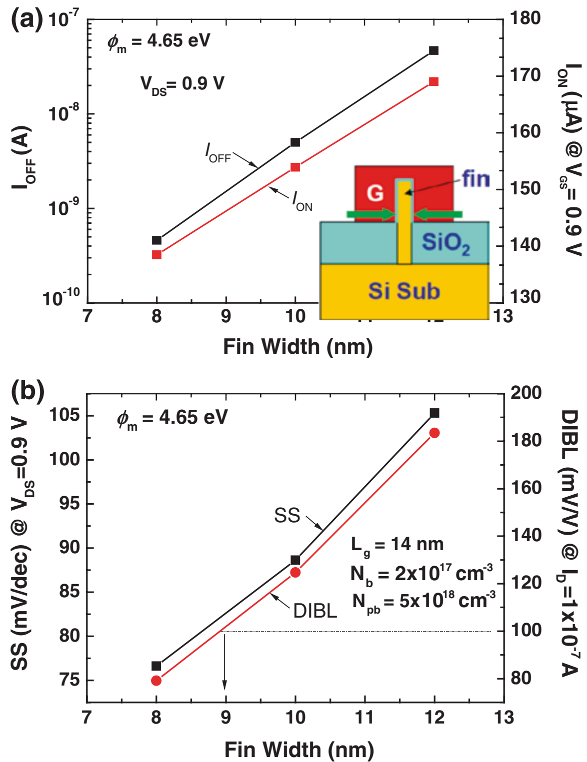


Fig. 10 I_{ON} and I_{OFF} behaviors according to the fin body width (**a**). *SS* and *DIBL* behaviors according to the width (**b**). The inset in figure (**a**) represents the cross-sectional view cut across the channel length. The *arrows* indicate the fin body width. The junction-to-junction length in this figure is 14 nm. The fin height and the source/drain junction depth are 100 nm, respectively. The I_{ON} was obtained at the V_{GS} and V_{DS} of 0.9 V, while the I_{OFF} was obtained at the V_{GS} of 0 V and V_{DS} of 0.9 V

DIBL. The W_{fin} of about 9 nm gives a *DIBL* of 100 mV/V at the given body concentration of 2×10^{17} cm^{-3}. If we increase the body doping concentration to 2.5–3 $\times$ 10^{17} cm^{-3}, the *DIBL* decreases, and the W_{fin} required to give a *DIBL* of 100 mV/V will be about 10 nm.

4 Design of 14 nm SOI FinFET

In Sect. 3, we investigated the key design factors of 14 nm bulk FinFETs. Now, we examine the device performance of 14 nm SOI FinFETs and compare the key properties of SOI FinFETs with those of bulk FinFETs.

4.1 *Effect of Fin Body Doping*

In SOI FinFETs, there is no need for the local doping required in bulk FinFETs to suppress punch-through between the source and drain, because it contains buried oxide (*BOX*) located under the fin body where the channel and source/drain are formed. The 3-D schematic view of the 14 nm SOI FinFET is shown in Fig. 11a. The geometry of the device is exactly the same as that of the bulk FinFET shown in Fig. 3, except for the buried oxide. The BOX is 300 nm thick. Figure 11b shows the *SS* and *DIBL* of 14 nm SOI FinFETs as a function of fin body doping

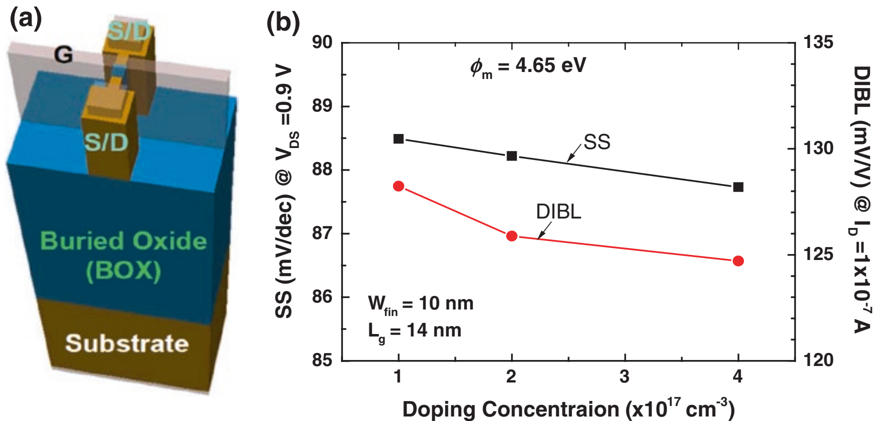


Fig. 11 Three-dimensional schematic view of the 14 nm SOI FinFET (**a**). Device geometry is exactly the same as that of the 14 nm bulk FinFET shown in Fig. 3. The only difference is the buried oxide (*BOX*) present in the SOI FinFET. The thickness of the *BOX* is 300 nm. *SS* and *DIBL* versus the doping concentration of the fin body (**b**). The junction-to-junction length in this figure is 14 nm. The fin height and the source/drain junction depth are 100 nm, respectively. The *SS* was obtained at the V_{DS} of 0.9 V

concentration (N_b). As the N_b increases from 1×10^{17} to 4×10^{17} cm^{-3}, the *SS* decreases from 88.5 to 87.7 mV/dec, while the *DIBL* decreases from 128.2 to 124.7 mV/V. These parameters change only slightly with the N_b, and quite similar to those of the 14 nm bulk FinFET in Fig. 8.

4.2 Effect of Junction-to-Junction Length

The effect of junction-to-junction length on the device performance of 14 nm SOI FinFETs is investigated in Fig. 12.

In Fig. 12, the body doping (N_b) is 1×10^{17} cm^{-3}. With examination of Fig. 12a, it can be seen that I_{ON} and I_{OFF} decrease significantly with increase in the length. Especially high changes in the I_{OFF} can be observed, by ~100 times with change in the length from 10 to 24 nm. At length of 14 nm, the I_{ON} at the V_{GS} and V_{DS} of 0.9 V is ~730 μA/μm, while the I_{OFF} at the V_{GS} of 0 V and V_{DS} of 0.9 V is 26 nA/μm. Note that only a drift and diffusion model was applied in this simulation. The I_{ON} and I_{OFF} of the SOI FinFET are quite similar to those of the bulk FinFET shown in Fig. 9a. Figure 12b shows the *SS* and *DIBL* characteristics of the 14 nm SOI FinFET according to the length. As the junction-to-junction

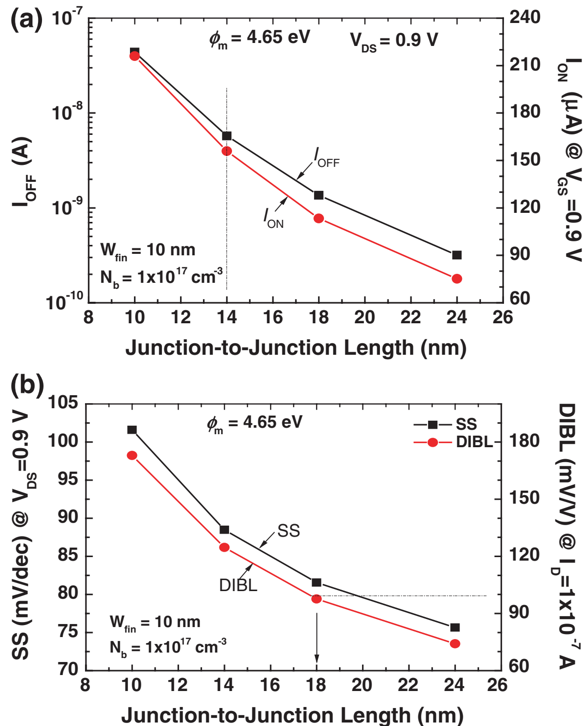


Fig. 12 I_{ON} and I_{OFF} behaviors according to the length between the source and drain (**a**). *SS* and *DIBL* behaviors according to the length (**b**). The I_{ON} was obtained at the V_{GS} and V_{DS} of 0.9 V, while the I_{OFF} was obtained at the V_{GS} of 0 V and V_{DS} of 0.9 V. The fin height and the source/drain junction depth are 100 nm, respectively. The *SS* was obtained at the V_{DS} of 0.9 V

length increases, the *SS* and *DIBL* are greatly improved. However, appreciable degradation of the I_{ON} and I_{OFF} occurs with increase in the length. If a *DIBL* of below 100 mV/V is maintained, the length needs to be larger than 18 nm when the body doping is 1×10^{17} cm^{-3}. If the reasonable body doping of 2–3 $\times$ 10^{17} cm^{-3} is considered, the length can be reduced to ~14 nm. Thus, the 14 nm SOI FinFET shows nearly the same performance as that of the 14 nm bulk FinFET.

4.3 *Effect of Fin Body Width*

As in the case of the bulk FinFET, the effect of fine body width is studied in the 14 nm SOI FinFET in this section. By decreasing the fin body width from 12 to 8 nm, the I_{ON} decreases by 18 % while the I_{OFF} decreases by ~100 times, as shown in Fig. 13a. Since the I_{OFF} is highly sensitive to the fin body width, the width should be controlled accurately to maintain a narrow I_{OFF} distribution. As mentioned in Sect. 3.4, the decrease of the total effective channel width is less than 2 % of the W_{fin}. The decrease in the I_{ON} can mainly be attributed to the decrease in V_{th} with decreasing W_{fin}. Conversely, the decrease in I_{OFF} with

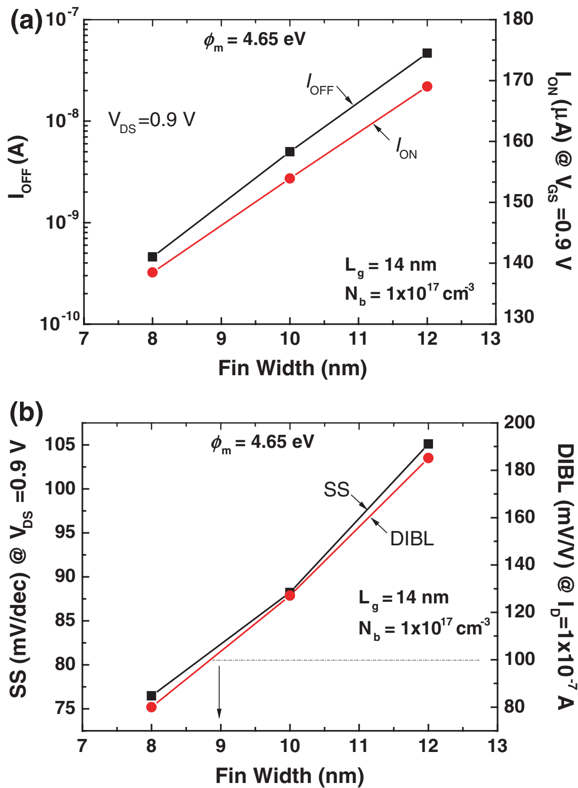


Fig. 13 I_{ON} and I_{OFF} behaviors according to the fin body width (**a**). *SS* and *DIBL* behaviors according to the width (**b**). The *inset* in figure **a** represents the cross sectional view cut across the channel length. The *arrows* indicate the fin width. The junction-to-junction length in this figure is 14 nm. The fin height and the source/drain junction depth are 100 nm, respectively. The I_{ON} was obtained at the V_{GS} and V_{DS} of 0.9 V, while the I_{OFF} was obtained at the V_{GS} of 0 V and V_{DS} of 0.9 V

decreasing W_{fin} can be explained by the accompanying decrease in V_{th} and increase in *SS*. The V_{th} defined at an I_D of 1 μA/μm is decreased by 0.14 V with a decrease in W_{fin} from 12 to 8 nm. Figure 13b shows the *SS* and *DIBL* according to the W_{fin}. As the W_{fin} lowers from 12 nm to 8 nm, the *SS* decreases from 106 to 77 mV/dec, while the *DIBL* decreases from 183 to 80 mV/V. Therefore, W_{fin} also has a significant effect on the *SS* and *DIBL*. The W_{fin} which gives a *DIBL* of 100 mV/V is about 9 nm at a given body concentration of 1×10^{17} cm^{-3}. If we increase the body doping concentration to $2.5–3 \times 10^{17}$ cm^{-3}, the *DIBL* decreases, and the W_{fin} of about 10 nm is required to provide a *DIBL* of 100 mV/V. All analyses done for the bulk FinFET are exactly the same as those for the SOI FinFET.

4.4 Comparison of I_D–V_{GS}s of Bulk and SOI FinFETs

In this section, we compare the I_D–V_{GS} curves of 14 nm bulk and SOI FinFETs. As mentioned above, they have exactly the same geometry, except that the bulk FinFET has local doping located under the source/drain junction depth, whereas the SOI FinFET contains buried oxide at that location instead of local doping. The fin height and width are 100 and 10 nm, respectively. The fin body concentration is uniformly doped with a concentration of 2×10^{17} cm^{-3}. As shown in Fig. 14, bulk and SOI FETs have exactly the same *I–V* characteristics, as depicted by circle and square symbols, respectively.

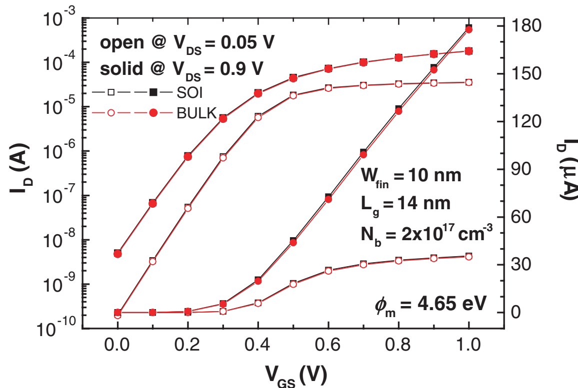


Fig. 14 I_D–V_{GS} curves of 14 nm bulk (*circle*) and SOI (*square*) FinFETs. The junction-to-junction length in this figure is 14 nm. The fin height and source/drain junction depth are 100 nm, respectively. Solid and open symbols represent the curves for the V_{DS} values of 0.9 and 0.05 V, respectively. The fin body is uniformly doped with a concentration of 2×10^{17} cm^{-3}. Gate oxide thickness is 1 nm

5 Parasitic Resistance and Capacitance in Bulk FinFETs

In this section, the parasitic resistance and capacitance of bulk FinFETs are discussed. The resistance is related to the contact resistivity between the source/drain region and the metal, as well as the contact area between them. The resistance in the source/drain regions has significant effects on the I_{ON} and speed characteristic. The capacitance is studied in terms of the junction and gate capacitance. The junction capacitance is formed between the source/drain regions to the substrate in the fin body. Since the speed of the device depends on the capacitance, examination of the capacitance is needed.

5.1 Effect of S/D Contact Resistance

As mentioned in Sect. 4, the width of the fin body in 14 nm FinFETs is about 10 nm. The source/drain and channel are formed in the fin body. The source/drain regions formed in the thin fin body are particularly likely to have high resistance due to diffusion resistance (R_{SD}) and contact resistance (R_{CO}).

Before starting the discussion on contact area and structure, the effects of the interfacial contact resistivity (ρ_c) on the drain current of the FinFET will be examined. Figure 15a shows a schematic top view of a FinFET, wherein the contact resistance and diffusion resistance of the source/drain are depicted. Such parasitic resistance is reflected in the equivalent circuit of a MOSFET, and causes degradation of the drain current and transconductance (g_m). Figure 15b shows the normalized drain current of a 20 nm FinFET at the given V_{GS} and V_{DS} of 0.8 V. Since the FinFET has a fin body width of 5 nm and contact length (l_c) of 50 nm, the contact area between the source/drain region and the metal electrode becomes $5 \times 20\ \text{nm}^2$. If the sheet resistivity of the diffusion resistance (ρ_{sd}) is too small to meet $l_c \ll \sqrt{(\rho_c/\rho_{sd})}$, the contact resistance (R_{CO}) is given by $\rho_c/(W_{fin} \cdot l_c)$. The drain current is normalized to the drain current occurring when the ρ_c is $0\ \Omega\ \text{cm}^2$. As ρ_c increases from $0\ \Omega\ \text{cm}^2$ to $1 \times 10^{-7}\ \Omega\ \text{cm}^2$, the drain current decreases by ~30 %, which means that ρ_c must be kept as low as possible in highly scaled MOSFETs.

Next, the drain current of a 14 nm bulk FinFET is investigated, according to ρ_c. The inset of Fig. 16 represents the 3-D structure of the 14 nm FinFET for 3-D device simulation. The fin body width (W_{fin}) is 10 nm, and the fin height (H_{fin}) is 100 nm. The source and drain regions on both ends of the fin body have a wide width (40 nm), which can be used to effectively reduce the source and drain parasitic resistance. The length of the wide source/drain regions is 40 nm, making the area of the top region $40 \times 40\ \text{nm}^2$. In this inset, the area of the metal electrodes formed on the top surface of the source/drain regions is $30 \times 30\ \text{nm}^2$.

Figure 16 shows the I_{ON} behavior of a 14 nm FinFET at the V_{GS} and V_{DS} of 0.9 V with variation of ρ_c from 0 to $1 \times 10^{-7}\ \Omega\ \text{cm}^2$. Decrease of the I_{ON} from

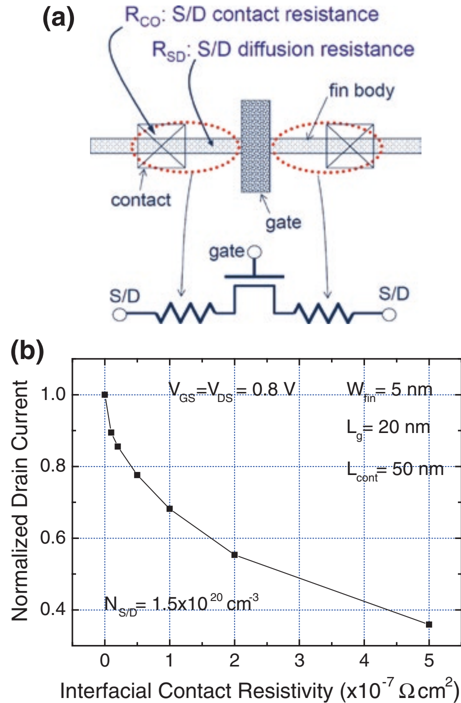


Fig. 15 Schematic *top view* of a FinFET and equivalent circuit including parasitic source and drain resistances (**a**). Normalized drain current of a 20 nm bulk FinFET versus the interfacial contact resistivity between the source/drain region and the metal electrode (**b**). Here, the fin body width (W_{fin}) is 5 nm and the contact length (l_c) is 50 nm. The source and drain regions are doped uniformly with a concentration of 1.5×10^{20} cm^{-3}

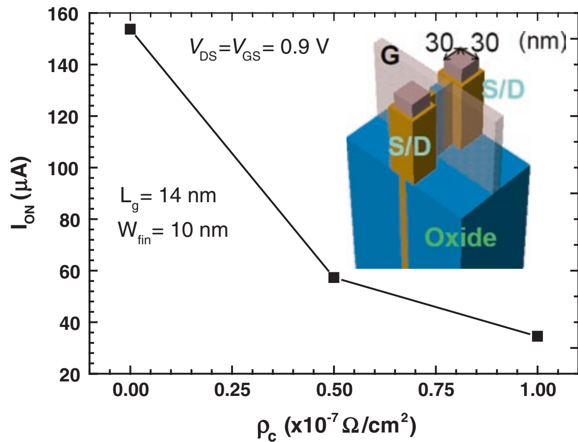


Fig. 16 I_{ON} behavior of 14 nm bulk FinFETs versus interfacial contact resistivity. The *inset* shows the 3-D view of a FinFET. The source and drain regions to provide the contact pad have a height of 100 nm and an area of 40×40 nm^2. Here, the fin body width (W_{fin}) is 10 nm, and the metal contact area on the pad is 30×30 nm^2

153.7 to 34.5 μA occurs as ρ_c increases from 0 to 1×10^{-7} Ω cm^2. The reduction of the current with the increase of ρ_c is remarkable (77.6 % decrease). It should be noted that ρ_c of 1×10^{-7} Ω cm^2 is popular in the industry. Thus, it is imperative to develop new materials and/or process methods to decrease the interfacial contact resistivity. Are there any methods to reduce the contact resistivity other than

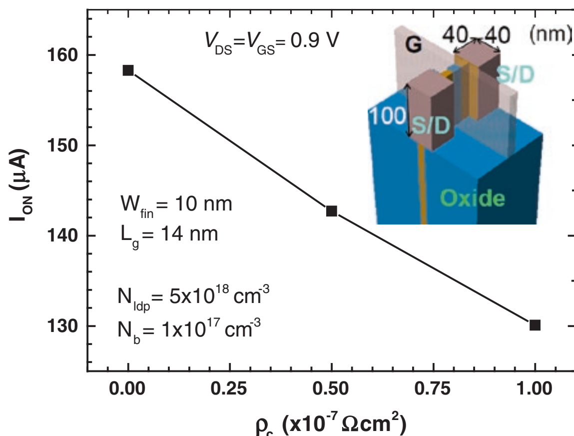


Fig. 17 I_{ON} behavior of 14 nm bulk FinFETs versus interfacial contact resistivity (ρ_c). The *inset* shows the 3-D view of a FinFET. The source and drain regions for provision of the contact pad have a height of 100 nm and an area of 40 × 40 nm^2. Here, the fin body width (W_{fin}) is 10 nm and the metal contact area on the source or drain is 13,600 nm^2 (3 × 40 × 100 nm^2 + 40 × 40 nm^2)

developing new materials and/or process methods? Increasing the contact area between the metal electrode and the source/drain may provide the answer to this question.

Figure 17 shows the I_{ON} behavior according to the ρ_c for a 14 nm bulk FinFET with a wide contact area (3 × 40 × 100 nm^2 + 40 × 40 nm^2 = 13,600 nm^2) between the metal electrode and the source (or drain). The FinFETs in Figs. 16 and 17 have the same doping and geometry, but the contact area in Fig. 17 is about 15 times wider than that examined in Fig. 16. The inset in Fig. 17 clearly illustrates the metal electrode, which covers the surfaces of three sides and the top of the wide source (or drain) region. As ρ_c increases from 0 to 1 × 10^{-7} Ω cm^2, the I_{ON} decreases from 158 to 130 μA (about an 18 % decrease). In an effort to increase the contact area, deposit of *Ni* on epitaxially grown source and drain regions via atomic layer deposition (ALD) was reported, after which nickel silicide (*NiSi*) was formed to provide a wide contact area [53]. In Fig. 18, the I_D–V_{GS} curves of 14 nm bulk FinFETs are compared in terms of the contact area between the metal electrode and the source/drain regions. The FinFETs with narrow (90 nm^2) and wide (13,600 nm^2) contact areas examined are exactly the same as those in Figs. 16 and 17, respectively. When the contact resistivity is 0 Ω cm^2, the I_{ON} is nearly the same, regardless of the contact area. However, a significant reduction in the I_{ON} occurs when the contact resistivity is 1 × 10^{-7} Ω cm^2. From these results, the importance of keeping the contact area between the metal electrode and the source (or drain) region as wide as possible can be understood, in addition to lowering the interfacial contact resistivity.

Now, let us systematically examine the effect of the contact area and structure. Figure 19 shows 3-D structures of 14 nm SOI FinFETs with exactly the same

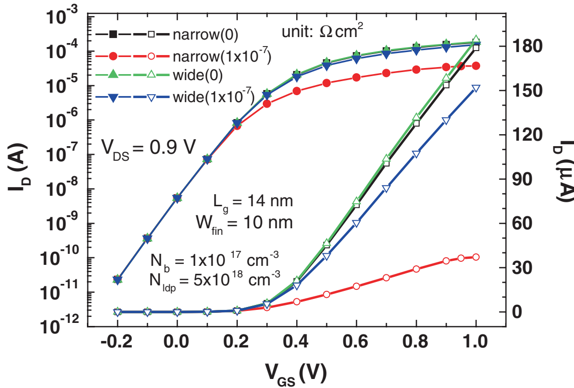


Fig. 18 I_D–V_{GS} curves of 14 nm bulk FinFETs with different contact areas between the metal electrode and the source (or drain) as a parameter of the interfacial contact resistivity (ρ_c). The wide and narrow contact areas are 90 and 13,600 nm^2, respectively

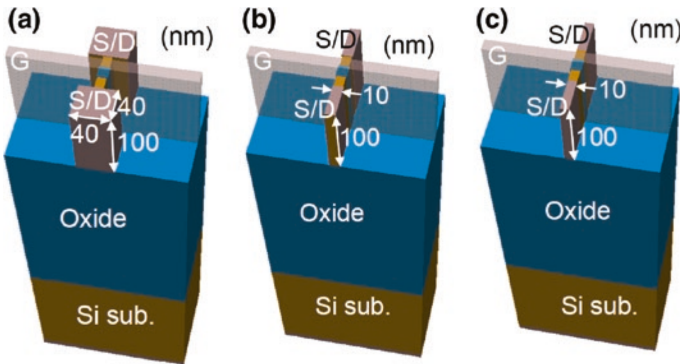


Fig. 19 3-D structures of 14 nm bulk FinFETs with different contact areas and structures. Figure **a** shows the wide metal contact formed on epitaxially grown source and drain regions. The contact area in figure **a** is 13,600 nm^2 (= 3 × 40 × 100 nm^2 + 40 × 40 nm^2). The wide metal contact of 9,400 nm^2 (= 10 × 40 nm^2 + 10 × 100 nm^2 + 2 × 40 × 100 nm^2) is formed on the source and drain regions without an epitaxial layer (**b**). In figure **c**, the source/drain regions have no epitaxial layer, and fully-silicided source and drain regions are formed 15 nm away from the gate electrode. Therefore, the contact area between the metal and the source (or drain) is 400 nm^2 (10 × 40 nm^2)

doping and device geometry, except for the source and drain regions. As mentioned in Fig. 14, both SOI and bulk FinFETs have nearly the same characteristics. Consequently, it is acceptable to consider the SOI FinFETs as bulk FinFETs when investigating the I_{ON} with the metal contact area and the source/drain

structure. Figure 19a has the same metal contact area and source/drain structure as that presented in the inset of Fig. 17. The epitaxial layer is formed on the fin source/drain regions at a width of 10 nm. The source/drain regions have a width of 10 nm from the edge of the gate electrode until 15 nm away from the gate, and a width of 40 nm from 15 nm away from the gate to 55 nm away, along the fins. The metal electrode for the contact of the source/drain region is assumed to cover the top and side surfaces of the wide source/drain region. Therefore, the contact area is 13,600 nm^2 (= 40 × 40 nm^2 + 3 × 100 × 40 nm^2). In Fig. 19b, the source/drain structure has no epitaxial layer, and the metal contact area is 9,400 nm^2 (= 10 × 40 nm^2 + 10 × 100 nm^2 + 2 × 100 × 40 nm^2). The source and drain regions in Fig. 19c have no epitaxial layer, and the source and drain regions 15 nm away from the gate are fully silicided. As a result, the contact area between the silicided region and the remaining source (or drain) region becomes 1,000 nm^2 (= 10 × 100 nm^2).

Figure 20 shows the I_D–V_{GS} curves of the 14 nm FinFETs illustrated in Fig. 19 as a parameter of the interfacial contact resistivity. The curves represented by squares, circles, and triangles correspond to the data of the source/drain structures of Fig. 19a–c, respectively. The I_D–V_{GS} curves among the three different arrangements of FinFETs are quite similar when the interfacial contact resistivity is 0 Ω cm^2. However, appreciable difference according to the contact structures can be seen at the contact resistivity of 1 × 10^{-7} Ω cm^2. The FinFET with the epitaxial layer and wide contact area shows the largest I_{ON} among the three devices, depicted by square symbols. The device with no epitaxial layer and a wide contact area displays 10 % degradation of the I_{ON} compared to that with an epitaxial layer and a wide contact area. In contrast, 71 % degradation of the I_{ON} can be observed for the FinFET with the fully silicided source/drain region. These data emphasize the importance of having a wide contact area between the metal and source/drain regions.

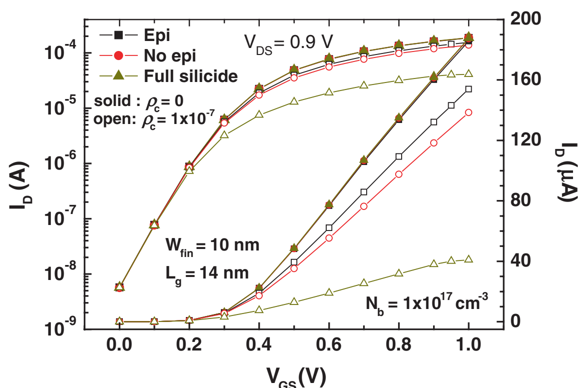


Fig. 20 I_D–V_{GS} curves of 14 nm bulk FinFETs with different contact areas and structures as a parameter of the interfacial contact resistivity (ρ_c). The curves corresponding to *square*, *circle*, and *triangle* symbols represent the data for the structures shown in Fig. 19a–c, respectively

5.2 *Junction and Gate Capacitances of Bulk FinFETs*

In addition to parasitic resistance, the parasitic capacitance is also important in determining the speed characteristic and power consumption. In this section, investigation of the source/drain to substrate and gate oxide capacitances is carried out.

Figure 21a shows a cross-sectional view cut along the channel length, wherein the contour of the hole concentration in the fin body at the V_{GS} and V_{DS} of 0 V is depicted. In this simulation, the source and drain regions are doped uniformly with a doping concentration of 5×10^{20} cm^{-3}, and the fin body has the uniform doping of 1×10^{17} cm^{-3}. It should be noted that the punch-through stopper (PTS) is

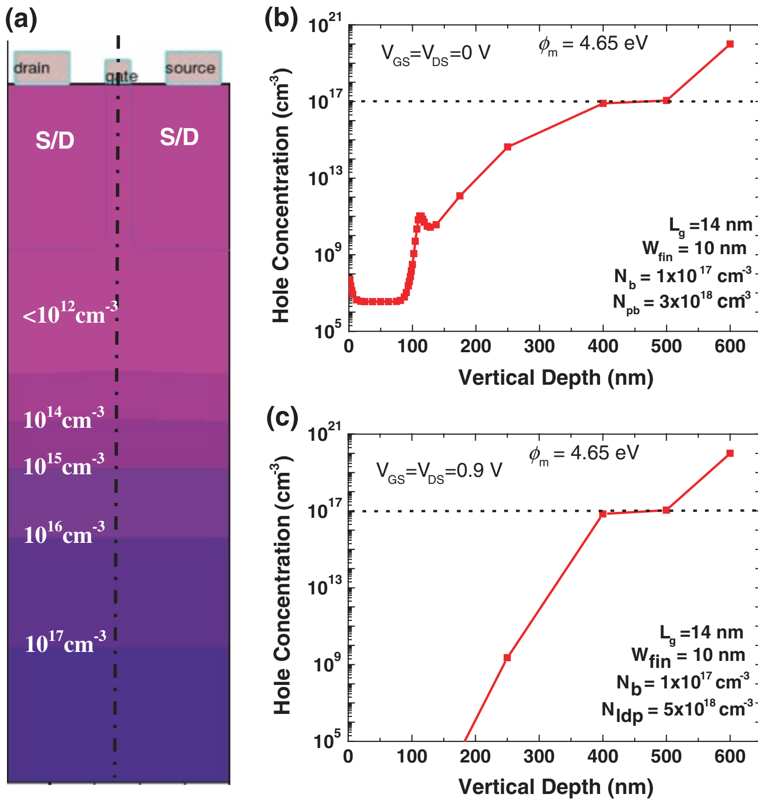


Fig. 21 Cross-sectional view of a fin body, cut along the channel length (**a**). The source/drain junction depth is 100 nm from the top of the fin body. Hole concentration profile, cut along the dashed–dotted line in figure (**a**), when both the V_{GS} and V_{DS} are 0 V (**b**) and 0.9 V (**c**), respectively. The fin body is uniformly doped with a concentration of 1×10^{17} cm^{-3}. The peak concentration of the local doping is 5×10^{18} cm^{-3}

located at 110 nm, with a peak p-type doping concentration of 5×10^{18} cm^{-3} and a standard deviation of 8 nm, for suppression of the punch-through between the drain and source. In this n-type FinFET, the fin body and PTS are doped with the same p-type impurity, resulting in increase of the impurity concentration of the fin body. In this situation, the junction capacitance between the source (or drain) and the substrate formed within the fin body is now discussed. One method to determine the depletion width at the junction is to check the hole concentration profile of the fin body. An example hole contour at equilibrium is shown in Fig. 21a. For quantitative evaluation of the hole profile, the hole concentration along the dash–dot line in Fig. 21a was extracted and prepared in Fig. 21b, c, where the profiles were obtained at equilibrium ($V_{GS} = V_{DS} = 0$ V) and given $V_{GS} = V_{DS} = 0.9$ V, respectively. In both figures (b) and (c), the region from 0 to 100 nm is the channel wherein the fin body is fully depleted. In Fig. 21b, the channel region has a hole concentration of less than ~10^8 cm^{-3} because of the work-function difference between the gate and body at equilibrium. The peak hole concentration is found near 110 nm due to the locally enhanced p-type doping for the PTS. Near 400 nm, the hole concentration becomes the doping level of the body (1×10^{17} cm^{-3}). In Fig. 21c, the fin body at a position less than 400 nm is fully depleted, and much lower hole concentration compared to that in Fig. 21b is observed, due to the given V_{GS} and V_{DS}. Note that the height of the total fin body protruding from the substrate is 400 nm in this simulation. Then, the depletion width under the source/drain regions becomes about 300 nm, which is equivalent to oxide with a thickness of 100 nm, considering the dielectric constant of Si and SiO_2. The junction capacitance of the bulk FinFET in Fig. 21 is similar to that of the SOI FinFET when 100-nm thick buried oxide is used. As mentioned in Sect. 6.2, the thickness of the buried oxide needs to be thinned down to ~20 nm to allow effective transfer of the heat generated in the channel to the substrate. From this point of view, the bulk FinFET can have smaller junction capacitance than the SOI FinFET, although SOI devices traditionally have the merit of low junction capacitance due to the thick buried oxide.

Figure 22 shows the simulated C_G–V_{GS} curves of 14 nm SOI FinFETs as a parameter of the buried oxide (BOX) thickness. As a reference, the C_G–V_{GS} curve of the 14 nm bulk FinFET is also depicted in open triangle symbols. Here, the source, drain, and substrate are grounded. The area of the gate electrode butted to the isolation oxide is 2,660 nm^2 (14×190 nm^2). The channel area covered by the gate is 2,940 nm^2 ($2 \times 14 \times 100$ nm^2 $+ 14 \times 10$ nm^2). As can be seen in the figure, decreasing BOX thickness increases the capacitance between the gate and the substrate. The gate oxide thickness is fixed at 1 nm. The SOI FinFET with a BOX thickness of 300 nm displays a quite similar C_G–V_{GS} curve to that of the bulk FinFET. As the BOX thickness decreases, the C_G increases at all gate biases. The C_G difference in the inversion region is quite small, but it becomes more appreciable in the depletion region due to a significant reduction in the gate oxide capacitance.

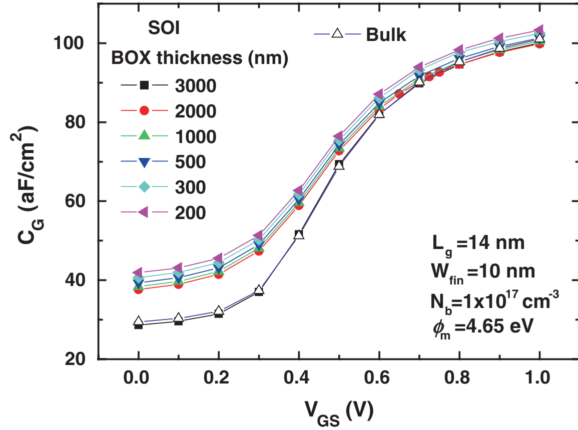


Fig. 22 Simulated gate capacitance (C_G) versus gate bias in 14 nm SOI FinFETs as a parameter of buried oxide (BOX) thickness. As a reference, the C_G–V_{GS} of the 14 nm bulk FinFET is also depicted by *open triangle* symbols

6 Current Fluctuation with Charge Trap and Temperature

In analog integrated circuit (IC) applications, drain current fluctuation with the trapping/detrapping of channel charges significantly affects the signal-to-noise ratio. The temperature in channels under operation can increase significantly. This results in changes in the key device parameters, including threshold voltage, mobility, and leakage, which causes problems in the ICs. In this section, we will study the fluctuation of drain current and temperature in the channel.

6.1 Drain Current Fluctuation with Single Charge Trap

The carriers in the channel can become trapped inside the gate oxide and detrapped into the channel randomly, leading to random fluctuation of the drain current. This kind of fluctuation adversely affects the amplification of very small signals. In this section, we examine how much drain current fluctuation occurs due to random trapping and detrapping of a carrier. Since the trap position along the channel width can cause different fluctuations in the drain current of bulk FinFETs, the effect of trap position should be determined. The position effect has partially been reported in SOI FinFETs [54]. Figure 23a shows a cross-sectional view across the fin body, wherein numbers 1, 2, 3, and 4 identify the trap position along the channel width. In this study, the channel length of the bulk FinFET is set to 22 nm because the current fluctuation in the bulk FinFETs is compared to that of the 22 nm planar MSOFET. Here, the fin body width is 8 nm and the gate oxide thickness is 1 nm. Positions 1, 2, 3, and 4 represent the trap position of the top center, top corner, and the center and bottom of the side, respectively. These traps are assumed to be located in the middle of the channel length. The trap depth (x_T), which is the distance from the interface between the gate oxide and the fin body to a position inside the gate oxide, is fixed at 0.5 nm.

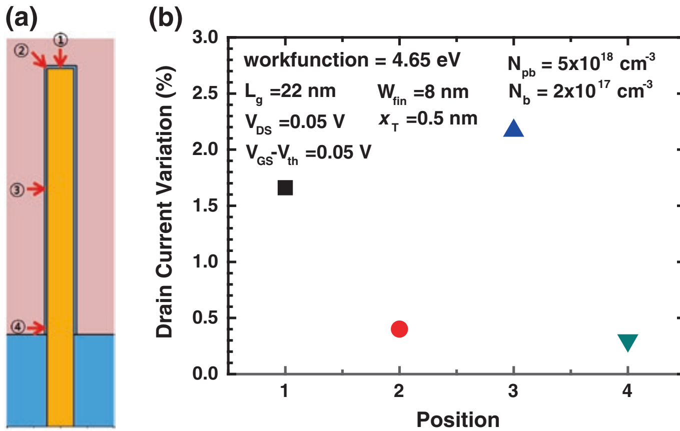


Fig. 23 Cross-sectional view of the fin body, cut across the channel length. The numbers in *circles* along the channel width of the FinFET represent the trap positions examined. The positions are assumed to be located in the *middle* of the channel between the source and the drain. Numbers *1* and *2* represent the positions of the *top center* and *top corner* traps, respectively. Numbers *3* and *4* represent the positions of the trap on the *center* and *bottom* of the side, respectively. The trap depth (x_T), which is defined as the distance from the interface between the gate oxide and the fin body for the channel, is fixed at 0.5 nm. Drain current variation according to trap position in the 22 nm bulk FinFET (**b**). The drain current fluctuation was obtained at the V_{GS}–V_{th} of 0.05 V

In Fig. 23b, the drain current variation according to trap position is shown at the V_{GS}–V_{th} of 0.05 V and the V_{DS} of 0.05 V. Trap position 3 gives the largest current fluctuation, because trapping of an electron in position 3 appreciably decreases the channel electron density near the trap, while partly decreasing the electron density on the opposite side (right side) of the channel. Accordingly, it seems that a trap located inside the gate oxide on one of the facing channels will result in larger effects to the other channel when the fin width is thinner. This phenomenon cannot be observed in planar channel MOSFETs. It should be noted that the threshold voltage of the bulk FinFET is mainly determined by the workfunction difference between the gate and the fin body, not by the channel doping ($1–3 \times 10^{17}$ cm^{-3}). Therefore, the current density along the channel width is uniform; no current crowding region is present in the channel.

Next, the dependence of drain current fluctuation on the fin body width is examined. Figure 24 shows the contours of the channel electron density on part of the cross-sectional views for three fin widths. Here, 'G' indicates the gate electrode. The fin widths, from left to right, are 6, 8, and 10 nm, respectively. It is assumed that a trap is located at the depth of 0.5 nm from the interface between the gate oxide and the fin body on the left side. Through this examination, it can be seen that the electron density in the channel on the right side can be affected more appreciably for thinner fin body width. For the given fin height of 100 nm and the channel length of 22 nm, the drain current variations for the fin body widths of 6, 8, and 10 nm are 2.29, 2.17, and 1.85 %, respectively. This

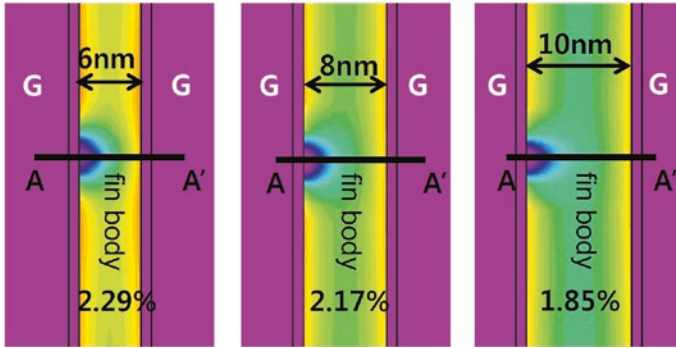


Fig. 24 Contours of electron density on the cross-sectional views cut across the fin body at the V_{GS}–V_{th} of 0.05 V and V_{DS} of 0.05 V. The fin body widths, from *left* to *right*, are 6, 8, and 10 nm, respectively. The trap depth (x_T) is 0.5 nm

demonstrates that thinner fin body gives rise to higher variation of the drain current. If the fin height is reduced, the variation will be increased.

In fabricated FinFETs, there may be a sort of percolation path which comes from random dopant fluctuation (RDF) [55] and/or work-function variation of the gate electrode [56]. Since the threshold voltage (V_{th}) of bulk FinFETs is mainly determined by the work-function difference between the gate electrode and the fin body, a percolation path cannot be generated by random dopant fluctuation. Planar channel MOSFETs can have percolation paths generated by the RDF and work-function difference. If the trap position mentioned above is aligned to the percolation path, an increase in the drain current variation can be expected. Figure 25 shows the percentage of drain current variation as a function of the distance between the percolation path and trap. Here, the trap depth (x_T) is 0.1 nm. When the distance between the percolation path and the trap is 0 nm (perfectly aligned), the variation is the largest, as expected. Note that the effective size of the

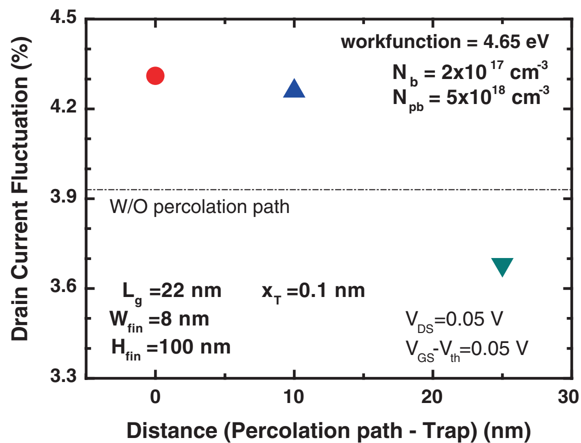


Fig. 25 Drain current fluctuation at the V_{GS}–V_{th} of 0.05 V and V_{DS} of 0.05 V as a function of the distance between the percolation path and trap. A distance of 0 nm means that the trap position corresponds to the percolation path

percolation path is 2 × 2 nm^2 in the cross-section cut across the fin body, located near the surface of the fin body while contacting the gate oxide. When the distance becomes 10 nm, the variation is slightly decreased, through still similar to that at the distance of 0 nm. The dashed–dotted line represents the current variation without a percolation path. The variations were larger with a percolation path. As the distance increases to 25 nm, the variation decreases appreciably, even dropping below that without percolation due to the higher drain current with the percolation path than without. In a planar (2-D) MOSFET with the same channel length and width, V_{th}, and bias conditions, the drain current variation is 5.7 % when the distance is 0 nm.

6.2 Device Temperature

When MOSFETs are turned on, a temperature increase in the channel should occur due to Joule heating. Conventional SOI MOSFETs have experienced heat

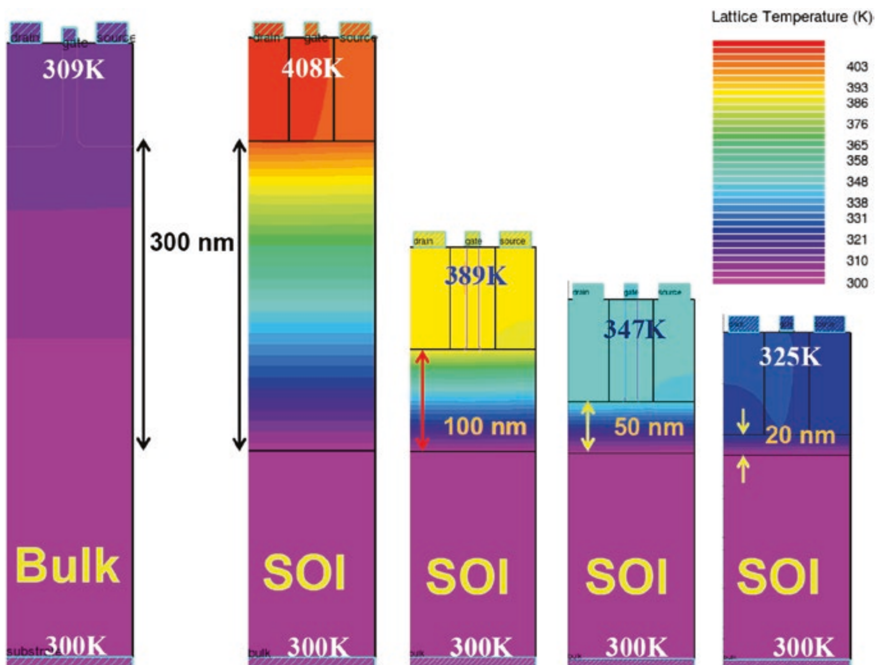


Fig. 26 Temperature contours in cross-sectional views of the fin body, cut along the channel length, when the 14 nm FinFETs are turned on. The temperature contours in SOI FinFETs are shown with BOX thickness from 300 to 20 nm, and are compared to the contour of the bulk FinFET (*leftmost*). The *arrows* in the cross-sections of SOI FinFETs indicate the thickness of the BOX

dissipation problems because of poor thermal conductivity of the buried oxide (SiO_2 0.8–1.4 W/m K) [57] compared to that of crystalline Si (~150 W/m K) [58] at room temperature. Thus, crystalline Si has over 100-fold higher thermal conductivity than amorphous SiO_2. If the temperature in the channel is not reduced, the carrier mobility will be decreased, leading to low drain current, and finally, low-speed operation. Here, we investigate the temperature generated in the channel of 14 nm bulk and SOI FinFETs. Figure 26 shows the temperature contour in cross-sectional views cut along the channel length of the fin body when the devices are turned on. In this simulation, the electrode for substrate contact is grounded and acts as a thermal contact, fixed at 300 K. Examination of the maximum temperatures displayed on top of the cross-sectional views reveals the maximum device temperature of 309 K in the 14 nm bulk FinFET (the left). In contrast, the 14 nm SOI FinFET with a BOX thickness of 300 nm (the second) displays a maximum temperature of 408 K. If the BOX thickness is reduced to 20 nm, the maximum temperature in the 14 nm SOI FinFET can be reduced to 325 K.

Thus, it is required that the BOX thickness be reduced to 20 nm and beyond to allow effective transfer of the heat generated in the channel to the substrate. In this case, increase of the parasitic capacitances between the gate, the source, and the drain to the substrate should be addressed.

Acknowledgments This work was supported by the Center for Integrated Smart Sensors, funded by the Ministry of Science, ICT & Future Planning, as the Global Frontier Project.

References

1. Lilienfeld JE (1930) Method and apparatus for controlling electric currents. U.S. Patent 1,745,175 (filed in 1926, issued in 1930)
2. Kahng D, Atalla MM (1960) Silicon-silicon dioxide field induced surface devices. Paper presented at the IRE solid-state devices research conference, Pittsburgh, PA, June 1960
3. Deal BE (1969) Method of making stable semiconductor devices. U.S. Patent no. 3,426,422, 11 Feb 1969
4. Semiconductor Industry Association (SIA) (2013) International technology roadmap for semiconductors, 2013 edn
5. Kedzierski J, Xuan P, Subramanian V, Anderson E, Bokor J, King T-J, Hu C (2000) A 20 nm gate-length ultra-thin body p-MOSFET with silicide source/drain. Superlattices Microstruct 28(5):445–452
6. Low T, Li FF, Shen C, Yeo Y-C, Hou YT, Zhu C, Chin A, Kwong DL (2004) Electron mobility in Ge and strained-Si channel ultrathin-body metal-oxide semiconductor field-effect transistors. Appl Phys Lett 85(12):2402–2404
7. Choi Y-K, Lindert N, Xuan P, Tang S, Ha D, Anderson E, King T-J, Bokor J, Hu C (2001) Sub-20 nm CMOS FinFET technologies. In: IEDM technical digest, pp 421–424
8. Park Tai-su, Yoon Euijoon, Lee Jong-Ho (2003) A 40 nm body-tied FinFET (OMEGA MOSFET) using bulk Si wafer. IEE Phys E 19(1):6–12
9. Doyle B, Boyanov B, Datta S, Doczy M, Hareland S, Jin B, Kavalieros J, Linton T, Rios R, Chau R (2003) Tri-gate fully-depleted CMOS transistors: fabrication, design, and layout. In: Technical digest of symposium on VLSI technology, pp 133–134

10. Wann CH, Noda K, Tanaka T, Yoshida M, Hu C (1996) A comparative study of advanced MOSFET concepts. IEEE Trans Electron Devices 43:1742–1753
11. Wong H-SP, Chan KK, Taur Y (1997) Self-aligned (top and bottom) double-gate MOSFET with a 25 nm thick silicon channel. In: IEDM technical digest, pp 427–430
12. Ferain I, Colinge CA, Colinge J-P (2011) Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transitors. Nature 479:310–316
13. Leobandung E, Jian G, Guo L, Chou SY (1997) Wire-channel and wrap-around-gate metal-oxide-semiconductor field-effect transistors with a significant reduction of short channel effects. J Vac Sci Technol B 15(6):2791–2794
14. Auth CP, Plummer JD (1997) Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's. IEEE Electron Device Lett 18(2):74–77
15. Kedzierski J, Nowak E, Kanarsky T, Zhang Y, Boyd D, Carruthers R, Cabral C, Amos R, Lavoie C, Roy R, Newbury J, Sullivan E, Benedict J, Saunders P, Wong K, Canaperi D, Krishnan M, Lee K-L, Rainey BA, Fried D, Cottrell P, Philip Wong H-S, Ieong M, Haensch W (2002) Metal-gate FinFET and fully-depleted SOI devices using total gate silicidation. In: IEDM technical digest, pp 247–250
16. Yang F-L, Chen H-Y, Chen F-C, Huang C-C, Chang C-Y, Chiu H-K, Lee C-C, Chen C-C, Huang H-T, Chen C-J, Tao H-J, Yeo Y-C, Liang M-S, Hu C (2002) 25 nm CMOS omega FETs. In: IEDM technical digest, pp 255–258
17. Kedzierski J, Fried DM, Nowak EJ, Kanarsky T, Rankin JH, Hanafi H, Natzle W, Boyd D, Zhang Y, Roy RA, Newbury J, Yu C, Yang Q, Saunders P, Witlets CP, Johnson A, Cole SP, Young HE, Carpenter N, Rakowski D, Rainey BA, Cotrell PE, Ieong M, Wong H-HP (2001) High-performance symmetric-gate and CMOS-compatible Vt asymmetric-gate FinFET devices. In: IEDM technical digest, pp 437–440
18. Rosner W, Landgraf E, Kretz J, Dreeskornfeld L, Schafer H, Stalele M, Schulz T, Hofmann F, Luyken RJ, Specht M, Hartwich J, Pamler W, Risch L (2004) Nanoscale FinFETs for low power applications. Solid-State Electron 48(10-11):1819–1823
19. Shang H, Chang L, Wang X, Rooks M, Zhang Y, To B, Babich K, Totir G, Sun Y, Kiewra E, Ieong M, Haensch W (2006) Investigation of FinFET devices for 32 nm technologies and beyond. In: Symposium on VLSI technology digest, pp 66–67
20. von Arnim K, Augendre E, Pacha C, Schulz T, San KT, Bauer F, Nackaerts A, Rooyackers R, Vandeweyer T, Degroote B, Collaert N, Dixit A, Singanamalla R, Xiong W, Marshall A, Cleavelin CR, Schrufer K, Jurczak M (2007) A low-power multi-gate FET CMOS technology with 13.9 ps inverter delay, large-scale integrated high performance digital circuits and SRAM. In: Symposium on VLSI technology digest, pp 106–107
21. Okano K, Izumida T, Kawasaki H, Kaneko A, Yagishita A, Kanemura T, Kondo M, Ito S, Aoki N, Miyano K, Ono T, Yahashi K, Iwade K, Kubota T, Matsushita T, Mizushima I, Inaba S, Ishimaru K, Suguro K, Eguchi K, Tsunashima Y, Ishiuchi H (2005) Process integration technology and device characteristics of CMOS FinFET on bulk silicon substrate with sub-10 nm fin width and 20 nm gate length. In: IEDM technical digest, pp 721–724
22. Wambacq P, Verbruggen B, Scheir K, Borremans J, De Heyn V, Van der Plas G, Mercha A, Parvais B, Subramanian V, Jurczak M, Decoutere S, Donnay S (2006) Analog and RF circuits in 45 nm CMOS and below: planar bulk versus FinFET. In: Proceedings of European solid-state device research conference technical digest, pp 53–56
23. Raskin J-P, Chung TM, Kilchytska V, Lederer D, Flandre D (2006) Analog/RF performance of multiple gate SOI devices: wideband simulations and characterization. IEEE Trans Electron Devices 53(5):1088–1095
24. Guo Z, Balasubramanian S, Zlatanovici R, King T-J, Nikolic B (2005) FinFET-based SRAM design. In: Proceedings of international symposium on low power electronics and design technical digest, pp 2–7

25. Kawauka H, Okano K, Kaneko A, Yagishita A, Izumida T, Kanemura T, Kasai K, Ishida T, Sasaki T, Takeyama Y, Aoki N, Ohtsuka N, Suguro K, Eguchi K, Tsunashima Y, Inaba S, Ishimaru K, Ishiuchi H (2006) Embedded bulk FinFET SRAM cell technology with planar FET peripheral circuit for hp 32 nm node and beyond. In: Symposium on VLSI technology digest, pp 86–87
26. Park T, Cho HJ, Chae JD, Han SY, Park D, Kim K, Yoon E, Lee JH (2006) Characteristics of the full CMOS SRAM cell using body-tied TG MOSFETs (bulk FinFETs). IEEE Trans Electron Devices 53(3):481–487
27. Cho IH, Park T-S, Choi SY, Lee JD, Lee J-H (2003) Body-tied doube-gate SONOS flash (omega flash) memory device built on bulk Si wafer. In: Proceedings of device research conference technical digest, pp 133–134
28. Hsu T-H, Lue H-T, Lai E-K, Hsieh J-Y, Wang Z-Y, Yang L-W, King Y-C, Yang T, Chen K-C, Hsieh K-Y, Liu R, Lu C-Y (2007) A high-speed BE-SONOS NAND flash utilizing the field-enhancement effect of FinFET. In: IEDM technical digest, pp 913–916
29. Gerardi C, Lombardo S, Cina G, Tripiciano E, Corso D, Ancarani V, Iacono G, Bongiorno C, Garozzo C, Barbe P, Costa GA, Coccorese C, Vecchio M, Rimini E, Melanotte M (2007) Highly manufacturable/low aspect ratio Si nano floating gate FinFET memories: high speed performance and improved reliability. In: Proceedings of non-volatile semiconductor memory workshop technical digest, pp 44–45
30. Hwang J-R, Lee T-L, Ma H-C, Lee T-C, Chung T-H, Chang C-Y, Liu S-D, Perng B-C, Hsu J-W, Lee M-Y, Ting C-Y, Huang C-C, Wang J-H, Shieh J-H, Yang F-L (2005) 20 nm gate bulk-FinFET SONOS flash. In: IEDM technical digest, pp 154–157
31. Cho ES, Kim T-Y, Cho BK, Lee C-H, Lee JJ, Fayrushin A, Lee C, Park D, Ryu B-I (2006) Technology breakthrough of body-tied FinFET for sub 50 nm NOR flash memory. In: Symposium on VLSI technology digest, pp 110–111
32. Lee CH, Yoon JM, Lee C, Yang HM, Kim KN, Kim TY, Kang HS, Ahn YJ, Park D, Kim K (2004) Novel body tied FinFET cell array transistor DRAM with negative word line operation for sub 60 nm technology and beyond. In: Symposium on VLSI technology digest, pp 130–131
33. Lee C, Yoon J-M, Lee C-H, Park JC, Kim TY, Kang HS, Sung SK, Cho ES, Cho HJ, Ahn YJ, Park D, Kim K, Ryu B-I (2004) Enhanced data retention of damascene-finFET DRAM with local channel implantation and < 100 > fin surface orientation engineering. In: IEDM technical digest, pp 61–64
34. Kim Y-S, Lee S-H, Shin S-H, Han S-H, Lee J-Y, Lee J-W, Han J, Yang S-C, Sung J-H, Lee E-C, Song B-Y, Chung D-J, Kim K, Lee W (2005) Local-damascene-FinFET DRAM integration with p+ doped poly-silicon gate technology for sub-60 nm device generations. In IEDM technical digest, pp 315–318
35. Lee D-H, Lee S-G, Yoo JR, Buh G-H, Yon GH, Shin D-W, Lee DK, Byun H-S, Jung IS, Park T-S, Shin YG, Choi S, Chung U-I, Moon J-T, Ryu B-I (2007) Improved cell performance for sub-50 nm DRAM with manufacturable bulk FinFET structure. In: Symposium on VLSI technology digest, pp 164–165
36. Eneman G, Hellings G, De Keersgieter A, Collaert N, Thean A (2013) Quantum-barriers and ground-plane isolation: A path for scaling bulk-FinFET technologies to the 7 nm-node and beyond. In: IEDM technical digest, Dec 2013, pp 320–323
37. Wu S-Y et al (2013) A 16 nm FinFET CMOS technology for mobile SoC and computing applications. In: IEDM technical digest, Dec 2013, pp 224–227
38. Gupta S, Moroz V, Smith L, Lu Q, Saraswat KC (2013) A group IV solution for 7 nm FinFET CMOS: Stress engineering using Si, Ge and Sn. In: IEDM technical digest, Dec 2013, pp 644–647
39. Guo W, Moroz V, Van der Plas G, Choi M, Redolfi A, Smith L, Eneman G, Van Huylenbroeck S, Su PD, Ivankovic A, De Wachter B, Debusschere I, Croes K, De Wolf I, Mercha A, Beyer G, Swinnen B, Beyne E (2013) Copper through silicon via induced deep out zone for 10 nm node bulk FinFET CMOS technology. In: IEDM technical digest, Dec 2013, pp 340–343

40. Kang CY, Sohn C, Baek R-H, Hobbs C, Kirsch P, Jammy R (2013) Effects of layout and process parameters on device/circuit performance and variability for 10 nm node FinFET technology. In: Symposium on VLSI technology digest, 2013, pp. 90–91
41. Maeda S, Ko Y, Jeong J, Fukutome H, Kim M, Choi J, Shin D, Oh Y, Lim W, Lee K (2013) 3 dimensional scaling extensibility on epitaxial source drain strain technology toward Fin FET and beyond. In: Symposium on VLSI technology digest, pp 88–89
42. Mitard J, Witters L, Loo R, Lee SH, Sun JW, Franco J, Ragnarsson L-A, Brand A, Lu X, Yoshida N, Eneman G, Brunco DP, Vorderwestner M, Storck P, Milenin AP, Hikavyy A, Waldron N, Favia P, Vanhaeren D, Vanderheyden A, Olivier R, Mertens H, Arimura H, Sonja S, Vrancken C, Bender H, Eyben P, Barla K, Lee S-G, Horiguchi N, Collaert N, Thean AV-Y (2014) 15 nm-W_{FIN} high-performance low-defectivity strained-germanium pFinFETs with low temperature STI-last process. In: Symposium on VLSI technology digest, pp 138–139
43. Waldron N, Merckling C, Guo W, Ong P, Teugels L, Ansar S, Tsvetanova D, Sebaai F, van Dorp DH, Milenin A, Lin D, Nyns L, Mitard J, Pourghaderi A, Douhard B, Richard O, Bender H, Boccardi G, Caymax M, Heyns M, Vandervorst W, Barla K, Collaert N, Thean AV-Y (2014) An INGaAs/InP auantum well FinFet using the replacement Fin Process integrated in an RMG flow on 300 mm Si substrates. In: Symposium on VLSI technology digest, pp 32–33
44. http://newsroom.intel.com/docs/DOC-2032
45. Hashemi P, Kobayashi M, Majumdar A, Yang LA, Baraskar A, Balakrishnan K, Kim W, Chan K, Engelmann SU, Ott JA, Bedell SW, Murray CE, Liang S, Dennard RH, Sleight JW, Leobandung E, Park D-G (2013) High-performance $Si_{1-x}Ge_x$ channel on insulator trigate PFETs featuring an implant-free process and aggressively-scaled Fin and gate dimensions. In: Symposium on VLSI technology digest, pp 18–19
46. Seo K-I et al (2014) A 10 nm platform technology for low power and high performance application featuring FINFET devices with multi workfunction gate stack on bulk and SOI. In: Symposium on VLSI technology digest, pp 14–15
47. Hashemi P, Balakrishnan K, Majumdar A, Khakifiroooz A, Kim W, Baraskar A, Yang LA, Chan K, Engelmann SU, Ott JA, Antoniadis DA, Leobandung E, Park D-G (2014) Strained $Si_{1-x}Ge_x$-on-Insulator PMOS FinFETs with excellent sub-threshold leakage, extremely high short-channel performance and source injection velocity for 10 nm node and beyond. In: Symposium on VLSI technology digest, pp 18–19
48. Matsukawa T, Fukuka K, Liu YX, Endo K, Tsukada J, Yamauchi H, Ishikawa Y, O'uchi S, Mizubayashi W, Migita S, Morita Y, Ota H, Masahara M (2014) Lowest variability SOI FinFETs having multiple V_t by back-biasing. In: Symposium on VLSI technology digest, pp 142–143
49. Colinge J-P (1991) Silicon-on-insulator technology: materials to VLSI. Kluwer Academic Publishers, Dordrecht
50. Lee J-H, Park T-S, Yoon E, Park JJ (2003) Simulation study of a new body-tied FinFETs (Omega MSOFETs) using bulk Si wafers. In: Proceedings of Si nanoelectronics technical digest, pp 102–103
51. http://www.chipworks.com/ko/technical-competitive-analysis/resources/blog/intels-22-nm-tri-gate-transistors-exposed/
52. http://www.goldstandardsimulations.com/news/blog_search/simulation-analysis-of-the-intel-22nm-finfet/
53. Metz MV, Datta S, Docay ML, Kavalieros JT, Brask JK, Chau RS (2008) Uniform silicide metal on epitaxially grown source and drain regions of three-dimensional transistors, Intel, US 7,425,500
54. Fan M-L, Hu VP-H, Chen Y-N, Su P, Chuang CT (2012) Analysis of single-trap-induced random telegraph noise on FinFET devices, 6T SRAM cell, and logic circuit. IEEE Trans Electron Devices 59(8):2227–2234
55. Asenov A (1998) Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D "atomistic" simulation study. IEEE Trans Electron Devices 45(12):2505–2513

56. Li Y, Hwang C-H, Li T-Y, Han M-H (2010) Process-variation effect, metal-gate work-function fluctuation, and random-dopant fluctuation in emerging CMOS technologies. IEEE Trans Electron Devices 57(2):437–447
57. Burzo MG, Komarov PL, Raad PE (2003) Thermal transport properties of gold-covered thin-film silicon dioxide. IEEE Trans Compon Packag Technol 26(1):80–88
58. Glassbrenner CJ, Slack GA (1964) Thermal conductivity of silicon and germanium from 3 oK to the melting point. Phys Rev 134(4A):A1058–A1069

Micro and Nanoelectromechanical Contact Switches for Logic, Memory, and Power Applications

Yong-Ha Song and Jun-Bo Yoon

Abstract This chapter describes a wide range of technologies for contact-based microelectromechanical system (MEMS)/nanoelectromechanical system (NEMS) switches and their applications to logic, memory, and power devices. The mechanical switches based on the MEMS/NEMS technology are considered as ideal switching components owing to their small, low power consumed, and scalable features, which is very suitable for ultralow power electronics. In this chapter, after providing a fundamental background ranging from the needs and operation principle of the MEMS/NEMS switches to the contact physics, we introduce individual features, excellent advantages, and critical issues of the MEMS/NEMS contact switches, respectively. Also, potentials of these devices in the next-generation logic gate, memory element, and power switching devices and main research flow including previous progresses and future research direction are presented and discussed.

Keywords MEMS/NEMS contact switch · Microswitch · Nanoswitch · Contact · Logic · Memory · Power

Y.-H. Song · J.-B. Yoon (✉)
Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea
e-mail: jbyoon@kaist.ac.kr

Y.-H. Song
e-mail: yhsong1@illinois.edu

Y.-H. Song
Department of Electrical and Computer Engineering, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA

© Springer Science+Business Media Dordrecht 2016
C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_3

1 Background

1.1 Why Miniaturized Mechanical Switches?

Less than a year after the telephone was invented by Alexander Graham Bell in 1876, automatic switching systems were required to selectively and automatically connect telephone lines. At first, a telephone operator worked to connect the telephone line manually. However, this procedure quickly became complicated and impractical. To address this problem, a magnetically operated automatic switch was developed, which utilized magnetic field induced by flowing current through coil to make mechanical contact and connect electrical signal. Since then, automatic switching systems with step-by-step (from 1910s) and crossbar (from 1930s) methods were actively researched [1]. This indicates that the mechanical switch is the origin of the switching device for automatic systems.

In the meantime, silicon-based switching devices have intensively developed since the transistor was invented. An aggressive scaling of the complementary metal-oxide-silicon (CMOS) devices has provided a significant stimulus for advances in the semiconductor industry to increase both the density and speed over the past four decades [2]. However, as semiconductor-based switching technologies advanced, a small, convenient, and high-performance mechanical switching device also attracted considerable attention because the silicon-based switching device had started to suffer from physical limitations such as short channel effects [3–5]. Also, the modern magnetic switches (relays) have their limitation with respect to size, power consumption, and scalability. Thus, various types of switching devices ranging from mercury switch, wafer switch, and reed switch to toggle switch have been developed to be utilized in many applications. However, these switches are considered impossible to replace the silicon-based switching devices so far. Is there any other solution for breaking the limitations in silicon-based switching devices?

In 1979, a micro-sized mechanical switch with an electrostatically actuated cantilever was first presented by Petersen [6]. Since then, microelectromechanical systems (MEMS)/nanoelectromechanical system (NEMS) switches have been considered as one of the ideal switching devices, because they potentially have both advantages of the mechanical switch and the silicon-based switch (as explained in detail in Sect. 2.2). Their main features are as follows. In general, the micro/nanomechanical switch is operated through a three-terminal configuration (a gate, a source, and a drain) similar to the conventional metal–oxide-semiconductor field-effect-transistor (MOSFET), as shown in Fig. 1a. Also, it is electrostatically actuated by moving a suspended (conducting) electrode in contrast to the silicon-based switch, which is operated by modulating the conductivity of a semiconducting channel. At the on/off-state, two contact parts (here, the source beam and the drain electrode) of the miniaturized mechanical switch make contact with each other and are isolated mechanically. Owing to the mechanical structure and

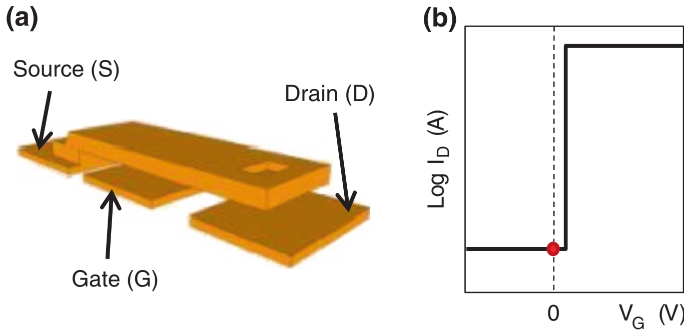


Fig. 1 **a** Representative MEMS/NEMS switch configuration **b** I–V characteristic of the MEMS/NEMS switch

operation mechanism, the MEMS/NEMS switches can possess excellent features of low power consumption, high on-current, and abrupt switching characteristics (Fig. 1b). That is, the MEMS/NEMS switch is a small, low power consumed, and scalable mechanical switching device. Thus, a considerable body of research on MEMS/NEMS switches has been reported for numerous applications such as telecommunications, logic operation, memory devices, and power switching.

1.2 MEMS/NEMS Contact Switch Principle

There are several mechanisms to actuate the MEMS/NEMS switching device such as electrostatic [7], thermal [8], magnetic [9], piezoelectric [10], etc. (various actuation mechanisms are utilized especially for MEMS switches and most NEMS switches are operated by electrostatic actuation). Among them, the electrostatic actuation method has been favorably employed, since it has a lot of advantages, including extremely low power consumption, relatively short switching time, and simplicity to scalability [11, 12]. To explain the electrostatic actuation, a simplified parallel plate design is used here.

As shown in Fig. 2, when a bias voltage is applied between two plates (typically, one plate is fixed to the bottom and the other plate is suspended and movable), positive and negative electric charges are induced at each side due to the potential difference, thus resulting in an electric field between them. These electric charges generate the electrostatic force, so the upper plate is gradually going downward, maintaining equilibrium condition (between mechanical restoring force originated from the stiffness of the spring and the generated electrostatic force) as shown in Fig. 2a. When the upper plate moves down further than a certain position by increasing the bias voltage (Fig. 2b), the electrostatic force is always higher than the restoring force resulting in sudden collapse of the upper

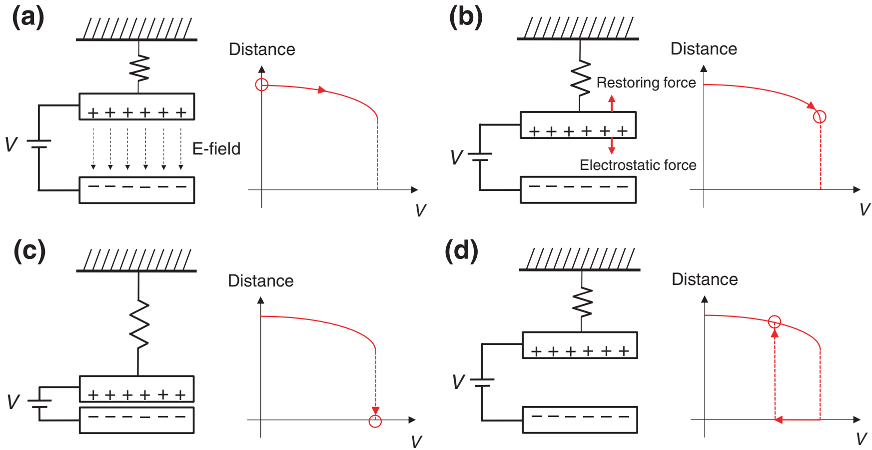


Fig. 2 Electrostatic actuation mechanism

plate (Fig. 2c). This is referred to as "pull-in" phenomenon. When the bias voltage decreases until the point where the restoring force is higher than the electrostatic force, the upper plate is detached and returns to the equilibrium condition ("pull-out" phenomenon as shown in Fig. 2d).

Nathanson et al. performed modeling of the electrostatically actuated parallel plate-type actuator for the first time [13]. When the upper plate of area A is actuated downward to move x by the external voltage V, the corresponding electrostatic and mechanical restoring forces are given by

$$F_{\text{electrostatic}} = \frac{\varepsilon A V^2}{2(h_o - x)^2} \tag{1}$$

$$F_{\text{restoring}} = kx \tag{2}$$

where ε is the permittivity of the surrounding medium, normally air, k is the stiffness of the spring, and h_o is the initial air gap, respectively. The electrostatic force and the restoring force increase in a quadratic and linear manner, respectively, as the displacement of the upper plate x increases. Figure 3 indicates a diagram showing the relationship between the electrostatic force and the spring (mechanical restoring) force. When the applied voltage is lower than the pull-in voltage, the electrostatic force makes equilibrium with the restoring force at a position x; whereas when the applied voltage is higher than the pull-in voltage, the electrostatic force is always larger than the restoring force at all points of x and the upper

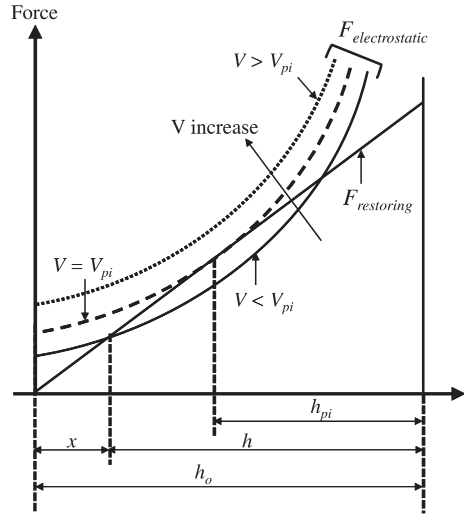


Fig. 3 Electrostatic and spring (mechanical restoring) forces acting on the movable plate [132]

plate therefore is pulled down to the bottom plate. Referring to Fig. 3, at $x = h_o - h_{\rm pi}$, the relationship between the electrostatic and restoring forces is as follows at the applied voltage, which is equal to the pull-in voltage,

$$F_{\rm electrostatic} = F_{\rm restoring} \tag{3}$$

$$\frac{{\rm d}F_{\rm electrostatic}}{{\rm d}x} = \frac{{\rm d}F_{\rm restoring}}{{\rm d}x}. \tag{4}$$

Using (1)–(4), the pull-in voltage and pull-in point can be obtained as follows:

$$h_{pi} = \frac{2}{3}h_0 \tag{5}$$

$$V_{pi} = \sqrt{\frac{8\,kh_0^3}{27\varepsilon A}}. \tag{6}$$

Figure 4 shows a typical structure of the electrostatically actuated MEMS/NEMS switch (cantilever type). At the off-state, the cantilever beam (i.e., source electrode) is separated from the metallic drain electrode so that no current can flow. At the on-state where the gate voltage is greater than the pull-in voltage (the source electrode is electrically grounded), the cantilever beam comes down and makes contact with the drain electrode, providing a conductive path for current to flow.

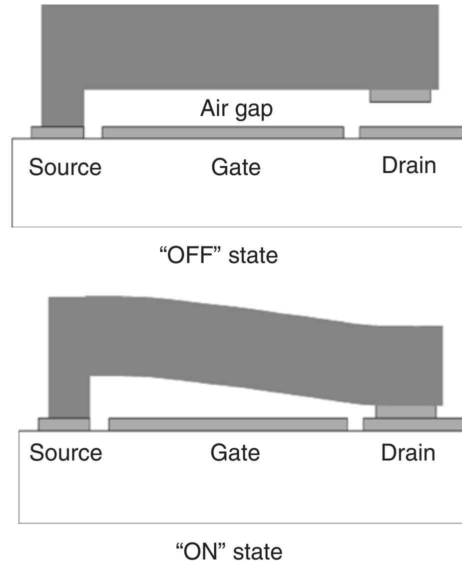


Fig. 4 Schematic diagrams of the electrostatically actuated MEMS/MEMS switch at the off and on states

1.3 Contact Physics

History

To characterize the MEMS/NEMS contact switches, it is of fundamental importance that contact physics should be understood, because many-core parameters including resistance, reliability, power handling, and voltage hysteresis are strongly determined by the contact property. However, it has been known that the prediction of the contact phenomena such as real contact area, contact deformation depending on the contact pressure, contact resistance, and generated heat at the micro/nanoscale regime are complicated and not completely discovered yet, because the contact surfaces consist of a tremendous number of nanometer scale contact asperities in the contact surfaces, as conceptually presented in Fig. 5. Also, as two metal surfaces make contact with each other, a very limited fraction of the wide macroscopic contact area ("apparent" contact area) establishes physical contact [14, 15]. This real ohmic contact area is referred to as an "effective" contact area.

Understanding of two metal contact surfaces when making a contact was first sought by Heinrich Hertz in 1882, by estimating contact deformation [16]. Although elastic contact (where the local stresses applied at the real contact asperities are lower than the plastic yield point) was assumed and surface roughness was ignored, the theory has been widely used in tribological contact theory due to its simplicity and reliability. In the 1960s, Holm developed the concept of a plastic deformation model (where the local stresses at the contact spots are sufficiently higher than the plastic yield point), yet the overall stresses at the entire contact

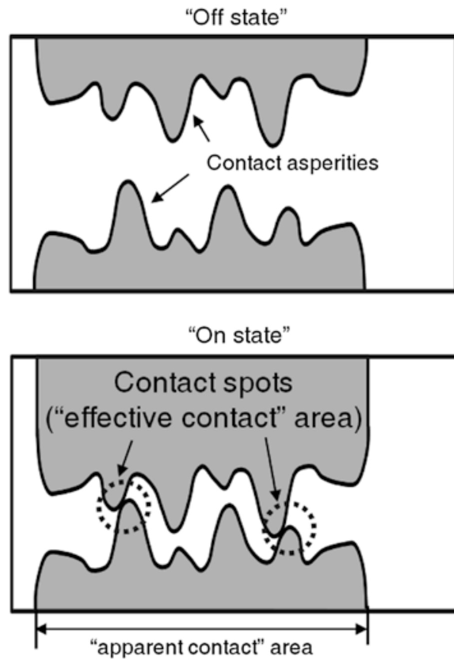


Fig. 5 Illustration of real contact surfaces with nm scale contact asperities

area were in the elastic regime [17]. Here, surface contamination was not ignored, and he insisted that high contact forces could break the contamination layer. The theory of Holm has been widely accepted and recognized by the research community of electric contact. In 1966, Greenwood and Williamson introduced a new asperity contact model, which is more accurate (closely related to the real contact surface) than Holm's [18]. In this contact theory, they substantiated how contact deformation depends on the surface topography and established the criterion for distinguishing the elastic contact deformation from the plastic contact deformation. They insisted that elastic deformation is uncommon from the viewpoint of engineering practice, and the contact asperities are more frequently plastically deformed. This work has been the foundation for a model by Chang, Etison, and Bogy (CEB). CEB suggested a multi-asperity surface for elastic-plastic deformation [19]. Since then, many research groups have assumed that the multiple contact asperities could be converted into a single effective contact area.

All the theories and models provide a good estimation of the real contact area. However, the MEMS/NEMS switching devices operate with very low contact force, thus always resulting in some variations from the theory due to the contact surface topography. Therefore, for an accurate approach, direct measurement of the contact topography together with simulation of the interfacial contact can also be a good candidate for estimating the real contact area [14].

Contact Mechanics

As mentioned before, the contact deformation of the MEMS/NEMS contact switches can be described as either elastic, plastic, or elastic-plastic, depending on whether the local stresses (pressures) are higher than the plastic yield point. In each case, the corresponding effective contact radius (important parameter that determines the contact resistance of the mechanical switch) is somewhat different. First, in the case of elastic deformation, the effective contact radius, r, can be derived if the contact spot is assumed to have a circular shape,

$$r = \sqrt[3]{\frac{3F_cR}{4E'}} \tag{7}$$

$$\frac{1}{E'} = \frac{1 - v_1^2}{E_1} + \frac{1 - v_2^2}{E_2} \tag{8}$$

where F_c is the contact force, R is the radius of curvature of the asperity, v_1 and v_2 are the Poisson ratios of the respective contact materials, and E_1 and E_2 are the elastic modulus of the respective contact materials.

Under the plastic deformation that commonly occurs when the contact force applied to the contact spot is higher than 200 μN [11], the effective contact radius becomes simpler and can be expressed as

$$r = \sqrt{\frac{F_c}{H\pi'}} \tag{9}$$

where H is the hardness of the contact material. Owing to its simplicity, Eq. (9) has been frequently used to estimate the contact resistance in MEMS community.

Since elastic-plastic deformation represents the transition region between the elastic and plastic deformations, it is more complicated in comparison to both elastic and plastic contact deformations. The elastic-plastic deformation is calculated under the assumption that some parts of the contact area are plastically deformed and surrounded by elastically deformed area [20]. The corresponding effective contact radius is

$$r = \sqrt{R\alpha\left(2 - \frac{\alpha_c}{\alpha}\right)} \tag{10}$$

where α is the contact asperity's vertical deformation and α_c is the contact asperity's vertical deformation when plastic deformation occurs.

Contact Resistance

On the basis of the effective contact radius calculated above, contact resistance of MEMS/NEMS switches can be calculated. It is also divided into two categories, depending on the method of electron transport through the contact spot (by means

of comparison between the size of real contact area and mean free path of electrons): diffusive and ballistic. In general, total contact resistance R_{CR} is the sum of a constriction resistance term, R_C, and a film resistance term, R_F [21]. The film resistance is frequently ignored because its value is sufficiently small in comparison to the constriction resistance. Thus, from now on, the contact resistance means the constriction resistance in this chapter. If the effective contact radius r is small compared to the electron mean free path length l_e of the contact material (ballistic), the resistance of the contact spot is dominated by the Sharvin mechanism [22], and it can be expressed as

$$R_S = \frac{4\rho l_e}{3\pi r^2} \tag{11}$$

where ρ is the resistivity of the contact material. On the other hand, if an effective contact radius is sufficiently larger than the electron mean free path length l_e of the contact material (diffusive), it is dominated by the Maxwell mechanism, and it can be expressed as

$$R_M = \frac{\rho}{2r}. \tag{12}$$

The overall (complete) form of the contact resistance derived by Wexler (considering both Sharvin and Maxwell mechanisms) is described as

$$R_C = f\left(\frac{l_e}{r}\right) R_M + R_S = \frac{1 + 0.83\left({l_e}/{r}\right)}{1 + 0.83\left({l_e}/{r}\right)} \frac{\rho}{2r} + \frac{4\rho l_e}{3\pi r^2} \tag{13}$$

where $f(l_e/r)$ is an interpolation function [23].

Contact Heating

To design MEMS/NEMS contact switches effectively, heat generated at the contact spots should also be considered, because it strongly affects switching performance such as current capability and contact reliability. In general, the current passing through the contact spots creates heat and the heat is dissipated by heat conduction through the bodies in the contact spots [21]. This heat affects (increases the temperature of) the contact asperities, and the contact temperature is related with the voltage drop across the contact, which is

$$V = 2\sqrt{2\int_{T_1}^{T_m} \lambda \rho \mathrm{d}T} \tag{14}$$

where V is contact voltage, λ is thermal conductivity, ρ is electrical resistance, and T_m and T_1 are the contact temperature and ambient temperature, respectively. Assuming that the electrical resistivity and thermal conductance vary little with the temperature, Eq. (14) yields the voltage-temperature relation [21], as follows:

$$T_m - T_1 = \frac{V^2}{8\lambda\rho} \tag{15}$$

Here, the electrical resistivity and thermal conductance can be replaced through the Wiedemann-Franz law [21, 24] such that,

$$\lambda\rho = LT \tag{16}$$

where L is the Lorenz constant (2.45×10^{-8} V^2 K^{-2}) and T is the absolute temperature. Assuming that the dimension of the real contact spots is larger than the electron mean free path in the contact material, the voltage-temperature relation in (15) can be expressed as

$$V^2 = 4L(T_m^2 - T_1^2) \tag{17}$$

That is, the contact voltage generates heat which causes the contact material to soften, melt, and even boil.

2 Microelectromechanical System (MEMS) Switches

This section details the general features, categories, strengths, and issues of the MEMS contact switch. In particular, the MEMS switch has been the most widely studied switch in radio frequency (RF) applications so far, because it has powerful strengths including high isolation, low power consumption, high linearity, and usefulness in a large bandwidth from DC up to tens of gigahertz [11, 25], and thereby the RF MEMS switches are well established and summarized by many other researchers [11, 26]. Unlike the previously published papers and books which concentrated on RF applications, here, we focus on the general and unique features of the MEMS contact switches used in various applications including not only RF but also logic, memory, and power applications.

2.1 MEMS Switch Types

As summarized in Table 1, MEMS switches can be classified into various categories, depending on actuation mechanism, movement of the suspended structure, type of the moving part, and number of terminals (electrodes). In terms of the actuation mechanism, electrostatic actuation has been the most preferred one because it features very low power consumption and fast response, thus being used in various micro-actuators such as comb-drive actuator and micro-motor [27, 28]. Thermal and magnetic actuations have been partially utilized for the micro-actuators that require large displacement and high force [8, 9]. Also, vertical movement for making a contact is commonly used since it leads to easy fabrication, high contact force, and small footprint [11]. In terms of the switch's moving part,

Table 1 MEMS switch categories

Classification	Category	Remark
Actuation mechanism	Electrostatic	Fast, low power consumed
	Thermal	Low voltage, slow
	Magnetic	High force, suitable for power switching
	Piezoelectric	Low voltage, complicated
	Combined actuations	High performance, complicated
Movement	Vertical	Widely used
	Lateral	Large size, one mask fabrication
	Torsional (tilted)	Low voltage, complicated
Moving structure	Membrane	Stable, low voltage, large area
	Fixed-fixed	High restoring force
	Cantilever	Widely used, simple
Number of terminals (electrodes)	3-terminal	Widely used, simple
	4-terminal	Not depend on the source voltage

the cantilever-type has been so far the most widely used structure owing to its simplicity, and at the same time, other structures such as membrane-type have also been continuously researched for enhancing performance ranging from contact resistance and power handling capability to reliable operation. Similar to the conventional MOSFET, a 3-terminal configuration is also common, and a 4-terminal type which consists of a gate, source, drain, and body (or counter-gate) electrode is selectively used for logic [29] and RF [30] applications for additional functions.

2.2 MEMS Switch Advantages

As mentioned before, MEMS switches have been actively researched since Peterson demonstrated the cantilever-type MEMS switch, owing to their strong advantages and ideal characteristics, as follows:

Ideally zero leakage current: Silicon-based switching devices such as MOSFET inevitably possess leakage currents owing to the finite subthreshold slope and gate-to-channel leakages even when the devices are at the off-state [3–5]. Whereas the mechanical structure of MEMS switches does not suffer from any leakage current, because separation of the two contact parts (source and drain electrodes) is mechanically ensured; thereby static power consumption can be completely zero.

Freedom in selecting substrate: MEMS switches are generally fabricated through a surface micromachining technique (it makes microstructures by deposition and etching of various structural layers on top of the substrate, in comparison to bulk micromachining where a substrate is selectively etched to fabricate microstructures) and therefore it can be manufactured regardless of the substrate materials. Also, typical MEMS switches can be stably fabricated under low

temperature process (<250 °C) [31, 32]. Thus, various kinds of substrates such as glass, ceramic, and even polymer substrates can be exploited. This feature accordingly results in low cost and good fabrication compatibility.

Insensitive to radiation and temperature changes: MEMS switches commonly consist of metal materials, and they do not use any electron and hole pairs residing in semiconductor materials. Therefore, their switching performance such as switching response and actuation voltage is seldom affected by radiation and temperature changes, compared to silicon-based electrical switches.

Low on-resistance: Current in the MEMS switch commonly flows through highly conductive metal materials, thus it potentially affords lower on-resistance. Although most of the on-resistance of the MEMS switch stems from the contact resistance, very low-level of contact resistance MEMS switches have recently been demonstrated [33]. This low on-resistance feature is suitable for telecommunications and power switching applications, from the viewpoint of power loss and consumption.

Perfectly abrupt switching behavior: The MEMS switch operates through mechanical movement (engaging and disengaging contacts mechanically), and thereby it can present sudden current increase/decrease at both on and off switching transitions, respectively. As mentioned, they have extremely low leakage current below the noise signal level of the measurement system at the off-state and low on-resistance (accordingly high on-current) at the on-state at the same time, thus leading to an outstanding on/off current ratio.

2.3 Actuation Voltage

Despite these strengths, there are also several issues in MEMS switches. One of the big issues of the MEMS switch is its high actuation (i.e., pull-in) voltage originated from the electrostatic actuation method that is most commonly employed. Generally, the actuation voltages of electrostatically actuated MEMS switches have been of the order of 40–120 V [34]. Accordingly, MEMS switches have been considered difficult to be utilized in many commercialized applications including telecommunication and portable devices. High actuation voltage of the electrostatic actuation is due to the fact that the generated electrostatic force is relatively small to operate the suspended (movable) structures, compared to the other forces achievable by thermal and magnetic actuations. In the beginning, the simplest scheme to solve this high actuation voltage issue was using a charge pumping circuit which can amplify the low applied voltage. Since Motorola adapted this configuration to the MEMS switch [35], this method has been widely used, especially in RF applications.

Also, a considerable body of research on structural approaches to lower the actuation voltage has been reported, as shown in Fig. 6. In general, to reduce the actuation voltage, low stiffness (low spring constant) of the spring constant k, small air gap h_o, large overlap area between two parallel plates (electrodes) A, are preferred, as presented in Eq. (6). Among them, many researches have been

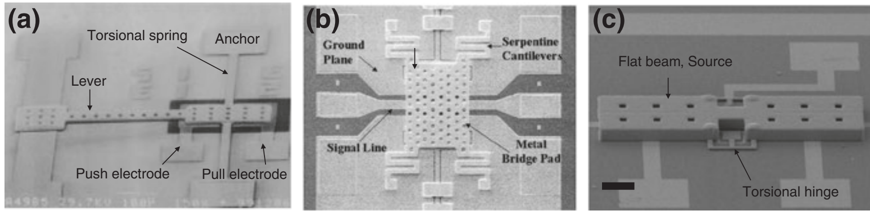


Fig. 6 SEM images of the MEMS switches for low operation voltage. **a** Torsion spring and leverage MEMS switch [30]. **b** Low stiffness spring MEMS switch [38] (Reproduced from Ref. [38] with permission from IEEE) and. **c** See-saw type MEMS switch with the stiction-recovery actuation [39]. (Reproduced from Ref. [30] with permission from IEEE)

conducted for reducing the spring stiffness. In 2000, a torsional spring structure was first presented to realize low actuation voltage, as shown in Fig. 6a [30]. In general, the torsional spring which uses a tilting motion has much smaller stiffness than those of conventional springs which use a bending motion such as the cantilever and fixed-fixed type springs [36]. Also, by using a leverage effect (effective torque increases when a point of applied electrostatic force is apart from the central point), the actuation voltage could be decreased greatly. By doing so, the minimum actuation voltage of 5 V with the two $300 \times 20 \times 1.4\ \mu m^3$ torsional springs was successfully demonstrated. In 2002, University of Illinois presented a MEMS switch with 15 V in actuation voltage through low stiffness spring structure and wide overlap area (Fig. 6b) [37, 38]. However, low spring stiffness is directly related with the low restoring force, which is difficult to get over the contact adhesion force and therefore may lead to permanent contact stiction. Recently, a see-saw type MEMS switch with stiction-recovery actuation was suggested [39]. Kim et al. demonstrated that contact stiction could be overcome by applying a sufficient voltage to the counter electrode like a see-saw (push-pull) actuation, although the MEMS switch was failed by the in-use stiction occurred due to low restoring force (Fig. 6c). Thus, very low spring stiffness and small air gap of 0.45 μm were designed, and thereby very low actuation voltage of 1.7 V was accomplished. This is the lowest pull-in voltage among electrostatic MEMS switches [40], and even lower than the actuation voltage (2.5 V) of a MEMS switch operated with piezoelectric actuation [10].

Apart from the many efforts for low operation voltage through the structural approach, a simple pre-charged electrode method was introduced in 2007 [41]. As shown in Fig. 7, Yang et al. placed an additional pre-charged plate between the top and bottom plates and electrically isolated the pre-charged plate from the bottom plate. Then, by electrically charging the pre-charged electrode with some amount of voltage, V_1 (after that, the pre-charged electrode is electrically disconnected), the pull-in voltage could be significantly reduced as much as the pre-charged voltage V_1, similar to the operation mechanism of the FLASH memory. One fundamental merit of the pre-charged method is that there is no need for reducing both the restoring and contact forces of the MEMS switch to decrease the actuation

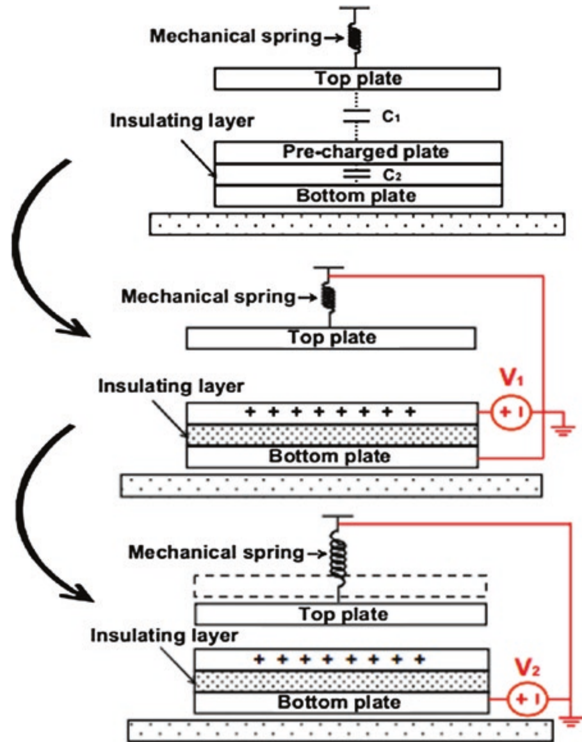


Fig. 7 Schematic diagram of the pre-charged electrode method [41]

voltage. Retention time of the pre-charges in both air and vacuum ambient was relatively short [42], thus a special charge-trapped layer such as electret [43, 44] is more appropriate and can increase the retention time up to the device lifetime.

Also, Ko et al. introduced an approach to replace the gap in surrounding medium between the two actuated plates (electrodes) for increasing permittivity ε [45]. That is, the actuation voltage was decreased as much as the ratio of the square root of permittivity by means of filling the gap with an insulating liquid (mineral oil) instead of an air—refer to Eq. (6). The actuation voltage measured in mineral oil was decreased from 58.2 V (measured in air) to 39.7 V, corresponding to a 31.8 % decrease.

As presented above, low actuation voltage of MEMS contact switches has been widely researched and developed, by means of various structural designs and charging pumping technique. However, low actuation voltage may generally lead to low contact force, thus accordingly resulting in unstable contact property and low contact reliability. Thus, it is noteworthy that the designer should also consider the proper contact force (in the case of Au-to-Au contact, over 80 μN of contact force is generally required for stable and reliable contact property [46]), when designing a low actuation voltage MEMS switch.

2.4 Reliability/Lifetime

Since the MEMS switch has been considered as a powerful alternative for silicon-based switching devices and bulky electromagnetic relays (EMRs) owing to its fancy and attractive performance, the most critical and major concern of MEMS switches has been contact reliability [11, 33]. In order to overcome this obstacle, numerous studies have focused on enhancing the endurance of MEMS switches. Among them, as a matter of fact, several superior research works were accomplished under low and medium power conditions. For example, Radant MEMS switch has reported an excellent lifetime, which exceeds 100 billion cycles [47, 48] and OMRON has demonstrated high reliability over 100 million cycles under hot switching conditions [49]. Nevertheless, to be utilized in long-term applications such as RF radar and other systems, the required lifetime should be up to several hundred billion cycles [26, 48] (in the case of logic applications, lifetime over 10^{14} cycles is necessary [50]). That is, long-term reliability of the MEMS switch is still challenging for product commercialization, and it is more severe especially at high power and hot switching conditions [51].

To address the reliability issue, thorough failure analysis is fundamental and therefore should take precedence. In general, contact failures are classified into two categories: stiction and contact resistance degradation during actuation. Stiction indicates a sticking effect caused by the surface adhesion force between two movable contact parts. This failure phenomenon occurs when the adhesion forces between the two contact surfaces in the MEMS switch is higher than the restoring force of the mechanical spring. Second, contact resistance degradation means that the contact resistance of the MEMS switch gradually increases with actuation cycles to a certain level. This may come from the thin organic film which contains carbon and hydrogen [52] and contact wear [46]. As a matter of fact, contact failure analysis is still under investigation because degradation mechanisms of the MEMS switch are complicated and cannot be explained by one phenomenon but many effects are involved simultaneously. The detailed contact failure modes are as follows.

Failure Mechanism

Wear: Wear is an erosion of the solid surface of the contact material by the repetitive action of the contacting solid surface. Although wear can be utilized in a constructive manner such as polishing and sharpening in some mechanical operations, it is generally considered an undesirable phenomenon in the MEMS contact switch. Generally, surface fatigue, abrasion, and corrosion are thought to be the roots of the wear phenomenon [53]. Also, wear includes mechanical pitting and hardening, which is one of the main causes of device failure through high contact resistance [46]. Especially, current-induced wear has been considered as the major failure mechanism in the high current level. This is because high current density at the contacts can easily bring about rapid joule heating, resulting in progressive and non-recoverable material deformation as can be seen in Fig. 8a.

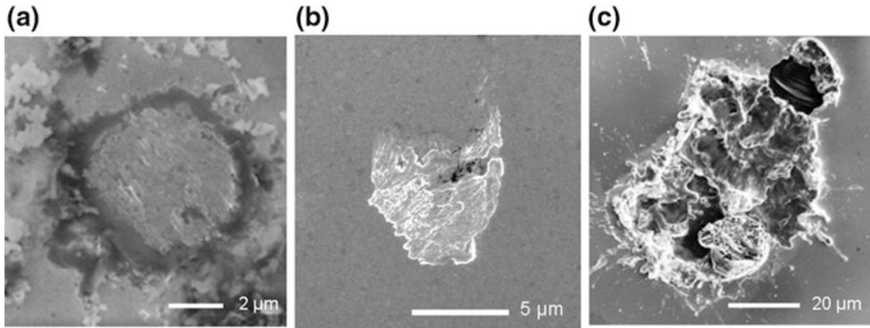


Fig. 8 Various contact failure mechanisms. **a** Wear [133]. **b** Material transfer [82], and **c** Arcing [133]

Material transfer: This indicates that small volume of the contact material at one contact part is transferred to the other contact part during switching operation. Accordingly, it leads to critical degradation in contact surfaces, thus affecting contact resistance and endurance very much. The material transfer phenomenon generally stems from various mechanisms: field evaporation, electromigration, bridge transfer, and arc transfer. Recently, Yang et al. demonstrated that field evaporation is the major cause for the material transfer through his unique setup based on a modified AFM [54, 55]. In addition, Poulain et al. showed that current flow was observed just before contact closure, thus resulting in heating the opposite contact part and evaporation of the anodic material [56]. From both cases, we recognize that high electric field between two contact parts could bring about severe material transfer during the switch "closing" and "opening" moments as can be seen in Fig. 8b.

Contamination film: In company with the mechanical deformations, contamination layers at the contact surfaces due to chemical reaction and oxidation are one of the main factors increasing contact resistance during switching operations. Therefore, gold has been widely chosen due to its high resistance to corrosion, but it is also susceptible to the mono-layers of hydrocarbons and water film. Thus, well-encapsulated (hermetic) packaging with proper ambient such as nitrogen is thought to be necessary for reliable contact endurance.

Arcing: Arcing phenomenon has been researched with the development of electric relays and switches in the twentieth century. The arcing phenomenon is defined as the electrical discharge between two contact parts at high electric-field conditions. At lower power level, the arcing process can be divided into metal and gas phase arcs. In the MEMS switch, only metal phase arc is possible owing to its small device size and microscale gap between the contacts. According to the theory, the real contact area is decreased as the contact parts are gradually separated at the off-state. At that time, the temperature of the contact asperities elevates rapidly up to boiling points of the contact materials. Consequently, the molten contact bridge gets evaporated and high electric field within the contact parts could lead to field emission of the electrons from the cathode to anode, as seen in Fig. 8c.

Accordingly, metal vapor fills the gap between the electrodes. If either the electric field or the electric emission is insufficient during the arcing procedure, the arcing is terminated immediately [57].

Stiction: Stiction failure in the MEMS contact switch comes from the increased surface adhesion force between two contact parts, compared with the restoring force from the movable spring. In case of high current conditions, the permanent contact stiction might originate from the increased effective contact area in the event of repeated plastic deformation during switching cycles [58] or micro-welding due to elevated temperature caused by successive opening and closing of the contact [59]. Also, when the devices are operated in humid environments, water films condensed on the contact surfaces generate capillary forces, and this can aggravate permanent contact stiction failure.

Research Flow

Together with analysis of contact failure mechanisms, numerous studies have concentrated on enhancing the reliability of the MEMS switches through structural, circuitry, and material approaches. Among them, several representative research works are introduced in this section. The first is the structural approach. In 2004, Rebeiz group at UCSD showed the parallel multi-contacts by placing multiple switching elements side-by-side, thus reducing current that flows through a single switching element (Fig. 9) [60]. By doing so, not only was overall insertion loss (contact resistance) decreased, but also contact endurance could be prolonged. Although this method may bring about large footprint and low yield due to the fact that all individual switching elements should operate simultaneously, the Radant MEMS switch also employed this parallel actuation scheme owing to its excellent performance [61].

Also, Chow et al. developed a ball-shape contact dimple array in 2007 (Fig. 10) [62]. By simply using the closely packed extrusion dimple array, field emission

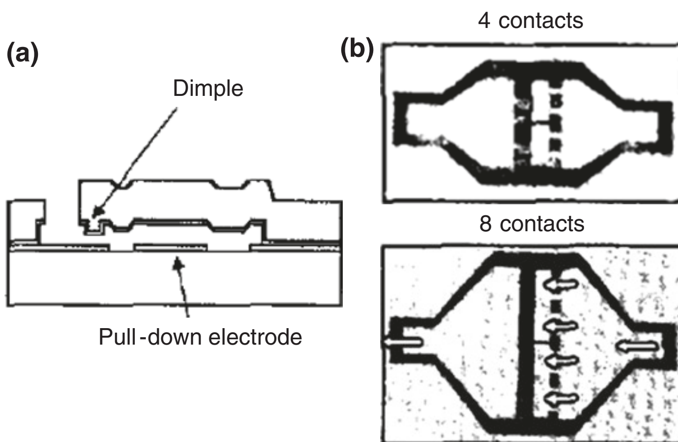


Fig. 9 **a** Cross sectional and **b** *top views* of the parallel-contact MEMS switch [60] (Reproduced from Ref. [60] with permission from IEEE)

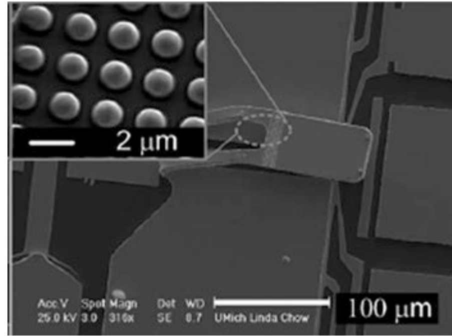


Fig. 10 SEM image of the MEMS switch with ball-shape contact dimple array [62] (Reproduced from Ref. [62] with permission from IEEE)

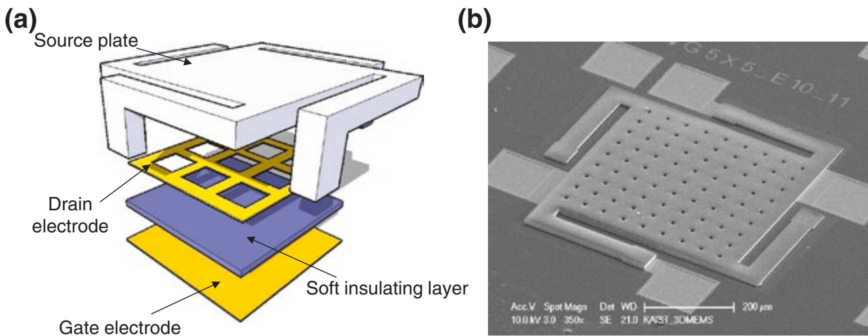


Fig. 11 **a** Schematic illustration and **b** SEM image of the stacked electrode MEMS switch with a soft insulating layer [33]

between two contacts decreases, which stems from electric-field screening effect [63]. At the same time, the ball-shape dimple leads to low adhesion force. Using the simple and easy design, they demonstrated switch lifetime in excess of 100 million cycles at 1 W in hot switching condition. More fundamentally, Song et al. suggested and demonstrated an extremely low contact resistance MEMS switch [33]. They insist that contact resistance should be minimized to reduce the generated heat at contact spots for high contact endurance, because heat generated during switching cycles is the most significant factor leading to failure in high current level [46, 64] as explained before. For very low contact resistance (see Eq. (9)), a stacked-electrode structure for high contact force over a few mN and a soft insulating layer for low effective hardness of the contact material were utilized simultaneously, as presented in Fig. 11. By doing so, it showed the lowest contact resistance of 4 mΩ among the reported MEMS switches and operated over 1 million cycles at 100 mA level. Recently, Rebeiz group at UCSD consistently developed a high contact force (over mN) MEMS switches having small footprints, throughout some minor design variations (Fig. 12) [65, 66]. Owing to

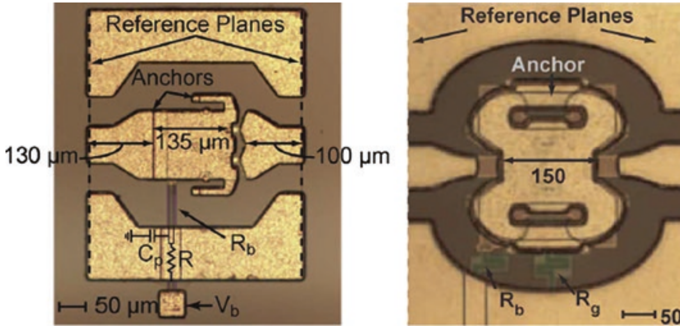


Fig. 12 Photomicrographs of the UCSD MEMS switches with high contact force [65, 66] (Reproduced from Ref. [65] with permission from IEEE)

its small size, low sensitivity to stress gradients, stresses, and ambient temperature can be additionally achieved. They demonstrated high reliability up to 100 million at 100 mW under hot switching condition.

Also, as a circuity approach, the University of Illinois employed a cold switching method (switching operation under a condition in which there is no electric field across the contacts as the switch is turned on or off), increasing switch lifetime up to 7 billion cycles [38]. Since then, the cold switching scheme has been widely used, especially in RF MEMS switches. Also, to reduce the inevitable electric field and arcing between two contact parts, arc-suppression circuit has also been employed [67] (which will be explained later in Sect. 5.3). Recently, a new method introducing a drain voltage-sustaining capacitor, which leads to electric field-less or arc-less switching operations have also been developed [68]. It just used a single capacitor to eliminate the voltage difference between contacts, thus being considered as a simple and effective method.

Many researchers have also studied contact materials to solve the reliability issue, by employing alternative contact materials having a mechanically harder property than pure gold which is the most commonly used contact metal in MEMS switches, as explained later. As presented in Fig. 13, many groups have developed harder materials such as CNT, metal alloys, and refractory metals (Ir, Ru, W, etc.) for the contact materials, and substantiated that they could achieve more reliable contact endurance [69–75]. This is due to the fact that hard metal has relatively high young’s modulus, hardness, and melting point, and thereby it can mitigate the contact failure stemming from wear and material transfer. Also, the hard contact material can enhance the contact reliability in terms of permanent contact stiction, because the hard contact materials showed lower adhesion force than that of the soft contact materials [76]. However, some hard materials such as platinum (Pt) should be carefully considered as a contact material, because it is known that Pt-group metals are susceptible to contamination and frictional polymerization [77, 78] which can lead to unstable and increased contact resistance during the switching cycles. Recently, dissimilar contact materials such as Au-to-Ru and

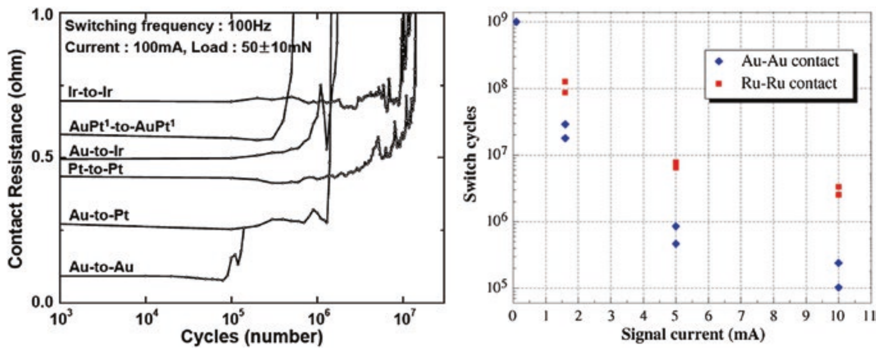
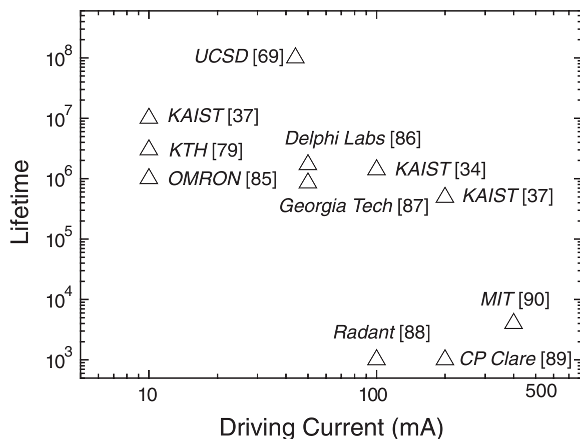


Fig. 13 Representative lifetime results demonstrating higher lifetime of the hard contact materials in contrast to that of the soft contact materials [75, 134] (Reproduced from Ref. [75, 134] with permission from IEEE)

Au-to-RuO_2 contacts have attracted attention, and demonstrated 1 billion cycles at relatively high current level [66] and up to 10 billion cycles at low current level [79]. Thus, a material such as Ru and Rh, which is still conductive even when it is oxidized and at the same time has hard contact property, is recently preferred as a promising contact material.

Although there have been active researches so far, the contact reliability issue is still challenging and the most critical issue in MEMS contact switches, especially in high current/power levels (hot switching condition), as shown in Fig. 14. This is mainly due to the fact that accelerated failure mechanisms under medium and high power levels are still unknown and not understood well. It is our opinion that development of innovate contact material or operation mechanism might be a key solution for enhancing lifetime up to 100 billion cycles at high power level, together with thorough contact failure analysis. Hermetic sealing (packaging), proper switch design for high contact force, and appropriate contact material selection are crucial.

Fig. 14 Lifetime results under hot switching condition with respect to the driving currents



2.5 Selecting Contact Material

The last major issue in the micro-contact switch is the selection of contact material, which is greatly and fundamentally related to the important switching performance criteria, not only contact reliability (as explained above), but also contact resistance and power handling capability. Table 2 lists several metal candidates for the contact material of MEMS switches (The values of material hardness and resistivity may be somewhat different with respect to the measurement setup and environment). Conventionally, gold has been widely used owing to its chemical inertness, low resistivity of roughly 2 μΩ cm, easy fabrication (can be deposited by evaporation, sputtering, and electroplating), and easily deformable property even under low contact force, which leads to low and stable contact resistance. Thus, contact physics for gold material have actively been researched in terms of adhesion force, asperity creep, durability, high and low current behavior, and contact resistance with respect to the applied contact force [46, 59]. However, Au-to-Au contacts showed relatively low hardness and melting point, thus vulnerable to material transfer and wear [59]. Also, it causes high adhesion force (0.3–2.7 mN [59]) and therefore easily results in permanent contact stiction when the restoring force is not large enough to detach the contacts. These features accordingly are primarily responsible for low contact endurance and low power handling capability. Owing to these drawbacks, as previously mentioned, many groups have tried to choose harder materials as contact material such as Rh, Ru, Ir, W, and Pt, demonstrating that hard contact materials offer more reliable contact endurance and high power handling [69–75]. However, as presented in Table 2, hard contact materials indispensably result in high contact resistance due to their relatively

Table 2 Contact material candidates

Contact material	Resistivity (μΩ cm)	Hardness (GPa)	Melting point (°C)	Contact force (μN)	Contact resistance (Ω)
Gold (Au) [46, 136, 137]	2.2	1–2	1060	50–200	0.05–0.2
AuNi (20 %) [138]	20.5	6.1	1040	150	1–1.8
Rhodium (Rh) [139]	9.3	9.75	1960	50–200	3–5
Ruthenium (Ru) [140]	13.8	15.28	2330	50–100	7–8
Iridium (Ir) [134]	15.7	17.4	2460	–	–
Tungsten (W) [141]	–	–	3420	150	<1000
Platinum (Pt) [47, 134, 142]	15.8	5.1	1770	140–200	3–4
CNT [143]	–	–	–	10	285 (on-resistance)
RuO_2 [142]	50	–	–	140	7–10

high resistivity and hardness. Thus, from the viewpoint of contact resistance, soft contact materials such as Au, Cu, and Ag are primarily preferred. To sum, there has been a significant trade-off between two critical performance criteria, contact resistance and reliability. Accordingly, designers have to select the appropriate contact materials in accordance with target applications and their proper usages.

There have been several researches dealing with these conflicting goals (to chase these two hares at once). First, dissimilar contact materials such as Au-to-Pt and Au-to-Ru contacts were employed [80, 81]. By doing so, contact reliability of the MEMS switch became higher than that with only the soft contact material, Au, and the contact resistance became lower than that with only the hard contact materials, Pt and Ru. In other words, they compromised on reliability and resistance which they achieved at intermediate-level contact resistance and reliability. In 2011, Patel et al. suggested to use hard contact material (Ru) with high contact force above mN-level to achieve both high endurance and low contact resistance [65] at the same time, showing contact resistance as low as 2–3 Ω at 90–100 V in the gate voltage, together with high contact endurance. Recently, a new method referred to as “complementary dual-contact” scheme (Fig. 15) was suggested and demonstrated to achieve low contact resistance as much as <0.3 Ω and high reliability as much as 10^5 cycles at 10 V–10 mA level (which is higher than that of a Pt-to-Pt contact switch) simultaneously [82]. The method used two sub-switches with soft (Au-to-Au) and hard (Pt-to-Pt) contact materials in a single switch. In

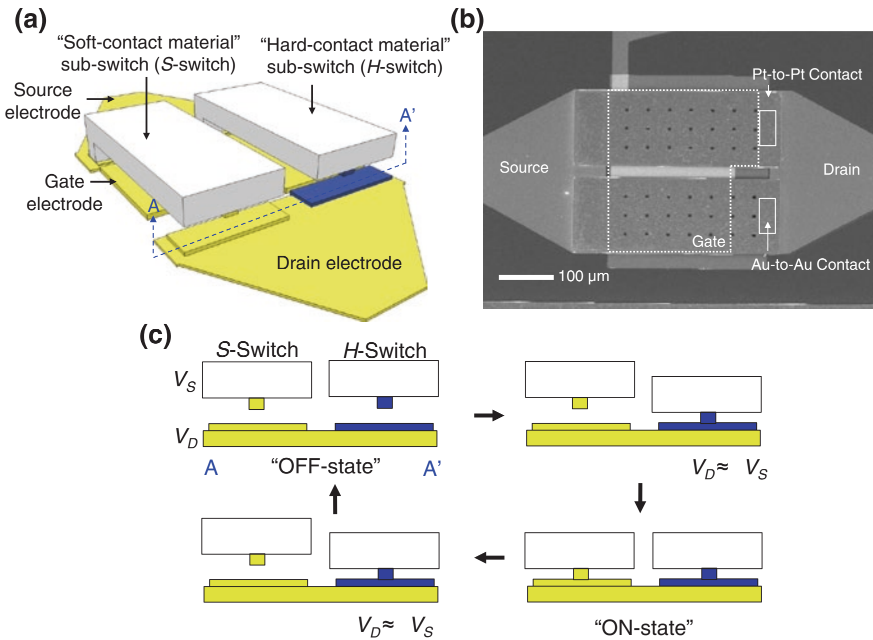


Fig. 15 **a** Schematic illustration and **b** SEM image and **c** operation principle of the “dual-contact” MEMS switch [82]

operation, the transition of the Au-to-Au contact happens exactly within the transition of the Pt-to-Pt contact. By doing so, the Au-to-Au contact can operate under no or little potential difference between the source and drain electrodes during contact "closing" and "opening" moments, thus obtaining more enhanced contact endurance. At the same time, the Pt-to-Pt contact can make up for its high contact resistance thanks to the Au-to-Au contact.

Recently, an advanced type of complementary dual-contact switch that utilizes zipping-technique has been demonstrated for a much smaller form factor and a simpler operation principle than the previous method [83].

2.6 Other Selected MEMS Switches

The Radant MEMS switch is one of the most representative MEMS contact switch because it is a well-known commercialized MEMS device with hermetic packaging (through wafer bonding method) for RF applications [47, 48, 84]. As presented in Fig. 16a, the structure is based on a cantilever beam and the gate electrode is at the end of the suspended beam. All structures consist of metal materials and copper sacrificial layer is employed. Now, the MEMS switches for low and high power applications are, respectively, commercialized. The device demonstrated the

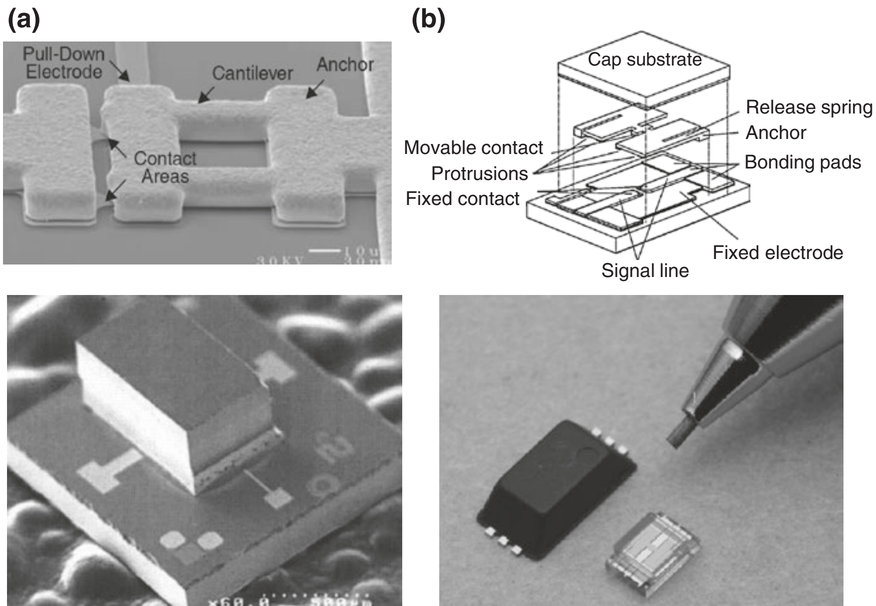


Fig. 16 **a** SEM image and packaging photograph of the Radant MEMS switch [47] (Reproduced from Ref. [47] with permission from IEEE). **b** Schematic illustrate and packaging photograph of the OMRON MEMS switch [49]

actuation voltage of 90 V, extremely high lifetime of 10^{11} cycles at 30 dBm in the RF signal power under the cold switching condition, contact resistance of roughly 2–4 Ω, and the maximum current handling capability of 1 A. Because the device size is not big (roughly 100 μm $\times$ 100 μm), it has relatively high switching speed of 10 μs.

Since 1999, OMRON has also concentrated on the MEMS contact switch, and finally succeeded in commercialization of the MEMS switching device [49, 85]. The switch is fabricated on the basis of a thick single crystal silicon membrane (residual stress and stress gradient are significantly low) and hermetically sealed twice. By etching the movable plate near the anchors, spring stiffness is reduced, as shown in Fig. 16b. The silicon membrane is roughly 1400 μm $\times$ 1400 μm. It showed actuation voltage of 30 V, contact resistance of under 1 Ω owing to high contact force over 1 mN, switching time of 0.3 ms. Also, the device is operated 10^8 cycles at 0.5 mA under the hot switching condition.

3 Nanoelectromechanical System (NEMS) Switches

This section describes the NEMS switch which is about a thousand times smaller than the MEMS switch. The NEMS switch has been considered promising to give revolutionary impact in various research areas such as relays, logic gates, memory devices, and sensors, as NEMS technology is dramatically advanced for future nanoelectronics. Here, advantages, fabrication methods, research flow, and current issues of the NEMS switches are discussed.

3.1 NEMS Switch Advantages

The NEMS switch has been considered as a promising device especially in low-power electronics for the following reasons. Similar to MEMS switches (as explained in Sect. 2.2), the NEMS switch also has unique advantages, because it is also a mechanically operated switching device; ideally zero leakage current, free from substrate kinds, insensitive in radiation and temperature changes, low on-resistance (as compared to the MOSFET), perfectly abrupt switching behavior. Moreover, owing to its extremely small device footprint, the NEMS switch can theoretically have lower actuation voltage, higher switching speed, lower power dissipation, and lower cost compared to the MEMS switch (most reported NEMS switches are actuated through electrostatic scheme because the electrostatic actuation has excellent scalability). For example, if a cantilever-type MEMS switch (length $L = 30$ μm, thickness $t = 1$ μm, air gap $g = 0.5$ μm) fabricated by poly-Si could be scaled down 100 times, it would get the pull-in voltage of 0.35 V and resonant frequency of 140 MHz theoretically, which is comparable to MOSFET in terms of dynamic power consumption and switching speed. Also, in case of being

utilized in logic applications, the NEMS switch requires fewer components to form logic gates, in comparison to that composed by CMOS components [86–88].

Among these advantages, an extremely low leakage current which can lead to very low static power consumption is the most attractive feature, and this is the main reason why NEMS switches receive such strong attention as a next-generation switching device. In the semiconductor-based switching device, a significant leakage current is the major issue as the device size decreases (more detailed descriptions are explained in the introduction to the next section), thus many researchers have utilized new materials with high-dielectric constant (high-*k*) gate dielectrics such as hafnium oxide (HfO_2), zirconium oxide (ZrO_2), alumina (Al_2O_3) and their silicate for replacing SiO_2 [89]. By doing so, the physical thickness of the gate dielectric layer can be thicker than that of SiO_2, thus bringing about several orders of magnitude lower leakage current from the gate to the channel. Also, structural approaches including a self-aligned double-gate MOSFET (FinFET) and a gate-all-around (GAA) transistor were extensively researched to minimize the short channel effect [90]. However, these methods cannot perfectly reduce the leakage current, and some side effects such as threshold voltage (Vth) instability and relatively poor device reliability are left unsolved. Owing to these facts, the NEMS switch with ideally zero leakage current (subthreshold slope lower than 60 mV/dec [91]) has an absolute strength over the semiconductor-based switching devices.

3.2 Fabrication Methods

There are two main streams in fabricating the NEMS switches: one is a top-down approach and the other is a bottom-up approach. The top-down approach means a process that starts from a large piece and subsequently uses finer tools for creating correspondingly smaller structures. It commonly employs a conventional lithography technique (in case of nano-device, E-beam lithography), thus resulting in CMOS-compatible fabrication process, high reproducibility, and high throughput. Also, the bottom-up approach means making larger and more organized systems by combining smaller components of atomic or an externally applied driving force. In other words, the bottom-up approach generally uses a self-assembly principle that exploits chemical or physical forces, assembling basic units (molecular or nanosized chemicals) in nanoscale to make larger structures. Most NEMS switches with the bottom-up scheme employ carbon nanotube (CNT) [91]. CNT or chemically synthesized nanowire NEM switch presents high density and high speed compared with those fabricated by the top-down approach (in general, most NEMS switching devices with the top-down approach are difficult to have high scalability up to a few nanometer scale level); whereas assembling the nanowires into arrays is difficult owing to random CNT growth. Having their respective strengths and weaknesses, both fabrication techniques are currently exploited to realize NEMS switches.

3.3 Research Flow

In the early stage, NEMS switches were fabricated through the bottom-up approach using CNT (or chemically synthesized nanowire) as a structural material. Rueckes et al. [92] first proposed the CNT-based nanomechanical memory with high switching speed, high density, and nonvolatility. Then, Kaul et al. [93], Dujardin et al. [94], and Cha et al. [95] reported the CNT-based NEMS switches that operate at low actuation voltages of 2, 2.9, and 3 V, respectively. Also, Lee et al. [96] developed the NEMS switch integrated with CNT-FET (field-effect-transistor) for a nonvolatile memory application. Jang et al. [97] proposed and characterized the CNT-based nanomechanical memory using vertically grown MWCNT (multi-wall CNT), demonstrating high density possibility. However, the existing NEMS switching devices on the basis of the CNTs growth are still very challenging to be realized, owing to various obstacles such as difficulty in controlling the CNT's position and managing population of each CNTs. Accordingly, excellent fabrication platform and process are required to conduct a parametric study and material optimization. As a matter of fact, NEMS switches based on the bottom-up approach are not likely to achieve reliable performance in the near future. More detailed researches on the CNT-based NEM switch are well described in a recent review-paper [91].

On the other hand, as technology for CMOS fabrication advances, NEMS switches with the top-down scheme, which can control the position of the device arrays precisely, have been actively sought. Therefore, in this chapter, various kinds of NEMS switches based on the top-down approach are mainly described. In 2008, Jang et al. introduced a cantilever type 2-terminal NEMS switch with a 15 nm thick suspension air gap (Fig. 17a) [98]. Because the air gap is fabricated with an accurate film deposition, a very stable and flat beam structure can be fabricated. Also, it employs the well-developed CMOS fabrication techniques and

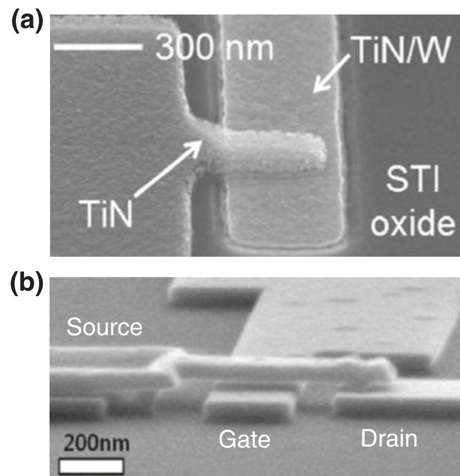


Fig. 17 SEM images of **a** 2-terminal and **b** 3-terminal KAIST NEMS switches [98, 99]

CMOS-compatible materials (TiN), thus facilitating high scalability and monolithic integration with CMOS devices. Although it is the first demonstration of very small air gap nanomechanical switch with the top-down approach, the device showed repetitive switching operation of over several hundred switching cycles by employing ultrathin oxide layer, where the lifetime is relatively long in contrast to other NEMS switches (most NEMS switches failed after a few cycles [91]). Subsequently, this work was followed by a 3-terminal NEMS switch with a 50 nm thick TiN cantilever and a 40 nm air gap, which was also demonstrated in KAIST [99], as presented in Fig. 17b.

Since then, various kinds of unique structures were devised for NEMS switches. Similar to the micro-scale mechanical switches, there was an effort to use the torsional spring for achieving low actuation voltage in NEMS switches by using focused ion beam (FIB) milling and silicon-on-insulator (SOI) substrate [100]. As shown in Fig. 18a, the structure consists of a cantilever beam (9 μm long and 1.5 μm wide) and a torsional spring of length of 2.4 μm and width 530 nm. By doing so, relatively low actuation voltage of 5.5 V was achieved. Also, You et al. suggested a U-shaped NEMS switch consisting of two silicon nanowires and a capacitive paddle (Fig. 18b) [101]. The dimension of the nanowires was 5 μm long with cross section of 90 nm by 90 nm. Owing to the large capacitive paddle for high electrostatic force and long and thin nanowires for low spring stiffness, the actuation voltage was 1–2 V. These two NEMS switches demonstrated the possibility of low actuation voltage, but it is noteworthy that this feature accordingly may lead to permanent contact stiction. Recently, sub-1 V (roughly

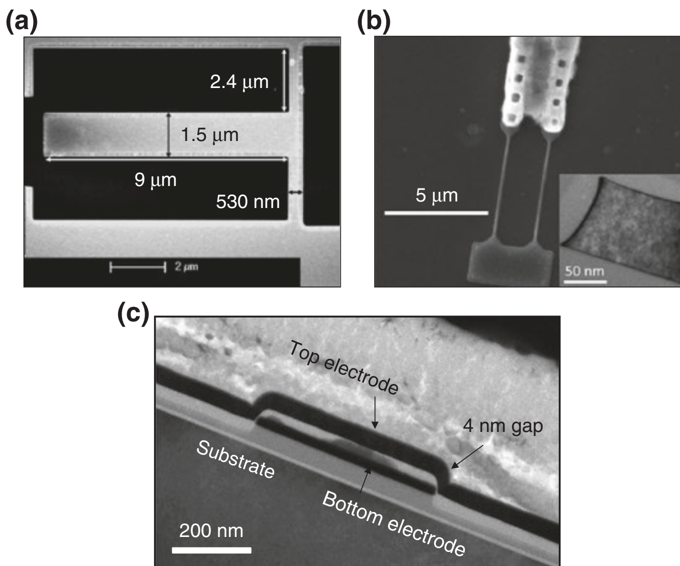


Fig. 18 SEM images of the NEMS switches for low operation voltage. **a** Torsional spring [100] and **b** U-shape [101] and **c** pipe-clip [102] NEMS switches

0.4 V) NEM switch was first demonstrated by Lee et al. [102]. They exploited a unique pipe clip structure to form an extremely small air gap of 4 nm, as shown in Fig. 18c. Generally, making a small air gap for NEMS switch is difficult because the small air gap leads to low restoring force in comparison to the contact adhesion force. In this research, by restricting the apparent contact area through only the curved region in the pipe clip structure, the van der Waals force that is responsible for the contact stiction can be minimized. This result (sub-1 V of operation voltage) represents a significant meaning that the NEMS switch can operate at an ultralow voltage similar to or even lower than that of the MOSFET, together with extremely low leakage current. That is, it demonstrated that the NEMS switch can be a suitable candidate for being used in ultralow power electronics.

Together with the low actuation-voltage NEMS switches, one of the strong merits of the miniaturized mechanical switch, which is its insensitiveness to harsh environments (high temperature ranging from 300 to 600 °C), was demonstrated using a cantilever-type NEMS switch made of silicon carbide (SiC) in 2010 [103]. At temperature higher than 300 °C, the silicon-based devices are difficult to be utilized owing to excessive leakage current which stems from the p–n junction degradation and thermionic emission [104]; whereas the NEMS device made of SiC has excellent thermal stability and mechanical robustness, thus making the device immune to high temperature. Also, because it is a laterally operated mechanical switch (Fig. 19a), all structures can be fabricated at once using a single lithography. The length, width, and air gap of the cantilever beam are 8 μm, 200 nm, and 150 nm, respectively. Here, high reliability over 2 billion cycles at 500 °C was demonstrated. Also, very thin and long SiC nanowire NEMS switch was suggested for low actuation voltages down to the level of 1 V and fast switching time under sub-microsecond (Fig. 19b) [105].

Apart from these researches, a mechanical and electrical transistor structure (METS) which consists of a movable nanowire (as a channel for FET) and two gate electrodes have been developed (Fig. 20) [88, 106, 107]. This new switching

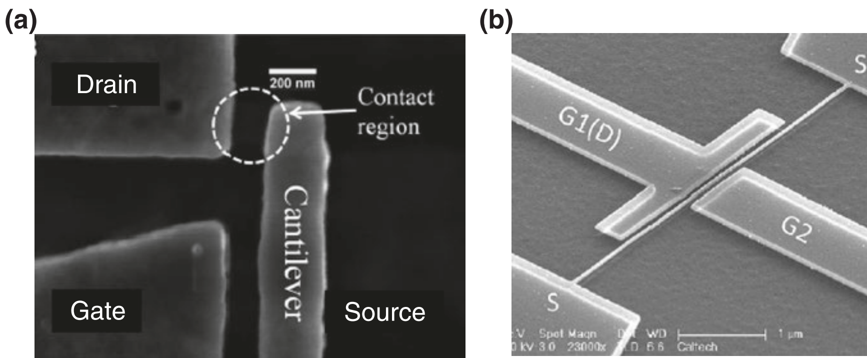


Fig. 19 SEM images of the SiC NEMS switches based on **a** the cantilever beam [103] (Reproduced from Ref. [103] with permission from the American Association for the Advancement of Science) and **b** the long nanowire [105]

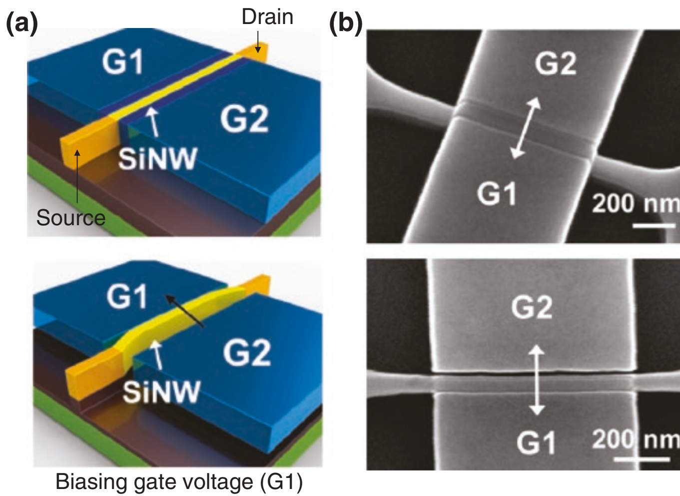


Fig. 20 **a** Schematic illustration and **b** SEM image of the mechanical and electrical transistor [88]

device is a combination of the FET and the NEMS switch. The movable nanowire posted at the drain and source electrodes make contact with one of the two separate gate electrodes by electrostatic force (flip-flop operation). As shown in Fig. 20, the gate electrode is mechanically apart from the silicon nanowire at the off-state, thus the off leakage current is very low, similar to that of NEMS switch. At the same time, when the suspended nanowire is actuated and made contact with the gate electrode, sudden current increase occurs and the current level becomes the same as that of the conventional MOSFET. The air gap could be reduced down to 2 nm, therefore it can achieve very low actuation voltage of sub 2 V by atomic layer deposition (ALD) [107]. It could be fabricated with full CMOS-compatible fabrication process. Also, combining both the mechanical and electrical switching devices demonstrated multifunctionality such as making a p–n junction through two contact parts and logic gates [88].

3.4 Current Issues

In contrast to the microscale mechanical switch, demonstrating the nanosized mechanical switch itself has been difficult so far, in terms of design, fabrication, and cost. Accordingly, switching performances of the currently reported NEMS switch are relatively immature. Being exploited in various applications including memory device and logic gates integrated with conventional CMOS, there are several challenging issues to be solved.

First, actuation voltage should be minimized down to sub-1 V. To employ the NEMS switch as a core device component in the ultralow power IC, reducing actuation voltage is important for reducing supply voltage (V_{DD}), which is directly related to dynamic power consumption. However, small air gap and low stiffness for sub-1 V lead to low restoring force, thus causing permanent contact stiction. As explained earlier, various structural approaches such as long and thin nanowire and torsional spring were suggested. Also, since Lee et al. recently demonstrated possibility of the ultralow actuation voltage NEMS switch, many researchers have actively focused on reducing actuation voltage down to sub-1 V with various designs [107, 108].

When utilizing NEMS switches in logic and memory applications, the switching speed should also be enhanced since existing CMOS-based devices are very fast. The reason why the switching speed of NEM switches is slower than that of the CMOS-based device is that there should be a mechanical movement of the NEMS switch structure during switching operation. As mentioned before, by scaling down the size of the NEMS switch, the corresponding resonant frequence (i.e., switching speed) can be increased. Selecting materials with smaller mass and higher spring stiffness is also a good method to increase switching speed. However, device designer should consider the trade-off between the switching speed and the actuation voltage since high spring stiffness for high resonant frequency accordingly leads to high actuation voltage. In nanoscale, owing to the mechanical settling time and contact adhesion force, the pull-out (switching-out) time of the NEMS switch may be longer than pull-in (switching-on) time.

Another severe problem of the NEMS switch is its contact reliability, which is the same in the MEMS contact switch. Especially, very low contact force of NEM switches (generally under 1 μN) leads to unstable contact and high contact resistance over 1 kΩ (it is difficult to achieve ohmic contact through piercing the thin adsorbed or contamination film layer). Also, for achieving stable contact, high drain voltage to break the shallow insulating layer is commonly applied. By doing so, the contact degradation becomes more severe because the air gap of the NEMS switch is extremely small, therefore even very small drain voltage could generate extremely high electric field between the two contact parts. We think that the most effective solution to this problem is using contact materials having high thermal conductivity, high melting point, and high mechanical robustness such as carbon-based material and SiC to suppress contact degradation.

Additionally, most reported NEMS switches have been demonstrated through two-terminal configuration. To favorably exploit the NEMS switch for replacing CMOS circuits and conventional memory devices, three-terminal structure should be considered in the near future. The devices also have to be designed and fabricated in light of CMOS compatibility and hermetic packaging.

4 Logic and Memory Applications

4.1 Introduction

In 1947, Shockely, Brattain, and Bardeen invented the transistor, which is the most primary type of silicon-based electronic devices, at the Bell laboratory. Since then, integrate circuits (IC) technology that monolithically integrates transistors, resistors, inductors, and capacitors in the same planar substrate, has been aggressively progressed together with the extensive development of micro/nanofabrication technology. Also, after creating complementary-metal–oxide-semiconductor (CMOS) which employs both p- and n-type MOSFETs to implement logic gates and other digital circuits [2], it is favorable for very large-scale integration (VLSI) owing to its excellent scalability, low power dissipation, and high speed. In particular, its superior miniaturization ability leads to tremendous performance improvement and cost reduction simultaneously. In other words, growth of IC industry has been dramatically accelerated with competitive scaling down tendency. This increasing device density tendency has been predicted by Moore's Law, which is well known as the number of transistors per IC becomes double every 18 months [109, 110].

However, scaling in the conventional CMOS and CMOS-based memory to increase both the integration density and the operation speed is rapidly facing a physical limitation. That is, many difficult challenges in continuing to scale down beyond sub-100 nm have been encountered so far, such as power dissipation, parasitic leakage currents, and short-channel effects [3–5]. These fundamental limits have led to pessimistic predictions of the "End of Moore's Law" for the semiconductor industry.

On the other hand, many barriers to further scaling of CMOS have introduced new technologies that overcome its physical limits: introduction of new approaches for logic and memory devices based on a nonsolid state device. Among them, mechanically operated switching device, the MEMS/NEMS switch (especially, NEMS switch due to its small size), can be a good solution for breaking the physical limit of the CMOS, thanks to the various reasons explained in Sect. 3.1 (particularly, quasi-zero leakage current and perfectly abrupt switching behavior).

4.2 Logic Applications

Among the CMOS' fundamental processes and physical parameters including lithography, power supply and threshold voltage, short-channel effect, gate oxide, high field effect, dopant number fluctuations, and interconnect delays [3], the most critical issue when the MOSFET is especially utilized in logic applications is power dissipation. The total power dissipation in the digital CMOS circuits can be summarized as follows [111]:

$$P_{\text{total}} = p_t(C_L \cdot V \cdot V_{\text{DD}} \cdot f_{\text{CK}}) + I_{\text{SC}} \cdot V_{\text{DD}} + I_{\text{OFF}} \cdot V_{\text{DD}} \tag{18}$$

The first term represents power from the switching component, where p_t is the probability that a power consuming transition occurs (activity factor), C_L is the load capacitance, f_{CK} is the clock frequency, V is the voltage swing (in most cases, it is the same as V_{DD}). The second term stems from the direct-path short circuit current I_{SC}, which arises when both NMOS and PMOS are turned-on, and current flows from power supply to ground. The final term is related to the leakage current (I_{OFF}). As shown in Eq. (18), the most effective way to reduce active (dynamic) power dissipation is to reduce V_{DD}. Accordingly, V_T should also be reduced to maintain high current and fast operation speed [50]. When we reduce V_T, however, off-state leakage current between the drain and source electrodes (I_{OFF}) would exponentially increase because I_{OFF} (which is directly related with passive power dissipation) increases with decreasing V_T, as follows [50]:

$$I_{\text{OFF}} \propto 10^{-V_T/S} \tag{19}$$

where S is the subthreshold swing. This means that there is a trade-off between active and passive power dissipations, and this makes IC design complicated. Thus, the miniaturized mechanical switch with ideally zero leakage current and infinite subthreshold swing regardless of the supply voltage can be considered attractive in logic applications, which can eliminate the trade-off between active (dynamic) and leakage (static) power dissipations.

Complementary logic gates including inverter, NAND, and NOR utilizing mechanical switches can be implemented by replacing both the p- and n-type MOSFETs with the MEMS/NEMS switches ("CMOS-equivalent"). For example, Fig. 21a presents the complementary mechanical logic inverter (basic unit of the logic gates) schematic [112] and the configuration is the same as that of the CMOS inverter. The source electrode of one mechanical switch (pull-up device) is

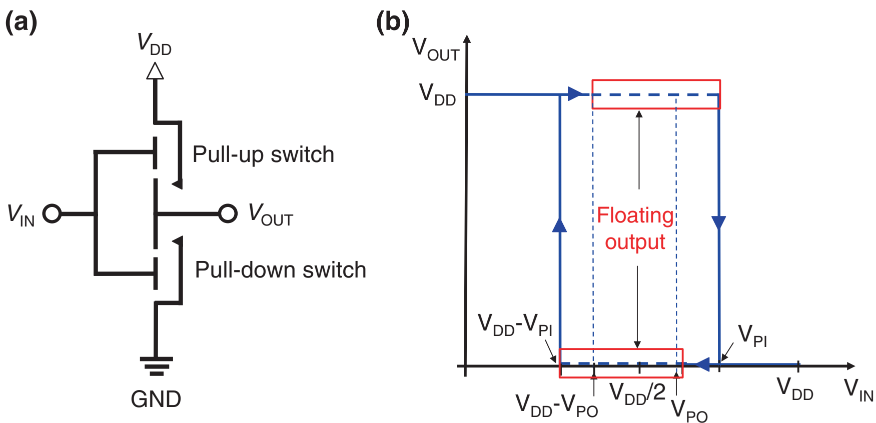


Fig. 21 **a** Mechanical logic inverter configuration. **b** Voltage transfer characteristic [112] (Reproduced from Ref. [112] with permission from IEEE)

biased to the supply voltage (V_{DD}), and that of the other mechanical switch (pull-down device) is connected to the ground. Both the gate and drain electrodes of the two mechanical switches are electrically connected, respectively. In terms of the voltage transfer characteristics (VTC), which is the key indicator to evaluate the complementary logic inverter, the mechanical logic inverter has unique features, as presented in Fig. 21b. When the input voltage (V_{IN}) is close to 0 V, owing to the voltage difference between the source and gate electrodes of the pull-up switch, the pull-up switch turns on, causing the output voltage (V_{OUT}) to be V_{DD}. As the input voltage increases and when the voltage difference reaches to the pull-out voltage (V_{PO}) of the pull-up switch, the pull-up switch is then detached from the output electrode and the output electrode becomes electrically floating. As the input voltage increases further than the voltage where the pull-down switch turns on (pull-in voltage (V_{PI}) of the pull-down switch), the output voltage becomes 0 V. As the input voltage is then swept back to the negative voltage, the pull-down switch is then turned off and the output voltage becomes electrically floating again. Finally, as the input voltage deceases close to ground, the pull-up switch is reattached, causing the output voltage to be V_{DD}. As can be seen in Fig. 21b, there is a voltage window due to the hysteresis of the mechanical switches, which is not good with respect to the voltage noise margin. However, the hysteresis can sufficiently be minimized by proper design of the mechanical switch which excludes the pull-in phenomenon [11] and by reducing the adhesion force between the two contact parts of the mechanical switch [82].

Next, a variety of MEMS/NEMS switch implementations and corresponding logic gates (i.e., inverter, NAND, NOR) for ultralow power electronics are discussed. Among them, a new logic switching device (referred to as a suspended-gate field-effect transistor, SGFET) utilizing an electrostatically actuated mechanical membrane (NEMS/MEMS switch) and a MOSFET simultaneously was suggested (refer to the METS in Sect. 3.3). Actually, the SGFET concept which combines both the electrical devices was first introduced 45 years ago [13] for gas and pressure sensing. In 2002 and 2005, it was theoretically conceived as an abrupt current switching device by Ionescu et al. [113] and Kam et al. [114] to demonstrate excellent electrical characteristics including high on/off current ratio, low power dissipation, and low leakage current. Then, an analytic model of the SGFET was developed by Akarvardar in 2008 [115]. Recently, the SGFET concept was demonstrated using a nanowire [107, 116], highlighting the potential of the nanoscale SGFET as an ultralow power logic element (Fig. 22). Basically, the SGFET is also a 3-terminal switching device (the same as M/NEMS switches) where the current from a source to a drain is modulated by the gate electrode. In Fig. 22, the channel is normally on at 0 V of the gate voltage owing to whole gas accumulation [116], and at the OFF-state, it comes to make contact with the gate electrode by applying a threshold (pull-in) voltage to the gate electrode.

Also, there were effective design considerations for the complementary mechanical logic gates, which inspired many other researchers later [112]. Figure 23a shows the lateral and vertical design rules for the complementary mechanical logic inverter. In both inverter designs, when high voltage (V_{DD}) is

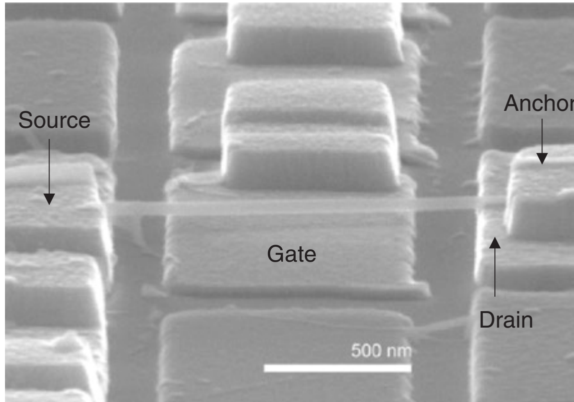


Fig. 22 SEM images of the Nanowire SGFET [116]

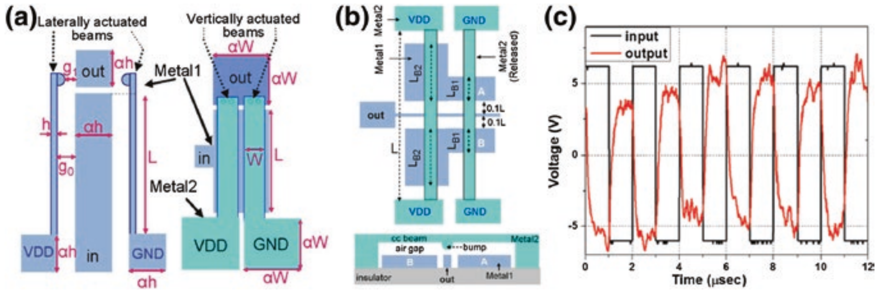


Fig. 23 Lateral and vertical designs for **a** logic inverter and **b** NAND gate [112] (Reproduced from Ref. [112] with permission from IEEE) **c** Input-output voltage waveform of the mechanical inverter 500 °C [103] (Reproduced from Ref. [103] with permission from the American Association for the Advancement of Science)

applied to the input electrode, a cantilever electrically connected to the ground (GND) would be operated to make contact with the output electrode, and vice versa. Especially, the laterally actuated inverter theoretically occupied only 0.03 $\mu m^{2,}$ that is, 1/3 of the area occupied by the minimum sized inverter in 45 nm CMOS technology (channel length of the CMOS is much smaller than the mechanical inverter, but contact part and active area of the CMOS require larger area than the mechanical inverter). Furthermore, they introduced a simple NAND gate design with a unique "dual-beam" scheme, as shown in Fig. 23b. By means of the laterally operated mechanical logic inverter design, Lee et al. [103] demonstrated the logic inverter with the SiC-based NEMS switch (as explained in Sect. 3.3), with $V_{DD} = 6$ V and $V_{SS} = -6$ V and actuation frequency of 500 kHz (Fig. 23c). Although the waveform of the output voltage is relatively rough wing to high contact resistance and high operation frequency, it has a strong meaning as

the first demonstration of the nanoscaled mechanical logic gate in harsh environment (at 500 °C) where the MOSFET cannot be operated.

Liu group has actively conducted various researches on mechanical logic gates for ultralow power ICs [29, 50, 117]. In terms of the mechanical switch design for the logic gates, they employed a four-terminal configuration where the switch on/off state is determined by the voltage difference between the gate and body electrodes (Fig. 24a, b). If $V_{GB} > V_{PI}$, the mechanical switch becomes the on-state (floating channel electrode makes contacts with both the drain and the source electrodes and current can flow between them), and if $V_{GB} < V_{PI}$, the mechanical switch becomes the off-state, and the channel and gate electrodes are electrically isolated by the gate oxide. There are two main reasons to employ the four-terminal configuration. First, when connecting multiple switches in series (it is sometimes desirable), the source voltage is not uniquely defined in the conventional three-terminal switches, thus operation voltage is variable. However, by using the four-terminal configuration, actuation voltage is not affected by the source and the drain electrodes. Also, by means of changing the body bias, the actuation (gate) voltage can be controlled because the switch on/off is determined by V_{GB}, and thereby actuation voltage <2 V can be achieved. Here, tungsten was used as the contact material and the device was successful to operate up to 10^9 cycles together with 10 kΩ of the contact resistance. As shown in Fig. 24c, d, they also demonstrated

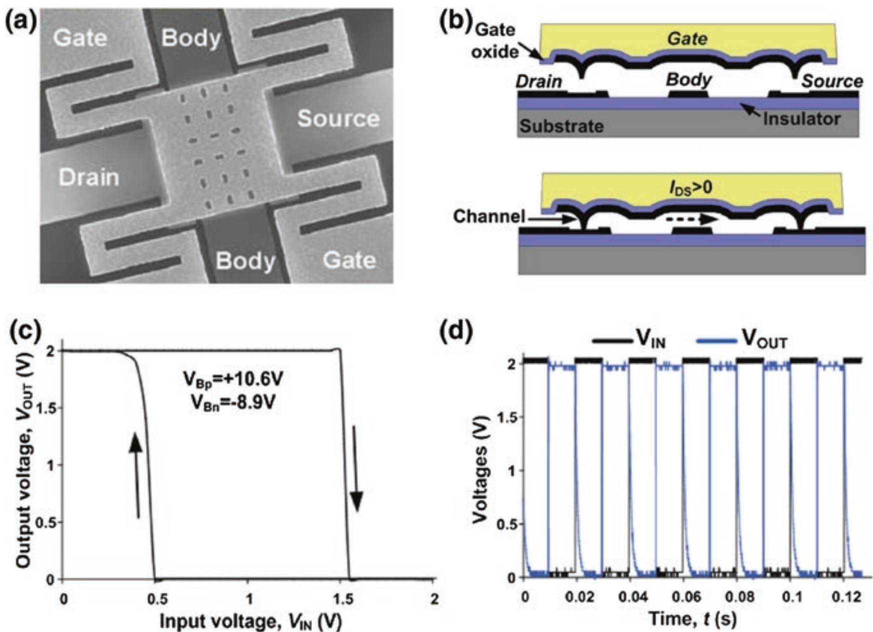


Fig. 24 **a** SEM image and **b** operation principle of the 4-termical MEMS switch (UC Berkeley). **c** VTC curve and **d** timing diagrams of UC Berkeley mechanical inverter [50] (Reproduced from Ref. [50] with permission from IEEE)

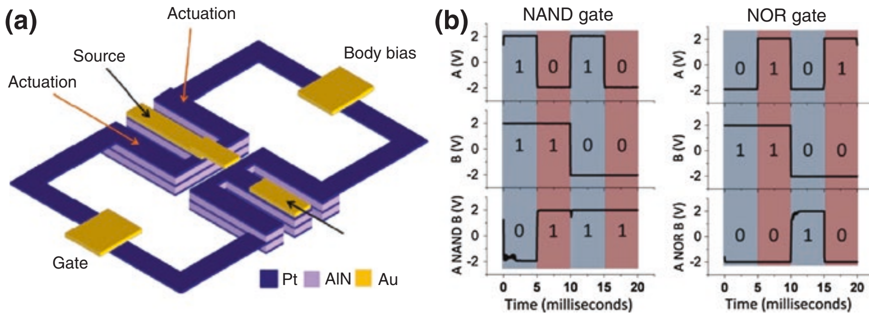


Fig. 25 **a** Schematic illustration of piezoelectrically actuated mechanical switch using the opposite body biasing scheme for logic application. **b** Measurement results for the NAND and NOR gates [118] (Reproduced from Ref. [118] with permission from IEEE)

the mechanical logic inverter circuit. With $V_{DD} = 2$ V, the symmetric VTC curve was achieved, but hysteresis ($V_{PI} - V_{PO}$) should be minimized for large noise margin.

In addition, Kim et al. demonstrated the mechanical logic gates, including an inverter and a mechanical latch using the see-saw type MEMS switch with stiction-recovery actuation (as explained in Sect. 2.3) [39]. Using these see-saw type mechanical switches, they demonstrated the mechanical logic inverter with $V_{DD} = 3$ V and a MEMS D-latch consisting of three mechanical switches with $V_{DD} = 5$ V. Interestingly, in contrast to all other mechanical logic gates with the electrostatic actuation, MEMS logic gates implemented through a piezoelectrically actuated mechanical switch (Fig. 25a) was also suggested [118]. The piezoelectric MEMS switch presented low actuation voltage of 1.5 V through opposite body biasing scheme which is similar to the four-terminal configuration. Through the "CMOS-equivalent" configurations using the piezoelectric MEMS switches, not only the complementary inverter but also the NAND and NOR gates were successfully demonstrated, as shown in Fig. 25b.

Although there were several demonstrations of the logic gates with the mechanical switches, it is still challenging for MEMS/NEMS switches to be implemented in logic applications, in contrast to the commercialized CMOS logic gates. To obtain competitive performance in comparison to that of the CMOS logic gates, mechanical switches should have low actuation voltage <1 V, low contact resistance <10 kΩ, high lifetime >10^{14} cycles, extremely small size, and switching speed at picosecond level [50]. Among them, low actuation voltage was achieved by various structural approaches (ex. torsional spring) and the four-terminal configuration (body-bias scheme). Actually, to achieve actuation voltage <1 V, pull-out condition of the mechanical switch is important; and hence, contact adhesion force which is directly related to the pull-out voltage should be minimized. Then, contact resistance specifications of the mechanical switch for logic applications are relatively high, which is different from that for RF applications (<1–2 Ω of contact resistance). Thus, hard contact materials which have relatively

high contact resistance and high reliability such as platinum and tungsten are more appropriate, compared to soft contact materials including gold, because the lifetime is a more severe issue in mechanical switches in logic gates. Also, most mechanical logic gates were substantiated with the microscale mechanical (MEMS) switches owing to their easy fabrication. Therefore, active demonstration of various nanosized mechanical logic gates is required. Lastly, Pott et al. compared the MEMS switch with the 65-nm scale MOSFET (based on specifications in the international technology roadmap for semiconductors (ITRS)) in terms of the operation voltage, layout area, and speed [50]. At this time, the most significant problem is switching speed of the mechanical switch, which is 4–5 orders of magnitude slower than that of the MOSFET. Thus, we think that it is also urgent for the mechanical logic gate to enhance the switching speed now.

4.3 Memory Applications

Along with the logic gates, memory device is one of the key applications where miniaturized mechanical switches can be utilized. Historically, a society requires computational memory tools to memorize and support the development of our civilization. As a society develops and gets larger, higher memory capacity is demanded. However, conventional memory technologies are facing many challenges including difficulties of miniaturization and guaranteeing of high reliability. Typically, scaling of the memory cells including dynamic random access memory (DRAM), static random access memory (SRAM), and FLASH memory is getting difficult owing to leakage current, immunity to process variations, and noise. Thus, various kinds of emerging memory devices are suggested to address these issues of conventional memory technologies, and one of the primary candidates is a mechanical memory using miniaturized mechanical switches.

In the beginning, a mechanical beam was utilized to demonstrate the concept of the mechanical memory devices, similar to the commonly employed mechanical toggle switches in our daily life (using beam buckling to make bistable conditions). Halg et al. developed a primitive mechanically operated nonvolatile memory [119], which stored the data using capacitance difference through the bistable motion of a metal bridge in the early 1990s, as shown in Fig. 26. This method leads to unlimited retention time and immunity to electromagnetic fields and mechanical shocks; whereas switching voltage was as high as around 30 V and the memory cell size was relatively large to be used as a memory device at that time.

Then, Cavendish-Kinetics developed a new mechanically operated nonvolatile memory (NVM) using a vertical flip-flop scheme [120, 121]. They used CMOS backend process. For program (state “1”), actuation voltage is applied to the bottom electrode and the cantilever beam makes contact with it (the bottom electrode). At this time, owing to its low stiffness and large adhesion force which stems from the metal-to-metal bonding and van der Waals force lead to holding

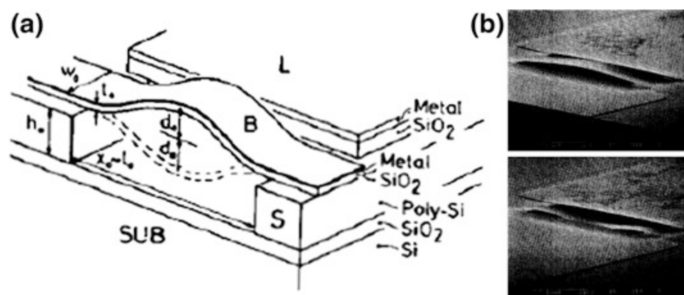


Fig. 26 **a** Schematic illustration and **b** SEM images of the mechanically operated non-volatile memory using bistable motion [119] (Reproduced from Ref. [119] with permission from IEEE)

the beam at the down state "1" through maintaining electrical connection between the bottom electrode and the cantilever beam. When making state "0" (erase state), sufficient voltage is applied to the top electrode to detach the contact (ceiling of the packaged cavity is used as the top electrode). With scaling of the device geometry, the switching speed could be reduced by 100 ns, which is faster time to program and erase than those of FLASH and silicon-oxide–nitride-oxide-silicon (SONOS) devices. Also, its unique stacked-electrode design naturally leads to hermetically sealed packaging. As a matter of fact, control and prediction of the contact adhesion force between two contact parts is not easy, but they guarantee it up to a certain point through theoretical modeling.

Jang et al. developed a nanoscale memory cell utilizing nanoelectromechanical switched capacitor structure, which used a vertically aligned multiwall carbon nanotube (MWCNT) in 2008 [97]. As shown in Fig. 27, one memory cell consists of a CNT (60 nm)-insulator (SiN_x, 40 nm)-metal (Cr, 30 nm)—CIM capacitor at

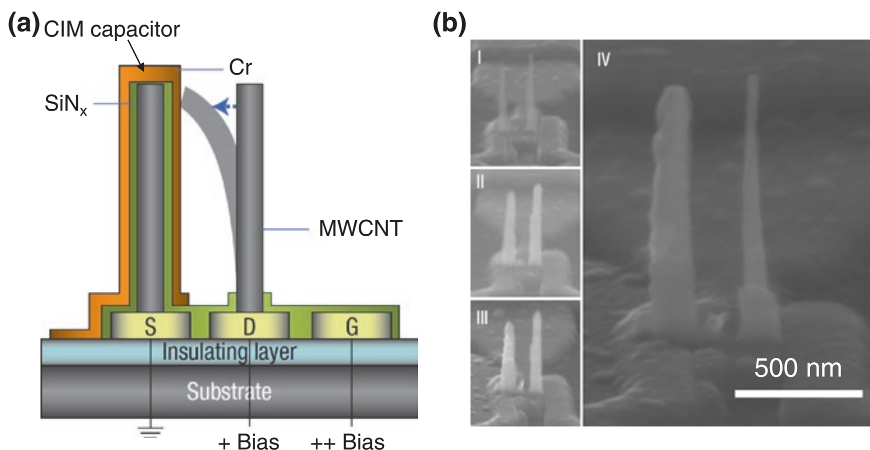


Fig. 27 **a** Schematic illustration and **b** SEM images including the fabrication process of the memory cell based on the nanoelectromechanical switched capacitor [97]

the position of the source electrode, a vertical drain electrode with MWCNT, and a gate electrode which has no nanotube. In operation, the gate and the drain electrodes are connected to the word and bit lines, respectively, and the source electrode is electrically grounded. When a constant positive voltage is applied to the bit line and at the same time a positive voltage higher than the bit line voltage is applied to the gate electrode, a repulsive electrostatic force between the gate and drain electrodes and an attractive electrostatic force between the drain and source electrodes make the nanotube at the drain electrode to be actuated and make contact with the metal (Cr) on the capacitor. At that time, a transient current flows to charge the CIM capacitor. The overall write-read scheme is the same as that of the conventional DRAM. Here, the proposed operation method uses repulsive force, not like the conventional electrostatic actuation utilizes attractive force. Nevertheless, its vertical MWCNT structure leads to a significantly small footprint, high switching speed (switching frequency, 62–750 MHz), and low power consumption compared to the conventional DRAM.

KAIST also suggested and demonstrated a new type of nonvolatile memory called mechanically operated random access memory (MORAM), which is composed of one mechanical switch and one storage capacitor [122]. Its configuration is similar to that of the DRAM, except the fact that the switching device (MOSFET) in DRAM is replaced by the mechanical switch, as shown in Fig. 28. Using the mechanical switch with ideally zero leakage current, the charge leakage path from the storage capacitor through the off-state mechanical switch does not exist, thus it can be operated like a nonvolatile memory (DRAM loses stored data in the storage capacitor owing to the inevitable leakage current of the MOSFET when the supply power is cut off).

To summarize, many innovative mechanically operated devices have been researched for replacing existing nonvolatile memory devices so far. Owing to their strengths, the mechanical memory devices can take a place in the ITRS

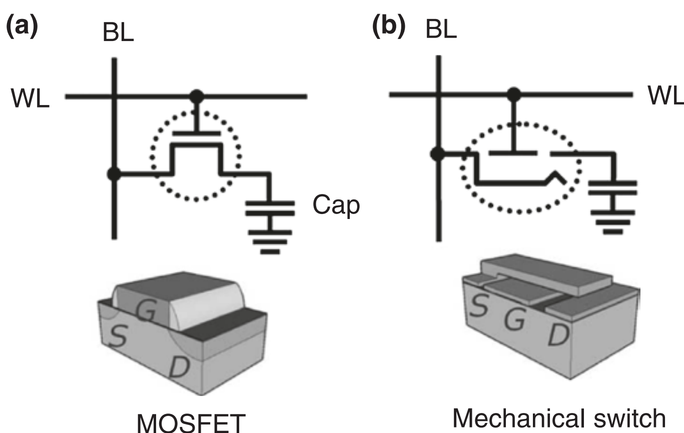


Fig. 28 Schematic diagrams and equivalent circuits of **a** the conventional DRAM and **b** new-type DRAM with the mechanical switch [122]

roadmap as potential alternatives. In general, to be the "next" memory device, it needs to possess the following attributes: fast access time, nonvolatility, infinite read/write/erase cycles, low power, a wide operating temperature range, scalability, low cost, and CMOS compatibility. Among them, the mechanical memory using the MEMS/NEMS switch can have the miniaturized mechanical switch's own strengths, thus possessing the strong points in terms of nonvolatility, extremely low power, insensitiveness to radiation, and temperature changes. Thus, although most MEMS/NEMS switch-based memory cells presented relatively low scalability compared with existing nonvolatile memory devices such as FLASH so far, we strongly believe that mechanically operated memory devices (also logic devices) have promising potential and may compete with conventional memory architectures by overcoming the technical challenges.

5 Power Applications

5.1 Introduction

In the field of power electronics, a switching device is one of the most important components because it is utilized in tremendous applications such as measurement equipment, batteries, and portable devices. Basically, power switches are divided into two categories, electromagnetic relays (EMRs) and solid-state relays (SSRs). The former (EMRs) is well known for its low on-resistance, high current capability, and low cost, whereas they are bulky and commonly require continuous coil current to hold the on-state. On the other hand, the latter (SSRs) has its own merits including high speed and easiness in on-chip integration. However, low-off resistance, polarity sensitivity, and generally high cost at high current are widely known demerits. Consequently, having their respective strengths and weaknesses, both switching devices separately keep up their market share depending on the specific applications.

Nowadays, these two main switching devices are facing several obstacles as current electronics trends are gradually changed. That is, as needs for the portable devices are increasing, the electronic device sizes are demanded to be minimized. At the same time, energy or power efficiency is also becoming very important, as inevitable environmental issues including environment-friendly green growth and restraining greenhouse gas emissions are on the rise. In this environment, the micro/nanosized mechanical switch can be a promising candidate for the switching device in power applications.

The miniaturized mechanical switches have various merits when they are applied to power switching applications. First, high power efficiency can be achieved. Electric devices based on semiconductor always have leakage currents such as subthreshold and gate leakages when the devices are at the off-state. However, mechanical structure of MEMS/NEMS switches does not suffer from any leakage current. Also, they are generally fabricated with highly conductive

metal, thus potentially affording much lower on-resistance. At the same time, they can operate with voltage driving method, thus power consumption is much lower than that of conventional EMRs which should flow the current continuously to keep "on-state". Also, batch fabrication of miniaturized mechanical switches can be realized by a silicon process, so they have additional advantages in terms of cost, size, and design in comparison to conventional EMRs. That is, keeping pace with the current electronics trends of miniaturization and high power efficiency, these strengths make MEMS/NEMS switches to be considered as one of next-generation switching devices.

5.2 Requirements for Power Switching

In terms of power switching, the major and important parameters are somewhat different from other applications such as RF, logic, and memory devices. First, power handling capability, which means the maximum input or output power that the mechanical switch can provide or handle without damage, should be maximized. However, the MEMS/NEMS contact switch is particularly vulnerable to this feature, because the concentrated heat at the contact spots is relatively high compared to the small effective or real contact area and micro-arc and material transfer easily occurs at high bias drain voltage, thus attacking the contact asperities [46].

On-resistance is also a critical factor since it is directly related to power efficiency. Thus, the contact resistance ought to be minimized, and the signal line (electrode) should have sufficient thickness and width to lower the total on-resistance. Electroplated structure is preferred to that from evaporation or sputter.

Next, another main parameter in power switching application is high stand-off voltage. It indicates a maximum drain-to-source voltage for the switch to withstand at the off-state. This value should be high for mechanical switches to be undamaged when high-voltage external noise or signal is suddenly (inevitably) applied between two contact parts at the off-state. In general, high voltage between the drain and source electrodes generates high electric field in the gap between two contact parts. This can cause the mechanical switch to fail by self-actuation (high electrostatic force between the drain and source electrodes causes the source or drain electrode to actuate) or by generating electrical breakdown of the air gap [36]. In any case, a sufficiently wide air gap is required to have high stand-off voltage in accordance with the Paschen curve [123–125].

Lastly, a certain level of lifetime at high power (current) condition should be definitely assured. Generally, the contact endurance of MEMS/NEMS switches under high current is relatively poorer than that of EMRs and SSRs (see Fig. 14). This is the primary reason why MEMS/NEMS switches are difficult to be exploited in power applications. Thus, current researches for the mechanical power switch are mainly concentrating on enhancing the contact reliability at high power levels.

5.3 Research Flow

In 1998, a fully integrated magnetically actuated MEMS switch (relay) was developed (Fig. 29a) [126]. By fabricating all structures of the device with micromachined technique, the switch does not require any assembly of a coil (batch fabrication is possible) and therefore it has smaller total size in contrast to other EMRs. Also, actuation power could be reduced to 33 mW through minimizing the coil design. The device demonstrated a current carrying capability of 1.2 A and a contact resistance of 22 mΩ. Although the device is magnetically operated, it has opened and closed at a frequency of 1 kHz owing to the relatively small switch dimension (the upper plate is 1.95 mm × 3.5 mm). Then, Wong et al. suggested an electrostatically actuated MEMS switch for power applications, in 2000 (Fig. 29b) [127]. They tried to employ the MEMS switch for the automotive vehicle. The device is operated with either electrostatic (20 V) or pneumatic (1200 Pa) actuations and showed low contact resistance of 14 mΩ and a current carrying capability of 400 mA. Although there were great efforts to use MEMS switches for telecommunication applications, these two works have strong meaning as the first attempt to apply the MEMS switch to power switching applications.

Following these works, Delphi lab has fabricated a copper blade cantilever-type micro relay with electrostatic actuation in 2002 [128]. The cantilever-type switch showed a relatively large size of 1.88 mm × 1 mm and Cu-to-Cu contact was utilized. Using the sectioned blade, as presented in Fig. 30a, the device can obtain better flexibility when making contact and thereby have multiple contact points. The device demonstrated switching speed of 50 μs and current capability of 1 A. It is also operated for more than 1.7 × 10^6 cycles at 50 mA of the current level. Furthermore, they employed arc-suppression circuit to enhance the power handling capability for the first time (Fig. 30b, c) [67]. By doing so, the relay could be

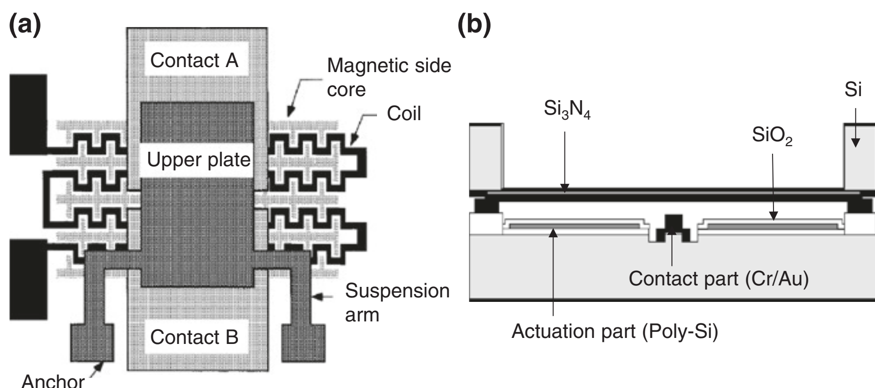


Fig. 29 Schematic illustrations of **a** the magnetically actuated MEMS relay [126] (Reproduced from Ref. [126] with permission from IEEE) and **b** the electrostatically (or pneumatically) actuated MEMS relay [127] (Reproduced from Ref. [126] with permission from IEEE)

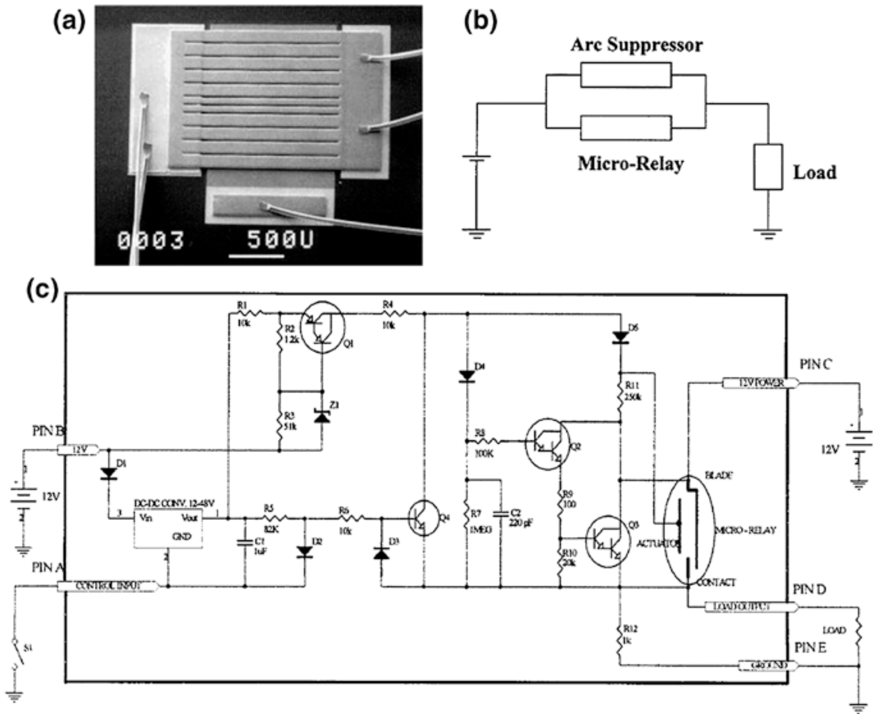


Fig. 30 Delphi lab MEMS switch. **a** SEM image. **b** Functional block diagram of the arc-suppression circuit. **c** Bipolar version of an arc-suppression circuit [67] (Reproduced from Ref. [67] with permission from IEEE)

protected from the arcing during relay "opening" and "closing" moments, thus the lifetime at 0.35 A was enhanced from 20 to 2000 cycles. Although the results are not sufficient to achieve their goal (specifications) for automotive application, they demonstrated relatively high reliability and power handling capability compared to previous MEMS switches.

A thermally actuated MEMS relay (switch) with bistable operation was developed in 2005 [129]. Because it is fabricated with bulk-micromachined technique (silicon wafer is etched by DRIE) and thermally actuated with high contact force, stable contact, high current carrying capability and low contact resistance can be achieved. As shown in Fig. 31a, when a thermal actuator #1 is bent downward, the mechanically pre-curved double beam snaps toward its second stable position, thus the movable crossbar makes contact with two contact bumps simultaneously. The switch demonstrated 2–3 A of current carrying capability and 60 mΩ of on-resistance. In terms of power efficiency and switching speed, thermal actuation is generally inferior to the electrostatic actuation. However, power consumption can be much reduced by bistable operation and its features of high contact force and large displacement definitely make it possible to have strong advantages in high power handling and stand-off voltage.

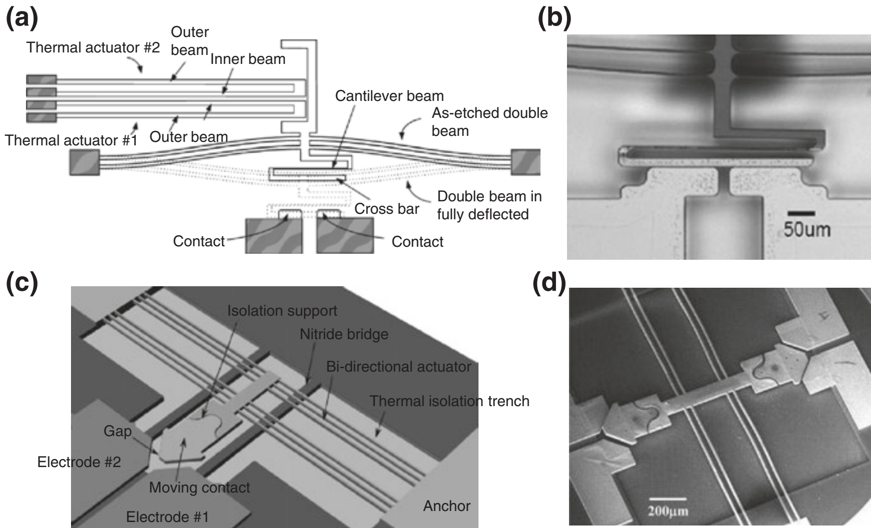


Fig. 31 **a** Schematic illustration and **b** SEM image of the thermally actuated power MEMS switch [129]. (Reproduced from Ref. [129] with permission from IEEE) **c** Schematic illustration and **d** SEM image of the bidirectional electrothermal and electromagnetic power MEMS switch [135] (Reproduced from Ref. [135] with permission from IEEE)

Also, Cao et al. developed an electrothermal and electromagnetic MEMS relay with liquid metal wetted contacts (Fig. 31c). Here, to lower the contact resistance and increase the power handling capability, a new method which fills the air gap with liquid metal (gallium alloy, melting point of −20 °C) when making contact was suggested. In order to do so, the surfaces of the contact are coated with liquid metal using the Galistan fluid deposition process. By doing so, the contact resistance is reduced as low as 0.015 Ω, and over 4–5 A of current can flow. However, the switching devices with the liquid metal contact eventually failed by "bridging," which means the liquid metal electrically connects between two contact parts even at the off-state, after ten switching cycles under high current condition, thus further research to address the issue is required.

Song et al. developed an electrostatically actuated MEMS relay suitable for power switching application in 2012 [36]. As presented in Fig. 32a, b, the device is based on the stacked-electrode structure for extremely low contact resistance under 5 mΩ (refer to Sect. 2.3). Furthermore, when a voltage is applied to the switch-off electrode at the off-state, the source plate is lifted up to form a large air gap of 10 μm by means of levering and torsional spring (Fig. 32c), thus resulting in high stand-off voltage of 360 V, which is the highest level among the electrostatically actuated MEMS relays. Also, for the first time, they demonstrated "resurrection" of the switch (when the device failed by permanent contact stiction, it can be revived by applying a voltage to the switch-off voltage to detach two contact parts each other), bring about a tenfold increase in its lifetime. The device

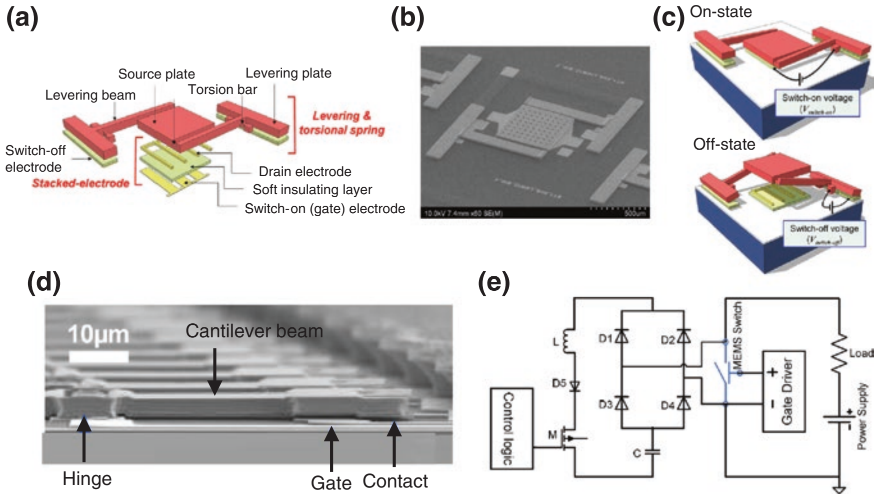


Fig. 32 Electrostatically actuated power MEMS switches. **a** Schematic illustration. **b** SEM image, and **c** operation principle of the MEMS relay from KAIST [36]. **d** SEM image and **e** arc-suppression circuit of the device from GE [130] (Reproduced from Ref. [130] with permission from IEEE)

showed lifetime of 5×10^5 cycles at 200 mA of current level in a hot switching condition and current carrying capability over 1 A.

GE (general electric) has developed a MEMS switch which possesses an array of microscale cantilever relays (beam width and length is about 50 μm) to overcome the power density limitation of a single relay [130, 131]. As expected, the multiple cantilever switches in parallel (up to 20 cantilever beam) can enhance the current carrying capability very much. Also, an arc-suppressor circuit consisting of diode bridges was used to open and close the contacts without contact surface degradation. By doing so, power handling capability close to 1 kW (60 V–13 A) was achieved. It demonstrated that the MEMS switch can be a revolutionary power switching device in low-to-medium power level by presenting dramatically higher power handling capability in comparison to other MEMS contact switches.

As can be seen in Table 3, lifetime at high current level and power handling capability are still lower than those of the conventional EMRs and SSRs, and therefore these parameters should be first enhanced in order to commercialize the MEMS switch for power applications. We believe that low contact resistance to generate less heat at the contact parts and appropriate heat sink to minimize heat are important for it. At the same time, studying contact materials which make it possible to obtain high current carrying capability is critical, and an additional arc-suppressor circuit is also necessary. As presented, technology development of the micro/nanomechanical switches for power switching applications looks slightly slow. However, we think that as the reliability and power handling capability are being enhanced, the possibility to apply the MEMS switch to power switching will be dramatically increasing, thus creating various potential applications.

Table 3 MEMS switches for power applications

	Actuation mechanism	Contact resistance (mΩ)	Stand-off voltage (V)	Power handling	Lifetime (hot switching)
UC Berkeley [135]	Thermal	15	200	4–5 A	10^3@10 V–200 mA
MIT (Wong J) [127]	Electrostatic pneumatic	14	300	400 mA	4×10^3@400 mA
Georgia Tech [126]	Magnetic	22.4	–	1.2 A	8.5×10^5@50 mA
KAIST [36]	Electrostatic	<5	360	1 A	5×10^5@200 mA
MIT (Qui J) [129]	Thermal	60	–	2–3 A	–
GE [130, 131]	Electrostatic	–	100–300	780 W	–

6 Conclusion

In this chapter, we introduced the miniaturized mechanical contact switch based on MEMS/NEMS technology with special emphasis on future logic, memory, and power switching applications. It has both advantages of the conventional magnetic relay and silicon-based switching device, including ideally zero leakage current, perfectly abrupt switching behavior, and insensitiveness to radiation and temperature, thus being considered as a powerful candidate for next-generation switching device for ultralow power applications. Despite the noteworthy progress that has been made with MEMS/NEMS switches, there are still significant challenges that need to be addressed in terms of the contact reliability, actuation voltage, switching speed, as well as the power handling capability, according to the respective applications. For overcoming the challenges, we need to make more effort to investigate fundamental solutions in terms of structural, circuitry, and material approaches. However, we believe that extensive ongoing research will gradually address weaknesses and issues of MEMS/NEMS switches and these efforts will eventually revolutionize the existing logic gate, memory device, and power switching element with the MEMS/NEMS switches.

Acknowledgments This work is supported by the Center for Integrated Smart Sensors (CISS) funded by the Ministry of Science, ICT & Future Planning as the Global Frontier Project. The authors would like to thank Kyeonghun Kim for his assistance.

References

1. http://telephonetribute.com/switches_survey_intro_chapter_1.html
2. Baker RJ (1998) CMOS: circuit design, layout and simulation. IEEE Press, New York
3. Sakurai T (2004) Perspectives of low-power VLSI's. IEICE Trans Electron 87(4):429–436

4. Davari B, Dennard RH, Shahidi GG (1995) CMOS scaling for high performance and low power-the next ten years. Proc IEEE 83(4):595–606
5. Nowak EJ (2002) Maintaining the benefits of CMOS scaling when scaling bogs down. IBM J Res Develop 46(2.3):169–186. doi:10.1147/rd.462.0169
6. Petersen KE (1978) Dynamic micromechanics on silicon: techniques and devices. IEEE Trans Electron Devices 25(10):1241–1250. doi:10.1109/T-ED.1978.19259
7. Receveur RAM, Marxer CR, Woering R et al (2005) Laterally moving bistable MEMS DC switch for biomedical biomedical applications. J Microelectromech Syst 14(5):1089–1098 doi:10.1109/JMEMS.2005.851843
8. Wang Y, Li Z, McCormick DT et al (2004) A low-voltage lateral MEMS switch with high RF performance. J Microelectromech Syst 13(6):902–911. doi:10.1109/JMEMS.2004.838395
9. Ruan M, Shen J, Wheeler CB (2001) Latching micro magnetic relays with multi-strip permalloy cantilevers. In: Proceedings of the IEEE international conference MEMS, Interlaken, Switzerland, pp 224–227 Jan 2001
10. Lee HC, Park JY, Bu JU (2005) Piezoelectrically actuated RF MEMS DC contact switches with low voltage operation. IEEE Microw Wirel Compon Lett 15(4):202–204. doi:10.1109/LMWC.2005.845689
11. Rebeiz GM (2003) RF MEMS: theory, design, and technology. Libre Digital. Wiley, New York
12. Senturia SD (2001) Microsystem design. Kluwer Academic Publishers, Boston
13. Nathanson HC, Newell WE, Wickstrom RA et al (1967) The resonant gate transistor. IEEE Trans Electron Devices 14(3):117–133. doi:10.1109/T-ED.1967.15912
14. Rezvanian O, Zikry MA, Brown C et al (2007) Surface roughness, asperity contact and gold RF MEMS switch behavior. J Micromech Microeng 17(10):2006–2015. doi:10.1088/0960-1317/17/10/012
15. Jensen B, Chow L, Huang K et al (2005) Effect of nanoscale heating on electrical transport in RF MEMS switch contacts. J Microelectromech Syst 14(5):935–946. doi:10.1109/JMEMS.2005.856653
16. Johnson KL (1985) Contact mechanics. Cambridge University Press, Cambridge
17. Holm R (1967) Electric contacts: theory and application, 4th edn. Springer, Heidelberg
18. Greenwood JA, Williamson JBP (1966) Contact of nominally flat surfaces. Proc R Soc Lond A, 295:219–232
19. Chang WR, Etison I, Bogy D (1987) An elastic-plastic model for the contact of rough surfaces. ASME J Tribol 109(2):257–263. doi:10.1115/1.3261348
20. Stronge WJ (2000) Impact mechanics. Cambridge University Press, Cambridge
21. Slade PG (1999) Electrical contacts, principles and applications. Horseheads, New York
22. Jansen AGM, van Gelder AP, Wyder P (1980) Point-contact spectroscopy in metals. J Phys C: Solid State Phys 13(33):6073–6118. doi:10.1088/0022-3719/13/33/009
23. Majumder S, McGruer NE, Adams GG et al (2001) Study of contacts in an electrostatically actuated microswitch. Sens Actuators, A 93(1):19–26. doi:10.1016/S0924-4247(01)00627-6
24. Jones LF (1957) The physics of electrical contacts. Clarendon Press, Oxford
25. Brown ER (1998) RF-MEMS switches for reconfigurable integrated circuits. IEEE Trans Microw Theory Tech 46(11):1868–1880. doi:10.1109/22.734501
26. Rebeiz GM, Muldavin JB (2001) RF MEMS switches and switch circuits. IEEE Microw Mag 2:59–71
27. Tang WC, Nguyen TH, Howe RT (1989) Laterally driven polysilicon resonant microstructures. Sens Actuators 20:25–32. doi:10.1109/MEMSYS.1989.77961
28. Yeh R, Hollar S, Pister KSJ (2002) Single mask, large force, and large displacement electrostatic linear inchworm motors. J Microelectromech Syst 11(4):330–336. doi:10.1109/JMEMS.2002.800937
29. Nathanael R, Pott V, Kam H et al (2009) 4-terminal relay technology for complementary logic. In: Proceedings of the international electron devices meeting, 2009 IEEE international, vol 1. Baltimore, MD, p 4 Dec 2009

30. Hah D, Yoon E, Hong S (2000) A low voltage actuated micromachined microwave switch using torsion springs and leverage. IEEE MTT-S Int Microw Symp Dig 48(12):157, 160
31. Tilmans HAC, Ziad H, Jansen H et al (2001) Wafer-level packaged RF-MEMS switches fabricated in a CMOS fab. In: IEEE international electron devices meeting, IEDM '01. Technical DIGEST. International, Washington, USA, 41.4.1–41.4.4, Dec 2011
32. Chang C, Chang P (2000) Innovative micromachined microwave switch with very low insertion loss. Sens Actuators, A: Phys 79(1):71–75. doi:10.1016/S0924-4247(99)00218-6
33. Song YH, Choi DH, Yang HH et al (2011) An extremely low contact-resistance MEMS relay using meshed drain structure and soft insulating layer. J Microelectromech Syst 20(1):204–212. doi:10.1109/JMEMS.2010.2100025
34. Majumder S, Lampen J, Morrison R et al (2003) MEMS switches. IEEE Instrum Meas Mag 6(1):12–15. doi:10.1109/MIM.2003.1184267
35. DeSilva AP, Vaughan C, Frear D et al (2001) Motorola MEMS switch technology for high frequency applications. In: IEEE Proceedings of the MEMS conference, Berkeley, CA, pp 22–24, Aug 2001
36. Song YH, Han CH, Kim MW et al (2012) An electrostatically actuated stacked-electrode MEMS relay with a levering and torsional spring for power applications. J Microelectromech Syst 21(5):1209–1217. doi:10.1109/JMEMS.2012.2198046
37. Becher D, Chan R, Hattendorf M et al (2002) Reliability study of low-voltage RF NEMS switches. In: Proceedings of the GaAs MANTECH conference, University of Illinois, USA, pp 54–57, 2002
38. Chan R, Lesnick R, Becher D (2003) Low-actuation voltage RF MEMS shunt switch with cold switching lifetime of seven billion cycles. J Microelectromech Syst 12(5):713–719. doi: 10.1109/JMEMS.2003.817889
39. Kim MW, Song YH, Yang HH et al (2013) An ultralow voltage MEMS switch using stiction-recovery actuation. J Micromech Microeng 23(4):7. doi:10.1088/0960-1317/23/4/045022
40. Kim MW, Song YH, Ko SD et al (2014) Ultra-low voltage MEMS switch using a folded hinge structure. Micro Nano Syst Lett 2:2. doi:10.1186/s40486-014-0002-y
41. Yang HH, Lee JO, Yoon JB (2007) Maneuvering Pull-in voltage of an electrostatic micro-switch by introducing a pre-charged electrode. IEDM technical digest, Washington, DC, pp 439–442, Dec 2007
42. Yang HH, Han CH, Lee JO (2014) Electrostatic micro-actuator with a pre-charged series capacitor: modeling, design, and demonstration. J Micromech Microeng 24(6). doi:10.1088/0960-1317/24/6/065012
43. Collins RE (1977) Measurement of charge distribution in electrets. Rev Sci Instrum 48:83–91
44. Kressmann R, Sesslerand GM, Gunther P (1996) Spacecharge electrets. IEEE Trans Dielectr Electr Insul 3(5):607–623. doi:10.1109/94.544184
45. Ko SD, Lee JO, Yang HH et al (2011) An insulating liquid environment for reducing adhesion in a microelectromechanical system. Appl Phys Lett 99(11):113516. http://dx.doi.org/+10.1063/1.3640228
46. Hyman D, Mehregany M (1999) Contact physics of gold microcontacts for MEMS switches. IEEE Trans Compon Packag Technol 22(3):357–364. doi:10.1109/6144.796533
47. Majumder S, Lampen J, Morrison R (2003) A packaged, high-lifetime ohmic MEMS RF switch. IEEE MTT-S Int Microw Symp Dig 3:1935–1938 (Philadelphia, PA, USA, June 2003)
48. Newman HS, Ebel JL, Judy D (2008) Lifetime measurements on a high-reliability RF-MEMS contact switch. IEEE Microw Wireless Compon Lett 18(2):100–102. doi:10.1109/LMWC.2007.915037
49. Seki T, Uno Y, Narise K (2006) Development of a large-force low-loss metal-contact RF MEMS switch. Sens Actuators, A: Phys 132(2):683–688
50. Pott V, Kam H, Nathanael R (2010) Mechanical computing redux: relays for integrated circuit applications. Proc IEEE 98(12):2076–2094. doi:10.1109/JPROC.2010.2063411

51. Toler BF, Coutu RA Jr, McBride JW (2013) A review of micro-contact physics for micro-electromechanical systems (MEMS) metal contact switches. J Micromech Microeng 23(10). doi:10.1088/0960-1317/23/10/103001
52. Hermance HW, Egan TF (1958) Organic deposits on precious metal contacts. Bell Syst Tech J 37:739–775
53. Stark B (1999) MEMS reliability assurance guidelines for space applications. Jet Propulsion Laboratory, California JPL Publication 99-1, Pasadena
54. Yang Z, Lichtenwalner D, Morris A (2010) Contact degradation in hot/cold operation of direct contact micro-switches. J Micromech Microeng 20(10):1–8. doi:10.1088/0960-1317/20/10/105028
55. Yang Z (2008) Contact material optimization and contact physics in metal-contact Microelectromechanical Systems (MEMS) switches. Ph.D. dissertation, repository. lib. ncsu. edu
56. Poulain C, Peschot A, Vincent M (2011) A nano-scale investigation of material transfer phenomena at make in a MEMS switch. In: Proceedings of the IEEE 57th holm conference on electrical contacts, Minneapolis, MN, USA, 1–7 Sept 2011
57. Chen ZK, Sawa K (1988) Effect of arc behavior on materials transfer: a review. IEEE Trans Comp, Packag Manufact Technol A 21:310–322
58. Adams GG, McGruer NE (2010) A review of adhesion in an ohmic microswitch. J Adhes Sci Technol 24(15–16):2571–2595. doi:10.1163/016942410X508154
59. Patton ST, Zabinski JS (2005) Fundamental studies of Au contacts in MEMS RF switches. Tribol Lett 18(2):215–230. doi:10.1007/s11249-004-1778-3
60. Nishijima N, Hung JJ, Rebeiz GM (2004) Parallel-contact metal-contact RF MEMS switches for high power applications. MEMS, 2004. 17th IEEE Int Microelectromech Syst Conf 781–784
61. Lampen J, Majumder S, Morrison R et al (2004) A wafer-capped, high-lifetime ohmic MEMS RF switch. Int J RF Microw Comput Aided Eng 14(4):338–344. doi:10.1002/mmce.20020
62. Chow LLW, Volakis JL, Saitou K (2007) Lifetime extension of RF MEMS direct contact switches in hot switching operations by ball grid array dimple design. IEEE Electron Device Lett 28(6):479–481. doi:10.1109/LED.2007.896811
63. Korotocov A, Huang YS, Tsai TY et al (2006) Effect of length, spacing and morphology of vertically aligned RuO_2 nano-structures on field-emission properties. Nanotechnology 17(13):3149–3153. doi:10.1088/0957-4484/17/13/012
64. Oberhammer J (2004) Novel RF MEMS switch and packaging concepts. Ph.D. dissertation. Roy Inst Technol, Stockholm, Sweden
65. Patel CD, Rebeiz GM (2011) RF MEMS metal-contact switches with mN-contact and restoring forces and low process sensitivity. IEEE Trans Microw Theory Tech 59(5):1230–1237. doi:10.1109/TMTT.2010.2097693
66. Patel C, Rebeiz GM (2012) A high-reliability high-linearity high-power RF MEMS metal-contact switch for DC-40-GHz applications. IEEE Trans Microw Theory Technol 60(10):3096–3112. doi:10.1109/TMTT.2012.2211888
67. Lee HS, Leung CH, Shi J et al (2002) Integrated microrelays: concept and initial results. J Microelectromech Syst 11(2):147–153. doi:10.1109/84.993449
68. Song YH, Kim MW, Joon Ann S et al (2013) Improvement of hot switching lifetime in MEMS DC switches using a drain voltage-sustaining capacitor. Solid-state sensors, actuators and microsystems (TRANSDUCERS & EUROSENSORS XXVII), 2013 transducers & Eurosensors XXVII: The 17th international conference on, Barcelona, pp 546–549, June 2013
69. Coutu RA, Reid JR et al (2006) Microswitches with sputtered Au, AuPd, Au-on-AuPt, and AuPtCu alloy electric contacts. IEEE Trans Compon Packag Technol 29(2):341–349. doi:10.1109/TCAPT.2006.875898
70. Schimkat J (1998) Contact materials for microrelays, MEMS 98. In: Proceedings, The Eleventh annual international workshop on Heidelberg, Germany, pp 190–194, Jan 1988

71. Hiltmann K, Schumacher A, Guttmann K et al (2001) New micromachined membrane switches in silicon technology. IEEE Trans Compon Packag Technol 25(3):397–401. doi:10.1109/TCAPT.2002.804611
72. Hiltmann K, Keller W, Lang W (1999) Micromachined switches for low electric loads. Sens Actuators, A: Phys 74(1):203–206
73. Coutu RA (2004) Electrostatic radio frequency (RF) micro-electromechnical systems (MEMS) switches with metal alloy electric contacts. Ph.D. dissertation, Air Force Inst Technol, Wright-Patterson AFB, OH
74. W. Jahn, (1990) Contact metals for reed switches. In: Proceedings of the 36th IEEE holm conference electron contacts, pp 53–58, Aug 1990
75. Ke F, Miao J, Oberhmmer J (2008) A ruthenium-based multimetalcontact RF MEMS switch with a corrugated diaphragm. J Microelectromech Syst 17(6):1447–1459
76. Schimkat J (1999) Contact measurements providing basic design data for microrelay actuators. Sens Actuators, A: Phys 73(1):138–143. doi:10.1016/S0924-4247(98)00264-7
77. Read MB, Lang JH, Slocum AH (2009) Contact resistance in flat thin films. In: Proceedings of 55th IEEE holm conference electron contacts, pp 303–309, July 2009
78. Abbott WH, Campbell WE (1978) Frictional polymer formation on precious metals—experimental observations. In: Proceedings of the 9th international electron contact phenomena conference, Chicago, IL, pp 359–363
79. Czaplewski DA, Nordquist CD, Patrizi GA et al (2013) RF MEMS switches with RuO2–Au contacts cycled to 10 billion cycles. J Microelectromech Syst 22(3):655–661
80. Kwon H, Park JH, Lee HC et al (2008) Investigation of similar and dissimilar metal contacts for reliable radio frequency micorelectromechanical switches. J Appl Phys 47(8):6558–6562. doi:10.1143/JJAP.47.6558
81. Liu B, Lv Z, He X et al (2011) Improving performance of the metal-to-metal contact RF MEMS switch with a Pt–Au microspring contact design. J Micromech Microeng 21(6):065038-1–065038-9. doi:10.1088/0960-1317/21/6/065038
82. Song YH, Kim MW, Lee JO (2013) Complementary dual-contact switch using soft and hard contact materials for achieving low contact resistance and high reliability simultaneously. J Microelectromech Syst 22(4):846–854. doi:10.1109/JMEMS.2013.2248125
83. Song YH, Kim MW, Seo MH et al (2013) A complementary dual-contact MEMS switch using a zipping technique. J. Microelectromech Syst 23(3):710–718. doi:10.1109/JMEMS.2013.2281835
84. http://www.radantmems.com/radantmems/switchoverview.html
85. Sakata M, Komura Y, Seki T et al (1999) Micromachined relay which utilizes single crystal silicon electrostatic actuator. In: Proceedings of the 12th IEEE international conference microelectromechanical systems MEMS', 99, Orlando, FL, USA, pp 21–24
86. Dadgour HF, Hussain MM, Banerjee K (2010) A new paradigm in the design of energy-efficient digital circuits using laterally-actuated double-gate NEMs. In: Proceedings of the 2010 ACM/IEEE international symposium on low power electron description, New York, NY, USA, pp 7–12
87. Dadgour HF, Hussain MM, Smith C et al (2010) Design and analysis of compact ultra energy-efficient logic gates using laterally-actuated double-electrode NEMS. In: Proceedings of the 2010 ACM/IEEE design automation conference Anaheim, CA, pp 893–896, June 2010
88. Choi SJ, Ahn JH, Han JW et al (2011) Transformable functional nanoscale building blocks with wafer-scale silicon nanowires. Nano Lett 11(2):854–859. doi:10.1021/nl104212e
89. Veloso A, Hoffmann T, Lauwers A et al (2007) Advanced CMOS device technologies for 45 nm node and below. Sci Technol Adv Mater 8(3):214–218. doi:10.1016/j.stam.2006.11.018
90. Hisamoto D, Lee WC, Kedzierski J et al (2000) FinFET a self aligned double gate mosfet scalable to 20 nm. IEEE Trans Electron Devices 47(12):2320–2325. doi:10.1109/16.887014

91. Loh OY, Espinosa HD (2012) Nanoelectromechanical contact switches. Nat Nanotechnol 7(5):283–295. doi:10.1038/Nnano.2012.40
92. Rueckes T, Kim K, Joselevich E et al (2000) Carbon nanotube-based nonvolatile random access memory for molecular computing. Science 289:94–97. doi:10.1126/science.289.5476.94
93. Kaul AB, Wong EW, Epp L et al (2006) Electromechanical carbon nanotube switches for high-frequency applications. Nano Lett 6(5):942–947. doi;10.1021/nl052552r
94. Dujardin E, Derycke V, Goffman MF et al (2005) Self-assembled switches based on electroactuated multiwalled nanotubes. Appl Phys Lett 87:193107
95. Cha SN, Jang JE, Choi Y et al (2005) Fabrication of a nanoelectromechanical switch using a suspended carbon nanotube. Appl Phys Lett 86:083105
96. Lee SW, Park SJ et al (2011) A fast and low-power microelectromechanical system-based non-volatile memory device. Nat Comm 2 doi:10.1038/ncomms1227
97. Jang JE, Cha SN, Choi YJ et al (2008) Nanoscale memory cell based on a nanoelectromechanical switched capacitor. Nat Nano 3:26–30. doi:10.1038/nnano.2007.417
98. Jang WW, Lee JO, Yoon JB et al (2008) Fabrication and characterization of a nanoelectromechanical switch with 15 nm thick suspension air gap. Appl Phys Lett 92:103110
99. Lee JO, Kim MW, Ko SD et al (2009) 3-terminal nanoelectromechanical switching device in insulating liquid media for low voltage operation and reliability improvement. IEDM Tech Dig 621–624
100. Xiang W, Lee C (2010) Nanoelectromechanical torsion switch of low operation voltage for nonvolatile memory applications. Appl Phys Lett 96:193113
101. Qian Y, Lou L, Tsai MJ et al (2012) A dual-silicon-nanowires based U-shape nanoelectromechanical switch with low pull-in voltage. Appl Phys Lett 100:113102
102. Lee JO, Song YH, Kim MW et al (2012) A-sub-1-volt-nanoelectromechanical-switching-device. Nat Nanotechnol 8:36–40. doi:10.1038/nnano.2012.208
103. Lee TH, Bhunia S, Mehregany M (2010) Electromechanical Computing at 500 °C with Silicon Carbide. Science 329(5997):1316–1318. doi:10.1126/science.1192511
104. Goetz J (2000) Sens Mag 17:20
105. Feng XL, Matheny MH, Zorman CA et al (2010) Low voltage nanoelectromechanical switches based on silicon carbide nanowires. Nano Lett 10(8):2891–2896. doi:10.1021/nl1009734
106. Han JW, Ahn JH, Kim MW et al (2010) Nanowire mechanical switch with a built-in diode. Small 6(11):1197–1200
107. Lee BH, Moon DI, Jang HJ et al (2014) A mechanical and electrical transistor structure (METS) with a sub-2 nm nanogap for effective voltage scaling. Nanoscale 7799–7804. doi:10.1039/C3NR06251A
108. Zaghloul U, Piazza G (2014) Sub-1-volt piezoelectric nanoelectromechanical relays with millivolt switching capability. IEEE Electron Device Lett 35(6):669–671. doi:10.1109/LED.2014.2318049
109. Moore GE (1995) Lithography and the future of Moore', s law. Proc SPIE 2437:2–17. doi:10.1117/12.209244
110. Zeitzoff PM, Chung JE (2002) Weighing in on digital logic scaling trends. IEEE Circ Devices Mag 18(2):18–27. doi:10.1109/101.994855
111. Chandrakasan AP, Sheng S, Brodersen RW (1992) Low-power CMOS digital design. IEEE J Solid State Circ E75-C(4):371–382
112. Akarvardar K, Elata D, Parsa R et al (2007) Design considerations for complementary nanoelectromechanical logic gates. IEDM 2007. IEEE international, Washington, DC, pp 299–302, Dec 2007
113. Ionescu AM, Pott V, Fritschi R et al (2002) Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal over-gate architecture. In: Proceedings of the international symposium on quality electronic design, pp 496–501

114. Kam H, Lee DT, Howe RT et al (2005) A new nano-electro-mechanical field effect transistor (NEMFET) design for low-power electronics. IEDM technical digest, Washington, DC, pp 463–466, Dec 2005
115. Akarvardar K, Eggimann C, Tsamados D et al (2008) Analytical modeling of the suspended-gate FET and design insights for low-power logia. IEEE Trans Electron Devices 55(1):48–59. doi:10.1109/TED.2007.911070
116. Kim H, Chen ZCY, Kwon S et al (2014) Three-terminal nanoelectromechanical field effect transistor with abrupt subthreshold slope. Nano Lett 14(3):1687–1691. doi:10.1021/nl5006355
117. Kam H, Nathanael R, Jeon J et al (2009) Design and reliability of a micro-relay technology for zero-standby-power digital logic applications. In: Proceedings of the IEEE international electron device meeting technical digest, Baltimore, MD, 1–4 Dec 2009
118. Sinha N, Jones TS, Guo A et al (2012) Body-biased complementary logic implemented using AlN piezoelectric MEMS switches. J. Microelectromech Syst 21(2):484–496. doi:10.1109/JMEMS.2011.2179015
119. Halg B (1990) On a micro-electro-mechanical nonvolatile memory cell. IEEE Trans Electron Devices 37(10):2230–2236. doi:10.1109/16.59913
120. Beunder MA, Kampen R, Lacey D et al (2005) A new embedded NVM technology for low-power, high temperature, rad-hard applications. In: Proceedings non-volatile memory technology symposium, Dallas, TX, pp 65–68, Nov 2005
121. Smith CG, Kampen R, Popp J et al (2007) Nano-mechanical cantilever arrays for low-power and low-voltage embedded nonvolatile memory applications. Proc SPIE 6464:646406. doi:10.1117/12.698964
122. Jang WW, Lee JO, Yang HH et al (2008) Mechanically operated random access memory (MORAM) based on an electrostatic microswitch for nonvolatile memory applications. IEEE Trans Electron Devices 55(10):2785–2789. doi:10.1109/TED.2008.2003052
123. Paschen F (1889) Ueber die zum Funkenübergang in Luft, Wasserstoff und Kohlensäure bei verschiedenen Drucken erforderliche Potentialdifferenz. Ann Phys 273(5):69–96
124. Townsend JS (1910) The theory of ionization of gases by collision. D. Van Nostrand Company, London Constable & Co. Ltd
125. Kunbardt EE, Luessen LH (1981) Electrical breakdown and discharges in gases—part A, fundamental processes and breakdown, Plenum Press, New York
126. Taylor WP, Allen MG, Dauwalter CR (1998) fully integrated magnetically actuated micromachined relays. JMEMS 7(2):181–191. doi:10.1109/84.679353
127. Wong JE, Lang JH et al (2000) An electrostatically-actuated MEMS switch for power applications. MEMS 2000 the thirteenth annual international conference on, Miyazaki, pp 633–638, Jan 2000
128. Lee H, Leung C, Shi J et al (2002) Electrostatically actuated copper-blade microrelays. Sens Actuators, A: Phys 100(1):105–113
129. Qui J, Lang JH, Slocum AH et al (2005) A bulk-micromachined bistable relay with U-shaped thermal actuators. J Microelectromech Syst 14(5):1099–1109. doi:10.1109/JMEMS.2005.856676
130. Li B, Keimel C, Claydon G et al (2011) Power switch system based on microelectro mechanical switch. In: Digest technical papers transducers'11 conference, Beijing, pp 675–678, June 2011
131. Keimel C, Claydon G, Li B et al (2011) Microelectromechanical-system (MEMS) based switches for power applications. In: IEEE industrial and commercial power systems technical conference 48(4):1163–1169. doi:10.1109/TIA.2012.2199949
132. Lee KB (2007) Closed-form solutions of the parallel plate problem. Sens Actuators, A 133(2):518–525
133. Gilbert KW (2009) Investigation into contact resistance and damage of metal contacts used in RF-MEMS switches. Accession Number: ADA511098

134. Kwon H, Choi DJ, Park JH (2007) Contact materials and reliability for high power RF-MEMS switches. In: Proceedings of the IEEE 20th international conference micro electro mechanical systems, Hyogo, pp 231–234
135. Cao A, Yuen P, Lin L (2007) Microrelays with bidirectional electrothermal electromagnetic actuators and liquid metal wetted contacts. J Microelectromech Syst 16(3):700–708. doi:10.1109/JMEMS.2007.893520
136. Majumder S, McGruer NE, Adams GG (2005) Adhesion and contact resistance in an electrostatic MEMS microswitch. In: Proceedings of the 18th IEEE international conference on micro electro mechanical systems, MEMS 2005, pp 215–218, Jan 2005
137. Majumder S, McGruer NE, Adams GG et al (2001) Study of contacts in an electrostatically actuated microswitch. Sens Actuators 93(1):19–26. doi:10.1016/S0924-4247(01)00627-6
138. Yang Z, Lichtenwalner DJ, Morris AS et al (2009) Comparison of Au and Au-Ni Alloys as contact materials for MEMS switches. J Microelectromech Syst 18(2):287–295. doi:10.1109/JMEMS.2008.2010850
139. Ma Q, Tran Q, Chou TKA et al (2007) Metal contact reliability of RF MEMS switches. Proc SPIE 6463:646–305. doi:10.1117/12.702177
140. Walker M, Nordquist C, Czaplewski D et al (2010) Impact of in situ oxygen plasma cleaning on the resistance of Ru and Au-Ru based RF microelectromechanical system contacts in vacuum. J Appl Phys 107(8):084509-1–084509-7
141. Chen Y, Nathanael R, Jeon JS et al (2012) Characterization of contact resistance stability in MEM relays with tungsten electrodes. J Microelectromech Syst 21(3):511–513. doi:10.1109/JMEMS.2012.2186282
142. de Boer MP, Czaplewski DA, Baker MS et al (2012) Design, fabrication, performance and reliability of Pt- and RuO_2-coated microrelays tested in ultrahigh purity gas environments. J Micromech Microeng 22(10):105027-1–105027-15. doi:10.1088/0960-1317/22/10/105027
143. Choi J, Lee JI, Eun Y et al (2011) Aligned carbon nanotube arrays for degradation-resistant, intimate contact in micromechanical devices. Adv Mater 23(19):2231–2236. doi:10.1002/adma.201100472

Part II
Systems and Circuits for Low Energy

Ultralow Power Processor Design with 3D IC Operating at Sub/Near-Threshold Voltages

Sandeep Samal and Sung Kyu Lim

Abstract The requirement for ultra low power and energy efficient systems is becoming more and more important with increase in the use of miniaturized portable devices and unsupervised remote sensor systems. Three-dimensional integration is an emerging technology which helps in reducing footprint as well as power. In this work, we carry out a detailed study of the combined benefits of 3D ICs and low-voltage supply designs to obtain maximum energy efficiency. We implement different types of circuits in conventional 2D and TSV-based 3D designs at different supply voltages varying from nominal to subthreshold voltages. The impact of 3D integration on these different types of circuits is analyzed. Our study is based on power and energy comparison of full GDSII layouts. Our study confirms that sub/near-threshold circuits indeed offer a few orders of magnitude power versus performance trade-off with further improvement due to 3D implementation. In addition, 3D designs reduce the footprint area up to 78 % and wire length up to 33 % compared with the 2D counterpart. Our studies also show that thermal and IR-drop issues are negligible in subthreshold 3D implementation due to its extreme low power operation. Lastly, we demonstrate the low-power and high-memory bandwidth advantages of many-core 3D subthreshold circuits.

Keywords 3D IC · Ultralow power (UVL) circuit · Subthreshold computing · Near-threshold computing · Smart sensor network · Computer-aided design (CAD)

S. Samal (✉) · S.K. Lim
Georgia Institute of Technology, Atlanta, GA, USA
e-mail: sandeep.samal@gatech.edu

S.K. Lim
e-mail: limsk@ece.gatech.edu

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_4

1 Introduction

One of the most effective ways of reducing the total power consumption of VLSI circuits is by scaling down the supply voltage. Previous works have shown that under optimum power supply voltage, a circuit can attain minimum energy consumption per operation, which is the primary goal for applications that require long battery life [16]. This supply voltage usually falls near the threshold voltage of the transistors. The digital gates can still function in subthreshold regions by utilizing subthreshold current, which has exponential dependence on gate voltage. Though the operating frequency reduces significantly, the energy efficiency improves and the longevity of battery is the major requirement for systems like sensor networks.

Low voltage and subthreshold operating conditions require the individual gates to be designed in a robust manner, which requires larger size of the transistors. For large designs, these result in area overhead over the nominal voltage designs. This results in a trade-off between extreme low power with good energy savings and silicon area. Since area is directly proportional to cost, it is desirable to have the same silicon usage but get low power and high energy efficiency. In full systems, the circuit may also need to interface with on-chip or off-chip memory, which further complicates low-voltage designs because reliable memory operation at low voltages requires additional area and design effort.

Three-dimensional ICs (3D ICs) are one of the most promising technologies that enable higher integration and further miniaturization with increase in memory bandwidth for off-chip memories, while reducing power dissipation and improving performance. Through-Silicon Vias (TSVs) have enabled individual dies to be stacked over one another reducing the long 2D wires into short 3D connections. Much shorter interconnects due to die stacking result in lower switching power dissipation compared to long interconnect designs. Memory stacked over logic integrates off-chip memories to on-chip. Studies have shown that a significant increase in memory bandwidth is obtained by moving to 3D [12]. The overall footprint of the design also reduces significantly further satisfying the needs of miniaturized designs. An important advantage of 3D integration is that it helps in integrating heterogeneous dies which use different technologies or have different supply voltages. There are many potential benefits of 3D stacking which, if used properly can help improve circuits and systems significantly. However, 3D ICs suffer from high power density leading to thermal problems and complicated power delivery issues, which have to be taken care of during design.

Through this work, we propose a new way to meet most of the requirements of low-power miniaturized designs using near/subthreshold circuits stacked with memory as well as mixed with memory in the same die. These result in low power and much smaller die footprint with increased memory bandwidth. The goal is to investigate and quantify the benefits and issues of using 3D integration of sub/near-threshold circuits using TSVs and identify what kind of ultralow power circuits can be improved by 3D design. Our studies are based on full GDSII

layouts and sign-off analysis of various types of designs, which vary from memory dominated to gate and interconnect dominated. The contributions of this work are as follows:

1. We design and investigate full-chip sub/near-threshold 3D circuits for various types of large circuits w/ and w/o memory. We implement special level-shifters for interfacing logic with memory when they operate at different supply voltages. We also carry out detailed variation studies for different supply voltages.
2. A full comparison of performance metrics of the 2D and 3D designs for different supply voltages varying from nominal to subthreshold voltages has been done analyzing the low supply voltage and 3D benefits.
3. The thermal and IR-drop issues that are generally the major drawbacks of 3D integration have been shown to have almost no adverse impact in low voltage designs due to extreme low power. This reduces the intrinsic voltage and temperature changes which are critical at low voltages.
4. We have further demonstrated 3D IC advantage with a many-core design to show massive savings on I/O power and increase in memory bandwidth.

The remainder of the chapter is organized as follows. In Sect. 2, we review the previous works related to low-voltage supply designs and design techniques. In Sect. 3, we elaborate the design procedure for the gates including sizing of transistors, obtaining the timing libraries, studying the supply and temperature variation effects on performance of cells, selection of essential circuits for reliable operation, and the full-chip designs. In Sect. 4, we analyze and compare the various design implementations in 2D and 3D for three different design cases which cover memory-dominated to interconnect-dominated designs and study the extent of 3D benefits for sub/near-threshold operation. We also show the full-chip thermal and static IR-drop analysis. In Sect. 5, we demonstrate more 3D benefits in reducing power and increasing memory bandwidth using a many-core design. Section 6 summarizes the design lessons and outlines for sub/near-threshold 3D IC design. Finally, we conclude our work in Sect. 7 along with the future work directions.

2 Review of Existing Works

Low-voltage designs have been a topic of research for a long time [1]. With advancement of technology, research related to sub/near-threshold design has gained much more momentum. Wang et al. designed an FFT processor operating at 180 mV [16]. They used 180 nm technology for their design with a threshold voltage of around 450 mV. As a follow-up of this work, Calhoun et al. studied the techniques for optimum sizing in the subthreshold regime for minimum energy of operation. They solved equations of total energy to provide analytical solutions for minimum energy operating conditions and also studied gate sizing impact on energy at subthreshold supplies.

In more recent works, Hanson et al. published their work on a full low-voltage processor with special techniques in memory design to reduce standby power to picowatt range [6]. Industry has also shown keen interest in near-threshold designs for obtaining the best energy efficiency [10]. This work discusses in detail the opportunities and challenges of near-threshold voltage designs which operate at maximum energy efficiency, i.e., maximum number of operations for unit energy dissipation. They give various real design examples and measured values to validate the great potential of near-threshold designs. Another 2D IC work on near-threshold operation designed a processor which runs at 1 MHz frequency, 0.4 V supply voltage, and consumes 79 μW [13].

With the advent of 3D IC technology with TSVs, there has been some important research work on using 3D implementation to save power and increasing memory bandwidth. A memory bandwidth of 63.8 GB/s is measured in two-tier 3D IC in 130 nm technology [12]. Jung et al. demonstrated 21.2 % power reduction in TSV-based 3D over 2D by implementing various techniques such as 3D floorplanning, block folding, metal layer usage control, and multi-Vth designs [7]. Though their work is focused on superthreshold voltage operation with nominal supplies, the techniques developed are important and can be used in low-voltage designs as well for getting further benefits from 3D IC implementation.

The first near-threshold 3D IC system was designed with 0.65 V for logic and around 0.87 V for SRAM [4]. The authors try to highlight the feasibility of thermal-constrained 3D IC designs by combining it with near-threshold architecture which also results in high energy efficiency. Our work in this chapter tries to explore the benefits of 3D IC implementation on various types of circuits and at various voltage supplies, which include subthreshold voltages. To our knowledge, there have been no previous works trying to explore subthreshold designs with 3D ICs. We use simple 3D design techniques to evaluate the performance and demonstrate the initial motivation of going for 3D IC low-voltage designs. By implementing more power saving techniques as discussed in the above works, we can see further improvements.

3 Cell Library Design Methodology

3.1 Issues Related to Sub/Near-Threshold Designs

At very low supply voltage, the logic gates become extremely sensitive to noise and there may be logic failure. Also, the strengths of NMOS and PMOS do not follow the same ratio as in nominal supply voltages. The standard cells may not function properly at subthreshold region. There can be several reasons for this behavior. First, the optimum ratio between NMOS and PMOS is mainly determined by the difference between mobility of electrons and holes for superthreshold design. But in subthreshold operation, the difference in threshold voltage determines the current variation between NMOS and PMOS. Using typical NMOS to PMOS ratio used in superthreshold operation may result in unsymmetrical voltage transfer

characteristic in subthreshold supplies. Second, the technology as well as the voltage and temperature variation increases the difficulty for subthreshold design. For superthreshold design, the on/off current ratio is typically more than 10^3. Although the technology causes some transistors to be stronger or weaker, the on-transistor still overwhelms the off-transistor. However, in the subthreshold region, only subthreshold leakage current can be used to switch the logic gates. The current relation becomes exponential with respect to threshold voltage. Therefore, any variation in technology may cause current to vary exponentially. Near-threshold operation is more stable than subthreshold in this respect. Third, many standard memory components such as standard 6T SRAM cannot work in subthreshold region since reduced noise margin (SNM) in subthreshold design causes signal integrity issues. The memory cells need to be modified to reach the energy minimum voltage [5, 6].

3.2 Sizing Techniques

To minimize process variation effects in our design, we used the GlobalFoundries 130 nm technology that has a nominal supply voltage of 1.5 V. We studied the transistor characteristics around the threshold voltage values (NMOS-0.53 V and PMOS-0.56 V). The spice simulations with supply varying from 0 to 0.8 V show that during superthreshold operation, NMOS is faster than PMOS as expected but under subthreshold conditions, PMOS is stronger.

To determine the minimum supply voltage to be used, we study the energy per cycle of a 20 inverter chain and find the optimum supply voltage at 0.3 V (Fig. 1). However, for reliable operation of larger cells like D-flip-flop, we set our supply voltage at 0.4 V after careful spice simulations. The flip-flop fails to operate below 0.35 V supply. The other two voltages are chosen to be 0.6 and 0.8 V. 0.6 V is chosen as it is

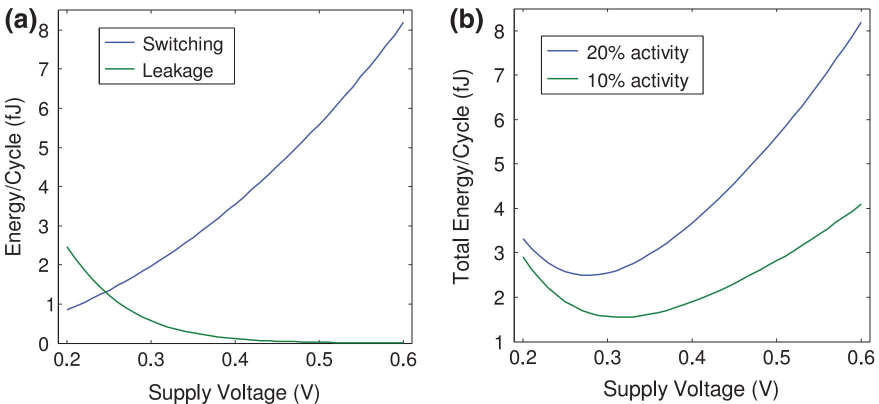


Fig. 1 Energy consumption per cycle for 20 inverter chain. **a** Switching and leakage energy. **b** Total energy

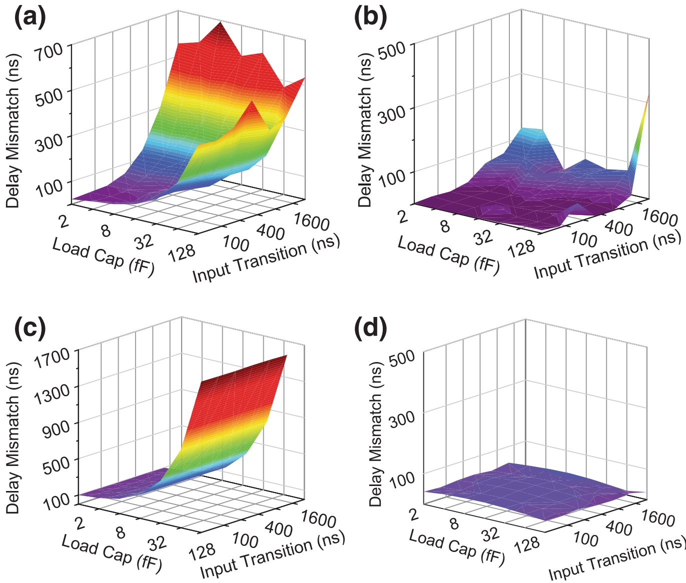


Fig. 2 Delay mismatch before and after sizing at 0.4 V supply. **a** INV before sizing. **b** INV after sizing. **c** DFF before sizing. **d** DFF after sizing

just greater than the specified threshold voltage of this technology. Memory is operated at 0.8 V and therefore the same voltage is also used for logic. In total, we have four different supply voltages varying from nominal 1.5 V to subthreshold 0.4 V.

The logic gates are sized accordingly, equalizing the strength of NMOS and PMOS to reduce propagation delay mismatch independently at these voltages [2]. Therefore, each gate has four implementations with different transistor sizes. The original sizes are used for 1.5 V operation. The boundary area of the cells is kept the same as nominal sizes to have a fair area comparison at different supplies. The wider standard cells are used as per the need of subthreshold operation. To save time and keep our focus on the objective targeted, we confine our standard cell library to the basic cells only, viz., Inverter, NAND, NOR, Buffer, D-flip-flop, and Multiplexer and avoid sizing of all gates for subthreshold operations. We use the same type of cells but resized for delay matching at each supply voltage, respectively. While sizing of standard cells, we use the load capacitance value of 16fF and input rise/fall time of 2 % of input switching period (Fig. 2).

3.3 Library Characterization

After sizing of the cells at the above-mentioned settings and completing the new cell layouts, we characterize our standard cell libraries with Cadence Encounter

Table 1 Standard cell power–delay product comparison for super-Vth (1.5 V) and sub-Vth (0.4 V) supply

	Power (μW)		Avg $t_{delay\ (ns)}$		Power x delay (fJ)	
Cell	Super-vth	Sub-vth	Super-vth	Sub-vth	Super-vth	Sub-vth
BUF	9.32	0.00041	0.164	323	1.53	0.131
DFF	14.83	0.00172	0.432	660	6.41	1/132
INV	7.49	0.00033	0.103	213	0.77	0.071
MUX	11.31	0.00078	0.207	595	2.34	0.461
NAND2	10.96	0.00070	0.122	290	1.34	0.202
NOR2	12.39	0.00080	0.132	293	1.64	0.235

Library Characterizer using the post-extracted spice netlist. These new libraries are used in the subsequent digital circuit designs. The cell power consumption and delay comparison for 1.5 V nominal voltage and 0.4 V subthreshold voltage for the, respectively, sized cells are shown in Table 1. These values are obtained after cell characterization corresponding to from intermediate input transition and load. There is a significant power reduction but at the expense of increase in propagation delays as we move from 1.5 to 0.4 V. The power–delay product reduces as we reduce the supply voltage. Since, we set our supply to be 0.4 V, the minimum energy per cycle supply voltage is not reached for all cells.

3.4 Variation Study

Since low voltage operation is highly sensitive to variation in supply and temperature, we study the standard cell library behavior with changes in these parameters and carry out Monte Carlo simulations to get a better picture of the impact of variation. We choose the inverter as our target standard cell and carry out the relevant simulations. We carry out Monte Carlo simulation with 10,000 runs. Gaussian distribution for threshold voltage and transistor dimensions is used with standard deviation of 50 mV for threshold voltage variation and 10 nm for dimension variation and all the variations are random. Figure 3 shows the inverter layouts at the different voltages with major differences highlighted and the 3D histograms of propagation delays. The x and y axes in the histograms are the high-to-low and low-to-high propagation delays, respectively. The histogram for 1.5 V is very compact (many red bars), even though they are not tall. The spread of delay variation increases as we move to lower voltages with almost 20 % maximum variation from mean for 0.4 V inverter.

Figure 4 further elaborates the delay distribution at nominal and subthreshold voltages. The red bars are for low-to-high propagation delay, i.e., when the PMOS is on and the blue bars are for high-to-low propagation delay when the NMOS is on. The left column is for nominal 1.5 V supply and the right is for subthreshold

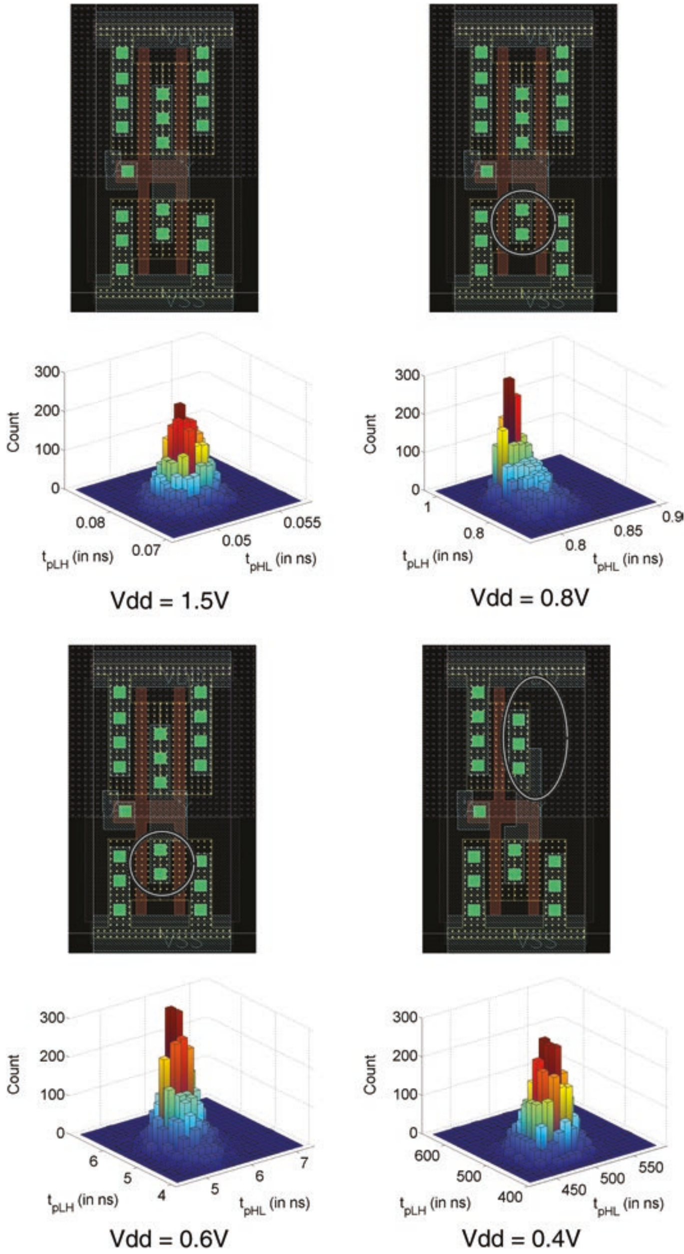


Fig. 3 Inverter cell layout with differences highlighted (*top*). Propagation delay 3D histograms after Monte Carlo simulations (*bottom*). Here t_{pHL} and t_{pLH} stand for high-to-low and low-to-high propagation delays, respectively

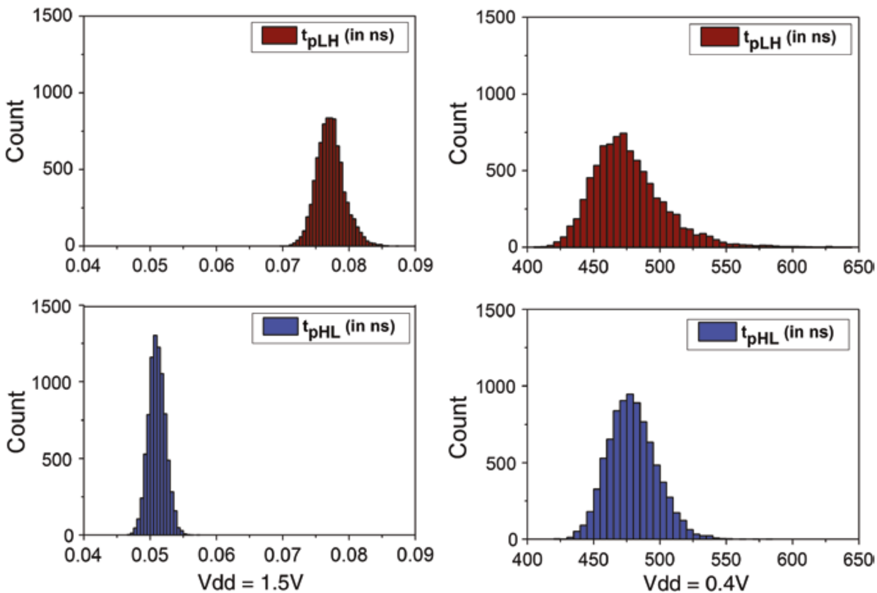


Fig. 4 Propagation delay variation comparison for Monte Carlo simulations of inverter at 1.5 and 0.4 V

0.4 V supply. As can be clearly seen, the distribution is more compact for the nominal supplies and for the high-to-low propagation delays when NMOS is on. There is a longer tail in the distribution for the low-to-high propagation delay compared to the high-to-low propagation delay at 0.4 V supply. The explanation for this can be found in the white circle in Fig. 3. The 0.4 V inverter has only one PMOS transistor but two NMOS transistors. Therefore, random variations in the NMOS transistors tend to average out and reduce the overall standard deviation of the high-to-low delay variation. However, there is no such averaging for low-to-high propagation delay since only one transistor is present. As a result, there is a longer tail and longer standard deviation. It is also important to note in these histograms that the mean propagation delays are similar for 0.8, 0.6, and 0.4 V inverters. This further verifies that appropriate gate sizing has been carried out to match the delays.

The performance changes in the subthreshold inverter with PVT variations are shown in Table 2. These results are obtained from spice simulation using 10µ input and 16fF load at output. The full-chip thermal and supply voltage variation analysis for one of the design cases presented in Sect. 4 is based on this standard cell performance variation studies.

Table 2 Effect of PVT variations on sub-Vth inverter at three corners (fast–fast, typical–typical, slow–slow)

Process corner	FF	TT	SS
Rise delay (ns)	239 (0.599×)	420 (1.0×)	1080 (2.57×)
Fall delay (ns)	251 (0.615×)	408 (1.0×)	1370 (3.35×)
Temperature (K)	323	298	273
Rise delay (ns)	309 (0.736×)	420 (1.0×)	503 (1.20×)
Fall delay (ns)	322 (0.789×)	408 (1.0×)	580 (1.42×)
Supply voltage (V)	0.44	0.40	0.36
Rise delay (ns)	336 (0.800×)	420 (1.0×)	542 (1.29×)
Fall delay (ns)	287 (0.703×)	408 (1.0×)	575 (1.41×)

3.5 Memory Designs

For practical applications, the requirement of memory is unavoidable. There exist a number of works on subthreshold memory design [5, 6]. However, we used commercial memory compilers to generate large memory macros at nominal supply voltages. For smaller memory sizes less than 1 KB, we design register files using our standard cells at the respective operating voltage to have very low power consumption.

The larger memory macros generated by commercial memory compilers uses standard 6T SRAM cells. We reduce the supply voltage of memory to 0.8 V, which is determined from spice simulations as the minimum voltage for reliable 6T cell SRAM operation for the same technology. The memory libraries are also scaled as per the scaling factors obtained from spice simulations at reduced voltage supply of 0.8 V. As mentioned earlier, this is the reason for choosing 0.8 V as one of the voltages for studying the 3D designs. However, there is a requirement of multiple voltage supplies of 0.4 V/0.6 V and 0.8 V for logic and memory, respectively, for a whole system design at these logic supplies. 3D helps us in this respect as we use memory macros in dies separate from logic die and, hence, they can have dedicated power supply without additional circuits.

Another critical issue is signal interfacing between logic and memory which require the use of level-shifters. As the voltage has to be raised from a sub/near-threshold voltage to a high voltage, we use a modified level-shifter circuit as shown in Fig. 5. Since the input is sub/near-threshold, the transistors are not strongly driven and the output may not change. Therefore, we add transistors in diode connection to control the pull-up current. The diode-connected transistors help to drive a larger current to switch the output successfully. The sizing of the transistors in this circuit has been done to satisfy the requirements and verified with spice simulations and the signals plot is also shown in the figure. These level-shifters are used at the address and data pin outputs of logic in our full-chip design. A new set of low-voltage cell libraries is obtained to accommodate the level-shifter timing delays inside the cell. The cells in these libraries are only used at the memory connections.

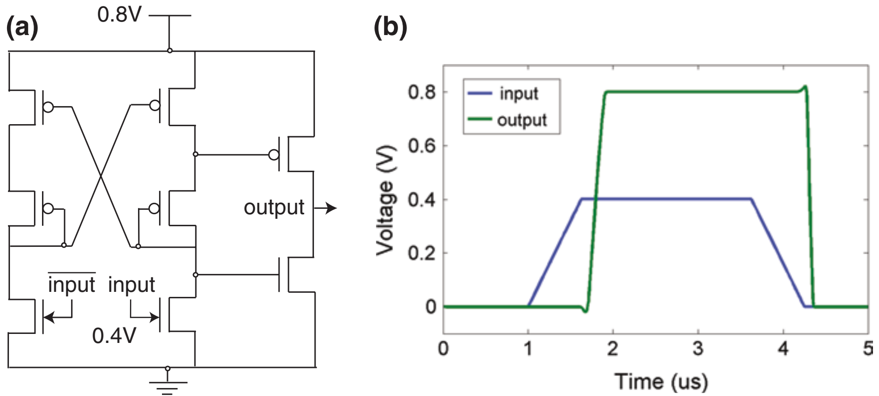


Fig. 5 Level-shifter circuit used for 0.4 V/0.6 V to 0.8 V shift **a** Schematic **b** Transient waveform

4 Full-Chip Design and Analysis

4.1 Design Flow

We study the impact of 3D designs and low voltages for various types of circuits. The different test cases we use are the 8052 microcontroller, LEON3 processor [14], and low density parity check (LDPC) circuit [15]. The rationale behind choosing these circuits is to have different kinds of test cases. While LDPC is pure logic design with very high degree of interconnection, 4-core LEON3 is heavily dominated by memory. The 8052 microcontroller uses internal RAM of 256 bytes and external RAM up to a maximum of 64 KB. It also has a ROM with size up to 64 KB. We implement two versions of 8052 with 16 KB and 64 KB external memory, respectively. All these designs are implemented in 2D and 3D. Memory is always operated at 0.8 V except for the nominal 1.5 V design when it is operated at 1.5 V itself.

We use Synopsys Design Compiler to obtain the netlist from the RTL and then use Cadence SoC Encounter to do the full-chip layout. Synopsys Primetime is used for timing and power analysis for 3D design. The different test cases are designed in different ways. For 8052 designs, based on the area of logic and memory, we completely separate both into different tiers for the 3D design. Since the system has input and output pins that are connected to the logic portion only, we put the logic and internal RAM on the bottom die (die0) and the external memory are stacked on top. Each 16 KB external RAM along with 16 KB external ROM makes one die. We use 60 TSVs of 5 µm in diameter to connect between two dies, and we set the TSV pitch to be 10 µm. The TSV parasitics are calculated accordingly [9]. For the LEON3 and LDPC designs we first carry out multi-way partitioning of the netlist using the best results from both flare [3] and hMETIS [8].

Then TSV insertion with true 3D placement is carried out using a 3D placer [11]. In 4-core LEON3 designs, the memory modules are spread in all the four tiers at the same 2D locations and they do not have any dedicated tier. For LDPC test case, two-way partitioning followed by 3D cell and TSV placement is done. The full-chip layouts for 64 KB 8052 designs are shown in Fig. 6 and the layouts for LDPC design are shown in Figs. 7 and 8.

We build 2D and 3D designs for all these test cases at four different supply voltages to compare the sub/near-threshold and 3D impact on the power, energy,

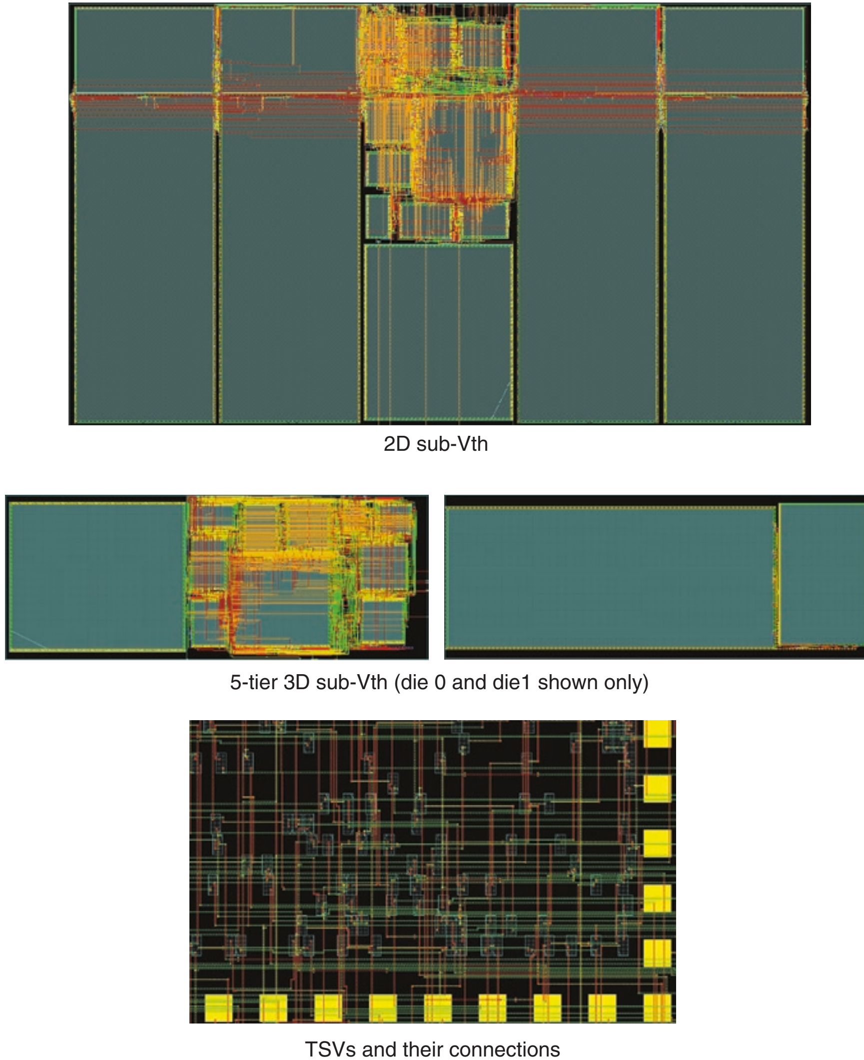


Fig. 6 Subthreshold layout with 64 KB external memory. The designs for die 1 to die 4 (= memory dies) in 3D design are almost identical

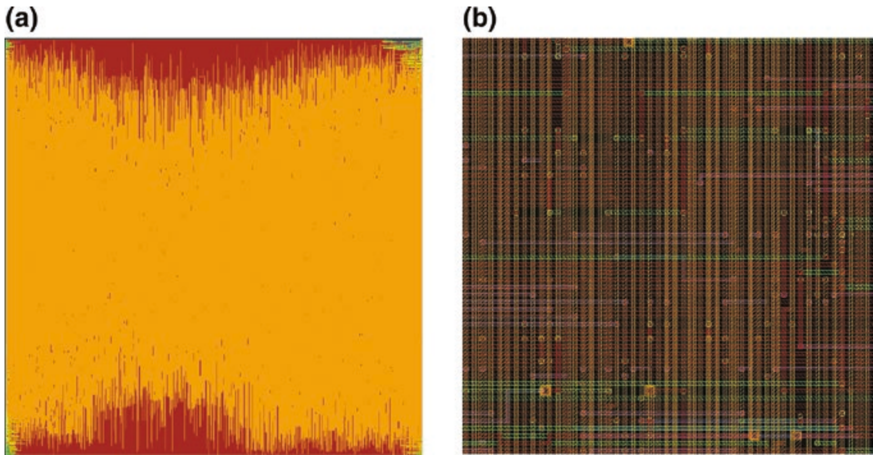


Fig. 7 Layouts of 2D IC LDPC design. **a** Full-chip. **b** Zoom-in shot that reveals routing congestion

and footprint area. The design specifications along with the number of tiers in 3D implementation are listed in Table 3.

We observe that by implementing the designs in 3D, we obtain up to 78 % reduction in the footprint area (for 64 KB 8052). The more the number of tiers in 3D, the greater the improvement in footprint area. Because of footprint reduction, the wires that are needed to connect the blocks in 8052 designs are shortened, and this results in reduction in the top-level interconnect wire length. The wire length saving is small in 16 KB design due to very few 3D connections. There is almost no impact of wire length saving in 4-core LEON3 design because memory dominates the design and 3D folding shows lesser benefits. However, in LDPC design which is heavily interconnect-dominated, we see up to 12 % wire length reduction by going to 3D.

4.2 Timing and Power Comparisons

Table 4 shows the detailed results of the different implementations at different supply voltages. We analyze each of the test cases individually and then try to summarize the design dependency of quantity of 3D advantages. The clock periods are kept the same for 2D and 3D designs in each case to have an iso-performance comparison. The total power reported is the sum of the internal, switching, and leakage powers. The memory power has been reported independently.

In 8052 designs, the clock frequency is up to 66.7 MHz for superthreshold design. Therefore, the internal power and the switching power are the main part of the total power consumption. Since both the logic and memory are under the same supply voltage and the switching activity for memory is very low, the memory

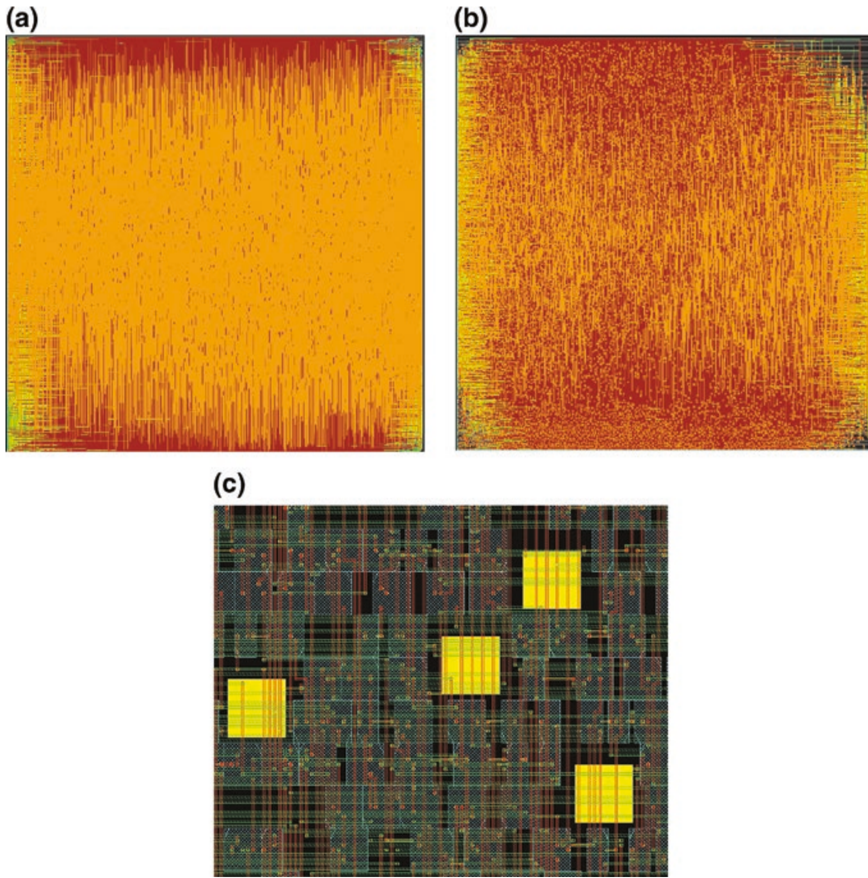


Fig. 8 Layouts of 3D IC LDPC design. **a** Top die. **b** Bottom die. **c** Zoom-in shot that reveals TSVs, logic cells, and interconnects

Table 3 Footprint area comparison for the different designs implemented

		Footprint area ($\mu m \times \mu m$)	No. of tiers in 3D
8052 (with 16 KB)	2D	940 × 1300	2
	3D	500 × 1300	
8052 (with 64 KB)	2D	2300 × 1300	5
	3D	500 × 1300	
LEON3	2D	3600 × 2400	4
	3D	1200 × 1800	
LDPC	2D	2100 × 2100	2
	3D	1430 × 1430	

Table 4 Timing and power results for the different designs cases

	1.5 V		0.8 V		0.6 V		0.4 V		$\frac{3D}{2D}$
	2D	3D	2D	3D	2D	3D	2D	3D	
8052 with 16 KB memory									
Target clk	15		150		1,600		50,000		1
Internal	4.713	4.589	0.1843	0.1775	0.00926	0.00927	0.0001714	0.0001713	1
Switching	5.893	5.690	0.1988	0.1841	0.01072	0.01061	0.0001322	0.0001303	0.99
Leakage	0.00595	0.00593	0.00098	0.00098	0.00092	0.00092	0.0008611	0.0008609	1
Total	10.6	10.3	0.3841	0.3616	0.0209	0.0208	0.001165	0.001163	0.99
PDP	159	154.5	57.62	54.24	33.5	33.28	58.25	58.15	0.99
Memory	0.7145	0.7141	0.0225	0.0216	0.0029	0.0028	0.0008861	0.0008861	1
8052 with 64 KB memory									
Target clk	15		150		1,600		50,000		1
Internal	6.214	5.30	0.249	0.2398	0.01504	0.01511	0.0001800	0.0001773	1.01
Switching	4.799	5.69	0.199	0.186	0.01105	0.01075	0.0001424	0.0001389	0.97
Leakage	0.02165	0.021	0.00346	0.00346	0.003387	0.00339	0.003334	0.00333	1
Total	11.03	11.01	0.4515	0.4293	0.0295	0.0293	0.003657	0.003649	0.99
PDP	165.45	165.15	67.725	64.395	47.2	46.88	182.85	182.45	0.99
Memory	0.7865	0.7841	0.08995	0.08635	0.01142	0.0111	0.003363	0.00336	0.97
LEON3									
Target clk	10		100		1,000		100,000		1
Internal	237	239.8	7.265	7.269	0.6199	0.6198	0.0056	0.0056	1
Switching	54.4	53.1	1.317	1.317	0.1229	0.1208	0.0003	0.0003	0.98
Leakage	0.069	0.069	0.010	0.010	0.0198	0.0098	0.0095	0.0095	1
Total	291.47	292.97	8.592	8.596	0.7525	0.7504	0.0154	0.0154	0.99

(continued)

Table 4 (continued)

	1.5 V		0.8 V		0.6 V		0.4 V		$\frac{3D}{2D}$
	2D	3D	2D	3D	2D	3D	2D	3D	
PDP	2914.7	2929.7	859.2	859.6	752.5	750.4	1543	1542	0.99
Memory	182.2	182.0	5.165	5.168	0.5268	0.5264	0.0146	0.0145	1
LDPC									
Target clk	10		400		4,000		100,000		1
Internal	37.53	34.43	0.413	0.474	0.0313	0.0303	0.00037	0.00035	0.95
Switching	216.9	199.1	1.463	1.247	0.0859	0.0796	0.00093	0.00084	0.90
Leakage	0.0034	0.0029	0.0004	0.0004	0.0002	0.0002	0.00009	0.00009	1
Total	254.5	233.5	1.876	1.721	0.1175	0.1101	0.00139	0.00127	0.91
PDP	2545	2335	750.4	688.4	470	440.4	139	127	0.91

Timing values are in ns, power (internal, switching, leakage, total, memory) in mW, and power–delay product (PDP) in pJ. The $\frac{3D}{2D}$ ratio is calculated at the best PDP settings

power is not a big portion of total power. By reducing the supply voltage and going to subthreshold computing, the design with 16 KB external memory shows 9099 times reduction in total power at the cost of 3333 times lower clock frequency. We use the symbol '×' from here for comparison. For certain low-power applications like sensors, the workload for each node is not heavy, and the performance of each computing node is not the major concern in most cases. Therefore, by reducing the supply voltage we can reduce the power and energy per cycle and ensure a longer battery life.

Since the external memory for subthreshold designs are working at 0.8 V, the larger the size of memory, the more the leakage, as is clear from the results. By reducing the clock frequency and supply voltage we can achieve a significant reduction in internal power and switching power. In the 64 KB external memory design, the internal and switching power is reduced by almost 35000× by changing from superthreshold design to subthreshold logic. But the leakage power is not directly related to clock frequency change, so we achieve 6.49× times saving in leakage power saving. The larger the memory size, the smaller the power savings. The 64 KB external memory design shows 3008× overall power reduction in contrast with the 9099× power saving for 16 KB external memory design.

The internal power and switching power is reduced by 1.5 and 2.5 %, respectively, in the subthreshold 3D design with 64 KB memory compared to subthreshold 2D design, and the total power is only reduced a little because the leakage power remains almost the same and it is the dominating part in total power. The best power–delay product is obtained at 0.6 V because memory leakage starts to dominate at 0.4 V. We can see from Table 4 that degradation of PDP is more severe for 64 KB designs compared to 16 KB designs, as there is more memory contributing to more leakage.

The 4-core LEON3 has huge memory content. Memory power is more than 60 % of the total power for superthreshold 1.5 V design. Therefore, even though there is some minor improvement in switching power in 3D implementation, the overall benefit is almost negligible. We also know that leakage power does not scale proportional to switching power. Therefore, as we reduce the supply voltage, dominance of memory increases further. Once again best power–delay product is obtained at 0.6 V which agrees with earlier studies on full processors [10].

For the LDPC design, we do not have any memory but only logic. Also, the design is highly interconnected and hence we see significant reduction in switching power in the 3D designs (Table 4). Since switching power is a major portion (85 %) of the total power, we observe 9 % reduction in total power which is a direct consequence of 10 % switching power savings. It is interesting to observe that the best power–delay product for this design is obtained at 0.4 V unlike the other test cases where it was at 0.6 V. This follows directly from the fact that leakage is negligible in LDPC due to absence of any memory and hence we have a similar minimum energy point as in the 20 inverter chain evaluated in Sect. 3.

4.3 3D Impact Versus Type of Design

The three design cases implemented at different voltages include memory-dominated designs as well as interconnect-dominated designs. As is clear from the results, 3D implementation has very less power improvement for memory-dominated designs, but a significant power reduction is observed for interconnect-dominated designs (LDPC). 3D also helps in the implementation of separate tiers with separate voltage supplies. The use of separate tiers for separate voltage domains simplifies the power delivery design where we will not require different voltage islands on the same die (8052 design case). Each die can have its own dedicated supply connection with the far tiers getting power through power/ground TSVs.

On the other hand, if the design is heavily dominated by memory (LEON3), it is not feasible to have all memory on single tier as there will be wastage of area in the logic tier. This necessitates the requirement of good mixed 3D placement. Bad placement can result in degradation of power with unnecessary long wires. But at the same time, near/subthreshold implementation is important for maximum energy efficiency for ultralow power remote applications, which also need to be made with minimum footprint area and therefore, 3D is helpful in that respect. Therefore, depending on the type of design and type of application, 3D IC implementation can be used to get the best possible benefits that are practically attainable.

4.4 Full-Chip Variation Study

As discussed in the previous section, subthreshold operation is highly sensitive to process, voltage, and temperature variations. We therefore analyze the 8052 subthreshold (0.4 V) design at different process corners and with temperature and voltage variations. The results for only logic variations are shown in Table 5. Only memory process variation effects are shown in Table 6. We observe that the design becomes faster at higher temperature unlike standard superthreshold circuits and consumes higher power. This is because subthreshold current increases exponentially with increase in temperature. The other variations affect the design performance and power consumption as expected. Since the critical path in our analysis is only through logic, variations in memory do not affect the timing performance.

4.5 Thermal Analysis

Since the 3D IC implementation stacks several dies into a single package, therefore heat dissipation is always a major concern. Also, the performance of low-voltage cells especially the subthreshold ones are very sensitive to temperature.

Table 5 Effect of PVT variations on subthreshold logic only: power numbers based on 20 kHz frequency operation

Process corner	FF	TT	SS
Longest path delay (ns)	9801	40062	327905
Core leakage power (μW)	0.185	0.037	0.021
Total core power (μW)	0.420	0.278	0.264
Temperature (°C)	50	25	0
Longest path delay (ns)	23316	40062	104595
Core leakage power (μW)	0.121	0.037	0.024
Total core power (μW)	0.364	0.278	0.266
Supply voltage (V)	0.44	0.4	0.36
Longest path delay (ns)	22424	40062	111500
Core leakage power (μW)	0.050	0.037	0.033
Total core power (μW)	0.347	0.278	0.230

Table 6 Effect of process variations on memory power only (Operating corners are obtained from memory compiler and scaled down)

Corner	FF/0.88 V/−40 °C	TT/0.8 V/25 °C	SS/0.72/125 °C
Leakage (μW)	3.12	3.3	6.52
Total (μW)	3.20	3.37	6.60

Therefore, we need to carefully simulate the thermal effects on our 3D designs. Since current tools cannot handle 3D designs properly, we use our in-house tools to build a thermal model for 3D IC and perform thermal simulation using ANSYS Fluent. First, we build a mesh for our chip and compute the thermal conductivity for each grid using layout information. Then we export power information from Primetime and build a power density map. Finally, we use Fluent to solve the thermal differential equations using the power density map and thermal conductivity information and obtain the temperature map of our design. In this simulation, we assume adiabatic boundary conditions on all four sides and the bottom side of the package, and the top side of the package is directly in contact with static air without any heat sink. The ambient temperature is 25 °C.

The temperature map for 2-die 8052 design with 16 KB memory is shown in Fig. 9. Since in superthreshold design, the memory power is only 7.3 % of the total power, therefore, the temperature of the chip is mainly determined by the blocks on the bottom die (Table 7). Therefore, the center of the blocks will usually have the highest temperature within that block. Also, since TSVs are made of copper, which has the highest thermal conductivity of all the materials on the chip, it can transfer heat quickly from the bottom die to the top die. Therefore, around the TSV arrays, the temperature is relatively low and that area becomes the coolest part of the full chip.

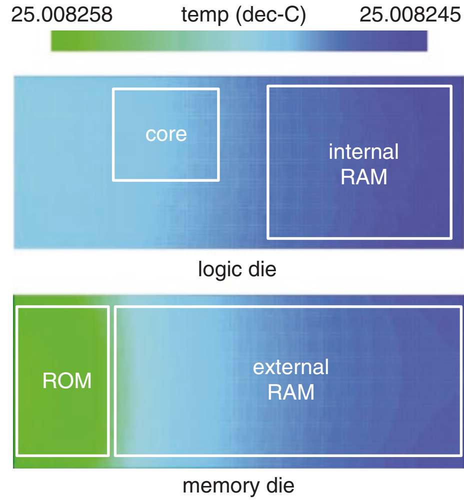


Fig. 9 Temperature map for subthreshold 3D design. Temperature gradient is negligibly small

Table 7 3D full-chip temperature and IR-drop (logic die) analysis

	Power density (mW/cm^2)		Max temp (°C)		Max static IR drop
	Die0	Die1	Die0	Die1	
3D Super-Vth	10410	750	97.964	97.821	26 mV
3D Sub-Vth	0.269	0.381	25.008	25.008	0.34 μV

By lowering the voltage supply and performing subthreshold computing, the power density on each die is significantly reduced. As a result, the maximum temperature increase from ambient temperature within the chip is reduced from 72.96 °C in superthreshold design to 0.00852 °C in subthreshold design. Also, since the memory has much larger power than the logic portion, the temperature on the bottom die is heavily affected by the top die. In this case, the ROM has the largest power density within the chip, so the maximum temperature appears on the part where ROM is placed. From the results, we can conclude that by stacking subthreshold circuits in 3D, we do not encounter serious thermal problems from within the chip. However, there will be external temperature effects on performance.

4.6 IR-Drop Analysis

We have shown that the performance of standard cells and, therefore, the full design is highly affected by supply voltage variation. This makes the power distribution

within the chip a critical design step. Although the external supply may not vary, the internal IR drop may result in reduced supplies to certain logic gates within the chip. To study this effect, we use a simple Power Distribution Network (PDN) for the 8052 design and analyze the static IR drop for subthreshold operation. The IR-drop issues will mainly affect the logic portion operating at subthreshold 0.4 V because the memory contains dies are operating at 0.8 V and have a dedicated PDN.

The blocks used in the top-level logic design have power rings on their boundaries. We use simple minimum PDN for top level with only rings at the die boundary in the top metal layers and use the Metal1 VDD and VSS rails to connect to the individual block rings as well as the top-level buffers and inverters used for timing closure. The power supply bump locations for IR-drop analysis are set at the four corners of the dies at the power ring intersections.

We use Cadence VoltageStorm for static IR-drop analysis. Detailed placement and layout information is used for IR-drop analysis to ensure exact calculations even within the hard blocks. Figure 10 shows the IR-drop map. The individual cell power consumption is obtained from Primetime simulations. We scale-up the initial power consumption of each cell by a factor of 10,000 to obtain accurate results of every minor IR drop and then scale down the voltage drop values back to original after obtaining the results from analysis. We observe that even for a minimal PDN design, the maximum static IR drop is only 0.34 μV in subthreshold design (Table 7). The values are small because the current drawn by each cell from the supply is subthreshold and, hence, very small. Therefore, we can conclude that IR drop is not an issue for our subthreshold 3D design.

As discussed earlier, another major advantage of going 3D is the integration of subthreshold and near-threshold (as in our work) or superthreshold dies without any additional circuits for separate PDNs. Since each die operates at its own supply, we can have dedicated power ground TSVs supplying the top tiers. The same design in 2D may require isolation and decoupling of the different supply networks.

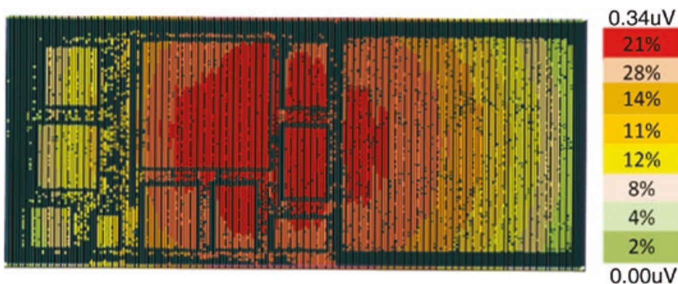


Fig. 10 IR-drop map for subthreshold logic tier

5 Power Benefits in Many-Core Designs

The power consumption discussed so far does not include I/O power. If we have an off-chip memory, the number of I/O pads will limit the bandwidth of memory access. As the I/O pads consume a huge amount of power, their count usually has an upper bound. However, when we use 3D integration, we can integrate off-chip memory to on-chip and get rid of the processor to memory I/O pads. We not only reduce the power consumption but also increase the memory bandwidth close to theoretical maximum. As the memory to processor connections are TSV-based in 3D design, they consume much less power than I/O pads. We study this feature quantitatively by implementing a many-core subthreshold design with off-chip memory and comparing it with the equivalent 3D implementation.

5.1 I/O Driver Design

The I/O pads provided in the standard library are large with complicated circuits in them and therefore consume a large amount of power and cannot be used for subthreshold circuits. They are meant for high performance in standard superthreshold circuits. Therefore, to have a reasonable quantitative power analysis of many-core implementation, we design our own I/O pads using level-shifter and buffers to drive a large load. We exclude ESD and other circuits from our simple design. The representative diagram for the output pad is shown in Fig. 11. We use a capacitive load of 5 pF to size the large buffer with spice simulations. The large load is representative of the pin capacitance and interconnect capacitance between the processor output pad to memory input pad for off-chip design. Level-shifters are required as the processor output is 0.4 V, while the memory operates at 0.8 V. We set the I/O supply voltage as 0.8 V. The total power dissipated by a single I/O pad is calculated from spice simulations to be 1.066 μW at 20 kHz with 5 pF load and 1 μs input slew. This is inclusive of switching power, cell power, and leakage power.

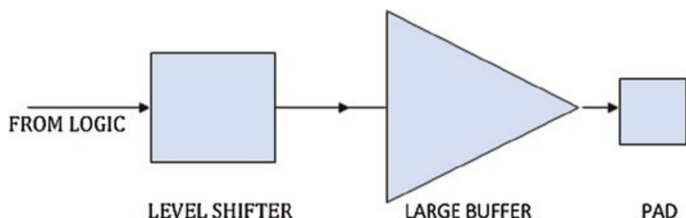


Fig. 11 I/O circuit for logic to off-chip memory connections in many-core 2D subthreshold design

5.2 *Power Saving in Many-Core Sub-Vth 3D Designs*

Using 3D implementation for the microcontroller design helps us put together many processors as per requirement in reduced area with reduced interconnect and hence lesser power. Since each of the cores is smaller in size in 3D, the inter-core connections become shorter and there is less loss of power in wires. As we move to lower technology processes, the wire length parasitic becomes more critical and contributes significantly to power. Also, we can achieve significant memory bandwidth increase by going to 3D.

For initial study, we used 128 cores in 2D and 3D designs and analyzed the area, wire length, processor to memory I/O power, and the memory bandwidth. For the 2D design we use a single large off-chip memory, while we use 5-tier stacking for 3D design. The many-core layouts are shown in Fig. 12. The comparison results are shown in Table 8. We use a two-channel off-chip 2D memory as our baseline for comparison and then analyze the benefits of 3D design. We observe that the reduction in wire power is proportional to the reduction in wire length. Also, the processor to memory I/O power, which is 18.5 % of the total power in 2D design, is completely removed in the 3D implementation. The theoretical maximum bandwidth also increases from 16 bits/cycle for 2D many-core to 1024 bits/cycle for 3D many-core compared to 8 bits/cycle for single core. Therefore, going for 3D designs will contribute significantly to power savings above the subthreshold savings with increase in memory bandwidth for specific applications.

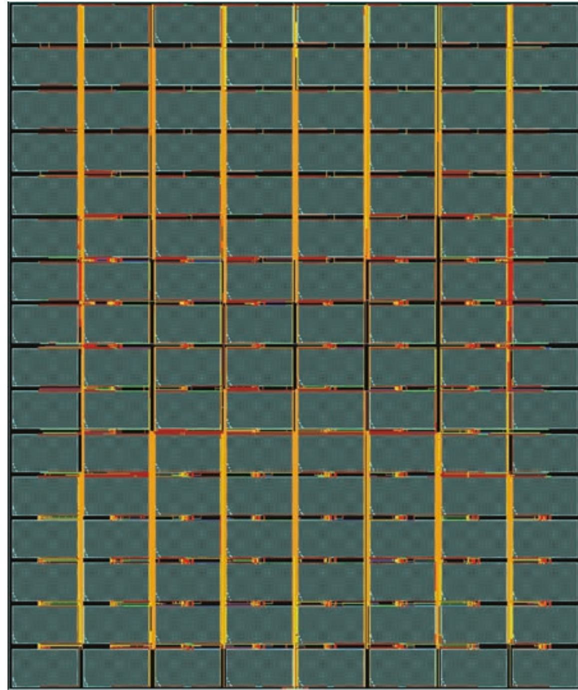
6 Design Lessons and Guidelines

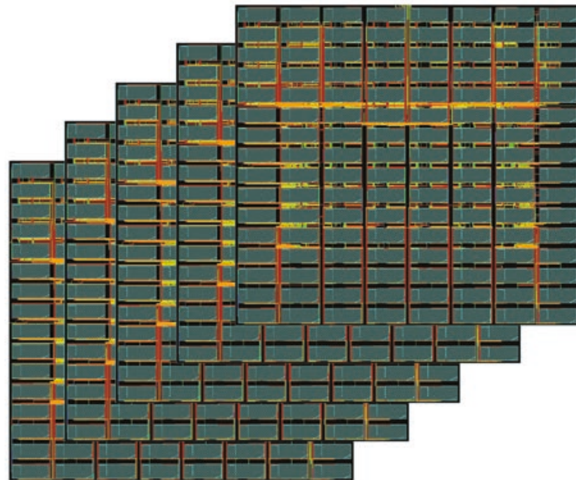
The design of low power sub/near-threshold gates involves a number of issues that need to be taken care of. Wider cells (2× or 4×) need to be used for low voltage operation for better performance in driving the same load as superthreshold circuits. Also, the narrow-width effect may be much more prominent in subthreshold operation of gates [17] and proper cell sizing is dependent on it.

While determining the optimum supply voltage for the design for maximum energy efficiency, we need to take care of proper functionality of all the gates. This is because the energy-minimum supply voltage for gates occurs at deep subthreshold region, where cells may fail to operate reliably, while the memory leakage contribution may shoot up at such low voltages. The D-flip-flop is the critical cell in our case whose correct operation determines the minimum supply voltage of 0.4 V. The cell characterization has to be done carefully based on sub/near-threshold operation-based input slew values.

For 3D sub/near-threshold design, we need to be careful about the thermal and IR-drop variations that the chip may encounter during operation. Subthreshold cells are highly sensitive to temperature and voltage variations. Since 3D stacking

Fig. 12 Many-core subthreshold layouts





causes thermal and power integrity issues, this is a critical part of the design. The TSV parasitic also need to be taken into account during full-chip timing and power analysis as they are a nonnegligible portion of total power dissipation, especially when the number of 3D connections is very high.

Table 8 Design comparison for subthreshold many-core implementation in 2D and 3D

Property	2D	3D (5-tier)
Area (cm×cm)	2 × 2.4 (100 %)	1.12 × 1.2 (28 %)
Wirelength (m)	54.70 (100 %)	23.94 (44 %)
Memory I/O power (μW)	104.5 (100 %)	0.768 (0.007 %)
Total power (μW)	564.5 (100 %)	460.8 (81.5 %)
Data connections	48 (I/O)	3072 (TSV)
Total connections	98 (I/O)	3108 (TSV)
Memory bandwidth (*bits/cycle*)	16 (100 %)	1024 (6400 %)

To reduce power and improve performance, circuit techniques like power gating, adaptive body biasing, or design of sub/near-threshold memory cells are very good options. We need to consider these in future designs to have further 3D benefits in low-voltage designs. In 3D sub/near-threshold design, it is sometimes preferable to use different supply cells in different dies in case we have a multiple supply design. The reason is that we can have dedicated power supply to the dies without major design issues. However, we need to design good level-shifter circuits for proper interfacing of signals with different voltage values and without any large delays. Another important factor is that the I/O pads used for low-voltage circuits need to be modified. Otherwise, the objective of low-power operation will be nullified by the power hungry I/O pads.

7 Conclusions

In this chapter, we explore and quantify the 3D IC benefits in ultralow power designs using sub/near-threshold circuits of different varieties. While logic circuits show excellent reduction in power consumption with good power–delay product improvement, memory contributes to maximum power in designs with memory because of its near-threshold region of operation and high leakage. The larger the memory in a design, the lesser the power savings. However, the footprint area reduction of up to 78 % is quite significant when we move to 3D and this plays a key factor in miniaturization of digital systems. We also carried out detailed variation studies at the cell level as well as the full-chip level. We showed that 3D IC implementation of sub/near-threshold circuits is free from internal thermal and IR-drop-related issues. The problem of dominating memory power can be improved significantly using special low-voltage memory cells [5]. This along with memory folding is one of the important steps to be considered in our future work. We have also demonstrated the idea of many-core design with increase in memory bandwidth and further reduction in power consumption due to the removal of processor to memory I/O pads. Therefore, 3D-stacked sub/near-threshold circuits with proper memory design approach present a major improvement for both ultralow power and miniaturization in processors.

Acknowledgments This work is supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT & Future Planning as the Global Frontier Project.

References

1. Burr JB, Shott J (1994) A 200 mV self-testing encoder/decoder using stanford ultra-low-power CMOS. In: ISSCC Dig. Tech. Papers, pp 84–85
2. Calhoun BH et al (2005) Modeling and sizing for minimum energy operation in subthreshold circuits. IEEE J Solid-State Circuits 40(9):1778–1786
3. Cong J, Lim SK (2004) Edge separability based circuit clustering with application to multi-level circuit partitioning. IEEE Trans Comput Aided Des Integr Circuits Syst 23(3):346–357
4. Fick D et al (2012) Centip3De: A 3930 DMIPS/W Configurable Near-Threshold 3D Stacked System with 64 ARM Cortex-M3 Cores. In: ISSCC Digest of Technical Papers, pp 190–191
5. Hanson S et al (2006) Ultralow-voltage, minimum-energy CMOS. IBM J Res Dev 50(4/5):469–490
6. Hanson S et al (2009) A low-voltage processor for sensing applications with picowatt standby mode. IEEE J Solid-State Circuits 44(4):1145–1155
7. Jung M, Song T, Wan Y, Lee Y-J, Mohapatra D, Wang H, Taylor G, Jariwala D, Pitchumani V, Morrow P, Webb C, Fischer P, Lim SK (2013) How to reduce power in 3d ic designs: a case study with OpenSPARC T2 Core. In: IEEE Custom Integrated Circuits Conference, pp 1–4
8. Karypis G, Aggarwal R, Kumar V, Shekhar S (1999) Multilevel hypergraph partitioning: applications in VLSI domain. IEEE Trans VLSI Syst 7(1):69–79
9. Katti G, Stucchi M, Meyer KD, Dehaene W (2010) Electrical modeling and characterization of through silicon via for three-dimensional ICs. IEEE Trans Electron Devices 57(1):256–262
10. Kaul H, Anders M, Hsu S, Agarwal A, Krishnamurthy R, Borkar S (2012) Near-threshold voltage (NTV) design opportunities and challenges. In: ACM Design Automation Conference, pp 1149–1154
11. Kim DH, Athikulwongse K, Lim SK (2009) A Study of through-silicon-via impact on the 3D stacked IC layout. In: IEEE International Conference on Computer-Aided Design, pp 674–680
12. Kim DH et al (2012) 3D-MAPS: 3D massively parallel processor with stacked memory. In: ISSCC Digest of Technical Papers, pp 188–189
13. Konijnenburg M et al (2013) Reliable and energy-efficient 1 MHz 0.4 V dynamically reconfigurable SoC for ExG applications in 40 nm LP CMOS. In ISSCC Digest of Technical Papers, pp 430–431
14. Leon3 Processor. http://www.gaisler.com/
15. OpenCores [Online], http://opencores.org/
16. Wang A, Chandrakasan A (2005) A 180-mV subthreshold FFT processor using a minimum energy design methodology. IEEE J Solid-State Circuits 40(1):310–319
17. Zhou J et al (2011) A 40 nm inverse-narrow-width-effect-aware sub-threshold standard cell library. In: Proceedings ACM Design Automation Conference, pp 441–446

Energy Harvesting from the Human Body and Powering up Implant Devices

Ross Kerley, Xiucheng Huang and Dong Sam Ha

Abstract This article reviews research activities at the system level on (i) energy harvesting with wearable devices from the human body and (ii) powering up implant devices. The first part reviews wearable devices to harvest energy from the human body for biomedical and portable devices. Harvestable human body energy sources can be classified into two categories, voluntary and involuntary. Voluntary sources are capable of providing high power levels, up to several watts, but are only available when the wearer is active. Involuntary sources are constantly available, but provide much smaller amounts of energy, of the order of milliwatts. The latter part of the article reviews research activities to power up devices implanted in the human body. There are two approaches to power up implant devices. The first approach is to harvest energy from the body or ambient sources. The energy sources are essentially limited to kinetic energy of the body, body heat, and solar. The second approach is to transmit power to implant devices wirelessly. The power level available to implanted devices through harvesting or power transmission is small, often of the order of microwatts.

Keywords Energy harvesting · Human body · Wearable · Implant · Power transmission

R. Kerley · X. Huang · D.S. Ha (✉)
Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA
e-mail: ha@vt.edu

R. Kerley
e-mail: rakerley@vt.edu

X. Huang
e-mail: xiucheng@vt.edu

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_5

1 Introduction

Technological advances in low-power electronics have increased the feasibility of powering them from ambient energy sources including solar, thermal, kinetic, and RF. Energy harvested from these sources can be used to charge or replace batteries in portable devices. A human body is a bountiful source of energy, which can be harvested to power biomedical devices such as body sensor networks, implants, and even long-range wireless sensor nodes. Harvesting body energy increases the operational life of medical devices or enables perpetual machines without the need for battery replacement or recharging.

A human body releases harvestable energy through two different mechanisms: heat and motion. Some body energy, such as motion from breathing and body heat, is released continuously, regardless of the person's intention. This kind of energy source, called involuntary, does not require a conscious action. In contrast, voluntary sources, such as walking, jumping, or other spontaneous activities, can generate a much larger amount of energy than involuntary sources, but depend on the human to intentionally provide the energy.

The first part of this article reviews recent research activities in the area of wearable human energy harvesters, which can be applicable to biomedical circuits. All of these devices harness the energy from body heat, voluntary, or involuntary motion. A wide variety of transducers including thermoelectric generators, electromagnetic generators, electrostatic generators, and piezoelectric transducers are considered to harness the energy.

Implantable electronic devices enable us to monitor the human body in real time, and their applications on the human body are expected to expand rapidly as illustrated in Fig. 1 [1]. Several key challenges must be addressed to realize these scenarios: implantable devices should be minimally invasive, completely biocompatible, and with low thermal dissipation and power autonomy. In particular, the power autonomy is critical, since replacement of batteries for implant devices is costly and may involve surgery. The capacity of batteries has increased continuously, but fails to meet the power requirement of most implant devices. Also, size constraints of implant devices may limit the use of batteries.

There are two promising approaches for the power autonomy of implant devices. The first approach is to harvest power from the human body or ambient sources, and the other is to transmit power wirelessly from external sources to implant devices. Harvestable energy sources for the first approach include knee, body, and head movement, heart, artery, muscle, body heat, glucose fuel cell, and solar. Wireless power transmission to implant devices can be accomplished by various means including inductive link, RF transmission, infrared, and ultrasonic.

Both fields, energy harvesting from human body and powering up implant devices, are rapidly growing owing to advancements in transducers, integrated circuit technology, and energy storage devices. Various energy sources and power transmission methods pose distinct technical challenges at various levels such as transducer, circuit, and system. Intensive research will address these problems, but

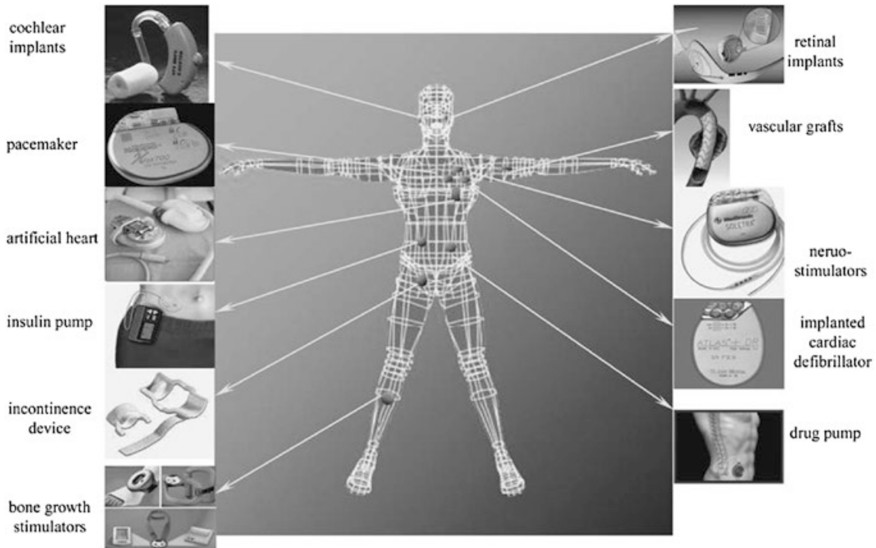


Fig. 1 Diversity of implantable device applications [1]. (Reproduced with permission from Springer)

it is still in an early stage. It is expected that the research in this area will be intensified in the next decade.

This chapter reviews the major recent research activities on the two topics: energy harvesting with wearable devices from the human body and powering up implant devices at the system level. The chapter is organized as follows: Sect. 2 describes energy harvesting from voluntary sources with wearable devices followed by involuntary ones. It is organized by the type of energy sources. Section 3 covers energy harvesting from the human body with implant devices, and it is also organized by the type of energy sources. Section 4 describes wireless power transmission to implant devices. Section 5 summarizes the chapter.

2 Wearable Devices

Voluntary human motion includes any intentional movement such as walking or moving an arm. This energy is available typically in large quantities, but not continuously. A major goal of energy harvesting from voluntary motion is to incur minimal impact on the wearer, which leads to harvesting only a small portion of the available energy. In many instances, research focuses on harvesting energy without incurring any additional metabolic cost. This section covers energy harvesting from the foot, knee, torso, and arm.

2.1 Foot

Foot-mounted kinetic energy harvesters can harvest from heel strikes, sole compression and, foot forward motion. Of these three, heel strikes generate the largest amount of energy. For a typical person walking two steps per second and a heel displacement of 5 cm, 70 W is expended [2]. The rubber soles of shoes absorb and dissipate mechanical energy as heat, while elastically returning some portion of the energy to walking or running.

2.1.1 Heel-Strike Generators

An electromagnetic heel-strike harvester uses a rack and pinion to convert the almost-linear motion of a heel into high-speed rotational motion [2]. The rotational motion spins an electromagnetic generator, which converts mechanical energy into electrical energy. The electromagnetic heel-strike harvester generates 1.8 W from a person walking two steps per second. The power management circuit consists of a DC converter supplying power through a full-wave rectifier to charge a 0.2 F capacitor. A 13 Ω load resistor and a cell phone are in parallel with the capacitor for their experiments. The cell phone extracts usable energy from the capacitor. The resistor dissipates energy in an attempt to reduce voltage ripple.

The PZT (lead zirconate titanate) dimorph device in Fig. 2 shows a heel-strike harvester installed inside the sole of a shoe [3]. The device is made of a PZT dimorph and a metal mid-plate. Though it is feasible to harvest several watts from walking, harvesting the level of energy would severely hamper the feel of a normal shoe. Shenck et al. harvested 8.4 mW from a heel-strike piezoelectric generator in a Navy work boot [3–5]. The generator was imperceptible to the wearer, but sufficient to power a radio identification transmitter.

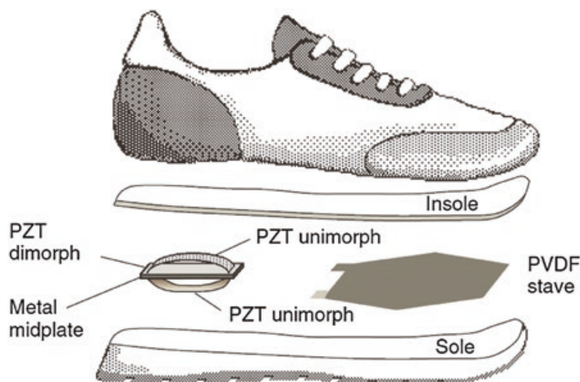


Fig. 2 Piezoelectric material placement inside a shoe [3]. (Reproduced with permission from IEEE)

2.1.2 Sole Compression/Decompression Generators

Piezoelectric materials generate electricity under mechanical strain. There are two different approaches to use piezoelectric materials in shoes for energy harvesting. The first approach replaces most or all of the heel rubber with a piezoelectric structure that has similar elasticity to the original rubber [3–5]. This causes a slight change in the “feel” of the shoe when walking. The second approach uses a flat sheet of piezoelectric material inside the shoes under the foot that compresses and decompresses every step [3, 5]. One example of this approach is use of a flexible polyvinylidene fluoride (PVDF) stave as shown in Fig. 2. The mechanical design is simple, and it offers potential for large energy harvesting. The shoe-sole piezoelectric harvester demonstrated produces an average of 1.3 mW [3, 5].

2.1.3 Foot Motion Generators

Research groups investigated the use of linear electromagnetic generators in shoes to harvest energy from the forward acceleration of a foot. The harvester proposed by Zeng et al. is shown in Fig. 3. It consists of stationary coils and a movable permanent magnet plate to induce current [6]. They used a boost converter with maximum power point tracking (MPPT) to charge a supercapacitor, while a cascaded buck converter charges a lithium ion battery. Their devices achieve an energy density of 8.5 mW/cm^3 [6]. In contrast, Stamenkovic et al. use a movable plate of coils and stationary permanent magnets and, simulation results show that it achieves the energy density of 5.379 mW/cm^3 [7].

2.2 Knee

A human knee is capable of generating a large amount of energy. However, energy harvesting from a knee should not increase the person’s metabolic cost, which incurs strain on the body and requires additional food intake. Both of them would be detrimental to the wearer’s comfort and health.

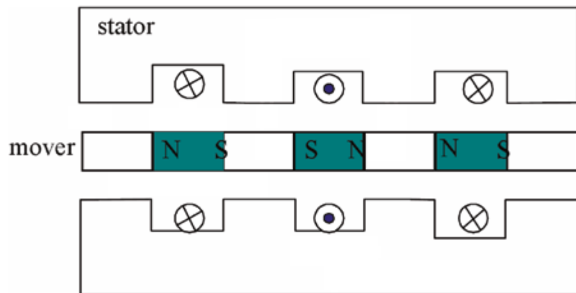


Fig. 3 Cross-sectional view of a shoe-mounted linear electromagnetic generator [6]. (Reproduced with permission from IEEE)

Donelan et al. focused on helping the knee decelerate without inhibiting acceleration [8]. Gears embedded in the electromagnetic generator increase rotational speed of the generator by several times, so that it produces a higher voltage to increase the system efficiency. In addition, it helps to produce a more stable voltage. A one-way clutch or ratchet allows energy harvesting from the knee only when the leg is extending. A controller measures the acceleration of the knee with the aid of a potentiometer, so that it harvests only when the knee is decelerating, which happens when the leg is almost fully extended. This regenerative braking ideally helps the knee and should not increase the wearer's energy consumption. In reality, a mechanical generator still costs metabolic energy during the period when the energy is not harvested. The energy harvester generates 4.8 W with only 5.0 W of additional metabolic energy cost. If the harvester harvests energy during the acceleration period from forward motion, the metabolic cost would be several times that of the harvested power [8].

Luciano et al. worked on a brushless electromagnetic generator to be embedded inside a prosthetic knee [9, 10]. This research focused on the optimal combination of the magnets and the coil. The generator has six magnets and one coil. As the knee bends, the six magnets pass over the coil in sequence. Preliminary test results at 1 Hz show 2 V peak voltage over a load of 1 MΩ, but average power is not provided [9, 10].

A knee-generator shown in Fig. 4 is based on piezoelectric bimorph cantilevers [11]. A bimorph generates maximum power when it oscillates at its resonance frequency, typically ranging from several tens to hundreds of Hz, which is far higher than the frequency of the knee motion. Plectra around the circumference of the device actuate bimorph cantilevers at a higher frequency than that of the knee motion. Simulations predict generation of up to 7 mW under steady walking. Theoretical and experimental results in [11] are based on a fixed resistor load. Usable electrical energy is lower when the loss of the power management circuit (for rectification and voltage regulation) is considered [11].

A dielectric elastomer is essentially a plate capacitor whose capacitance changes with the distance between the two plates and distortion of the dielectric material. Stretching the material converts mechanical energy into electric field energy. A dielectric elastomer generator could be attached to clothing on the front

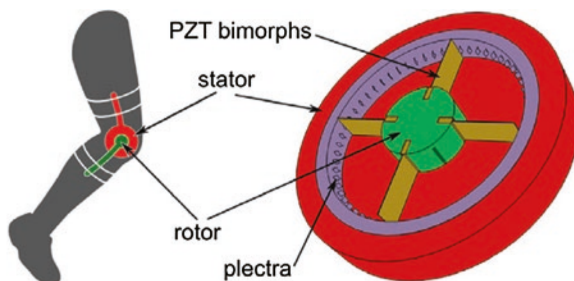


Fig. 4 Knee-mounted piezoelectric generator plucks multiple bimorphs to operate at resonance [11]. (Reproduced with permission from M. Zhu)

or back of the knee and cause very little discomfort to the wearer. Regular walking or running flexes the elastomer to generate electricity. The authors reported 10–50 μJ of energy harvested per knee extension [12].

2.3 Torso

Intensive research has been conducted on harvesting energy from human torso motion during walking or running [13–15]. An electromagnetic generator was developed to produce electricity from low-frequency human motion [13]. The device, shown in Fig. 5, consists of a magnetic ball contained in a nonconductive cage. A wire is wrapped around the cage, so that the moving ball induces AC voltage in the coils. The device can be placed anywhere on the human body such as a backpack or a pocket. It was shown that two sets of coils close to the ends of the cage (which is the right one in the figure) are more effective than one coil at the equator (the left one in the figure). The optimal number of the turns, ball diameter, and ball-to-cavity diameter ratio were determined for the particular test scenarios [13]. The electromagnetic generator achieves a power density of 0.5 mW/cm^3 with optimal parameters, but efforts to make the generator smaller dramatically reduced its power density [13].

Unlike the cage type, a linear electromagnetic generator shown in Fig. 6 has a tube with wire coils and a free-sliding magnetic mass inside the tube. Either fixed magnets or springs are attached at the ends of the tube, which make the sliding mass to oscillate [14, 15]. Two prototypes were tested on a shaker and a torso. The first prototype with fixed magnets at both ends generates 1.86 mW, while running with an oscillation frequency of 2.75 Hz. The second with only one fixed magnet

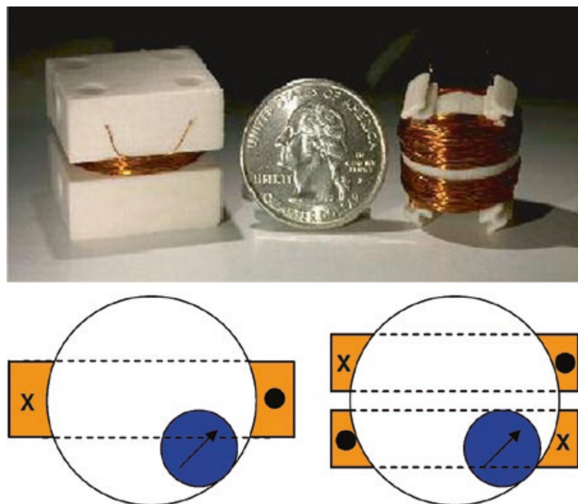


Fig. 5 Two spherical generator prototypes. The equator-wrapped coil is on the left and the offset-wrapped coil is on the right [13]. (Reproduced with permission from D.P. Arnold)

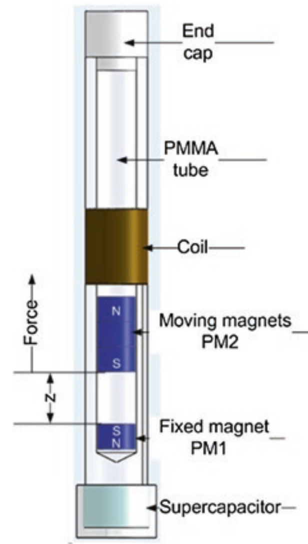


Fig. 6 Linear electromagnetic generator to convert kinetic energy into electricity [14]. (Reproduced with permission from Trans Tech Publications)

generates 2.46 mW. One of the prototypes, the paper does not specify which, attached to a torso generates average 0.9 mW for 1 h of walking, which is sufficient to power a wearable wireless sensor node [15].

Fujita et al. investigated vibration energy harvesting from forward and backward motion of the waist, wrist, and ankle using an electrostatic generator [16]. An electrostatic generator is similar to a dielectric elastomer generator, where moving electrodes convert the motion into electric field energy between the electrodes. Researchers mounted a prototype electrostatic energy harvester on the waist of a user and logged the amount of energy harvested. Over the course of an 8-minute walk across a college campus, the peak power was 0.5 nW, average power 61 pW, and the total energy extracted 32 nJ [16].

2.4 Arm

The energy level from an arm varies greatly from finger movement to forearm flexing. Typical arm movements are intermittent and low frequency, which necessitate unique transducers [17–20].

The piezoelectric shell transducer in Fig. 7 was investigated to increase the slope and magnitude of the generated impulse voltage. The piezoelectric shell structure was formed by attaching PVDF film to a curved thin polyester film, which provides a stiff structure for the piezoelectric shell generator [17]. As force is applied to a shell structure, it resists bending until it reaches a point and then all of a sudden bends quickly to generate a sharp impulse. A sharp impulse increases

Fig. 7 Plastic shell with a piezoelectric transducer [17]. (Reproduced with permission from IEEE)

the peak voltage. A higher voltage, even if it may carry the same amount of energy per pulse, reduces the power loss across diodes and resistance in the current path [17]. Test results show that the shell structure generates a peak impulse of 30 V, while a flat PVDF of only 5 V. When eight shell structures were inserted into an arm-cuff worn around a person's elbow, it exhibited peaks of 20 V, while flat PVDF films caused only 10 V peaks.

Impact-based generators were also considered to sharpen the impulse and increase the peak voltage. An impact-based generator moves a mass at a slow speed, which strikes against piezoelectric transducers. The energy of the moving mass is transferred to the transducer in a short period and then the transducer resonates after the impact. A mathematical model predicts that a 1 cm^3 device is capable of generating 40 μW on the wrist of a walking person [18]. Renaud et al. later presented a prototype generator consisting of two piezoelectric cantilevers, a free-moving mass, and a case. Such a generator could be worn on a wrist to power electronic devices. Electricity is generated by the impact of the mass on the piezoelectric cantilevers at each end of the frame. The device generates 47 μW in a resistive load at rotation of 30 rpm and 600 μW under 10 cm linear displacement at 10 Hz [19].

A piezoelectric wire, often bonded to a flexible substrate, can generate electricity by stretching and contracting. A single wire generator (SWG) is a piezoelectric wire manufactured on a polymer film. An SWG can harvest energy from low-frequency deformation such as air flow and human motion [20]. Yang et al. demonstrated SWGs harvesting energy from a finger and a live hamster. The finger-mounted SWG produces peaks of 20 mV and 150 pA from minute finger movements [20] (Fig. 8).

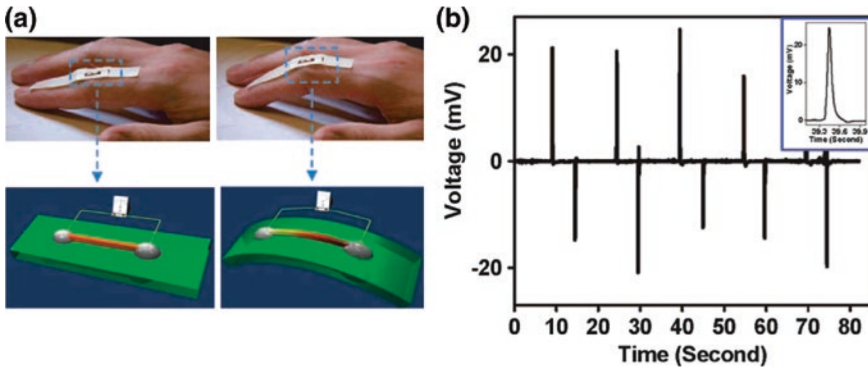


Fig. 8 Using an SWG to harvest energy from a human index finger. **a** Mechanical design and flexing. **b** Open circuit voltage [20]. (Reproduced with permission from American Chemical Society)

2.5 Breathing

Chest expansion due to breathing provides a high amount of energy. An added benefit of harvesting from respiration is that the frequency of input power is the rate of respiration, which eliminates the need for a separate respiratory rate monitor. Padasdao et al. demonstrated a prototype energy harvester strapped around a chest. It converts chest expansion into linear motion initially and then into high-speed rotational motion using a gearbox to drive an electromagnetic generator [21]. The generator is a modified servomotor, which includes a gearbox and sturdy case. The concept of the approach is shown in Fig. 9. Chest expansion from breathing pulls on the chest strap, which causes an electromagnetic generator to rotate. The performance of the prototype was measured for normal, fast, and very slow breathing. The electromagnetic generator produces 1.4 V peak to peak and 84.8 mA peak to peak or average power of 15 mW under the normal breathing condition. The high rotational speed obtained with gears attributes high efficiency to the prototype [21].

Shahhaidar et al. also developed a prototype energy harvester, where the approach is essentially the same as in the above [22]. The prototype is composed of two DC motors, a gearbox, springs, and a chest belt. The springs store energy in each expansion and keep the belt tight when the chest contracts. Despite the

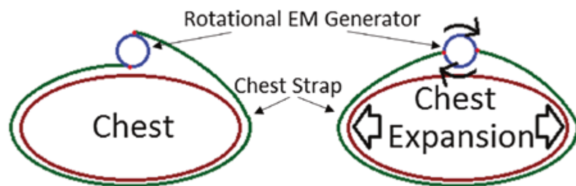


Fig. 9 Top view of chest and electromagnetic generator harvesting from respiration [21]. (Reproduced with permission from IEEE)

gearbox, the generator suffers from low rotational speed and produces 100 mV at the load. According to estimations, 0.2 Hz breathing can generate up to 30 mW. The prototype presented was able to harvest 2 mW [22].

2.6 Body Heat

Human bodies lose a tremendous amount of energy daily as heat, but harvesting a good amount of energy from body heat is challenging. The challenge stems from the fact that temperature gradient between the body and the surrounding air is small, typically 1°–2°. It is due to lack of air flow around the thermoelectric generator (TEG) worn on a body as well as a relatively small temperature difference between the body and the atmosphere. A TEG converts thermal energy into electrical energy because a temperature gradient applied to a p–n junction causes current to flow due to the Seebeck effect. Energy harvesting research from the human body focuses on three major areas: TEG design, low-voltage power electronics, and system design [23–30]. This section covers only the system design.

Several research groups focused on TEG systems to test their effectiveness and comfort for wearable electronics [25, 26]. Leonov et al. tested different textile configurations and situations to measure the system performance [30]. Figure 10 depicts the four different textile configurations, in which the power generated by a TEG (denoted as "1" in Fig. 10a) was measured. The configuration in Fig. 10a has an inner TEG metal plate touching the wearer's skin, and an outside metal plate exposed to the outside air. It would be the most effective

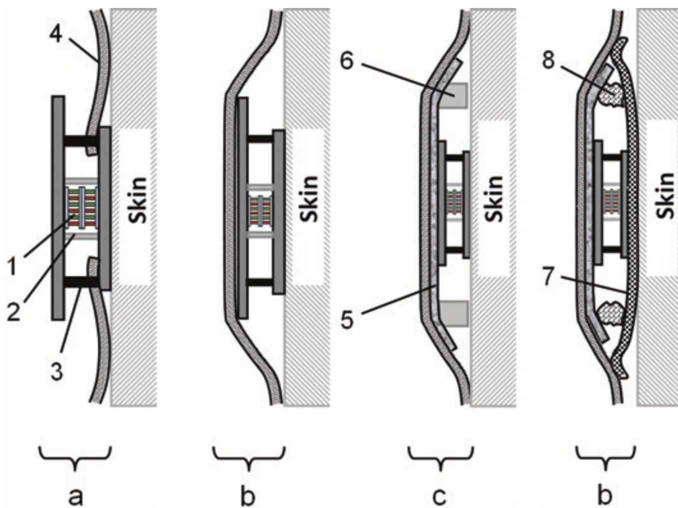


Fig. 10 Four different configurations to integrate TEGs in garments [30]. (Reproduced with permission from IEEE)

in terms of the temperature gradient, but least comfortable for the wearer. The configuration in Fig. 10b covers the outside metal plate with cotton. This causes only 5 % reduction in power, as long as there is no air gap between the cotton and the TEG. The configuration in Fig. 10c adds a carbon fiber heat-spreading layer between the cotton and the TEG. The carbon layer increases performance by approximately 30 %. The configuration in Fig. 10d adds another cotton layer between the skin and the TEG. It greatly increases comfort and completely hides the TEG, but slightly reduces performance. Fourteen units adopting the second configuration generated an average power of 4 mW while walking in a 17 °C environment [30].

3 Implant Devices

This section reviews recent research activities on energy harvesting with implant devices. Potential energy sources for harvesting include kinetic energy from the human body, body heat, fuel cells, and ambient sources specifically solar. It describes research activities for each energy source type in the following.

3.1 Knee

A knee can be exposed to force up to three times higher than the body weight [31, 32]. A typical transducer used for knee bending is a piezoelectric element or electromagnetic generator. Platt et al. investigated the use of three piezoelectric elements inside the orthopedic implants [31, 32]. A knee implant is shown in Fig. 11. The femoral component is a highly polished hard surface attached to the femur. Relative motion between the femoral and bearing surfaces allows the knee to function. The tibial tray supports a low-friction polyethylene bearing surface. Force applied by the femoral component is applied to the bearing surface and then to the three piezoelectric stacks on the distribution plate. When 900 N of force is applied over one piezoelectric stack (whose dimension is 10 × 10 × 20 mm) placed inside the prototype in a laboratory setup, it generates up to 1.6 mW of raw power, a total of 4.8 mW for the three stacks. The height of the piezoelectric stacks is 20 mm, which significantly increases the thickness of the tibial tray. Consequently, the quantity of tibial and femoral bone loss also increases during a knee arthroplasty.

Almouahed et al. investigated a new generation of a knee implant shown in Fig. 12 through modeling and experimental trials [33, 34]. The knee implant is more sophisticated than the earlier one in the above, and it uses four smaller piezoelectric elements of dimension 10 × 10 × 4 mm. The power generated is estimated to be about 1 mW for a single piezoelectric element with 50 kΩ resistive load, and the total power generated would be about 4 mW. Finally, Chen et al. reported PZT ceramics (of dimension 5 × 5 × 18 mm) and an associated circuit

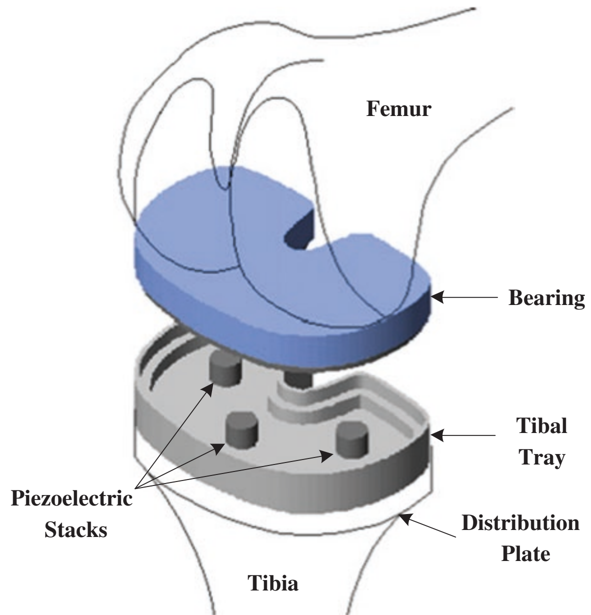


Fig. 11 Knee implant with three piezoelectric stacks [31, 32]. (Reproduced with permission from IEEE)

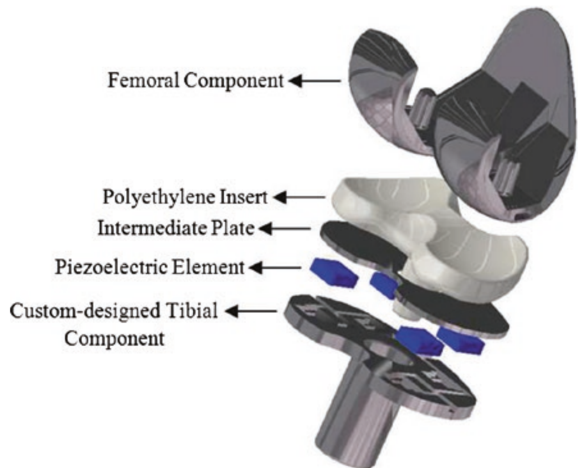


Fig. 12 Knee implant with four piezoelectric stacks [33, 34]. (Reproduced with permission from IEEE)

applicable to orthopedic implants [35]. Four piezoelectric elements can generate about 1 mW of power, which is sufficient to power a wireless monitoring system.

The above works adopt piezoelectric elements to generate electric energy from knee bending [31–35]. Luciano et al. investigated use of a miniaturized electromagnetic generator shown in Fig. 13, which can be implantable in a human knee prosthesis [9, 36]. Figure 13a depicts the upper portion of the knee (femur), in which permanent magnets are installed. Figure 13b shows a coil at the top of the

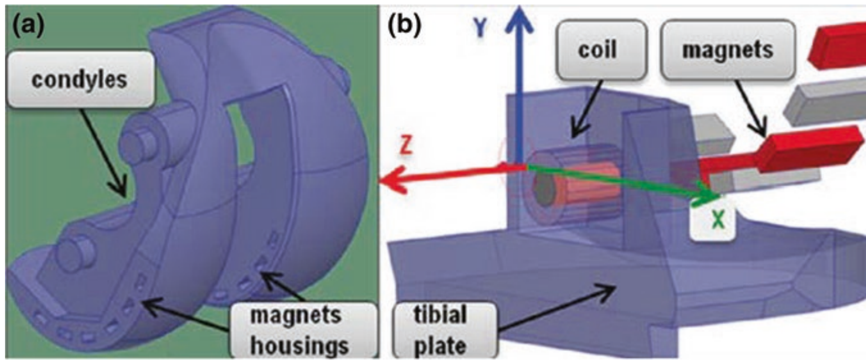


Fig. 13 Structure of the electromagnetic generator for knee prosthesis [9, 36]. (Reproduced with permission from IEEE)

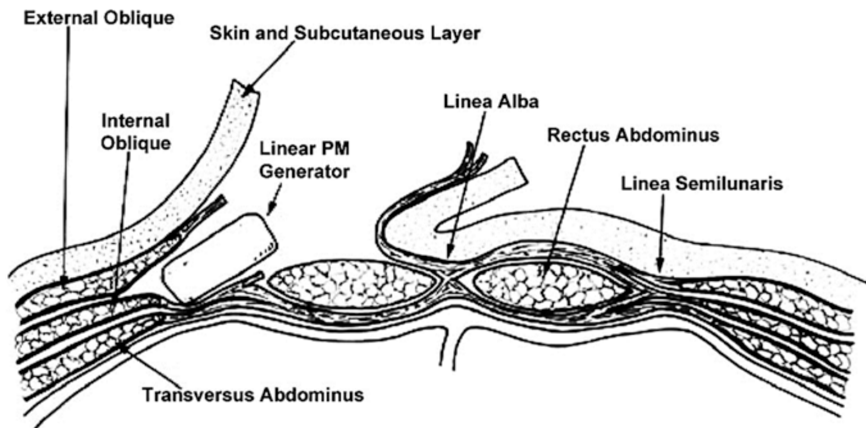


Fig. 14 Pocket for the linear permanent magnetic generator [37]. (Reproduced with permission from IEEE)

tibia or the lower part of the knee. When the knee flexes, the magnets move relative to and induce current in the coil. The system with a power conditioning circuit produces output energy of about 22.1 μJ in 7 s for a gait frequency of 1.02 Hz emulated with an electric motor.

3.2 Body Movement

Nasiri et al. investigated a linear permanent magnet generator implanted in the abdominal muscles as shown in Fig. 14 [37]. The linear generator consists of two layers of permanent magnets and one layer of coils. It generates power from

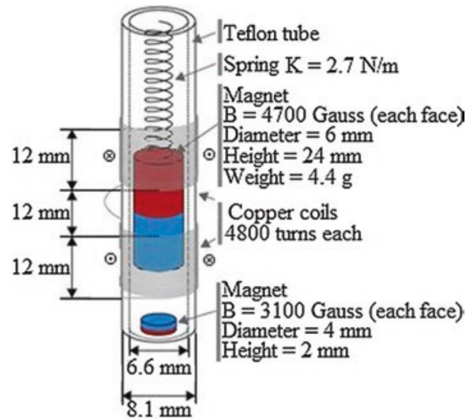


Fig. 15 Electromagnetic generator implantable in a hip prosthesis [38]. (Reproduced with permission from Elsevier)

multidirectional body movements. The generator was tested under different conditions including movements at different frequencies and amplitudes, resonance frequency test, and a walking test. The magnetic generator generates 1.1 mW of power at the resonant frequency of near 0.3 Hz. The peak output voltage drops from 1.5 V at the resonant frequency to 0.7 V at 0.1 Hz and 0.9 V at 6 Hz. It harvested 9 mJ from 22.5 s of walking motion.

Morais et al. reported a nonlinear electromagnetic generator implantable in a hip prosthesis [38]. As shown in Fig. 15, the generator consists of a Teflon tube with one or two external coils and a neodymium magnet attached to a spring. A magnetic brake is attached at the bottom of the tube to avoid collisions of the magnet against the bottom of the Teflon tube itself. This device intends to harvest energy from the human gait to power a telemetric system inserted in a smart hip prosthesis implant for early detection of loosening and implant failure. The device with a power management circuit is able to power the telemetric system for 9.2 s after charging for 34.8 s from a walking speed of 1.3 Hz.

3.3 Heart

A system based on a piezoelectric generator is reported in [39], which harvests energy from heartbeats to power cardiac implant devices. The output power follows the acceleration spectrum of heartbeats as shown in Fig. 16. The achievable power level and design parameters are determined from a spectral analysis to about 100 μW before electronics. The system is about 15 × 7 × 5 mm^3, which fits all the components as well as accommodates the proof mass travel range.

Martin et al. present a more comprehensive design, fabrication, and tests of a microspiral-shaped piezoelectric energy harvester and its associated microfabricated packaging that collects energy from ordinary blood pressure variations in

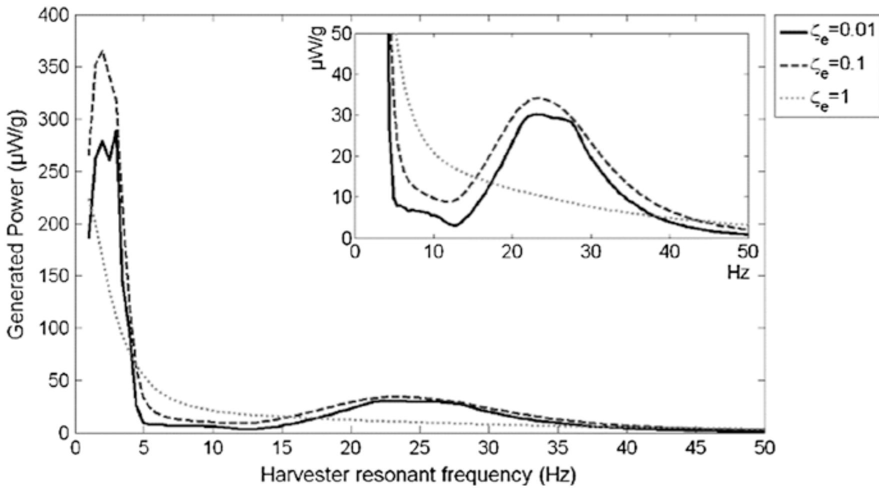


Fig. 16 Simulated output power per gram of proof mass for a piezoelectric generator as the transducer [39]. (Reproduced with permission from IEEE)

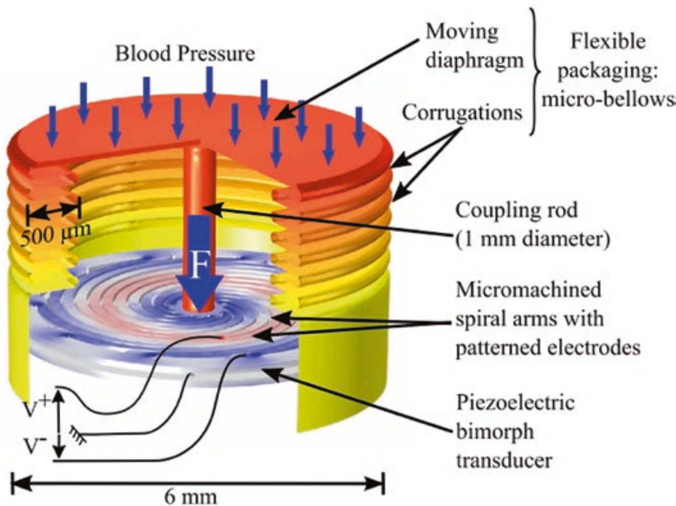


Fig. 17 Schematic of the implantable packaged energy harvester scavenging energy from blood pressure variation [40]. (Reproduced with permission from IEEE)

the cardiac environment [40], as shown in Fig. 17. A prototype of 10-μm-thin ultra-flexible electrodeposited micro-bellows (6 mm diameter, 21 mm^3 volume) is a new type of implant packaging. It enables direct blood pressure harvesting and permits high-efficiency energy transfer to a transducer operating in a quasi-static mode and hence adaptable and unaffected by frequent heartbeat frequency

changes. Spiral-shaped piezoelectric transducers are introduced for their flexibility and large incoming mechanical energy. Nontrivial optimal electrodes placement and best spiral design parameters are studied and discussed. Three types of spiral prototypes (11 mm^3 volume each) with double-sided microstructured electrode patterns are presented and characterized. A power of 3 $\mu J/cm^3$/heartbeat and a transduction efficiency of 5.7×10^{-3} have been obtained for the best design at 1.5 Hz.

An electrostatic generator harvesting energy from ventricular wall motion was developed to power a cardiac pacemaker perpetually [41], and it employs a honeycomb structure. A test setup was constructed with an accelerometer placed on a dog's heart. The same amount of acceleration sensed from the heart drives the generator. The dog's heart beats at 180 per minute, and the average power recorded for more than 2 h is 36 μW.

3.4 Artery

Potkat et al. investigated an implantable energy harvester from the expansion and contraction of an artery [42]. A PVDF thin film embedded within a flexible, self-curling medical-grade silicone cuff converts the expansion and contraction of the artery into electrical energy. The cuff is rolled around a latex tube mimicking an artery as shown in Fig. 18. A peak voltage of 1.2 V, a maximum instantaneous power of 16 nW, and an average power of 6 nW were measured. The microfabricated version of this implantable device would generate more than 1.0 μW for in vivo tests.

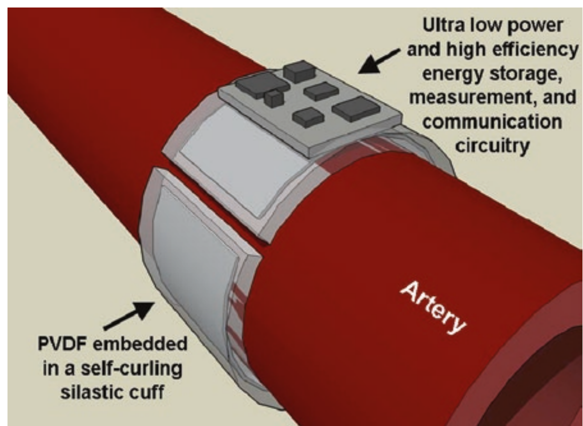


Fig. 18 PVDF thin film energy harvester for the artery [42]. (Reproduced with permission from IEEE)

3.5 Muscle

Lewandowski et al. reported an implantable piezoelectric generator to harvest energy from muscles [43]. As shown in the conceptual block diagram in Fig. 19, the device is attached in series with a muscle tendon. The system targets individuals with extensive paralysis, where the electrically stimulated muscle would not interfere with natural muscle contractions or activities. Electrically stimulated muscle contractions exert force on the piezoelectric generator producing a charge, which is collected in energy storage circuitry and used to power the stimulator and other loads. It is argued that the required power to artificially excite a muscle is minimal in comparison with the power generated by the muscle and hence such a device is justified. Experimental results indicate that a small PZT stack prototype ($5 \times 5 \times 18$ mm^3) generates 80 μW of power under application of force of 250 N.

3.6 Body Heat

Yang et al. exploited the thermal gradient between a skin surface and the inner body and harvested thermal energy with an implanted TEG as shown in Fig. 20 [44]. They performed numerical simulations on three-dimensional bioheat transfer of a human body with implanted TEGs in different depths and configurations.

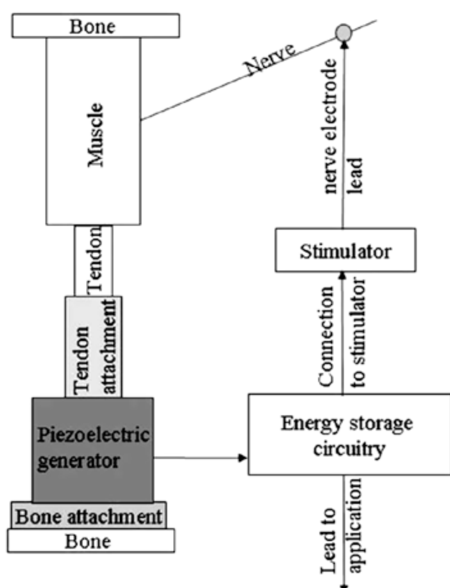


Fig. 19 Conceputal block diagram for an implanted piezoelectric generator for muscles [43]. (Reproduced with permission from Springer)

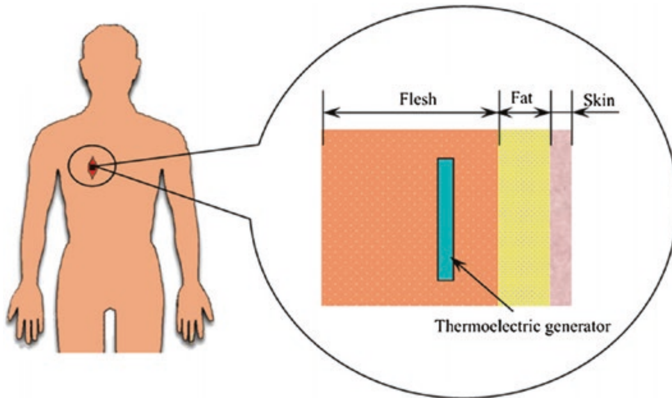


Fig. 20 Thermal energy harvesting with an implanted TEG and the model of a tissue [44]. (Reproduced with permission from J. Liu)

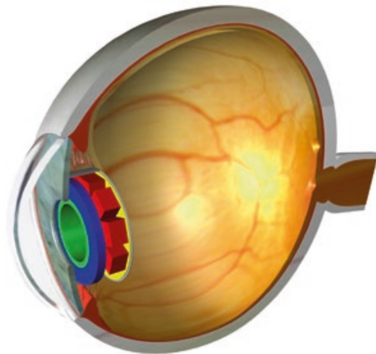


Fig. 21 Implanted artificial accommodation system within an eye [45]. (Reproduced with permission from IEEE)

To increase generation of energy, they also proposed and evaluated several approaches such that there is intentional cooling and heating of the skin surface. In their in vivo experiment, a TEG was implanted in the abdomen of a rabbit. The temperature difference stabilized at around 1.3° after 260 s with the TEG voltage of about 5 mV. When the rabbit skin surface was covered with an ice water bag, the temperature difference increased up to 5.5 °C and the TEG voltage increased to 25 mV.

Nagel et al. investigated an implanted TEG to power an artificial accommodation system shown in Fig. 21 [45]. They considered characteristics of high-performance thermoelectric materials and the temperature distribution within the human eye to estimate power generated by a TEG as a part of the artificial accommodation system. It is shown that power ranges from 4.96 μW in the worst case to 24.4 μW in the best case.

3.7 Glucose Fuel Cell

A fuel cell is an electrochemical device that generates current through the reaction of two chemicals flowing into it—the fuel on the anode site and the oxidant on the cathode site. Implantable fuel cells that use glucose as a reactant are probably the most studied biofuel cells, due to the high availability of glucose in body fluids. Glucose fuel cells can be divided into two groups: (1) abiotically catalyzed and (2) enzymatically catalyzed. The former group utilizes nonbiological catalysts such as noble metals or activated carbon. The latter group, instead, uses enzymes, such as glucose oxidase or laccase, as catalysts to enable the electrode reactions. These devices are summarized in [46]. Figure 22 shows one example of flow-through type fuel cells intended for blood stream implantation. During in vitro experiments, glucose fuel cells abiotically catalyzed can generate up to 50 μW.

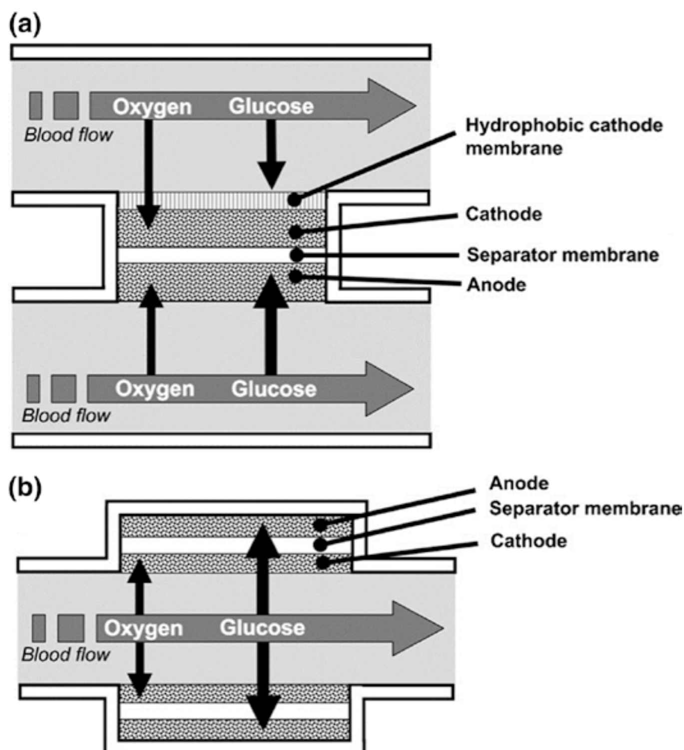


Fig. 22 Simplified schematics of flow-through type fuel cells intended for bloodstream implantation. **a** Fuel cell with hydrophobic cathode membrane. **b** Concentrically arranged fuel cell with oxygen-selective cathode catalyst [46]. (Reproduced with permission from Elsevier)

3.8 Inner Ear

Bandyopadhyay et al. investigated endocochlear potential (EP), the 70–100 mV electrochemical biopotential inside the mammalian ear shown in Fig. 22 [47]. Due to the anatomical constraints inside the inner ear, the total extractable power from the EP is limited close to 1.1–6.25 nW. An nW boost converter is used to increase the input voltage (30–55 mV) to a higher voltage (0.8–1.1 V) usable by CMOS circuits in the sensor. The power management unit can sustain itself, and a duty-cycled ultralow-power load while extracting power from the EP of a live guinea pig. The power management unit circuits have been implemented on a 0.18 μm CMOS process (Fig. 23).

3.9 Solar Energy

All the implant devices described above harvest human body energy, which comes from the food intake. In contrast, solar energy is external and hence harvesting solar energy does not affect the body. Outdoor solar energy provides the highest

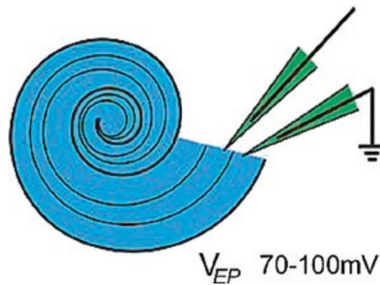


Fig. 23 Endocochlear potential [47]. (Reproduced with permission from IEEE)

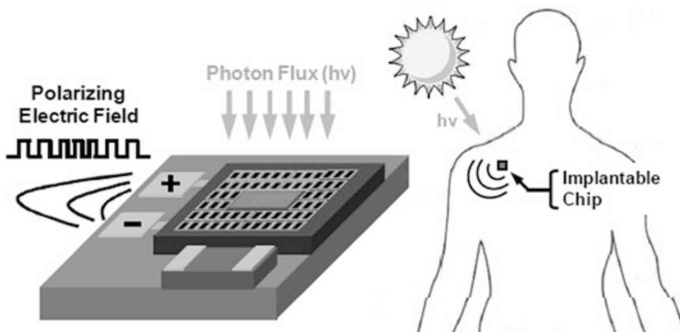


Fig. 24 Solar energy harvester [48]. (Reproduced with permission from IEEE)

power density among ambient energy sources, but the light penetration is greatly reduced by the human body tissue. A solar energy harvester for implant devices gathers the energy within the therapeutic window wavelengths, where the optical absorption is small [48]. Implanted subcutaneous photovoltaic cells shown in Fig. 24 can harvest microwatts of power in bright ambient light condition [48].

Ayzian et al. present an energy-autonomous, photovoltaic-driven, and MRI-compatible CMOS implantable sensor, as shown in Fig. 25 [49]. On-chip P +/N-well diode arrays are used as CMOS-compatible photovoltaic cells to harvest power from the light that penetrates into the tissue. In this 2.5 mm by 2.5 mm microwatt integrated system, the in vivo physiological signals are first measured by using a subthreshold ring oscillator-based sensor. The acquired data is then modulated into a frequency-shift keying signal, and finally transmitted neuromorphically to the skin surface by using a pair of polarized electrodes.

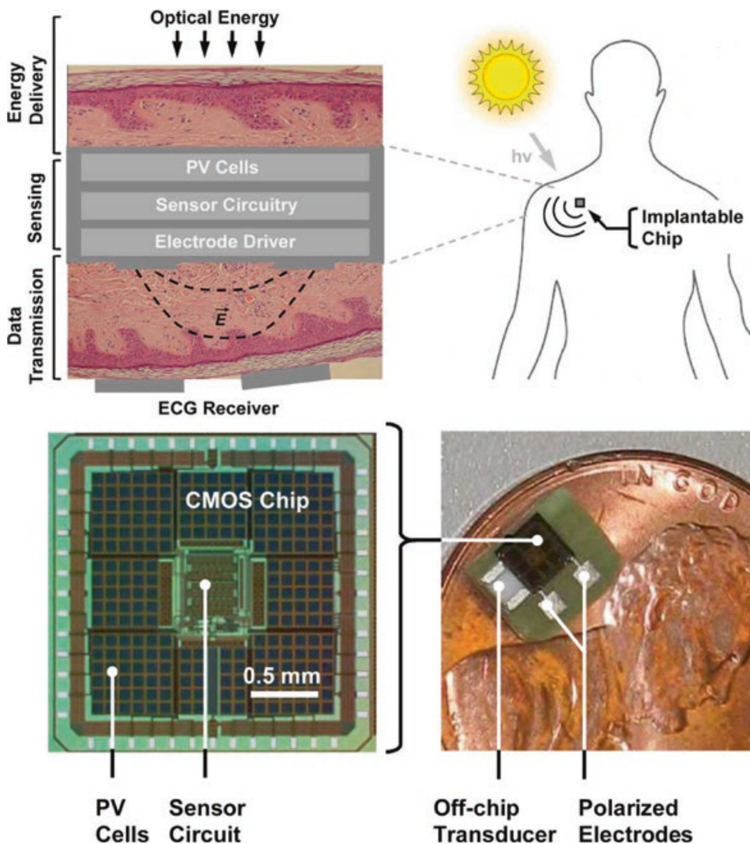


Fig. 25 A photovoltaic-driven energy-autonomous CMOS implantable sensor [49]. (Reproduced with permission from IEEE)

4 Wireless Power Transmission to Implant Devices

Wireless power transmission to implant devices can rely on various means including inductive link, RF transmission, infrared, and ultrasonic and is described in this section.

4.1 Inductive Link

The use of inductive links to power implant devices has been actively investigated in the past decade. An inductive link consists of two coils: a primary coil and a secondary coil. The primary coil placed outside the body generates a variable magnetic field, which causes the change of the magnetic flux for the secondary coil planted underneath the skin. Power is transferred wirelessly through the body tissue. The ratio of the energy captured by the secondary coil to the energy transmitted by the primary ranges between 0.01 and 0.1. The simplified model of the inductively coupled system is shown in Fig. 26. The left side of this model represents the outside components of the system and the primary coil, while the right side includes a basic model of the implanted system. Here, R_1 represents the parasitic resistance in the coil, C_1 is the tuning capacitance used to raise the coil voltage, and R_L is the load on the system. The primary coil L_1 is driven by an RF amplifier supplying current i_1 at frequency ω.

Most inductive link works reported utilize frequencies of the order of a few megahertz or lower. This frequency range minimizes the power absorption by the tissues, yielding higher transmission efficiency [50–52]. Sauer et al. in [50] present a telemetry chip shown in Fig. 27, which is powered by inductive coupling. The two coils are used to transmit both power and data. The chip fabricated in 0.5 μm CMOS technology supplies 1.7 mA at 3.3 V, over a distance up to 25 mm between coils.

Lenaerts et al. investigated an inductive link to power a wireless camera capsule for noninvasive visual inspection of the small bowel (small intestine) [51]. Up to 150 mW of usable dc power can be delivered to the capsule for the entire duration of its travel along the gastric track. Figure 28 shows parts of the power receiver. The outer dimensions of the power receiver compartment inside the capsule are 10 × 13 mm. The power efficiency of the link is measured to be 1 % under worst-case geometrical conditions.

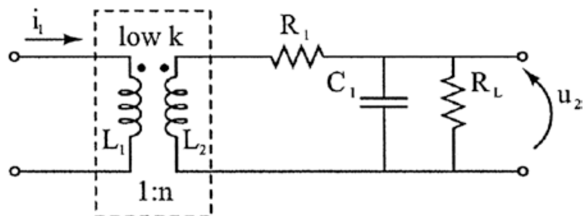


Fig. 26 Simplified model of the inductively coupled system

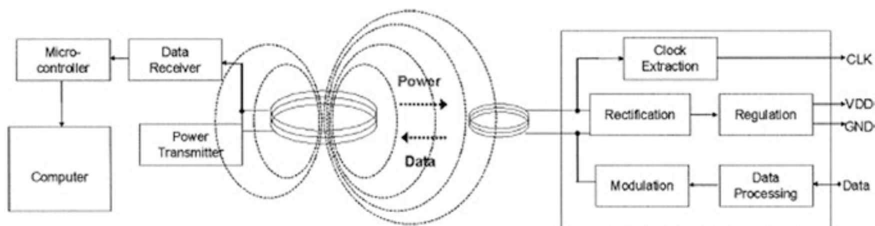


Fig. 27 An implant sensor system powered through inductive link [50]. (Reproduced with permission from IEEE)

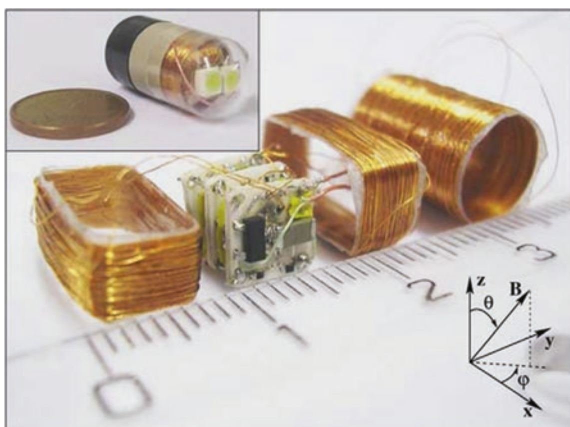


Fig. 28 The disassembled power receiver: three orthogonal coils and receiver electronics [51]. (Reproduced with permission from Elsevier)

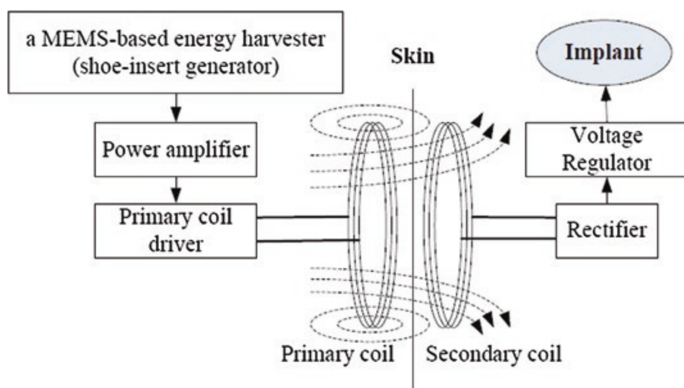


Fig. 29 Basic structure of the transcutaneous energy transmission system [52]. (Reproduced with permission from IEEE)

Niu et al. in [52] investigated a microelectromechanical system (MEMS)-based spiral piezoelectric energy harvester, which harvests kinetic energy from patient's shoes during walking. The harvested energy supplies power to an implant device, specifically an artificial heart, through inductive link as shown in Fig. 29. The regulated output power can reach up to 1 μW.

4.2 RF Transmission

Cao et al. investigated a device for gastroesophageal reflux disease (GERD) monitoring [53]. The conceptual design and sensor system are shown in Fig. 30. The system consists of an implantable, batteryless, and wireless transponder with integrated impedance and pH sensors. The transponder implant with the size of 0.4 × 0.8 × 3.8 cm harvests RF energy to operate dual-sensor and load-modulation circuitry. The external reader can store the data in a memory card and/or send it to a base station wirelessly. The device and system were tested at the bench, in a mannequin and in live pigs, which demonstrates the functionality, feasibility, accuracy, and reliability in detecting various stimulated reflux episodes.

Faul et al. developed a wireless implant lens system powered by RF and is shown in Fig. 31 [54]. The power for the implant system is derived from the 915 MHz RF signal transmitted by the reader, and the RF signal is converted into a regulated DC voltage with two-stage half-wave rectifier. A small antenna fits into an area less than 10 × 10 mm, and the system has an operating range of nearly 100 cm.

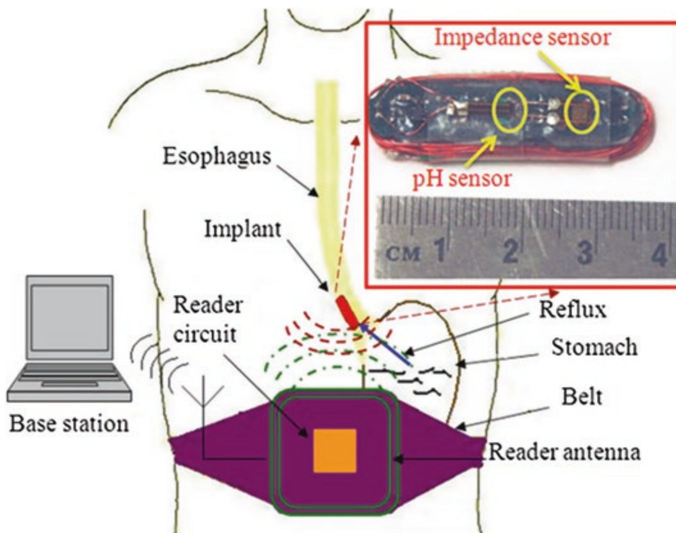


Fig. 30 Passive wireless system using dual sensors in a single capsule to monitor GERD symptoms [53]. (Reproduced with permission from IEEE)

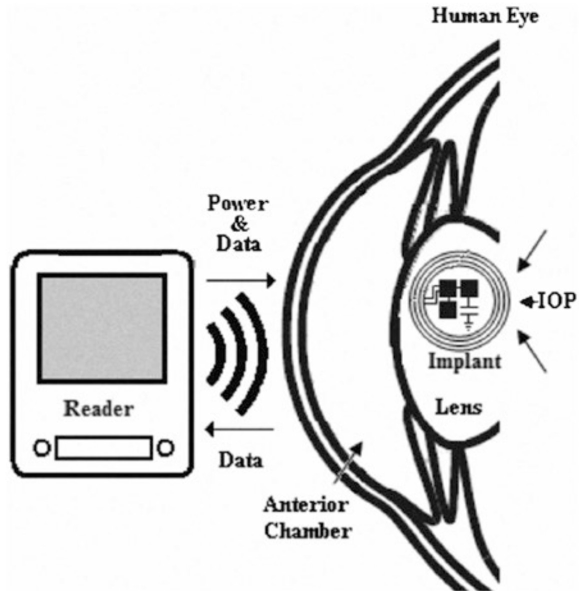


Fig. 31 External reader interfacing an intraocular (implanted) intraocular pressure sensor [54]. (Reproduced with permission from IEEE)

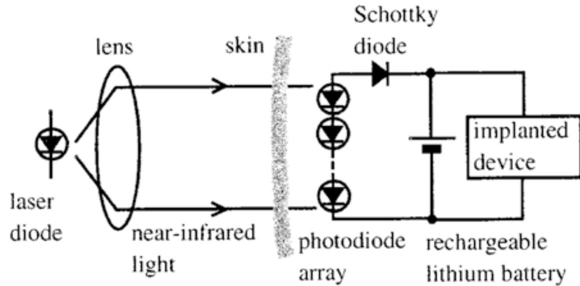


Fig. 32 Schematic diagram of the near-infrared power supply system [55]. (Reproduced with permission from IEEE)

4.3 *Infrared*

Goto et al. exploited the infrared spectrum to transmit power to implanted devices [55]. The device shown in Fig. 32 can supply power ranging from hundreds of microwatts to a few milliwatts to implanted devices. When illuminated with the power density of 22 mW/cm^2 for 17 min, the photodiode array generates sufficient power for a 20 μA cardiac pacemaker to operate for 24 h. The photothermal effect at the photodiode array raises the temperature at the irradiated skin by 1.4 °C for the power density, which does not pose any safety issues.

4.4 Ultrasonic

Ozeri et al. considered ultrasonic traveling waves to transfer power toward the receiver, whose energy is reconverted by the implanted receiver into electrical energy [56]. The recommended operating frequency is 100–1000 kHz. Piezoelectric generators convert the acoustic vibration energy into electrical power for implanted electronics devices as reported by [57–61]. Kim et al. investigated a piezoelectric energy harvester for low-frequency components of musical vibrations [57], which powers a wireless interrogation device. The system operates in two phases. A piezoelectric cantilever converts the sound vibration into electrical power and charges a capacitor. The stored charge is dumped into an LC tank, forcing it to oscillate at the resonance frequency and emit the energy to an outside receiver. A prototype transponder shown in Fig. 33 was tested using a PZT cantilever with a mechanical resonant frequency of 435 Hz encapsulated in a glass capsule (length = 40 mm, diameter = 8 mm) along with a rectifier circuitry and a storage capacitor. RF pulses can be picked up at distances of up to 7 cm without the tight requirement on alignment between the receiver and the transponder coils.

Shaul et al. proposed an ultrasonic transcutaneous energy transfer, which uses an external transmitting transducer attached to the skin surface facing an implanted receiving transducer [58]. The harvested power charges a capacitor during the charging period until it reaches a DC threshold voltage. Upon reaching the threshold level, the device starts to operate, while the capacitor being discharged. The implant sensing system shown in Fig. 34 consists of a small (2.2 × 2.3 mm) varactor, together with 10 turns of 30 gage wire wound over the implanted piezoelectric transducer.

Anthony et al. present design and characterization of a MEMS-based energy harvester with target applications including implanted biomedical sensors and actuators [59]. The harvester converts ultrasonic waves from an external transmitter to mechanical excitation and then the vibration of a central mass structure

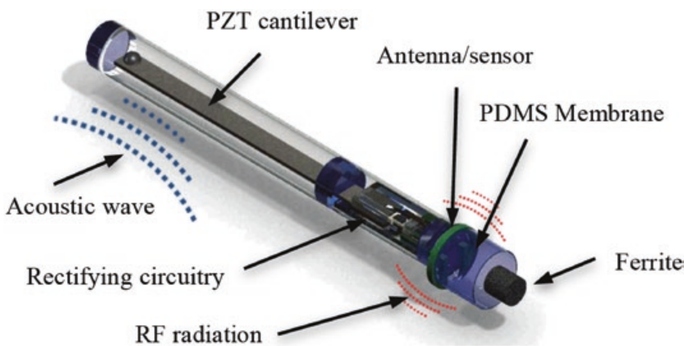


Fig. 33 3-D drawing of implanted inductive pressure sensor with acoustic energy harvesting [57]. (Reproduced with permission from IEEE)

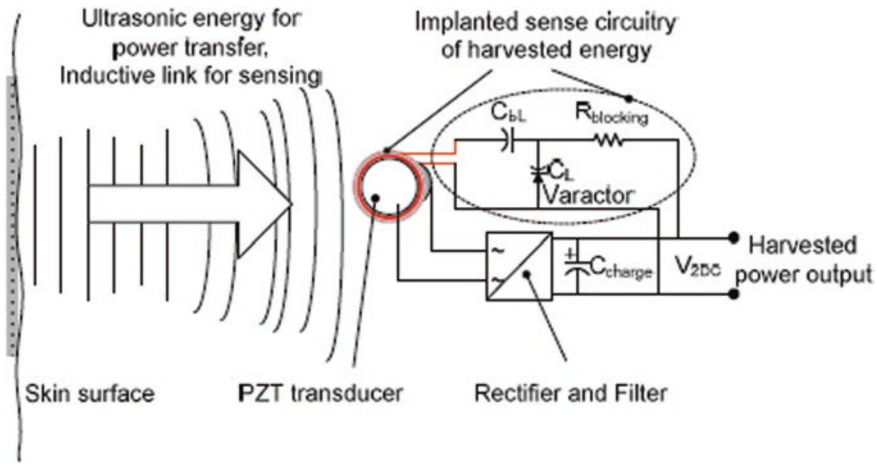


Fig. 34 Ultrasonic transcutaneous energy transfer system [58]. (Reproduced with permission from IEEE)

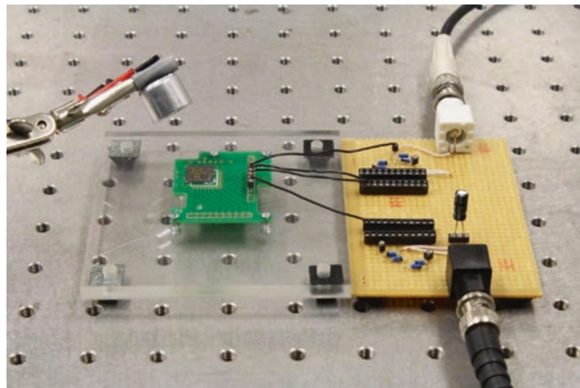


Fig. 35 Ultrasonic transcutaneous energy transfer system [59]. (Reproduced with permission from IEEE)

into electrical energy with an electrostatic transducer. The device features a novel 3 degrees of freedom design, which enables the harvester in any orientation to produce energy. The harvester is fabricated using a conventional silicon-on-insulator MEMS process. An example of the experimental setup is shown in Fig. 35. The experimental results show that the system is able to generate 24.7, 19.8, and 14.5 nW via the device's x-, y-, and z-axis, respectively, over 15 s.

Francesco et al. present the design of an ultrasound link for deep implanted medical devices [60]. To avoid biological side effects such as cavitations, the operating frequency of 1 MHz is set for the system, and a class-E power amplifier as the front-end circuit drives external transducers. A novel synchronous rectifier is proposed to maximize energy extraction from the implanted transducer. The link

efficiency is characterized in water for a transmitter–receiver distance of 105 mm, and a system efficiency of 2.3 % is achieved without using any phantom material. The link efficiency drops to 1.6 % when a phantom material is used.

5 Summary

The first part of this article reviews recent research activities on wearable energy harvesters, in which the target energy sources are limited to human bodies. The latter part of the article reviews the means to power implant devices. An implant device can be powered by harvesting energy from the human body, ambient sources, or through power transmission from external sources. Harvestable human body energy includes voluntary motion, involuntary motion, and heat. Each type of energy harvesting faces different technical challenges.

5.1 Energy Harvesting with Wearable Devices

Harvestable energy sources of a human body with wearable devices include foot, knee, torso, chest, and body heat. A comparison of power harvested from different body parts or types of energy (kinetic and thermal) is shown in Table 1. The power figure in the table is the maximum power obtained for each harvesting method, i.e., the highest power reported in a paper for the transducer type. The table indicates that electromagnetic generators are most effective for

Table 1 Summary of human energy harvesting

Source	Harvesting method	Power
Foot	Rotational electromagnetic harvester in heel [2]	1.8 W
	Linear electromagnetic harvester in shoe [6, 7]	830 mW
	Piezoelectric in heel [3–5]	8.4 mW
	Piezoelectric in sole [3, 5]	1.3 mW
	Electrostatic on ankle [16]	61 pW
Knee	Electromagnetic deceleration only [8]	4.8 W
	Piezoelectric pinwheel [11]	7 mW
	Dielectric elastomer [12]	25 μW
Torso	Spherical electromagnetic [13]	1.44 mW
	Linear electromagnetic [14, 15]	2.46 mW
	Thermoelectric [30]	4 mW
Arm	Impact-based piezoelectric [18]	40 μW
Chest	Respiratory rotary electromagnetic [21]	15 mW
Wrist	Thermoelectric [22–25]	250 μW

kinetic energy harvesting, followed by piezoelectric transducers of the order of milliwatts and dielectric elastomers of the order of tens of microwatts down to picowatts. Thermoelectric generator harvesting varies greatly depending on the environment and application ranging from hundreds of microwatts to a few milliwatts.

5.2 Powering Implant Devices

Implant devices can be powered through energy harvesting or transmission of power from external sources. Harvestable energy sources to power up implant devices include knee, heart, artery, muscle, body heat, and solar. Table 2 compares the maximum power reported for each harvesting method. The piezoelectric harvesters seem most promising in terms of the power produced. In particular, the harvesters in [31–35] generate power greater than 1 mW. However, these devices intended for the knees require significant force to generate power. Electromagnetic harvesters produce lower power than the above-mentioned piezoelectric devices, but they do not require large forces, and hence are applicable for any body part. The electrostatic harvesters produce much lower power than electromagnetic harvesters and are limited to low-power devices. Thermoelectric harvesters should be implanted close to the skin surface, where the temperature gradient is large. Large difference in the reported power is due to different sizes and thermal gradients used for the testing.

Table 2 Comparison of various harvesting techniques for implant devices

Source	Transducer	Power	Size
Knee	Piezoelectric [31–32]	4.8 mW	10 × 10 × 20 mm
	Piezoelectric [33–34]	7.2 mW	10 × 10 × 4 mm
	Piezoelectric [35]	1 mW	5 × 5 × 18 mm
	Electromagnetic [9,36]	22.1 μJ	
Body	Electromagnetic [37]	1.1 mW	1.7 in^3
	Electromagnetic [38]	1.9 mJ	3.76 cm^3
Heart	Piezoelectric [39–40]	100 μW	15 × 7 × 5 mm
	Electrostatic [41]	36 μW	50 × 30 × 30 μm
Artery	Piezoelectric [42]	1 μW	
Muscle	Piezoelectric [43]	80 μW	5 × 5 × 18 mm
Body heat	Thermoelectric [44]	25 mV	5 × 5 × 10 mm
	Thermoelectric [45]	24.4 μW	58.9 mm^2
Fuel cell	Glucose fuel cell [46]	50 μW	
Inner ear	None [47]	1 μW	25 mm^2
Solar	Photovoltaic [48]	1 μW	25 mm^2
	Photovoltaic [49]	30 nW	2.5 × 2.5 mm

Table 3 Comparison of various power transmission techniques for implant devices

Source	Harvesting/Transmission method	Power	Size
Inductive link	Coils coupling [50]	5.6 mW	0.1 mm^2
	Coils coupling [51]	150 mW	10 × 13 mm
	Piezoelectric and coils coupling [53]	1 μW	
RF Transmission	RF Transmission [53]		4 × 8 × 38 mm
	RF Transmission [54]		10 × 10 mm
Infrared	Photovoltaic [55]	46 mW	2.1 cm^2
Ultrasonic	Piezoelectric [57]	50 mV	8 × 8 × 40 mm
	Piezoelectric and RF transmission [58]	50 mW	
	Electrostatic transducers [69]	20 nW	
	External transducers [60, 61]	25 mW	0.114 mm^2

Implant devices can be powered up through various power transmission means including inductive link, RF transmission, infrared, and ultrasonic. The delivered power level of wireless power transmission methods is sensitive to the distance between the transducer and the implant. Table 3 compares various methods covered in this article. The inductive link power transmission reported in [51] delivers largest power, but its size is large. The infrared method may cause the skin temperature to increase, which should be carefully examined for safety and long-term effects. Overall, each method has unique characteristics and technical challenges to address.

5.3 Final Remarks

Modern research in energy harvesting from the human body is a broad field and is making steady progress in several areas. Most energy harvesting research focus on increasing the power level to harvest, and it will continue and be accomplished by improving transducers and power management circuits. Wearer's comfort during energy harvesting is another important topic for wearable energy harvesters. Transducer design for comfort relies on flexible materials and unobtrusive hardware. Increased comfort can also be attained by using wireless sensor networks instead of wired sensors.

Implant devices will be pervasive as the wellbeing of humans becomes ever important. Powering implant devices is a major issue for current and future implant devices. Energy harvesting and wireless power transmission will be pursued to address the problem. Another important issue for energy harvesters and power transmission devices for implant devices is health hazard, and it should be addressed before deployment of any such devices.

Acknowledgments This work is supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT and Future Planning as the Global Frontier Project.

References

1. Wei X, Liu J (2008) Power sources and electrical recharging strategies for implantable medical devices. Front Energy Power Eng Chin 2:1–13
2. Verma NK, Singla P, Roy A (2012) Energy harvesting by foot-propelled battery charger using shoe-model. Adv Mater Res 488–489:1268–1273
3. Shenck NS, Paradiso JA (2001) Energy scavenging with shoe-mounted piezoelectrics. IEEE Micro 21(3):30–42
4. Orecchini G, Yang L, Tentzeris MM, Roselli L (2011) Wearable battery-free active paper printed RFID tag with human-energy scavenger. In: Microwave Symposium Digest (MTT), 2011 IEEE MTT-S International, pp 1–4
5. Kymissis J, Kendall C, Paradiso J, Gershenfeld N (1998) Parasitic power harvesting in shoes. In: Second International Symposium on Wearable Computers, Digest of Papers, pp 132–139
6. Zeng P, Chen H, Yang Z, Khaligh A (2011) Unconventional wearable energy harvesting from human horizontal foot motion. In: Applied Power Electronics Conference and Exposition (APEC), 2011 Twenty-Sixth Annual IEEE, pp 258–264
7. Stamenkovic I, Milivojevic N, Zheng C, Khaligh A (2010) Three phase linear permanent magnet energy scavenger based on foot horizontal motion. In: Applied Power Electronics Conference and Exposition (APEC), 2010 Twenty-Fifth Annual IEEE, pp 2245–2250
8. Donelan JM, Li Q, Naing V, Hoffer JA, Weber DJ, Kuo AD (2008) Biomechanical energy harvesting: generating electricity during walking with minimal user effort. Science 319(5864):807–810
9. Luciano V, Sardini E, Serpelloni M, Baronio G (2012) Analysis of an electromechanical generator implanted in a human total knee prosthesis. In: Sensors Applications Symposium (SAS), 2012 IEEE, pp 1–5
10. Luciano V, Sardini E, Serpelloni M (2014) An Electromechanical generator implanted in human total knee prosthesis. In: Baldini F, D'Amico A, Natale CD, Siciliano P, Seeber R, Stefano LD, Bizzarri R, Andò B (eds) Sensors. Springer, New York, pp 25–30
11. Pozzi M, Zhu M (2011) Plucked piezoelectric bimorphs for knee-joint energy harvesting: modelling and experimental validation. Smart Mater Struct 20(5):055007
12. Lai H, Tan CA, Xu Y (2011) Dielectric elastomer energy harvesting and its application to human walking. ASME Conference Proceedings 2011(54884):601–607
13. Bowers BJ, Arnold DP (2009) Spherical, rolling magnet generators for passive energy harvesting from human motion. J Micromech Microeng 19(9):094008
14. Sun K, Liu GQ, Xu XY (2011) No-Load Analysis of Permanent Magnet Spring Nonlinear Resonant Generator for Human Motion Energy Harvesting. Appl Mech Mater 130–134:2778–2782
15. Saha CR, O'Donnell T, Wang N, McCloskey P (2008) Electromagnetic generator for harvesting energy from human motion. Sens Actuators, A 147(1):248–253
16. Fujita T, Onishi T, Fujii K, Kanda K, Maenaka K, Higuchi K (2012) Evaluation of the human vibration for autonomous power source. In: World Automation Congress (WAC), pp 1–4
17. Yang B, Yun K-S (2011) Efficient energy harvesting from human motion using wearable piezoelectric shell structures. In: Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS), 2011 16th International, pp 2646–2649
18. Renaud M, Sterken T, Fiorini P, Puers R, Baert K, van Hoof C (2005) Scavenging energy from human body: design of a piezoelectric transducer. In: The 13th International Conference on Solid-State Sensors, Actuators and Microsystems, Digest of Technical Papers, TRANSDUCERS '05, vol 1, pp 784–787

19. Renaud M, Fiorini P, van Schaijk R, van Hoof C (2009) Harvesting energy from the motion of human limbs: the design and analysis of an impact-based piezoelectric generator. Smart Mater Struct 18(3):035001
20. Yang R, Qin Y, Li C, Zhu G, Wang ZL (2009) Converting biomechanical energy into electricity by a muscle-movement-driven nanogenerator. Nano Lett 9(3):1201–1205
21. Padasdao B, Boric-Lubecke O (2011) Respiratory rate detection using a wearable electromagnetic generator. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp 3217–3220
22. Shahhaidar E, Boric-Lubecke O, Ghorbani R, Wolfe M (2011) Electromagnetic generator: as respiratory effort energy harvester. In: IEEE Power and Energy Conference at Illinois (PECI), pp 1–4
23. Wang Z, Leonov V, Fiorini P, Van Hoof C (2007) Micromachined thermopiles for energy scavenging on human body. In: Solid-State Sensors, Actuators and Microsystems Conference, 2007, TRANSDUCERS 2007, International, pp 911–914
24. Wang Z, Leonov V, Fiorini P, Van Hoof C (2009) Realization of a wearable miniaturized thermoelectric generator for human body applications. Sens Actuators A 156(1):95–102
25. Jung S, Lauterbach C, Strasser M, Weber W (2003) Enabling technologies for disappearing electronics in smart textiles. In: IEEE International Solid-State Circuits Conference (ISSCC), Digest of Technical Papers, vol 1, pp 386–387
26. Jo SE, Kim MK, Kim MS, Kim YJ (2012) Flexible thermoelectric generator for human body heat energy harvesting. Electron Lett 48(16):1013–1015
27. Koplow M, Chen A, Steingart D, Wright PK, Evans JW (2008) Thick film thermoelectric energy harvesting systems for biomedical applications. In: 5th International Summer School and Symposium on Medical Devices and Biosensors (ISSS-MDBS), pp 322–325
28. Zhang F, Zhang Y, Silver J, Shakhsheer Y, Nagaraju M, Klinefelter A, Pandey J, Boley J, Carlson E, Shrivastava A, Otis B, Calhoun B (2012) A batteryless 19 μW MICS/ISM-band energy harvesting body area sensor node SoC. In: IEEE International Solid-State Circuits Conference (ISSCC), Digest of Technical Papers, pp 298–300
29. Ramadass YK, Chandrakasan AP (2011) A Battery-Less thermoelectric energy harvesting interface circuit with 35 mV startup voltage. IEEE J Solid-State Circuits 46(1):333–341
30. Leonov V (2013) Thermoelectric energy harvesting of human body heat for wearable sensors. IEEE Sens J 13(6):2284–2291
31. Platt SR, Farritor S, Haider H (2005) On low-frequency electric power generation with PZT ceramics. IEEE/ASME Trans Mechatronics 10(2):240–252
32. Platt SR, Farritor S, Garvin K, Haider H (2005) The use of piezoelectric ceramics for electric power generation within orthopedic implants. IEEE/ASME Trans Mechatronics 10(4):455–461
33. Almouahed S, Gouriou M, Hamitiouche C, Stindel E, Roux C (2011) The use of piezoceramics as electrical energy harvesters within instrumented knee implant during walking. IEEE Trans Mechatron 16(5):799–807
34. Almouahed S, Gouriou M, Hamitiouche C, Stindel E, Roux C (2011) Design and evaluation of instrumented smart knee implant. IEEE Trans Biomed Eng 58(4):971–982
35. Chen H, Liu M, Jia C, Wang Z (2009) Power harvesting using PZT ceramics embedded in orthopedic implants. IEEE Trans Ultrason Ferroelectr Freq Control 56(9):2010–2014
36. Luciano V, Sardini E, Serpelloni M, Baronio G (2014) An energy harvesting converter to power sensorized total human knee prosthsis. J IOPscience Meas Sci Technol 25(2)
37. Nasiri A, Zabalawi SA, Jeutter DC (2011) A linear permanent magnet generator for powering implanted electronic devices. IEEE Trans Power Electon 26(1):192–199
38. Morais R, Silva NM, Santos PM, Frias CM, Ferreira JAF, Ramos AM, Simoes JAO, Baptista JMR, Reis MC (2011) Double permanent magnet vibration power generator for smart hip prosthesis. J Sens Actuators Phys 172(1):259–268
39. Deterre M, Boutaud R, Dalmolin R, Bosseau S, Chaillout JJ, Lefeuvre E, Gergram ED (2011) Energy harvesting system for cardiac implant applications. In: Proceeding of Design, Test, Integration and Packaging of MEMS (DTIP), pp 387–391

40. Deterre M, Lefeuvre E, Zhu Y, Woytasik M, Boutaud B, Molin RD (2014) Micro blood pressure energy harvester for intracardiac pacemaker. J Microelectromech Syst 23(3):651–660
41. Tashiro R, Kabei N, Katayama K, Tsuboi F, Tsuchiya K (2002) Development of an electrostatic generator for a cardiac pacemaker that harnesses the ventricular wall motion. Artif Organs 5:239–245
42. Potkat JA, Brooks K (2008) An arterial cuff energy scavenger for implanted microsystems. In: Proceedings of Bioinformatics and Biomedical Engineering (ICBBE), pp 1580–1583
43. Lewandowski BE, Kilgore KL, Gustafson KJ (2007) Design considerations for an implantable, muscle powered piezoelectric system for generating electrical power. Ann Biomed Eng 35(4):631–641
44. Yang Y, Wei X, Liu J (2007) Suitability of a thermoelectric power generator for implantable medical electronic devices. J Phys D Appl Phys 40:5790–5800
45. Nagel JA, Sieber I, Gengenbach U, Guth H, Bretthauer G, Guthoff RF (2010) Investigation of a thermoelectric power supply for the artificial accommodation system. In: Proceeding of Applied Sciences in Biomedical and Communication Technologies (ISABEL), pp 1–5
46. Kerzenmacher S, Ducre J, Zengerle R, Stetten FV (2008) Energy harvesting by implantable abiotically catalyzed glucose fuel cells. J Power Sources 182(1):1–17
47. Bandyopadhyay S, Mercier PP, Lysaght AC, Stankovic KM, Chandrakasan AP (2014) A 1.1 nW Energy-Harvesting System with 544 pW Quiescent Power for Next-Generation Implants. IEEE J Solid-State Circuits 49(12):2812–2824
48. Ayazian S, Hassibi A (2011) Delivering optical power to subcutaneous implanted devices. In: Proceeding of IEEE Engineering in Medicine and Biology Society (EMBC), pp 2874–2877
49. Ayazian S, Akhavan VA, Soenen E, Hassibi A (2012) A photovoltaic-driven and energy-autonomous CMOS implantable sensor. IEEE Trans Biomed Circuits Syst 6(4):336–343
50. Sauer C, Stanacevic M, Cauwenberghs G, Thakor N (2005) Power harvesting and telemetry in CMOS for implanted devices. IEEE Trans Biomed Circuits Syst 52(12):2605–2613
51. Lenaerts B, Puers R (2007) An inductive power link for a wireless endoscope. Biosens Bioelectron 22(7):1390–1395
52. Niu Q, Wang L, Dong T, Yang H (2009) Application of MEMS-based energy harvester for artificial heart wireless energy transmission. In: Proceeding of Computing, Communication, Control, and Management (CCCM), pp 38–41
53. Cao H, Landge V, Tata U, Seo Y, Rao S, Tand S et al (2012) An implantable, batteryless, and wireless capsule with integrated impedance and pH sensors for gastroesophageal reflux monitoring. IEEE Trans Biomed Eng 59(11):3131–3139
54. Faul A, Turner M, Naber J (2011) Implantable wireless microsystems for the measurement of intraocular pressure. In: Proceeding of IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Aug 2011, pp 1–4
55. Goto K, Nakagawa T, Nakamura O, Kawata S (2001) An implantable power supply with an optical rechargeable lithium battery. IEEE Trans Biomed Eng 48(7):830–833
56. Ozeri S, Shmilovitz D (2010) Ultrasonic transcutaneous energy transfer for powering implanted devices. Ultrasonics 50(6):556–566
57. Kim A, Maleki T, Ziaie B (2012) A novel electromechanical interrogation scheme for implantable passive transponders. In: Proceedings of IEEE MEMS, pp 31–34
58. Shaul O, Boaz S, Doron S (2012) Non-invasive sensing of the electrical energy harvested by medical implants powered by an ultrasonic transcutaneous energy transfer link. In: Proceeding of IEEE Industrial Electronics (ISIE), pp 1153–1157
59. Fowler AG, Moheimani SOR, Behrens S (2014) An omnidirectional MEMS ultrasonic energy harvester for implanted devices. J Microelectromech Syst 23(6):1454–1462
60. Mazzilli F, Lafon C, Dehollain C (2014) A 10.5 cm ultrasound link for deep implanted medical devices. IEEE Trans Biomed Circuits Syst 8(5):738–750
61. Mazzilli F, Thoppay PE, Praplan V, Dehollain C (2012) Ultrasound energy harvesting system for deep implanted-medical-devices (IMDs). In: Proceeding of IEEE International Symposium on Circuits and Systems (ISCAS), pp 2865–2868

Reconfigurable Photovoltaic Array Systems for Adaptive and Fault-Tolerant Energy Harvesting

Naehyuck Chang, Massoud Pedram, Hyung Gyu Lee, Yanzhi Wang and Younghyun Kim

Abstract This chapter introduces a reconfigurable photovoltaic (PV) cell array for adaptive and fault-tolerant energy harvesting in view of component modeling, architectures, properties, and reconfigurable algorithms for partial shading and fault tolerance. On top of traditional PV cell array-based energy harvesting research, the dynamically reconfigurable PV cell array gives additional significant benefits in both efficiency and cost. This is a representative example of how electronics design automation contributes to various problems in other domains.

Keywords Photovoltaic system · Solar cell · Reconfiguration · Efficiency · Fault tolerance

N. Chang (✉)
Korea Advanced Institute of Science and Technology, Daejeon, Korea
e-mail: naehyuck@kaist.ac.kr

M. Pedram · Y. Wang
University of Southern California, Los Angeles, CA, USA
e-mail: pedram@usc.edu

Y. Wang
e-mail: yanzhiwa@usc.edu

H.G. Lee
Daegu University, Daegu, South Korea
e-mail: hglee@daegu.ac.kr

Y. Kim
Purdue University, West Lafayette, IN, USA
e-mail: yhkim1@purdue.edu

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_6

1 Introduction

Due to increasing demand for energy sources and environmental concerns about fossil fuels as well as self-sustainable and maintenance-free operations, there has been a growing demand for energy harvesting power supply systems. There are ranges of energy harvesting sources from microwatt to megawatt or higher power capacity applications. Photovoltaic (PV) energy generation has received significant attention due to relatively higher energy density from milliwatts to megawatts among various energy harvesting power sources. The PV cell efficiency ranges from 20 to 30 % in commercial products and exhibits even a higher efficiency in the lab. PV cell power generation is largely affected by environmental factors such as time of day, season, weather, etc. However, short-term PV cell power generation stability is superior to other renewable power sources such as windmills. PV power generation does not require moving parts subject to wear and tear, noise, and vibration, etc., which is crucial for portable and long-term applications. This chapter focuses on PV power generation for small power applications below a kW order.

Thanks to extensive research efforts on PV power generation technologies, various scales of PV systems have been deployed for practical applications, such as PV power stations, solar-powered vehicles, and solar power heating and lighting appliances. This chapter presents the power models of PV energy harvesting system subcomponents, maximum power point tracking (MPPT), and maximum power transfer tracking (MPTT), reconfigurable PV cell arrays for partial shading and fault tolerance. This chapter also introduces a light-weight PV energy harvesting system that does not require power converters or energy storage elements while it performs MPPT.

1.1 Structure of PV Energy Harvesting Systems

The PV energy harvesting systems are comprised of a PV cell array, a load device, and a power converter between the PV cell array and the load device. There is an optional energy storage element between the PV cell array and the load device, and a power converter between the energy storage element and load device. Figure 1 illustrates two simplified PV energy harvesting system examples. Figure 1a does not include an energy storage element, and thus the load device is directly powered by the PV cell. There is a power converter between the PV cell array and the load device because it is generally hard to match the PV cell array output voltage with the load device supply voltage. Assuming that load device current is independent of the solar irradiance and temperature, which is a typical situation, the PV cell array output current is independent of the solar irradiance and temperature accordingly. It is crucial to control the PV cell array output current so that the PV cells output power is maximized. This is the so-called MPPT, which is

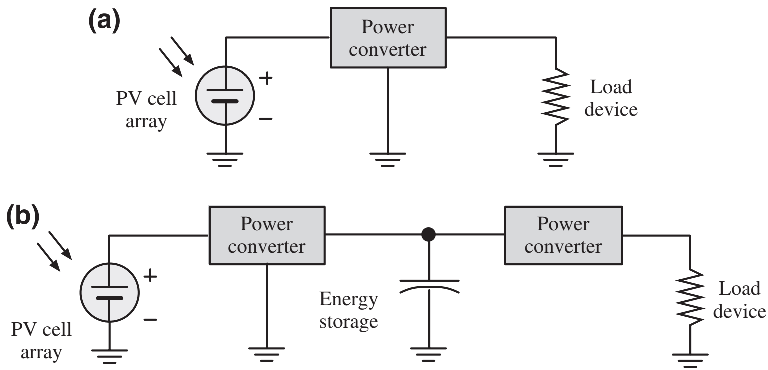


Fig. 1 Simplified diagrams of PV energy harvesting systems. **a** A PV energy harvesting system without an energy storage. **b** A PV energy harvesting system with an energy storage

explained in Sect. 2.3.1. To make a long story short, it is hard to realize the MPTT with the setup of Fig. 1a due to the mismatch between the load current and the maximum power point (MPP) PV cell array current. This chapter introduces a new MPTT with the architecture of Fig. 1a in Sect. 4.

In order to maintain the operating point at the MPP regardless the load current, we commonly use the architecture in Fig. 1b for the MPTT of the PV cell array though the load device does not have to operate when there is no solar irradiance. The energy storage element in Fig. 1b is, of course, useful to make the PV energy harvesting system functional even if there is solar irradiance interruption. Most energy storage elements exhibit variable terminal voltage by the state of charge and the load current. The nominal voltage of batteries is determined by the battery chemistry, and the terminal voltage of a capacitor is a linear function of the state of charge, which does not match with the load device power supply voltage in general. Therefore, there should be another power converter between the energy storage element and load device.

The power converter between the PV cell array and energy storage element keeps tracking of the PV cell array MPP current. In other words, the charging current to the energy storage element is not constant at all times unlike a battery charger plugged into the wall. The charging current keeps track of the MPP current of the PV cell array around the clock. We call this power converter an *MPP charger*. The power converter between the energy storage element and the load device is commonly a voltage regulator. We prefer to use switching-mode DC-DC converters both for the MPP charger and the voltage regulator for high conversion efficiency. Section 2.2 covers power efficiency and models of switching-mode DC-DC converters.

The energy storage element should provide a desired nominal voltage, nominal capacitance, minimum cycle efficiency, minimum cycle life, energy density, power density, form factor, cost, etc. Rechargeable batteries are commonly used considering these requirements. Lithium-ion batteries are preferred for high-performance

systems, and lead-acid batteries are considered for cost-efficient systems. One of the primary design goals of PV energy harvesting systems is low maintenance cost. The batteries in the PV energy harvesting systems become a primary cause of the limited lifetime. Electric double-layer capacitors (supercapacitors) are promising replacements of rechargeable batteries, thanks to their virtually unlimited cycle life. There are distinct shortcomings of supercapacitors, but their advantages often override their drawbacks in particular applications.

1.2 Design Consideration and Runtime Management

A PV cell is a semiconductor device that produces relatively low open-circuit and MPP voltages. The MPP current is a function of the PV cell area, and there are cost-effective sizes of the PV cell for mass production. Aside from a very small, low-power applications, the load device supply voltage and current are generally higher than a PV cell MPP voltage and current. This makes the PV energy harvesting systems use an array of PV cells instead of a single PV cell. The power converter, MPP charger, is connected to both ends of the PV cell array, which is named as a string inverter (converter) architecture. The string inverter architecture is cost effective but subject to degradation due to partial shading. This chapter gives a closer look into the partial shading problem and provides novel online PV cell array reconfiguration; a PV energy harvesting system, a typical example that requires a *cross-layer optimization*. The dynamically reconfigurable PV cell array must come with a novel management algorithm to maintain its maximum efficiency. Section 3 introduces a dynamically reconfigurable PV cell array architecture and algorithms to combat partial shading and temperature variation. This section also covers fault-tolerant features of the PV cell array.

Power converters are the primary factor of efficiency degradation. The energy storage element is subject to efficiency loss as well. Batteries' cycle efficiency and limited rate capability result in significant power loss. Supercapacitors show a very high cycle efficiency, but their severe terminal voltage variation by the state of charge makes it difficult for the power converters to maintain a high efficiency. Most of all, the power converters and energy storage elements are bulky, heavy, and expensive, which seriously discourage to implement low cost, tiny energy harvesting systems. Section 4 introduces a breakthrough MPPT method without the power converters and energy storage element for low-power PV energy harvesting applications.

The architecture in Fig. 1a does not give a freedom in the charge management of PV energy harvesting systems. The PV cell array harvests solar energy, and the load device uses the harvested energy immediately. However, Fig. 1b has an energy storage element, and thus energy harvesting and consumption can be independent of each other. In other words, energy harvesting can be done following the MPPT, and the load device energy consumption can be determined by its workload (user demand) and the power management policy. The average power coming into the

energy storage element should not be smaller than the power coming out from it at all times to avoid service interruption. Determination of the energy storage element is typically constrained by its minimum capacity when it comes to the battery-based energy storage. A larger size of the battery shows better cycle efficiency due to lower internal impedance and higher rate capability, a longer cycle life due to less number of cycles, but its cost, volume, and weight increase. On the other hand, a supercapacitor-based energy storage makes the design consideration more complicated. Section 2.3.2 introduces a new concept of MPPT, called MPTT, that jointly optimizes the power efficiency of the PV cell array and the power converter.

2 Efficiency of Photovoltaic Cell Energy Harvesting Systems

In this section, we first present a PV cell power model and a switching converter power model, which are two dominant components to the overall energy efficiency of solar energy harvesting systems. Next, we discuss and compare two operational techniques to maximize the output power of PV cells. Lastly, we discuss how the partial degradation of a PV cell array due to shading or permanent fault affects the performance of the array.

2.1 PV Cell Modeling

The basic building block of a PV array is a PV cell. The PV cells exhibit highly nonlinear voltage-current (V-I) output characteristics (curves) that change with the solar irradiance level. Figure 2a shows the PV cell V-I output characteristics

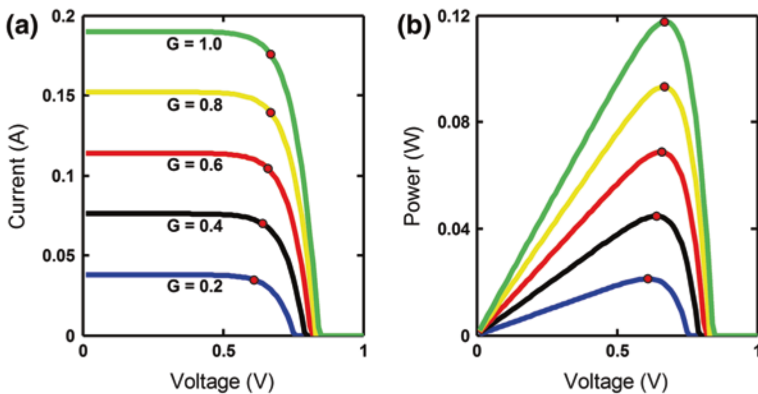


Fig. 2 V-I and V-P output characteristics of a PV cell and MPPs

under different solar irradiance levels. Figure 2b shows the corresponding voltage-power (V-P) output characteristics. The red dots in Fig. 2 denote the maximum power points (MPPs) of a PV cell where the PV cell achieves the maximum output power for the given solar irradiance level. Notations used in this section are listed in Table 1.

Let V^{pvc} and I^{pvc} denote the output voltage and current of a PV cell, respectively. The PV cell equivalent circuit model is shown in Fig. 3 with the V-I output characteristics given by

$$I^{\text{pvc}} = I_L - I_d - I_{\text{sh}} = I_L(G) - I_0(T) \cdot (e^{(V^{\text{pvc}}+I^{\text{pvc}} \cdot R_s) \cdot \frac{q}{AkT}} - 1) - \frac{V^{\text{pvc}} + I^{\text{pvc}} \cdot R_s}{R_p} \tag{1}$$

where

$$I_L(G) = \frac{G}{G_{\text{STC}}} \cdot I_L(G_{\text{STC}}) \tag{2}$$

and

$$I_0(T) = I_0(T_{\text{STC}}) \cdot (\frac{T}{T_{\text{STC}}})^3 \cdot e^{\frac{qE_g}{Ak} \cdot (\frac{1}{T_{\text{STC}}} - \frac{1}{T})} \tag{3}$$

Parameters in (1)–(3) are defined as follows: G is the solar irradiance level; T is the cell temperature; q is the charge of the electron; E_g is the energy bandgap; and k is the Boltzmann's constant. STC stands for standard test condition in which the irradiance level is 1000 W/m^2 and the cell temperature is 25 °C. The parameters listed above are either physical constants, environment-related value or

Table 1 Notations used in Sect. 2.1

V^{pvc}	Output voltage of a PV cell
I^{pvc}	Output current of a PV cell
$I_L(G)$	The photo-generated current at solar irradiance G
$I_0(T)$	Dark saturation current at temperature T
R_s	PV cell series resistance
R_p	PV cell parallel resistance
A	The diode ideality factor
q	Charge of an electron
E_g	The energy bandgap

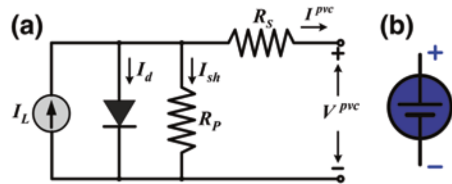


Fig. 3 Electronic circuit equivalent model (**a**) and symbol (**b**) of a PV cell (Reproduced with permission from Wang et al. [22])

configuration parameters. There are still five unknown parameters, commonly not provided by manufacturers, to be determined. These five parameters are the key that is capable of analytically describing the characteristics of a PV cell:

- $I_L(G_{\text{STC}})$: the photo-generated current at standard test condition.
- $I_0(T_{\text{STC}})$: dark saturation current at standard test condition.
- R_s: PV cell series resistance.
- R_p: PV cell parallel (shunt) resistance.
- A: the diode ideality factor.

We extract the unknown parameters from measured PV cell V-I curves at various irradiance levels and temperatures, in which each V-I curve is measured using data acquisition equipment under one specific environmental condition (G, T). Therefore, our parameter extraction method is not confined to only the parameters at STC. Instead, it extracts $I_L(G_0)$ and $I_0(T_0)$ flexibly under any environmental condition (G_0, T_0) from the measured data. Subsequently, the corresponding parameters at STC, $I_L(G_{\text{STC}})$ and $I_0(T_{\text{STC}})$, can be determined using (2) and (3). We apply the proposed combined parameter extraction method on the measured PV cell V-I curves [1]. Significant reduction (on average 8X) in root mean square (RMS) fitting error can be observed compared with the conventional method which only considers some specific points.

2.2 Power Converter (Charger) Power Model

Figure 4 shows the model of a PWM (pulse width modulation) buck-boost switching converter, which is used as the charger in the proposed PV system. The input ports of the charger are connected to the PV panel/array, whereas the output ports are connected to the EES element. The charger regulates the operating point of the PV panel by controlling the charger's input voltage, i.e., the PV output voltage (and then the PV panel output current is automatically determined by its V-I characteristics.) Notations used in this section are listed in Table 2. We denote the input voltage, input current, output voltage, and output current of the charger by V_{in}, I_{in}, V_{out}, and I_{out}, respectively. Depending on the relationship between V_{in} and

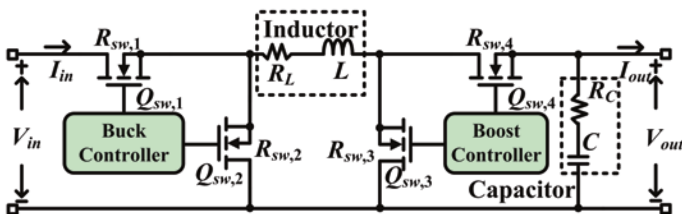


Fig. 4 Power converter electronic modeling (Reproduced with permission from Wang et al. [22])

Table 2 Notations used in Sect. 2.2

V_{in}, I_{in}	Input voltage and current of converter
V_{out}, I_{out}	Output voltage and current of converter
η_{conv}	Efficiency of the converter
P_{conv}	Power loss of the converter
D	PWM duty ratio
ΔI	Maximum current ripple of converter
f_s	The switching frequency
$I_{controller}$	Current of the microcontroller
R_L, R_C	Internal series resistances of inductor and capacitor
$R_{sw,i}$, $Q_{sw,i}$	Resistance and gate charge of the ith MOSFET

V_{out}, the charger operates in one of the two possible operating modes: the buck mode when $V_{in} > V_{out}$ and the boost mode otherwise [2, 3].

Power conversion is not free. Converting the voltage level involves nonzero amount of power loss. The overall power loss includes conduction losses by parasitic resistances of circuit components, switching losses by parasitic capacitances of switching devices, power consumption of the controller circuit, and so on. The power conversion efficiency η_{conv} is defined as:

$$\eta_{conv} = \frac{V_{out} \cdot I_{out}}{V_{in} \cdot I_{in}} = \frac{V_{in} \cdot I_{in} - P_{conv}}{V_{in} \cdot I_{in}} \tag{4}$$

where P_{conv} denotes the power loss in the converter/charger. This power loss is not constant, but varies depending on the input and output voltages and the amount of power that is transferred through the converter, and thus, the power conversion efficiency is also a variable. The power conversion efficiency is a critical factor because it determines the amount of harvested PV energy that can be ultimately transferred to the storage system (and to be used later.) In general, the power conversion efficiency will be maximized when (i) the input voltage and output voltage are close to each other, and (ii) the output current is within a certain desirable range.

We develop the converter power model based on the power model for buck switching converter provided in [2]. When the charger/converter is operating in the buck mode, its power loss P_{conv} is given by:

$$P_{conv} = (I_{out})^2 \cdot (R_L + D \cdot R_{sw,1} + (1 - D) \cdot R_{sw,2} + R_{sw,4}) \tag{5}$$

$$+ \frac{(\Delta I)^2}{12} \cdot (R_L + D \cdot R_{sw,1} + (1 - D) \cdot R_{sw,2} + R_{sw,4} + R_C) \tag{6}$$

$$+ V_{in} \cdot f_s \cdot (Q_{sw,1} + Q_{sw,2}) + V_{in} \cdot I_{controller} \tag{7}$$

where $D = V_{out}/V_{in}$ is the PWM duty ratio and $\Delta I = V_{out} \cdot (1-D)/(L \cdot f_s)$ is the maximum current ripple; f_s is the switching frequency; $I_{controller}$ is the current of the microcontroller of the charger; R_L; and R_C are the internal series resistances of the inductor L and the capacitor C, respectively; $R_{sw,i}$ and $Q_{sw,i}$ are the turn-on resistance and gate charge of the ith MOSFET switch shown in Fig. 4, respectively.

The charger power loss P_{conv} in the boost mode is given by:

$$P_{conv} = \left(\frac{I_{out}}{1-D}\right)^2 \cdot (R_L + D \cdot R_{sw,3} + (1-D) \cdot R_{sw,4} + R_{sw,1} + D(1-D)R_C) \tag{8}$$

$$+ \frac{(\Delta I)^2}{12}(R_L + D \cdot R_{sw,3} + (1-D)(R_{sw,4} + R_C) + R_{sw,1}) \tag{9}$$

$$+ V_{out} \cdot f_s \cdot (Q_{sw,3} + Q_{sw,4}) + V_{in} \cdot I_{controller} \tag{10}$$

where $D = 1 - V_{in}/V_{out}$ and $\Delta I = V_{in} \cdot D/(L \cdot f_s)$.

The power dissipation of the charger is minimized when (i) the input voltage and the output voltage of the charger are close to each other and (ii) the output current of the charger is within a certain range. Let $I_{out} = \text{Chg_Out_}I(V_{in}, I_{in}, V_{out})$ denote the function that calculates I_{out} based on V_{in}, I_{in}, and V_{out}.

2.3 Mitigating the Output Variation of PV Modules

2.3.1 Maximum Power Point Tracking

As discussed in Sect. 2.1, the output current of a PV cell is a function of the output voltage of the PV cell; hence, the resultant output power also varies by the voltage. As a result, the voltage-dependent output power is maximized only at a certain operating point, which is called the MPP. Since the V-I curve of a PV cell varies by irradiation level on the cell and the temperature of the cell, the MPP of the PV cell also changes by these factors. In order to generate the maximum power from PV cells, the operating point should be dynamically adjusted to remain at their MPP adaptive to the changes in the irradiance level and temperature. This technique for finding the MPP and adjusting the operating point to it is called the MPPT, and it is mandatory for realizing energy-efficient PV systems.

There are numerous research efforts on MPPT techniques [4]. A brute force technique is sweeping all the voltage levels from short-circuit to open-circuit to find the optimal voltage level. However, it takes a long time to find the optimal voltage, and it suffers from low power generation during the sweeping operation. Another simple yet effective technique is utilizing the precharacterized V-I curve

of the PV cell to predict the irradiance level and find the optimal voltage. Other MPPT techniques perform feedback control of the system to adjust the operating point. For example, perturb and observe (P&O) technique, or hill-climbing technique, makes a slight increase or decrease (perturbation) in the PV cell voltage to see which direction increases the output power (observe), and change the operating voltage in that way. Incremental conductance techniques utilize the property that $\mathrm{d}P^{\mathrm{pvc}}/\mathrm{d}V^{\mathrm{pvc}} = 0$ at the MPP. It compares the incremental conductance ($\mathrm{d}I^{\mathrm{pvc}}/\mathrm{d}V^{\mathrm{pvc}}$) and the PV cell's instantaneous conductance ($I^{\mathrm{pvc}}/V^{\mathrm{pvc}}$) to determine whether to increase or decrease V^{pvc}. Surveys on various MPPT techniques can be found from [5, 6].

2.3.2 Maximum Power Transfer Tracking

While the MPPT techniques are essential to maximize the power obtained from PV cells, it may not be the true optimal solution from the perspective of the whole system. The energy eventually consumed by a load or stored in an energy storage device is the output of the power converter, not the output power of the PV cells. As presented in Sect. 2.2, the power conversion efficiency of a power converter is not 100 %, but it is variable depending on the input and output conditions. Since the output of PV cells becomes the input of the power converter, the operating point of the PV cells affects the efficiency of the power conversion. By the use of an MPPT technique, we can maximize only the input power to the converter, which, however, does not guarantee the maximum output power from the converter. In addition, the output of the converter, which is a load device or an energy storage, is also subject to significant variation. For example, the terminal voltage of a supercapacitor changes linearly proportional to its state of charge. Therefore, the system-level energy-optimal design and operation of a PV system must take account of the variable efficiency.

Figure 5 is a conceptual graph that shows how the conversion efficiency affects the power harvested from a PV cell. The output power from the PV cell, which is the input power to the converter is maximum when the PV output voltage is at (A), and this point is the MPP. However, due to the variation of the conversion efficiency as denoted by the dotted curve, the converter output power, which is

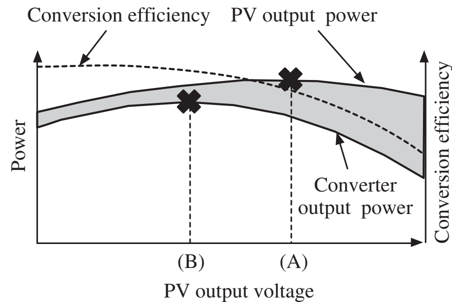


Fig. 5 PV output power and converter output power variation by PV output voltage. Shaded area denotes the power loss in the converter

the amount that we can actually utilize at the load, has a different maximum point, (B). This point is called the maximum power transfer point, or the MPT point, and we can maximize the net power by operating the PV cell at this point. The techniques for dynamically adjusting the operating point to the MPT point is called the maximum power transfer tracking (MPTT) [7]. The MPTT requires considerations not only on the characteristics of the PV cell, but also on the characteristics of the power converter, the load, and the storage device, all together.

In this context, some recent research efforts have recognized the impact of the power conversion efficiency and proposed novel circuits for maximizing the output power [8–11]. These output-maximizing converters are proposed to address the suboptimality of the conventional MPPT power converters. They have a feedback control loop that monitors the output of the power converter. Unlike the MPPT power converters that monitor the PV cell output, these converters monitor the converter output and adjust the operating point, e.g., the operating frequency of a charge pump.

More recently, MPTT-aware system design optimization techniques have been proposed. These techniques recognize the impact of components optimization in design time to the energy efficiency, such as the configuration of PV cell array or supercapacitor array. In [7], it is shown that the amount of energy harvesting may vary greatly depending on the PV array configuration and the capacitance of the supercapacitor. They proposed an optimization framework that finds the most cost-efficient PV array configuration and the most energy-efficient capacitance for a given requirement on the amount of energy. A technique proposed in [12] involves both a design-time optimization and a runtime reconfiguration technique. In design time, several supercapacitors of different capacitances are evaluated to find the best capacitance that maximizes energy efficiency. During operation, the array of the supercapacitors is dynamically reconfigured to change the terminal voltage and effective capacitance adaptively. In [13], a cross-layer optimization method is introduced, which derives the optimal design parameters such as PV cell silicon thickness, PV cell array configuration, and charge pump stages and frequency.

2.4 *Partial Shading and PV Cell Faults*

In reality, the solar irradiance levels received by PV cells in a PV system may be different from each other when a portion of PV modules is in shadow, and such a phenomenon is known as the partial shading effect. For example, moving clouds cause partial shading for stationary applications. On the other hand, shadows from nearby objects (e.g., buildings, trees, and poles) produce partial shading for PV systems on hybrid electric vehicles, which is much more severe as vehicles are moving through shaded or lighted regions. In these cases, the partial shading pattern may be quite regular (i.e., like a block), and we call this case *block shading*. Partial shading may also result from fallen leaves or dust on the PV modules [14], or other aging effects of PV modules. In this case, the partial shading pattern may

be randomized, and we call this case *random shading*. Moreover, PV cell faults, which make the output of the faulty PV cell to zero, have similar impact as the partial shading effect.

PV cells generally have different MPPs under the partial shading effect. Partial shading not only reduces the maximum output power of the shaded PV cells, but also makes the lighted or less-shaded PV cells that are connected in series with the shaded ones to deviate from their MPPs. In other words, the PV cells cannot simultaneously operate at their MPPs. With partial shading, the maximum output power of a PV module becomes much lower than the sum of the maximum output power values of all the individual PV cells in the PV module.

We demonstrate that the partial shading effect or PV cell faults may significantly degrade the output power level of a PV module with a fixed $n \times m$ configuration. We use a PV module with a 2×2 configuration as an example. As shown in Fig. 6, the PV module consists of two series-connected PV groups, and each PV group consists of two parallel-connected PV cells. The PV cell at the bottom right is completely shaded (with no solar irradiance, or has a PV cell fault) while the rest of PV cells receive the solar irradiance under the standard test condition i.e., G_{STC} W/m^2. Since only one PV cell out of four is shaded, the ideal setup should exhibit the PV module output power degradation of 25 % compared to the same PV module without any shading. However, the actual PV module output power degradation is much larger than 25 %.

We plot in Fig. 7 the V-I characteristics of the PV module under partial shading (or PV cell fault). Curve 1 corresponds to the V-I output characteristics of the bottom PV group with the shaded PV cell, whereas Curve 2 corresponds to the V-I output characteristics of the top PV group. Curve 2 has a higher current value than Curve 1 at the same voltage value. Curve 3 is the V-I output characteristics of the PV module, which is directly derived from Curves 1 and 2 since the PV module is a series connection of the two PV groups. Note that we assume that each PV cell is integrated with a bypass diode to protect the PV cell from reverse bias operation under partial shading [15] when we derive Curve 3.

We compare the end-to-end V-P output characteristics of the partially shaded PV module (or PV module with PV cell fault) with the same PV module without

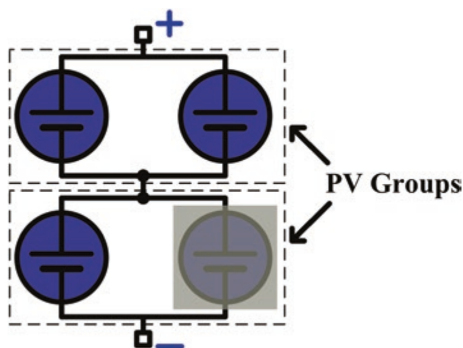


Fig. 6 An example of partial shading on a PV module with four PV cells (Reproduced with permission from Wang et al. [22])

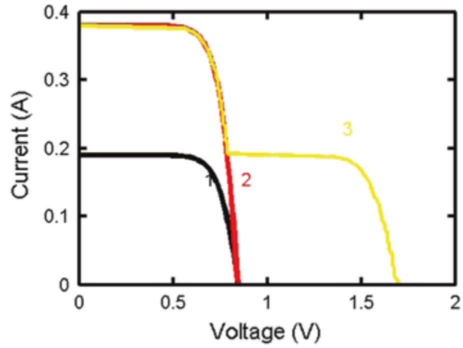


Fig. 7 V-I characteristics of the PV module under partial shading

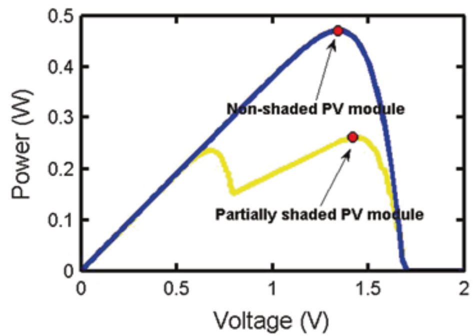


Fig. 8 V-P characteristics of the PV module under partial shading

shading in Fig. 8. The red dots in Fig. 8 show the MPPs. The maximum output power of the partially shaded PV module (or PV module with PV cell fault) is about 56 % of that of the same PV module without shading (or PV cell fault). As a result, one shaded (or faulty) PV cell degrades the PV module output power by as much as 44 %, which establishes the significance of the effect of partial shading effect (or PV cell fault).

In addition, partial shading (or PV cell fault) may result in multiple power peaks in the V-P output characteristics of a PV module, as also can be seen in Fig. 8. Therefore, the MPPT (or MPTT) techniques must be modified in order to dynamically track a global optimum operating point instead of a local optimal one [16, 17]. This is because the existing MPPT or MPTT techniques such as the perturb and observe method rely on the unimodality assumption about the V-P output characteristics of the PV module. The modified MPPT or MPTT techniques increase the complexity of the PV system control circuitry.

The modified MPPT techniques may restore part of the power loss due to partial shading, but they cannot fully utilize the lighted PV cells due to the deviation from their MPPs caused by the shaded cells. On the other hand, PV module reconfiguration techniques, which have the potential of fully exploiting the MPPs of both lighted and shaded PV cells in a partially shaded PV module, can help maintain the output power level of a PV system under partial shading. Various

PV reconfiguration techniques have been proposed, which are different from each other in terms of the system structure and control approach that they employ [18–20]. However, they suffer from one or more of the following limitations:

1. To compensate the power loss from shaded PV cells, many extra PV cells are needed for performing reconfiguration according to the shading pattern.
2. There is a lack of systematic and scalable structural support or effective control mechanism.
3. Variations in the conversion efficiency of the charger or inverter at different operating points are overlooked, which may result in a sizeable degradation in the overall energy conversion efficiency.

3 Reconfigurable Photovoltaic Cell Array

In this section, we introduce a reconfigurable PV cell array architecture and its operation. Reconfigurable PV cell arrays have several benefits when compared with fixed (nonreconfigurable) PV cell arrays. First, the output voltage and current of PV cell arrays can be adjusted so that the efficiency of the power converter is maximized. Second, the effects of partial shading can be mitigated. Third, faulty cells can be dynamically excluded from PV cell arrays without service interruption. We first introduce the switch network architecture of the reconfigurable PV cell array, and then, we describe its benefits in the rest of this section.

3.1 Reconfigurable Switch Network Architecture

In most cases, the output of a single PV cell can generate no more than a few watts. Therefore, it is typical to connect multiple PV cells to build a PV cell array to obtain a higher output voltage and/or a higher output current. Stacking PV cells in series increases the output voltage, while multiple PV cells in parallel increases the output current. It is important to properly design the number of series and parallel connections to obtain desired voltage and current levels. However, the optimal configuration that produces the maximum power changes by irradiance level, temperature, storage device state of charge, and so forth, as we discussed in Sect. 2.3.2. Therefore, instead of fixing the configuration, it is beneficial to be able to dynamically change the configuration during operation as needed. In [21], a switch network architecture for dynamic array reconfiguration is introduced. It is originally proposed for the reconfiguration of an energy storage array, e.g., supercapacitor array or battery array, but it can be leveraged for the reconfiguration of PV cell arrays in the same manner.

Figure 9 shows the reconfigurable switch network architecture for a PV cell array that consists of N PV cells, $C_{1,\ldots,N}$. First $N-1$ cells, from C_1 to C_{N-1}, have

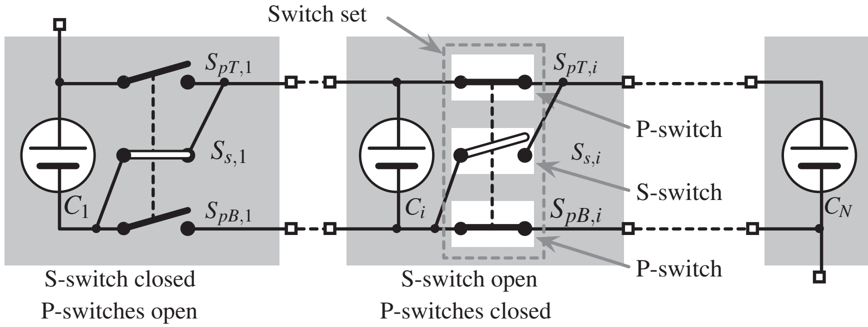


Fig. 9 Reconfigurable switch network architecture [21]

a set of three switches each. One switch set is composed of one series switch (S-switch) and two parallel switches (P-switches). The S-switch of the ith cell is denoted by $S_{s,i}$, and the two P-switches of the ith cell are denoted by $S_{pT,i}$ (top parallel switch) and $S_{pB,i}$ (bottom parallel switch).

There are a few rules for the operation of the three switches in a switch set:

- Two paired P-switches should be open or closed at the same time.
- Either S-switch or P-switches pair should be closed at any moment.
- S-switch and P-switches pair should not be closed at the same time.

More formally, for $i = 1, 2, \ldots, N-1$,

$$x_{p,i} + x_{s,i} = 1, \tag{11}$$

where the binary variables $x_{p,i}$ and $x_{s,i}$ are defined as

$$x_{s,i} = \begin{cases} 0 & \text{if } S_{s,i} \text{ is open,} \\ 1 & \text{if } S_{s,i} \text{ is closed,} \end{cases} \tag{12}$$

$$x_{P,i} = \begin{cases} 0 & \text{if } S_{pT,i} \text{ and } S_{pB,i} \text{ are open,} \\ 1 & \text{if } S_{pT,i} \text{ and } S_{pB,i} \text{ are closed.} \end{cases} \tag{13}$$

One and only one of $x_{s,i}$ and $x_{p,i}$ should be 1 at any moment. If both $x_{s,i}$ and $x_{p,i}$ are 1 at the same time, C_i is short circuited. On the other hand, if both $x_{s,i}$ and $x_{p,i}$ are 0 at the same time, C_i is disconnected from C_{i+1}.

Consecutive PV cells that are connected in series form a PV group. All the PV cells in the same PV group are connected in series; hence, the current through these PV cells is all identical. Multiple PV groups are connected to their adjacent PV groups in parallel. Therefore, the terminal voltage of all the PV groups becomes identical.

3.2 Balanced Configurations

A PV cell array configuration that all the PV groups have the same number of PV cells is called a *balanced configuration*. If the solar irradiance onto every PV cell is uniform, the output voltage and current of each PV cell are also uniform in a balanced configuration. As balanced configurations are the most basic configurations of a PV array, we first discuss their properties in this subsection. Then we discuss imbalanced configurations, where the number of PV cells in each PV group may differ, in the following subsection.

The flexibility of the reconfigurable switch architecture introduced in Sect. 3.1 allows any arbitrary balanced configuration. The number of feasible balanced configurations that we can make using N PV cells is equivalent to the number of divisors of N. For example, the number of balanced configurations of a 10-cell PV cell array ($N = 10$) is four: 10×1, 5×2, 2×5, and 1×10, where $m \times n$ is the m -series and n-parallel balanced configuration. Note that the number of divisors of 10 is four: 1, 2, 5, and 10.

Using the binary variables $x_{s,i}$ and $x_{p,i}$ defined in Sect. 3.1, the balanced configurations of a $m \times n$ PV cell array are obtained by switching operations which obey the following rule:

$$x_{s,i} = \begin{cases} 1 \text{ if } i = n \cdot k \text{ where } k = 1, 2, \ldots, m-1, \\ 0 \text{ otherwise,} \end{cases} \tag{14}$$

$$x_{p,i} = 1 - x_{s,i}. \tag{15}$$

Assuming that the solar irradiance is uniform across the whole PV cell array, the output voltage V_{pvm} and the output current I_{pvm} of a balanced $m \times n$ PV cell array are:

$$V_{\text{pvm}} = m \times V_{\text{pvc}}, \tag{16}$$

$$I_{\text{pvm}} = n \times I_{\text{pvc}}. \tag{17}$$

Figure 10 is the balanced configurations that we can make using a 4-cell PV cell array ($N = 4$). It shows three balanced configurations: 4×1, 2×2, and 1×4. For example, the 2×2 configuration consists of two PV groups, each of which is composed of two PV cells connected in parallel with a pair of P-switches, and the two groups are connected in series with a S-switch.

Recall that the amount of energy harvested from a PV cell array may vary depending on its configuration even if the number of PV cells is the same. Finding the optimal configuration was a part of the optimization framework introduced in [7], but this was a static design-time decision. However, the optimal configuration that maximizes the power delivered to the load or storage device keeps changing as we discussed in 2.3.2. We can leverage this PV cell array reconfiguration architecture to further improve the energy delivery by dynamically reconfiguring the PV array in runtime.

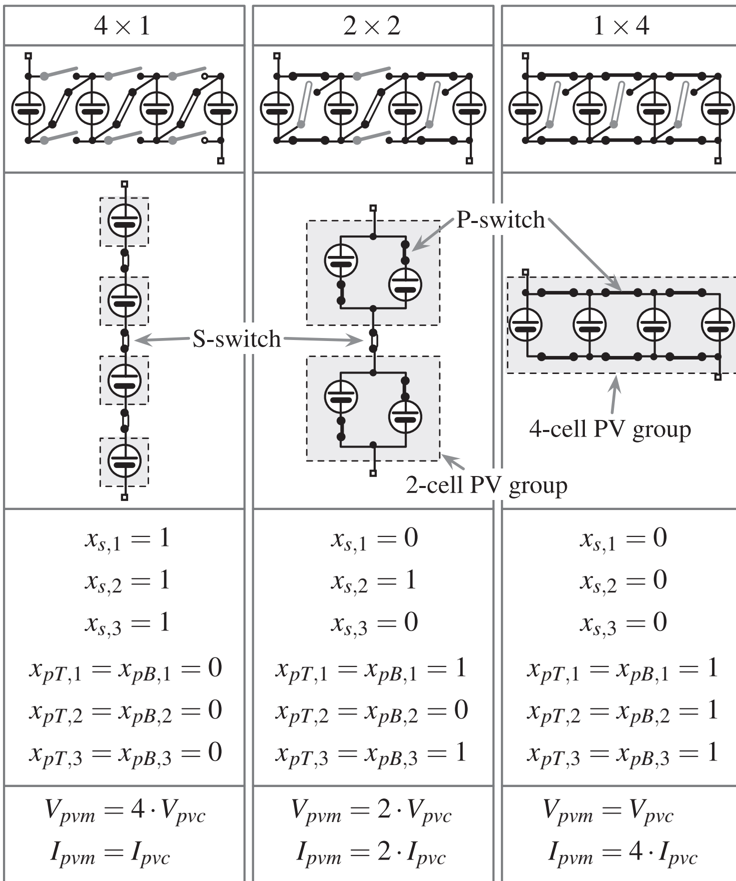


Fig. 10 Balanced configurations of a four-cell PV cell array ($N = 4$)

3.3 Imbalanced Reconfiguration to Combat Partial Shading

In general, a reconfigurable PV array (module) with N PV cells can achieve imbalanced reconfiguration, i.e., it can have a arbitrary number (less than or equal to N) of PV groups, each with arbitrary number of PV cells with consecutive IDs. Figure 11 is an example of PV module reconfiguration. The first four PV cells are connected in parallel to form PV Group 1; the next three PV cells form PV Group 2; and the last five PV cells form PV Group 3. These three PV groups are series-connected by the S-switches of the fourth and the seventh PV cells.

A reconfigurable PV module consisting of N PV cells may include an arbitrary number (less than or equal to N) of PV groups. The number of parallel-connected PV cells $r_j(> 0)$ in the jth PV group should satisfy

$$\sum_{j=1}^{g} r_j = N, \tag{18}$$

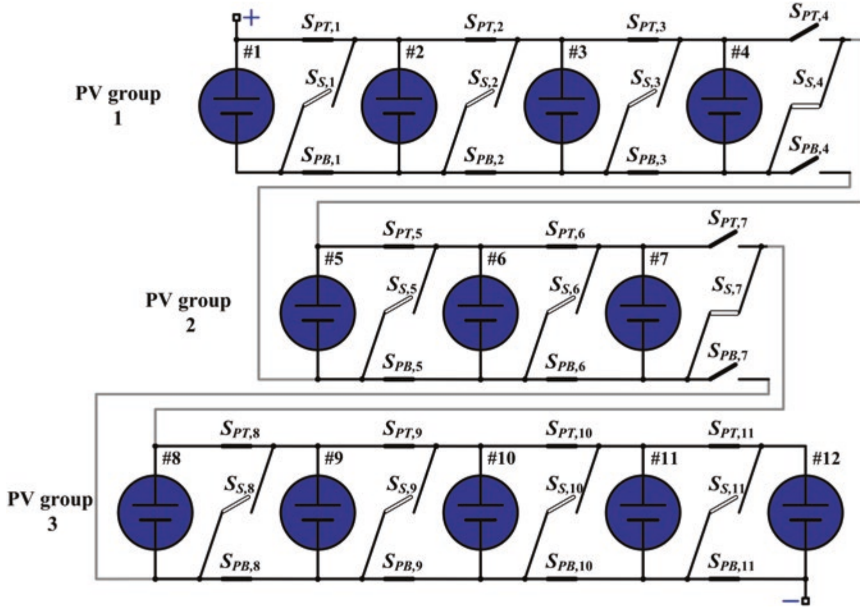


Fig. 11 An imbalanced reconfiguration example (Reproduced with permission from Wang et al. [22])

where g is the number of PV groups. This configuration can be viewed as a partitioning of the PV cell index set $\mathbf{A} = \{1, 2, 3, \ldots N\}$, where the elements in $\mathbf{A}$ denote the indices of PV cells in the array. This partitioning is denoted by subsets $\mathbf{B}_1$, $\mathbf{B}_2$,..., and $\mathbf{B}_g$ of $\mathbf{A}$, which correspond to the g PV groups comprised of r_1, r_2, ..., and r_g PV cells, respectively. The subsets $\mathbf{B}_1$, $\mathbf{B}_2$, ..., and $\mathbf{B}_g$ satisfy

$$\cup_{j=1}^{g} \mathbf{B}_j = \mathbf{A} \tag{19}$$

and

$$\mathbf{B}_j \cap \mathbf{B}_k = \emptyset, \quad \forall j, k \in \{1, 2, \ldots, g\}, \quad j \neq k \tag{20}$$

The indices of PV cells in group j must be smaller than the indices of PV cells in group k for any $1 \leq j < k \leq g$ due to the structural characteristics of the reconfigurable PV array, i.e., $i_1 < i_2$ for $\forall\, i_1 \in \mathbf{B}_j$ and $\forall\, i_2 \in \mathbf{B}_k$ satisfying $1 \leq j < k \leq g$. A partitioning satisfying the above properties is called an *alphabetical partitioning*.

Figure 12 shows the architecture of a PV system with a reconfigurable PV array (module), equipped with the reconfiguration architecture. The input and output ports of the charger are connected to the PV module and a supercapacitor array, respectively. The charger regulates the operation of the PV module by regulating its output voltage. The output current of the PV module is automatically determined based on its V-I characteristics. We adopt a software-based MPTT

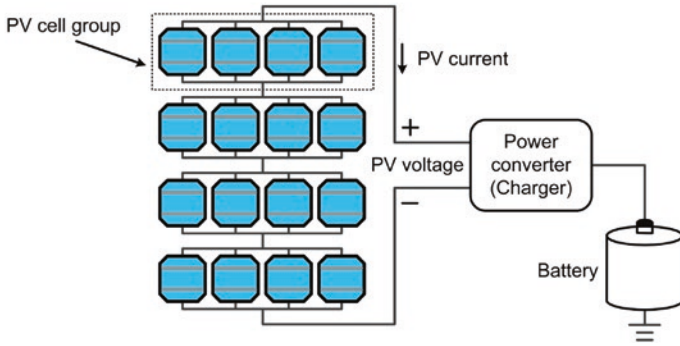


Fig. 12 The system structure of the reconfigurable PV system (Reproduced with permission from Lin et al. [23])

technique in the proposed PV system. It employs the perturb & observe (P&O) algorithm to maximize the charger output current through regulating the output voltage of PV module.

For the ith PV cell, the relationship between output voltage V_i^{pvc} and output current I_i^{pvc} depends on the solar irradiance G_i as given by PV cell characteristics. We obtain G_i of each PV cell using onboard solar irradiance sensors. The PV cell temperature has a relatively minor effect on the V-I characteristics.

The output voltage and current of the PV module are denoted by V_{pv} and I_{pv}, respectively. The power consumption of the charger is P_{conv}. The terminal voltage of the supercapacitor is V_{cap}, and the charging current of the supercapacitor is I_{cap}. We have

$$V_{\mathrm{pv}} \cdot I_{\mathrm{pv}} = P_{\mathrm{conv}} + V_{\mathrm{cap}} \cdot I_{\mathrm{cap}}. \quad (21)$$

We provide a formal problem statement for the PV module reconfiguration (PMR) problem in the following.

PMR Problem Statement: Given G_i of each ith PV cell and V_{cap}, **find** the optimal configuration of the PV module and the optimal operating point $(V_{\mathrm{pv}}, I_{\mathrm{pv}})$, **such that** I_{cap} is maximized. The objective is equivalent to maximizing the PV system output power.

We propose the near-optimal PV module reconfiguration algorithm comprised of a kernel algorithm and an outer loop. The kernel algorithm finds the optimal number of PV cells in each PV group with a given group number g such that the PV module MPP power is maximized. The kernel algorithm is based on dynamic programming with polynomial time complexity. The outer loop determines the optimal g value.

With details shown in [22], the kernel algorithm relies on the observation that the MPP voltages of a PV cell under different solar irradiance levels are very close to each other, but the corresponding MPP currents vary significantly. Let $V_{\mathrm{pvc}}^{\mathrm{MPP}}$ denote the solar irradiance-independent MPP voltage of a PV cell, and let $I_{\mathrm{pvc}}^{\mathrm{MPP}}(G)$

denote the MPP current as a function of the solar irradiance G. The kernel algorithm maximizes the estimated PV module MPP power as given by

$$g \cdot V_{\text{pvc}}^{\text{MPP}} \cdot \min_j \sum_{i \in \mathbf{B}_j} I_{\text{pvc}}^{\text{MPP}}(G_i) \tag{22}$$

Or equivalently, it maximizes $\min_j \sum_{i \in \mathbf{B}_j} I_{\text{pvc}}^{\text{MPP}}(G_i)$. Optimal substructure property with details shown in [22] applies to this problem. Hence, we apply dynamic programming method as the basis of the kernel algorithm to solve this problem. Details of the kernel algorithm are provided in [22].

In the outer loop, we first execute the kernel algorithm to find an optimal configuration for a given group number g value. After that, we calculate the estimated I_{cap} under such configuration using the charger characteristics. In the calculation of the estimated I_{cap}, we estimate the PV module output voltage and current by $g \cdot V_{\text{pvc}}^{\text{MPP}}$ and $\min_j \sum_{i \in \mathbf{B}_j} I_{\text{pvc}}^{\text{MPP}}(G_i)$, respectively. We find the optimal g value that maximizes the estimated I_{cap}, and the corresponding configuration subsequently. Details are provided in [22].

3.4 *Fault Tolerance*

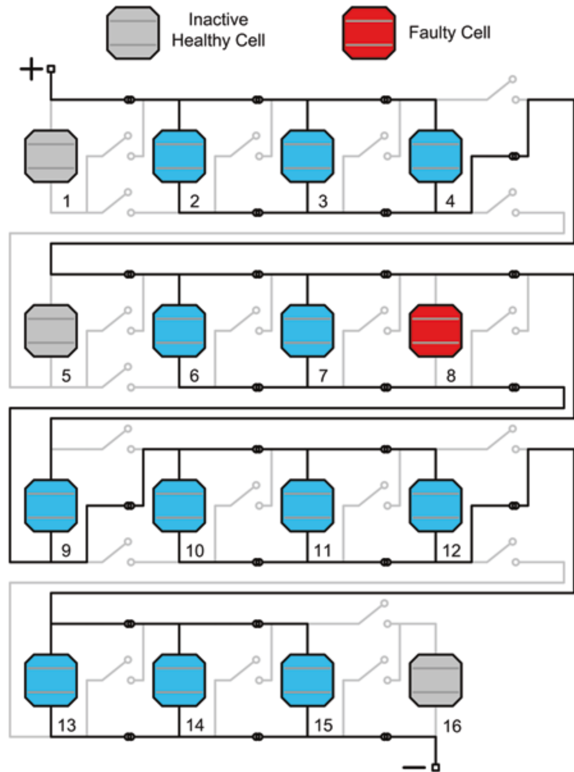
3.4.1 Reconfiguration for Fault Detection and Fault Bypassing

The objective of fault detection is to identify any PV cell faults in the PV panel. Fault bypassing aims at forming a new PV panel configuration to minimize the output power loss caused by PV cell faults [23]. Consider a PV panel assuming a $N \times M$ configuration during normal system operation. For fault detection, we may need to form a $k \times M$ PV panel configuration ($k \leq N$) with a selected set of PV cells and measure their combined output power to determine whether a PV cell fault exists in this portion of PV panel. During fault bypassing, we may need to form a $N_{\text{opt}} \times M_{\text{opt}}$ PV panel configuration to improve the system output power, where the faulty PV cells, and perhaps some healthy PV cells are excluded from the new configuration $N_{\text{opt}} \times M_{\text{opt}} < N \times M$.

Inactivating even healthy PV cells may become necessary in some cases of fault bypassing. For example, suppose we have 42 PV cells and one of them is faulty. We cannot form a good configuration with a prime number of 41 healthy PV cells such that the MPP voltage of this configuration matches with the battery voltage. Therefore, we would like to inactivate one healthy PV cell to have 40 active cells. Then we are able to perform a 5×8 or 8×5 configuration to increase the PV system output power. The inactivated healthy PV cell may be used later if we encounter an additional PV cell fault.

We use Fig. 13 as an example to illustrate how to effectively control the ON/OFF states of the switches for fault detection and fault bypassing. PV cells 1, 5, and 16 are inactive healthy PV cells, and PV cell 8 is a faulty PV cell. Figure 13

Fig. 13 An example of PV reconfiguration for fault detection and fault bypassing (Reproduced with permission from Lin et al. [23])



shows a 4 × 3 PV panel configuration formed with the remaining PV cells. The faulty PV cell 8 is open-circuited and inactive outside our control. The healthy PV cells can be isolated in either of two cases: (i) they are located between two PV cell groups; (ii) they are at the leftmost or the rightmost position in the electrical connection of the PV panel. In Fig. 13, an example of the first case is PV cell 5, whereas examples of the second case are PV cells 1 and 16.

The fault detection and bypassing algorithms for fault-tolerant PV systems are proposed based on the reconfigurable PV panel structure. The fault detection algorithm can identify a PV cell fault with logarithmic time complexity or determine the nonexistence of a PV cell fault in O(1) time. The fault bypassing algorithm determines the optimal configuration of a PV panel, such that the PV system output power degradation due to PV cell faults can be minimized.

The fault detection algorithm is executed every Δt units of time. Δt must be much smaller than the average fault occurrence time interval, which is in the order of days or months, so that we can safely assume that at most one PV cell fault occurs during each time interval and the fault detection algorithm only needs to detect at most one newly occurring fault at each execution. The fault detection algorithm first compares the actual PV panel output power with the theoretical output power of the PV panel without any faults. If the difference is smaller

than a prespecified error threshold, then there will exist no new fault, and the fault detection algorithm will terminate in O(1) time. Otherwise, the fault detection algorithm will continue to find the fault, and the fault bypassing algorithm will be executed. In most cases, the fault detection algorithm will confirm the nonexistence of a new fault. Therefore, the computational overhead of the fault detection and bypassing algorithms is small. In practice, we have found that Δt can be set to an hour.

3.4.2 Fault Detection Algorithm

The basic step of the fault detection algorithm is the *Fault Existence Checking* algorithm, which determines whether a PV cell fault exists in a set of $k \times M$ PV cells. We track the maximum output power of the $k \times M$ PV panel configuration using the charger in the PV system at Step 3 of the Fault Existence Checking algorithm. In reality, k must be larger than or equal to a threshold value $K_{\min}$ such that the output voltage of the $k \times M$ PV panel configuration is high enough to properly drive the charger. This means that the Fault Existence Checking algorithm cannot run on a PV panel configuration smaller than $K_{\min} \times M$.

The fault detection algorithm has two steps: first, determine which row the faulty PV cell is located at (row search); second, determine which column the faulty PV cell is located at (column search). To find the location of the potential faulty PV cell in the $N \times M$ PV panel, we first run the Fault Existence Checking algorithm on the whole PV panel. If it is confirmed that no PV cell fault exists, the fault detection algorithm will terminate. Otherwise, the fault detection algorithm will continue to find the location of the PV cell fault as explained next. Detailed procedure is described in Algorithm 1.

We use Fig. 14 to demonstrate how the row search and column search proceed. In this example, $N = 4$, $M = 4$, and $K_{\min} = 2$. For row search, we bisect the PV panel into the first two rows (A1) and the remaining two rows (A2). We run the Fault Existence Checking algorithm on A1 and find out that A1 contains a faulty PV cell. Then we bisect A1 into the first row (B1) and the second row (B2). The size of B1 is smaller than $K_{\min} \times M$. Therefore, we form a $K_{\min} \times M$ (2×4) configuration from B1 along with the third row, which has been confirmed to contain only healthy PV cells, and subsequently, run the Fault Existence Checking algorithm on this configuration. We determine that B1 does not contain the faulty PV cell, and therefore, the faulty PV cell is within B2. Now, we have located the row containing the faulty PV cell.

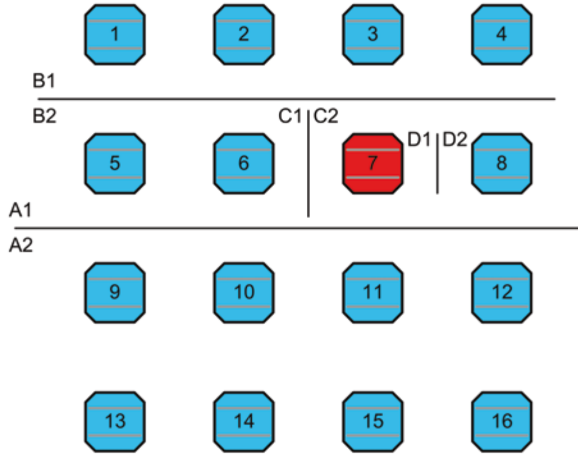


Fig. 14 Illustration of the fault detection algorithm (Reproduced with permission from Lin et al. [23])

Algorithm 1: Finding the Location of the PV Cell Fault

1 $\mathbf{V_{com}} \leftarrow$ the $N \times M$ PV module; $k \leftarrow N$
2 **while** $k > 1$ **do**
3     $\mathbf{V_{part1}} \leftarrow$ the first $k_1 = \lfloor k/2 \rfloor$ rows in $\mathbf{V_{com}}$
4     $\mathbf{V_{part2}} \leftarrow$ the rest $k_2 = k - \lfloor k/2 \rfloor$ rows in $\mathbf{V_{com}}$
5   **if** $k_1 \geq K_{min}$ **then**
6         $R \leftarrow$ Run Fault Existence Checking on $\mathbf{V_{part1}}$
7   If $R = 1$: $\mathbf{V_{com}} \leftarrow \mathbf{V_{part1}}$; $k \leftarrow k_1$
8   Else: $\mathbf{V_{com}} \leftarrow \mathbf{V_{part2}}$; $k \leftarrow k_2$
9   **else**
10         $\mathbf{V_b} \leftarrow K_{min} - k_1$ rows of PV cells that have been confirmed to be healthy PV cells
11         $R \leftarrow$ Run Fault Existence Checking on $\mathbf{V_{part1}}$ along with $\mathbf{V_b}$
12   If $R = 1$: $\mathbf{V_{com}} \leftarrow \mathbf{V_{part1}}$; $k \leftarrow k_1$
13   Else: $\mathbf{V_{com}} \leftarrow \mathbf{V_{part2}}$; $k \leftarrow k_2$
14 $\mathbf{V_{com}}$ is now the row containing the faulty PV cell
15 **while** $k > 1$ **do**
16     $\mathbf{V_{part1}} \leftarrow$ the first $k_1 = \lfloor k/2 \rfloor$ cells in $\mathbf{V_{com}}$
17     $\mathbf{V_{part2}} \leftarrow$ the rest $k_2 = k - \lfloor k/2 \rfloor$ cells in $\mathbf{V_{com}}$
18     $\mathbf{V_b} \leftarrow K_{min} \times M - k_1$ PV cells that have been confirmed to be healthy PV cells
19     $R \leftarrow$ Run Fault Existence Checking on $\mathbf{V_{part1}}$ along with $\mathbf{V_b}$
20   If $R = 1$: $\mathbf{V_{com}} \leftarrow \mathbf{V_{part1}}$; $k \leftarrow k_1$
21   Else: $\mathbf{V_{com}} \leftarrow \mathbf{V_{part2}}$; $k \leftarrow k_2$
22 Return $\mathbf{V_{com}}$

For column search, we bisect B2 into PV cells 5 and 6 (C1) and PV cells 7 and 8 (C2). We run the Fault Existence Checking algorithm on C1 along with PV cells 3, 4, and 9–12 that are confirmed healthy. We pick these healthy PV cells, because in this way we can form a 2 × 4 configuration, with PV cells 7 and 8 bypassed between the first PV cell group (PV cells 3–6) and the second PV cell group (PV

cells 9–12). We find out that C1 does not contain the faulty PV cell, and therefore, C2 contains the faulty PV cell. We bisect C2 into PV cell 7 (D1) and PV cell 8 (D2). We form a 2×4 configuration from D1 along with PV cells 4–6 and 9–12, and run the Fault Existence Checking algorithm on this configuration. We confirm that the faulty PV cell is PV cell 7, and thereby, conclude the column search.

3.4.3 Fault Bypassing Algorithm

The fault bypassing algorithm determines the optimal configuration of a PV panel, such that the PV system output power loss due to PV cell faults is minimized. We need to decide (i) the number of active healthy PV cells S, and (ii) the optimal PV panel configuration $N_{\text{opt}} \times M_{\text{opt}}$. Let us denote the number of factors of S by $F(S)$. The maximum output power of a PV panel is approximately proportional to S. And a S value is preferred if $F(S)$ is larger, since we have more choices of PV panel configurations with this S value.

Algorithm 2: Fault Bypassing Algorithm of a PV Panel

```
    Input: G, T, and V_bat
    Output: N_opt and M_opt
1   I_bat^max ← 0
2   Determine the set of candidate S values
3   for each S value in the candidate set do
4   |   for N' from 1 to S do
5   |   |   if M' = S/N' is an integer then
6   |   |   |   Calculate I_bat from M', N', V_bat and Chg_Out_I function
7   |   |   |   Update I_bat^max, N_opt, M_opt if I_bat > I_bat^max
```

Assume that the PV panel has $N \times M$ PV cells and L PV cell faults have been identified so far. Therefore, we have $S_{\max} = N \times M - L$. First, we determine a set of candidate S values in ascending order, which satisfies $S + F(S) \geq S_{\max} + F(S_{\max})$. There are $F(S)$ possible configurations using S active healthy PV cells. Among these configurations, there exists an optimal configuration that provides maximum PV system output power $P_{\max}(S)$. Based on the PV cell and charger model, we find the optimal S_{opt} value by ternary search on the set of candidate S values, such that $P_{\max}(S_{\text{opt}})$ is the maximum achievable PV system output power. The optimal PV panel configuration is determined accordingly. Detailed procedure is described in Algorithm 2.

4 Storage- and Converter-Less Energy Harvesting

Energy storage elements as well as power converters have been necessary components for performing an MPPT. However, as described, these components seriously limit the design and implementation of load devices in many aspects including the weight, form factor, cost, maintainability, etc. In addition, using those components may result in additional energy loss while storing the energy and converting the voltage level as described in the previous sections. So elimination of those components gives numerous benefits for designing load devices. This section mainly focuses on introducing a breakthrough MPPT method without neither power converters nor energy storage elements for low-power PV energy harvesting applications.

4.1 Principle of Operation

There are obvious obstacles in eliminating power converters and energy storage elements from the conventional energy harvesting devices. First, the MPP current and voltage should be exactly matched with the operating ranges of the load devices without using those two components. However typical electronic components including microprocessors, in general, are unable to fulfill this requirement by themselves. Second, the load devices should be functional even after power interruption when the PV energy is not strong enough to operate it. The basic principle of storage- and converter-less energy harvesting is to provide the harvested energy directly to the computing node by exploiting a fine-grained dynamic power management (DPM).

As shown in Fig. 15, the MPP voltage of PV cell is maintained within a narrow range regardless of the solar irradiance. This means that the PV cell produces

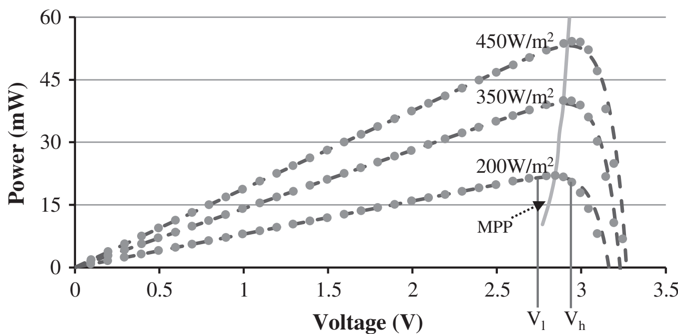


Fig. 15 Power-Voltage curve of a 4.5 × 5.5 cm^2 PV cell under different solar irradiance (Reproduced with permission from Wang et al. [24])

almost-constant voltage output, which is a main role of power converters, as long as we keep the track of the MPP current. Among the several ways to keep the track of the MPP current, fast enough DPM makes the PV cell keep the MPP as if the current of the load device is a DC current as long as the average current is the same to the MPP current. However, this may not be feasible if we do not carefully consider DPM overheads because fast DPMs may also bring nonnegligible energy and timing overheads during the state changes, which finally decreases the overall energy efficiency. So the proper control of the DPM frequency considering these overheads is necessary for maximizing the energy efficiency.

Storage-less operation does not directly affect the operation during a day even at very weak solar irradiance thanks to a very fine-grain DPM in a few hundreds microseconds. Although the load devices become unavailable after sunset, this may not be a serious problem in the applications where the operation is only required during the daytime.

4.2 Storage- and Converter-Less MPPT

Figure 16 illustrates the architecture of storage- and converter-less energy harvesting system proposed in [24]. Only single load switch with a DPM control logic is used in between the PV cell and load device. The DPM control logic continuously turns on and off the load switch depending on the level of the PV cell output voltage. Once the PV cell output voltage is higher than the upper threshold, V_h, the load device is turned on. Then, the PV cell output voltage decreases as the load devices are connected to the PV cell and the load current is higher than the MPP current. When the PV cell voltage becomes lower than the lower threshold, V_l, the MPP controller turns off the load switch and the PV cell output voltage recovers and goes back to V_h. This control policy maintains the PV cell output voltage within $[V_l, V_h]$.

The control process acts like a MPP tracking with the PV cell current steering as the load switch turns on and off faster than the cutoff frequency of the PV cell. This makes the PV energy harvesting system operational without power converters or an energy storage device while performing the MPPT.

The decoupling capacitor, C_{decoup} on the load side maintains power integrity of the load devices like normal devices. On the other hand, the bulk capacitor, C_{bulk}, connected in parallel with the PV cell extends the time constant of the PV cell so

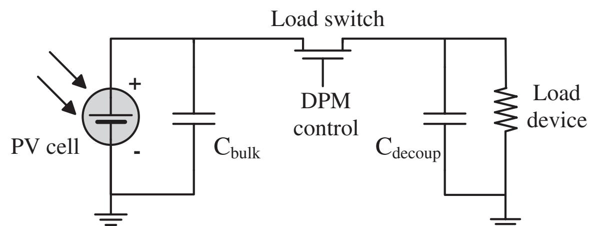


Fig. 16 Storage- and converter-less PV energy harvesting system

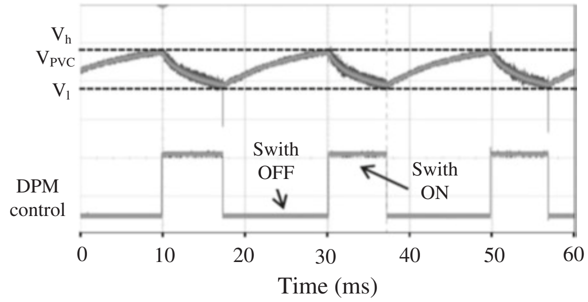


Fig. 17 Voltage changes depending on the DPM control (Reproduced with permission from Wang et al. [24])

that the PV cell's time constant may match with the feasible DPM period of the load devices. So the sizing of bulk capacitor plays an important role because it determines the DPM time granularity, voltage drop during charge sharing between the bulk capacitor and decoupling capacitor, and the time and energy overhead of the DPM scheme. The DPM period, T_{DPM}, can be estimated by

$$T_{\text{DPM}} \approx (\frac{V_h - V_l}{I_{\text{pvc,mpp}}} + \frac{V_{\text{mid}} - V_l}{I_{\text{on}} - I_{\text{pvc,mpp}}})C_{\text{bulk}}. \tag{23}$$

where V_{mid} is the voltage after the charge sharing, $I_{\text{pvc,mpp}}$ is the output current at the MPPs of the PV cell, and I_{on} is the load current when the load device is on. V_{mid} is simply calculated from the size of C_{bulk} and C_{decoup}, V_h, and V_l.

So the larger bulk capacitor, the longer DPM period. If the bulk capacitor size is small, the PV cell output voltage decreases fast as the load current discharges the capacitor. If the bulk capacitor is large, the PV cell output voltage decrease slowly.

Figure 17 shows the voltage changes observed on the PV cell during the operation of the load device. Under the control of the DPM control signal, the PV cell output voltage swings within the MPP windows of $[V_h, V_l]$.

4.3 Overhead Analysis

Though storage- and converter-less energy harvesting architecture significantly improves the energy efficiency by providing the harvested energy from the PV cell directly to the load device, it also brings nonnegligible energy overheads which are not used for executing the task but used for changing the power states or preparing the task execution before/after power state changes.

For example, the task cannot be executed until all the components used in the load device are waked up from the power-off state even after the load switch is on. For guaranteeing the reliable device operation, the task also cannot be executed during the preparation of the power state changes from on state to off state just before turning off the load switch. This is a nonavoidable DPM overhead in the

storage- and converter-less architecture. If T_{DPM} is set too small, DPM overhead may not be negligible. As described in Eq. 23, the period of DPM is mainly controlled with the size of C_{bulk}, $I_{pvc,mpp}$, and configuration of threshold values.

Another overhead of the storage- and converter-less architecture is the energy loss during the charge sharing between C_{bulk} and C_{decoup} when the load switch is turned on. The amount of energy loss is controlled by varying the size of C_{bulk} and C_{decoup}. However adjusting the size of C_{decoup} is not desirable because it maintains power integrity of the load device. The larger bulk capacitor, the less energy loss during the charge sharing. This means that the energy loss is minimized if we set the size of C_{bulk} as large as possible. However, this is not a simple problem because the size of C_{bulk} is tightly coupled with the period of the DPM as well. In addition, the larger capacitor makes it harder to track MPP of PV cell, which directly affects the quality of MPPT, though it lowers energy loss during the charge sharing as well as DPM overhead. So adjusting the size of bulk capacitor considering the application characteristics is important for maximizing the energy efficiency of the energy harvesting systems.

5 Conclusions

Solar energy harvesting is one of the most promising renewable energy generation that directly produces electric energy from solar irradiance. Despite its range of advantages, partial shading and cell fault may significantly degrade the whole array performance. This chapter introduced a systematic design practice of a dynamically reconfigurable photovoltaic (PV) cell array. The design framework introduced in this chapter significantly enhances the state-of-the-art PV energy harvesting systems both in performance and cost thanks to the cross-layer optimization.

Acknowledgments This work is supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT & Future Planning as the Global Frontier Project.

References

1. Lee W, Kim Y, Wang Y, Chang N, Pedram M (2011) Versatile high-fidelity photovoltaic module emulation system. In: Proceedings of the ACM/IEEE international symposium on low power electronics and design (ISLPED)
2. Choi Y, Chang N, Kim T (2007) DC–DC converter-aware power management for low-power embedded systems. IEEE Trans Comput Aided Des Integr Circ Syst 26(8):1367–1381
3. Wang Y, Kim Y, Xie Q, Chang N, Pedram M (2011) Charge migration efficiency optimization in hybrid electrical energy storage (HEES) systems. In: Proceedings of the ACM/IEEE international symposium on low power electronics and design (ISLPED)
4. Hohm DP, Ropp ME (2003) Comparative study of maximum power point tracking algorithms. Prog Photovoltaics Res Appl 11(1):47–62

5. Esram T, Chapman P (2007) Comparison of photovoltaic array maximum power point tracking techniques. IEEE Trans Energ Convers 22(2):439–449
6. Subudhi B, Pradhan R (2013) A comparative study on maximum power point tracking techniques for photovoltaic power systems. IEEE Trans Sustain Energ 4(1):89–98
7. Kim Y, Wang Y, Chang N, Pedram M (2010) Maximum power transfer tracking for a photovoltaic-supercapacitor energy system. In: Proceedings of the ACM/IEEE international symposium on low power electronics and design (ISLPED), pp 307–312
8. Shao H, Tsui C-Y, Ki W-H (2009) The design of a micro power management system for applications using photovoltaic cells with the maximum output power control. IEEE Trans Very Large Scale Integr VLSI Syst 17(8):1138–1142
9. Shao H, Tsui C-Y, Ki W-H (2010) Maximizing the harvested energy for micro-power applications through efficient MPPT and PMU design. In: Proceedings of the Asia and South Pacific design automation conference (ASP-DAC), pp 75–80
10. Lu C, Park SP, Raghunathan V, Roy K (2010) Efficient power conversion for ultra low voltage micro scale energy transducers. In: Proceedings of the conference on design, automation and test in Europe (DATE), pp 1602–1607
11. Kim S, No K-S, Chou P (2011) Design and performance analysis of supercapacitor charging circuits for wireless sensor nodes. IEEE J Emerg Sel Top Circ Syst (JETCAS) 1(3):391–402
12. Kim S, Chou P (2012) Size and topology optimization for supercapacitor-based sub-watt energy harvesters. IEEE Trans Power Electron 28(4):2068–2080
13. Mungan ES, Lu C, Raghunathan V, Roy K (2012) Modeling, design and cross-layer optimization of polysilicon solar cell based micro-scale energy harvesting systems. In: Proceedings of the ACM/IEEE international symposium on low power electronics and design (ISLPED), pp 123–128
14. Sulaiman SA, Hussain H, Leh NN, Razali MSI (2011) Effects of dust on the performance of PV panels. World Acad Sci Eng Technol 58:588–593
15. Swaleh MS, Green MA (1982) Effect of shunt resistance and bypass diodes on the shadow tolerance of solar cell modules. Sol Cells 5:183–198
16. Patel H, Agarwal V (2008) Maximum power point tracking scheme for PV systems operating under partially shaded conditions. IEEE Trans Industr Electron 55:1689–1698
17. Nguyen TL, Low K (2010) A global maximum power point tracking scheme employing DIRECT search algorithm for photovoltaic systems. IEEE Trans Industr Electron 57(10):345634
18. Nguyen D, Lehman B (2008) An adaptive solar photovoltaic array using model-based reconfiguration algorithm. IEEE Trans Industr Electron 55(7):2644–2654
19. Velasco-Quesada G (2009) Electrical PV array reconfiguration strategy for energy extraction improvement in grid-connected PV systems. IEEE Trans Industr Electron 56(11):4319–4331
20. Chaaban MA (2010) Adaptive photovoltaic system. In: Proceedings of the annual conference on IEEE industrial electronics society (IECON)
21. Kim Y, Park S, Wang Y, Xie Q, Chang N, Poncino M, Pedram M (2011) Balanced reconfiguration of storage banks in a hybrid electrical energy storage system. In: Proceedings of the international conference on computer-aided design (ICCAD), pp 624–631
22. Wang Y, Lin X, Kim Y, Chang N, Pedram M (2014) Architecture and control algorithms for combating partial shading in photovoltaic systems. IEEE Trans Comput Aided Des 33(4):917–930
23. Lin X, Wang Y, Pedram M, Kim J, Chang N (2014) Designing fault-tolerant photovoltaic systems. IEEE Des Test 31(3):76–84
24. Wang C, Chang N, Kim Y, Park S, Liu Y, Lee HG, Luo R, Yang H (2014) Storage-less and converter-less maximum power point tracking of photovoltaic cells for a nonvolatile microprocessor. In: Proceedings of the Asia and South Pacific design automation conference (ASP-DAC), pp 379–384

Low-Power Circuit Techniques for Efficient Energy Harvesting

Jaeha Kim, Myeong-Jae Park, Joonseok Yang and Wootaek Lim

Abstract This chapter presents the analysis and designs of efficient energy harvesting circuits interfacing two example energy generators: thermoelectric generator and piezoelectric generator. First, the key characteristics of these energy generators are described and the design criteria for the optimal performance of the basic energy harvester circuits are derived. For instance, the condition to minimize the start-up voltage of a blocking oscillator-based thermoelectric energy harvester and the condition to maximize the energy transfer efficiency of a full-bridge rectifier-based piezoelectric energy harvester are derived. Second, new circuit techniques that can overcome the limitations of the basic energy harvester circuits are introduced: the dual-path rectifier, bias-flip rectifier, and switched capacitor array. These circuit techniques either reduce energy losses, converting the voltage into a suitable range, or maintain impedance matching for the highest energy transfer.

Keywords Energy harvesting · Thermoelectric generator · Piezoelectric generator · Ultra-low-power circuit design

1 Introduction

There are increasing demands for microenergy harvesters as the interest for Internet-of-Things (IoT) applications emerge. In these applications, ubiquitous wireless sensors collect information regarding the environment such as temperature and humidity, structural heath of buildings and bridges, or human body conditions such as blood pressure and heart rate and communicate the information among one another via wireless links. One of the challenges lies in powering these

J. Kim (✉) · M.-J. Park · J. Yang · W. Lim
Department of Electrical and Computer Engineering, Seoul National University,
1 Gwanak-Ro, Gwanak-Gu, Seoul 151-744, Republic of Korea
e-mail: jaeha@snu.ac.kr

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_7

sensor devices, as it is impractical to periodically replace or recharge batteries for numerous wireless sensors distributed across a large area or implanted inside human bodies. Energy harvesting is a promising solution to this challenge as the device can continue its operation without recharging the battery or even without the battery itself.

There are many energy sources available for harvesting. For instance, photovoltaic cells harvest energy from light and piezoelectric generators do so from mechanical vibration. It is also possible to harvest energy from temperature difference using thermoelectric generators (TEGs) and from ambient RF signals using rectennas.

It is noteworthy that each energy harvesting device has different characteristics, which must be properly accounted for in order to collect their energy with the highest efficiency. For instance, a photovoltaic cell generates a DC voltage with a finite series resistance, and the maximum power transfer occurs when the load resistance is equal to this source resistance. A thermoelectric generator also outputs a DC voltage, but its magnitude is very small (a few tens of millivolts). A piezoelectric generator generates an AC current with a capacitive output impedance.

This chapter showcases low-power integrated circuit designs that can harvest energy from two energy sources: thermoelectric generators and piezoelectric generators. For all microenergy harvester circuits, including these two examples, it is imperative that the circuits themselves dissipate ultra-low-power consumption. The input power for the energy harvesting source is typically of the order of 1–100 μW and the circuits must dissipate lesser power than this input power in order to accrue net positive energy into the energy storage. Other desired characteristics include the ability to start from an energy-depleted state, the ability to operate robustly with unsteady power supplies, low energy loss during start-up or wake-up transients, etc. The remainder of this chapter will address these topics in detail.

2 Thermoelectric Energy Harvesting Circuits

Thermoelectric generators (TEGs) harvest energy from ambient heat and offer promising ways to supply power to body-worn electronics and industrial sensors. TEGs are usually in the form of a sheet that can be attached to a high-temperature surface (e.g., body skin or machine surface) facing the air with the lower room temperature. TEGs generate a voltage that is proportional to this voltage difference. Also, the effective series resistance of TEGs is typically very low, making TEGs nearly-ideal voltage sources. However, for wearable body sensors, the available temperature differences are only 1–5 K, with which the thermoelectric generators (TEGs) can generate only 15–75 mV.

Therefore, the challenge in harvesting energy using TEGs lies in converting this low DC voltage into a higher voltage, storing the energy while maintaining high transfer efficiency, and minimizing the power dissipation of all the circuits in operation.

2.1 Blocking Oscillator

A blocking oscillator is frequently used to convert a low DC voltage from a thermoelectric generator to a reasonably high-swing AC signal, which can then be converted into a high DC voltage via rectification [1]. Also, since the blocking oscillator can start oscillation as soon as the input voltage is higher than a few tens of millivolts, it has the ability to collect energy and recover from a fully energy-depleted state. This *cold-start* capability is particularly important in applications where the wireless sensors may experience low input energy conditions for an extended period of time, and hence can exhaust all the energy stored in the internal battery or capacitor.

Figure 1 shows an energy harvester circuit employing a blocking oscillator and voltage-doubling rectifier for thermoelectric generators. The oscillator consists of a normally-on transistor device and a high-turn-ratio transformer. The transistor can be either a junction field-effect transistor (JFET) or a depletion-mode MOSFET. The circuit is basically a negative-transconductance oscillator in which the parallel inductor and capacitor at the transistor's gate terminal (V_G) form a tuned LC-tank and the transistor restores the tank energy loss via a magnetic coupling between the two inductors L_S and L_P. Being a normally-on device, the transistor can provide small but nonzero transconductance (g_m) even when the DC bias voltage at V_G is at 0. The transformer magnifies this g_m value with a high-turn ratio (e.g., 100:1) and inverts its polarity to negative.

The circuit can start and sustain the oscillation when the negative transconductance has magnitude larger than the net loss conductance of the tank (g_{loss}). To analyze this condition, Fig. 2 shows the small-signal circuit model of the blocking oscillator. The resistor r_O is the finite output resistance of the transistor and R_L and C_G are the total resistance and capacitance seen at V_G, respectively. R_L is determined mainly by the series resistance of the secondary coil (L_S) and C_G is determined by the series capacitance C_S. The series resistance of the primary coil (L_P) is assumed negligible compared to r_O and R_L. Since the DC bias on the transistor's gate voltage (V_G) is fixed at 0 V, its transconductance (g_m) and output resistance

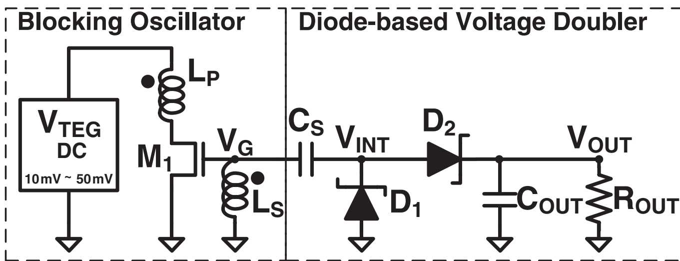


Fig. 1 A transformer-based blocking oscillator and a diode-based voltage doubler as an energy charger harvester circuit for thermoelectric generators (TEGs)

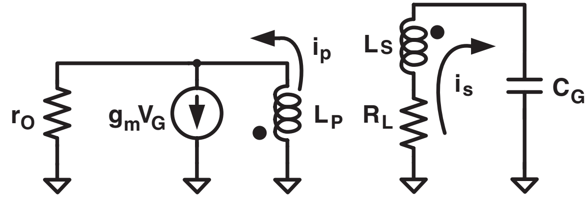


Fig. 2 The small-signal model of the blocking oscillator

(r_O) are affected mainly by the drain bias and hence the input voltage from the TEG device (V_{TEG}).

Applying the Kirchhoff current and voltage laws, a set of ordinary differential equations (ODEs) governing the inductor currents i_P and i_S can be derived as the following, where $M = k\sqrt{L_P \cdot L_S}$:

$$i_P = \frac{1}{r_o}\left(M\frac{\mathrm{d}}{\mathrm{dt}}i_S - L_P\frac{\mathrm{d}}{\mathrm{dt}}i_P\right) + g_m V_G \tag{1}$$

$$i_S R_L + L_S\frac{\mathrm{d}}{\mathrm{dt}}i_S - M\frac{\mathrm{d}}{\mathrm{dt}}i_P + \frac{1}{C_G}\int i_S dt = 0 \tag{2}$$

This set of ODEs can also be written in a state-space representation as below where the parameters α, β, γ, δ, ε and θ are defined as in (4)–(9).

$$\begin{pmatrix} i'_P \\ i''_P \\ i'_S \\ i''_S \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & \alpha & \beta & \gamma \\ 0 & 0 & 0 & 1 \\ 0 & \delta & \epsilon & \theta \end{pmatrix} \begin{pmatrix} i_P \\ i'_P \\ i_S \\ i'_S \end{pmatrix} \tag{3}$$

$$\alpha = \frac{-r_o L_S}{L_P L_S - M^2} \tag{4}$$

$$\beta = \frac{g_m r_o L_S - M}{C_G\left(L_P L_S - M^2\right)} \tag{5}$$

$$\gamma = \frac{-M R_L}{L_P L_S - M^2} \tag{6}$$

$$\delta = \frac{-M r_o}{L_P L_S - M^2} \tag{7}$$

$$\epsilon = \frac{M g_m r_o - L_P}{C_G\left(L_P L_S - M^2\right)} \tag{8}$$

$$\theta = \frac{-R_L L_P}{L_P L_S - M^2} \tag{9}$$

For the circuit to start an oscillation, the system described in (3) must be unstable and oscillatory. By applying the Routh–Hurwitz stability criterion, one can derive the following inequality condition for starting an oscillation:

$$\alpha\theta - \epsilon - \gamma\delta + \frac{\alpha\epsilon - \beta\delta}{\alpha + \theta} < 0 \tag{10}$$

Or, equivalently:

$$\frac{R_L C_G}{M} + \frac{L_P\left(R_L L_P + k^2 r_o L_S\right)}{M r_o (R_L L_P + r_o L_S)} < g_m \tag{11}$$

Assuming that the magnetic coupling coefficient k is sufficiently close to 1 and the term $R_L C_G/M$ is negligible, the start-up condition in (11) can be further simplified to a very intuitive expression:

$$n g_m r_o > 1 \tag{12}$$

where n denotes the turn ratio of the transformer, i.e., $n = \sqrt{L_S / L_P}$. To start the oscillation, the product of the transformer's turn ratio (n) and transistor's intrinsic gain ($g_m r_o$) must be greater than 1.

The start-up condition in (12) can help determine the required transformer turn ratio and transistor size that achieve a desired minimum input voltage level (V_{TEG}). For instance, the intrinsic gain ($g_m r_o$) is set by the DC bias on the drain node and hence the TEG input voltage (V_{TEG}). Depending on its value, one can either adjust the transistor's size (W_1) or transformer's turn ratio (n) to satisfy (12). In general, the intrinsic gain is lower with the lower drain bias and the larger W_1 or n is required to start the oscillation at a lower V_{TEG}. However, as W_1 or n increases, either the capacitance on the gate node (C_G) or the transformer loss (R_L) will increase, making the term $R_L C_G/M$ in (11) nonnegligible and therefore limiting the minimum V_{TEG} possible.

2.2 Voltage-Doubling Rectifier

The diode-based voltage-doubler circuit in Fig. 1 consists of a series capacitors (C_S) and two rectifying diodes (D_1 and D_2) and converts the AC signal of the oscillator output (V_G) into a DC voltage (V_{OUT}) stored on the output capacitor (C_{OUT}). If V_G swings between $-V_{G,neg}$ and $+V_{G,pos}$, the internal node V_{INT} will swing between 0 and $V_{G,neg} + V_{G,pos}$ and the diode D_2 will transfer charge to C_{OUT} whenever V_{INT} exceeds V_{OUT}. Therefore, at steady states, V_{OUT} will reach $V_{OUT,final} = V_{G,neg} + V_{G,pos}$.

Due to the finite output resistance of the oscillator, the size of the series capacitance C_S affects both the final output voltage level ($V_{OUT,final}$) and the energy harvesting efficiency (η). Here, efficiency η is defined as the ratio between the total energy transferred to the output capacitor and the total energy collected from the input power source (e.g., TEG device), while the output capacitor is charged from a fully-depleted level (i.e., 0 V) to the final level ($V_{OUT,final}$). Basically, the larger C_S lowers the effective load impedance presented to the oscillator, increasing the amount of energy transferred to C_{OUT} per cycle but making it difficult to sustain the oscillation. Therefore, a tradeoff between the output level ($V_{OUT,final}$) and efficiency (η) is expected when determining C_S.

Figure 3 shows the simulated final output voltage level ($V_{OUT,final}$) and energy harvesting efficiency (η) for different C_S values. For the TEG open-circuit voltage (V_{TEG}) of 35 mV, Fig. 3a shows that $V_{OUT,final}$ initially increases with C_S as more energy is transferred, but may decrease when excessive C_S reduces the oscillation amplitude. Figure 3b shows the existence of the optimal C_S that yields the peak efficiency η. Figure 3a and b suggests that there exists an optimum C_S that achieves the highest $V_{OUT,final}$ and η. For instance, when $V_{TEG} = 35$ mV and $R_{LOAD} = 5$ MΩ, C_S of 6 pF can achieve $V_{OUT,final}$ of 6.25 V and η of 48 %.

To increase $V_{OUT,final}$ further without degrading the efficiency η, it is also possible to use an energy harvester circuit employing two voltage-doubler circuits as shown in Fig. 4. Basically, the two voltage-doubler circuits work in a complementary fashion, each producing $V_{G,neg} + V_{G,pos}$, and $-V_{G,neg}-V_{G,pos}$, respectively, increasing $V_{OUT,final}$ up to 2×. However, since the effective load impedance presented to the oscillator is lower, slight degradation in the energy harvesting efficiency η is expected.

Figure 5 compares the simulated $V_{OUT,final}$ and η of the charger with the dual-path voltage-doubler to those of the basic charger in Fig. 1. For instance, at $V_{TEG} = 35$ mV, $V_{OUT,final}$ increases from 6.25 to 8.8 V, while η drops from 48 to 46 %. Also, the total energy stored on C_{OUT} increases by 1.98× since the energy is

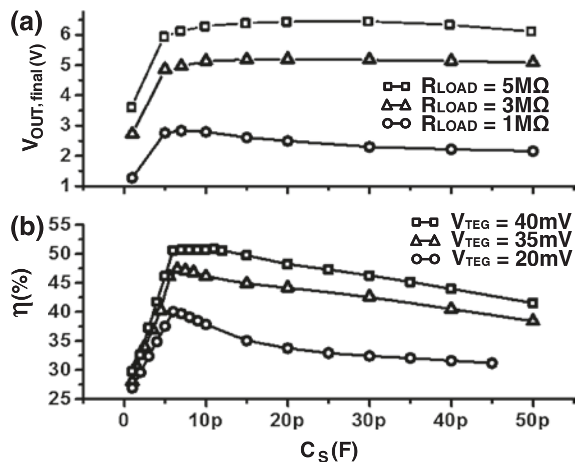


Fig. 3 The simulated final output voltage ($V_{OUT,final}$) and energy harvesting efficiency (η) of the blocking oscillator-based energy harvester in Fig. 1 for different sizes of C_S

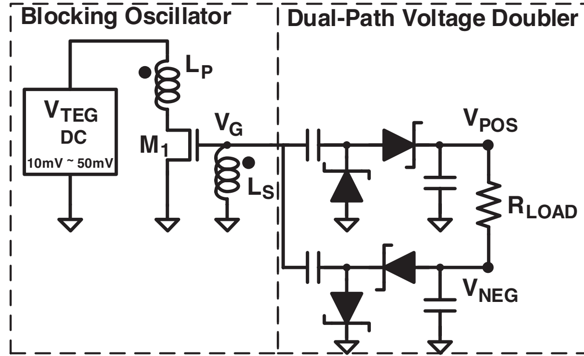


Fig. 4 A TEG energy harvester circuit with a dual-path voltage doubler

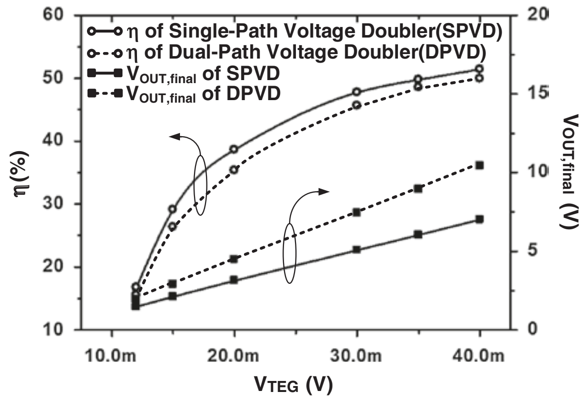


Fig. 5 The simulated final output voltage ($V_{OUT,final}$) and energy harvesting efficiency (η) of the TEG energy harvester circuit employing a dual-path voltage doubler

proportional to V_{OUT}^2. It is clear that the charger with the dual-path voltage-doubler is advantageous in collecting the higher voltage and more energy at the cost of minor degradation in the energy efficiency.

2.3 Switched Capacitor Storage

Often, the output from the described energy harvester is not suitable to charge a rechargeable battery directly. Or, it might be desirable not to use a battery at all, using capacitors as energy storages instead [2–4]. For instance, eliminating batteries can enable small form factors and low costs. However, the main challenge with charging a capacitor instead of a battery is that its voltage can vary widely with the amount of energy stored. For example, the TEG energy harvester must be able to operate and maintain high transfer efficiencies while the output voltage varies over a wide range.

Using a switched capacitor array can address this challenge. This switched capacitor array can change its configuration from series to parallel and keep the voltage seen by the energy harvester within a narrower range to maintain a high transfer efficiency. The capacitor configuration has to be administered by an ultra-low-power controller continuously monitoring the capacitor voltage level, where a pseudo-dynamic, low-crowbar-current comparator design keeps its power overhead low. It has been demonstrated that the overall EH system with four external 10 μF storage capacitors can harvest 500 μJ of energy in 10 min from a 1.4 cm^2 TEG experiencing a 3 °C temperature difference.

Figure 6 shows a switched capacitor array of four unit capacitors connected to the blocking oscillator-based energy harvester explained before. The capacitor array has three possible configurations: all four capacitors connected in series, two sets of two series capacitors in parallel, and all four in parallel. For each configuration, the effective capacitance is $C/4$, C, and $4C$, respectively, where C denotes the unit capacitance. In case of charging the capacitors from a fully-depleted state, the array is initially configured as $C/4$ and is subsequently switched to C and then to $4C$ each time the storage voltage V_{STO} reaches a certain level, e.g., 5 V. It improves the initial charging rate by 16× and keeps V_{STO} close to the optimal range of 3.5–4 V at which the charger exhibits the peak power transfer efficiency. On the other hand, when the buck converter discharges the capacitors, the array is switched back from $4C$ to C and then to $C/4$ as V_{STO} decreases and reaches 2 V. It allows a subsequent buck converter with a valid input range of 1.8–5 V to deliver 99 % of the energy stored in the capacitors and extend the sensor's operation time by 18 % compared to the case without reconfiguration.

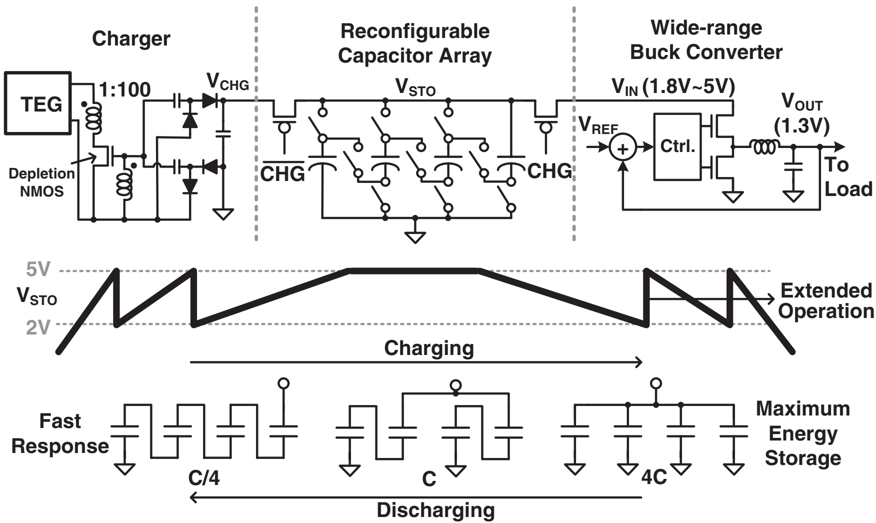


Fig. 6 A switched capacitor array to maintain the high energy transfer efficiency while charging and discharging a capacitor-based energy storage

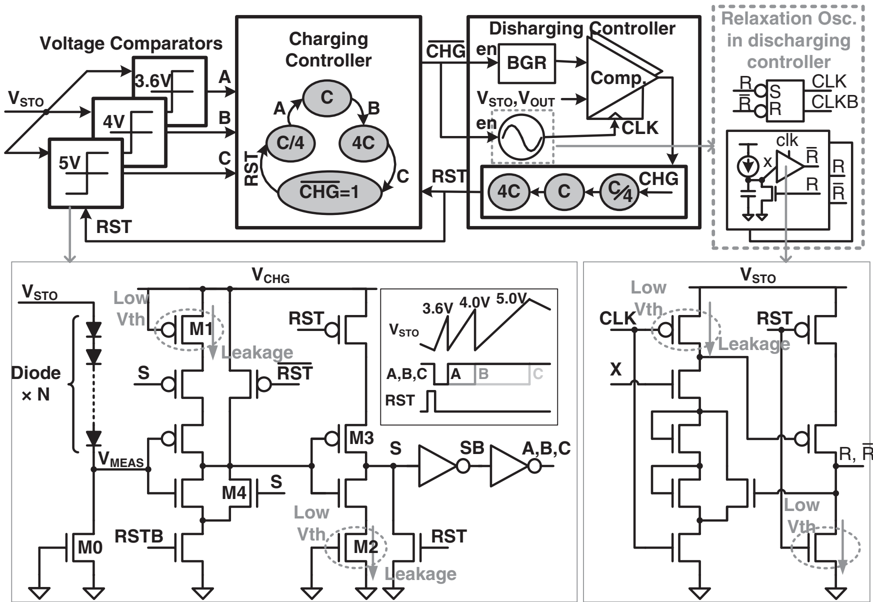


Fig. 7 The configuration controller circuit for the switched capacitor array

It is imperative that the controllers that administrate the capacitor configuration dissipate ultra-low-power consumption. To achieve this, the charging controller illustrated in Fig. 7 operates as an asynchronous FSM with a set of continuous-time voltage comparators detecting the crossing of a slowly-increasing V_{STO} ramp across different threshold levels set by the turn-on voltage drop of the series diodes. A typical issue with using a static comparator for this purpose is that the circuit can dissipate a large crowbar current, especially when the input is close to the threshold. Cascading pseudo-dynamic buffers of which the low-threshold devices M1 and M2 operate in subthreshold conduction mode can limit these crowbar currents while V_{MEAS} is below the threshold. On the other hand, when V_{MEAS} rises above the threshold, a positive feedback is initiated by the transistors M3 and M4, making sharp transitions for the outputs S and SB. A similar low-crowbar-current comparator is used for realizing a 34 nW, 5 kHz relaxation oscillator described in [5]. It provides a clock for the discharging controller implemented as a synchronous FSM, which periodically monitors V_{STO} and V_{OUT} against the reference voltages generated by a bandgap reference circuit.

3 Piezoelectric Energy Harvesting Circuits

Recently, there is an emerging demand for wireless remote switches that can operate without any batteries [6, 7]. Such batteryless wireless switches (BWS) can eliminate the need to distribute wires within buildings and recharge/replace batteries, and save installation and maintenance costs of various Internet-of-Things (IoT) systems (Fig. 8). One way to realize such a BWS is to harvest energy from the mechanical energy of pressing the button on the remote controller using a piezoelectric generator and operate a wireless transmitter using this energy [8, 9]. In particular, the advantage of piezoelectric generators (PGs) compared to magnetic-coil-based generators is their small size and light weight. Moreover, recent developments of piezoelectric material such as lead magnesium niobate–lead titanate (PMN–PT) greatly improved the energy generation efficiency of PGs.

This section will focus on the optimal condition for harvesting the maximum amount of energy from PGs using a passive full-bridge rectifier in Fig. 9, particularly when the mechanical vibration is single-pulsed rather than continuous (e.g., a button press). It will be shown that in an ideal condition, the passive bridge rectifier can harvest the maximum energy when the storage capacitance (C_{STOR}) is three times the internal capacitance (C_{PZ}) of the PG source. This analysis is adopted from [10].

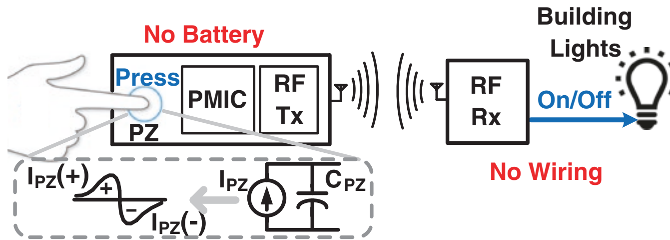


Fig. 8 Batteryless wireless switch using a piezoelectric generator as an energy source

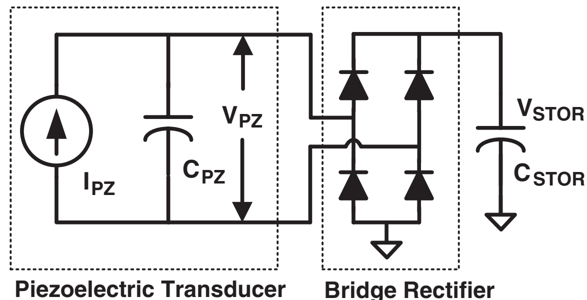


Fig. 9 A full-bridge rectifier harvesting energy from a piezoelectric generator (PG). *Reproduced by permission of the Institution of Engineering and Technology* [10]

3.1 Maximum Energy Transfer Condition for Full-Bridge Rectifier-Based Energy Harvester

In Fig. 9, a PG is connected to a passive bridge rectifier and a storage capacitor, C_{STOR}. The PG can be modeled as a current source producing a sinusoidal pulse along with a shunt capacitance C_{PZ} [11]. The PG's internal capacitance is typically low, and the output impedance is high.

Suppose the PG's sinusoidal pulse current $I_{PZ}(t)$ takes the following expression:

$$I_{PZ}(t) = \begin{cases} I_{P0}\sin\,\omega t & 0 \le t \le T \\ 0 & t > T \end{cases} \tag{13}$$

where I_{P0} is the peak current amplitude and ω is the angular frequency of the pulse. $I_{PZ}(t)$ charges mainly its internal capacitance C_{PZ} and the storage capacitor C_{STOR} via the bridge rectifier. Furthermore, if the rectifier is ideal with zero turn-on voltage (V_D) and zero on-resistance (R_{on}), the rectifier will transfer charge from C_{PZ} to C_{STOR} whenever $|V_{PZ}|$ is greater than $|V_{STOR}|$.

It has been shown in [10] that the resulting voltage on the capacitor $V_{STOR,1}$ after the first half-period of the current pulse, i.e., $0 \le t \le T/2$, is

$$V_{STOR,1} = \frac{2}{C_{PZ} + C_{STOR}} \cdot \frac{I_{P0}}{\omega} \tag{14}$$

And after the second half-period ($T/2 \le t \le T$), the voltage $V_{STOR,2}$ becomes

$$\begin{aligned} V_{STOR,2} &= \left| \frac{1}{C_{PZ} + C_{STOR}} \cdot \int_{t_1+T/2}^{T} I_P(t)\mathrm{d}t \right| \\ &= \frac{1}{C_{PZ} + C_{STOR}} \cdot \frac{I_{P0}}{\omega}(1 + \cos\,\omega t_1) \end{aligned} \tag{15}$$

$$t_1 = \frac{1}{\omega}\cos^{-1}\left(\frac{C_{STOR} - 3C_{PZ}}{C_{STOR} + C_{PZ}}\right) \tag{16}$$

where $t_1 + T/2$ is the time when the rectifier starts conducting again.

Figure 10 illustrates the waveforms of V_{PZ}, $|V_{PZ}|$, and V_{STOR} along with the input current IP during the first and second half-periods. During the first half-period, the rectifier stays on and the current I_{PZ} charges both C_{PZ} and C_{STOR} simultaneously. When the second half-period begins, the current I_{PZ} becomes negative and V_{PZ} starts to decrease from its peak value, $V_{STOR,1}$. Since $|V_{PZ}|$ is initially smaller than $|V_{STOR}|$, the rectifier turns off and C_{STOR} maintains its voltage. However, when the negative current further discharges C_{PZ} so that V_{PZ} finally becomes negative and its magnitude $|V_{PZ}|$ exceeds $|V_{STOR}|$, the other pair of diodes in the rectifier turns on and resumes the charge transfer from C_{PZ} to C_{STOR}.

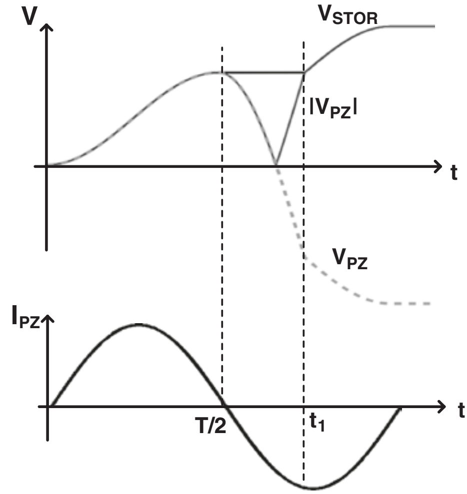


Fig. 10 Charging waveforms of V_{PZ} and V_{STOR} when the piezoelectric generator produces a single sinusoidal pulse. The diodes of the rectifier are assumed ideal. (*Reproduced by permission of the Institution of Engineering and Technology* [10])

The final output voltage after the pulse period V_{STOR} is the sum of $V_{STOR,1}$ and $V_{STOR,2}$:

$$\begin{aligned} V_{STOR} &= \frac{1}{C_{PZ} + C_{STOR}} \frac{I_{P0}}{\omega} \left(2 + \left(1 + \frac{C_{STOR} - 3C_{PZ}}{C_{STOR} + C_{PZ}} \right) \right) \\ &= \frac{I_{P0}}{\omega} \frac{4C_{STOR}}{(C_{PZ} + C_{STOR})^2} \end{aligned} \tag{17}$$

And the final energy stored in C_{STOR}, E_{STOR}, is

$$E_{STOR} = \frac{1}{2} C_{STOR} V_{STOR}^2 = \frac{8I_{P0}^2}{\omega^2 C_{PZ}} \frac{\left(C_{STOR}/C_{PZ} \right)^3}{\left(1 + C_{STOR}/C_{PZ} \right)^4} \tag{18}$$

For given values of I_{P0}, ω, and C_{PZ}, the amount of available energy depends on $x^3/(1 + x)^4$, where x is the ratio between C_{STOR} and C_{PZ}. It can be shown that the expression has a global maximum at $x = 3$. In other words, the maximum energy E_{max} can be obtained when $C_{STOR} = 3C_{PZ}$, which is

$$E_{max} = \frac{8I_{P0}^2}{\omega^2 C_{PZ}} \frac{3^3}{(1+3)^4} = \frac{27}{32} \cdot \frac{I_{P0}^2}{\omega^2 C_{PZ}} \tag{19}$$

However, this maximum energy value is likely to degrade when the diodes in the rectifier have nonzero turn-on voltage V_D and turn-on resistance R_{on}. For instance, when the diodes have nonzero V_D, the rectifier will turn on only when $|V_{PZ}|$ is greater than $|V_{STOR}|$ by V_D. This results in the reduction in the output voltage after the first half-cycle ($V_{STOR,1}$) and also the delay in the rectifier turn-on time during the second half-cycle (t_1). Therefore, the nonzero V_D will shift V_{STOR}

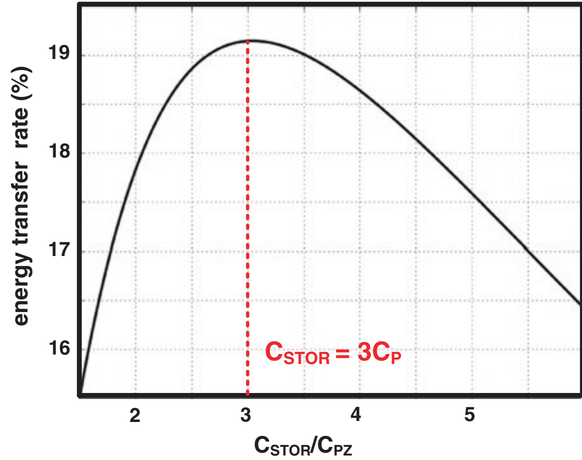


Fig. 11 Simulated transferred energy versus C_{STOR}/C_{PZ} ratio assuming a constant V_D model for the rectifying diodes. (*Reproduced by permission of the Institution of Engineering and Technology* [10])

curve in Fig. 10 both rightward and downward. Furthermore, when the diodes have nonzero R_{on}, their on-resistance will cause additional voltage drops across the conducting diodes, resulting in the lower final output voltage V_{STOR}.

Figure 11 shows the simulated plot of the transferred energy normalized to the maximum energy available using the constant V_D model. The result confirms that the maximum point is at the ratio of 3.

3.2 Bias-Flip Rectifier and Switched Capacitor Array

However, the piezoelectric energy harvester with a basic full-bridge rectifier described in the previous subsection can experience energy losses when the input current inverts its polarity and discharges the internal capacitance (C_{PZ}), or when the input capacitance of the harvester circuit is not matched to $3 \cdot C_{PZ}$. To recover these energy losses, it is possible to combine a bias-flip rectifier and a 6:1 series-parallel switched capacitor (SC) converter as shown in Fig. 12.

The bias-flip rectifier with a switched inductor L_B is modified from [12] and reduces the loss caused by the remaining energy in the internal capacitance C_{PZ}. In other words, when the polarity of the input current from the piezoelectric generator I_{PZ} changes from positive to negative, the switch S_1 turns on to form a resonant circuit between L_B and C_{PZ} and inverts the voltage stored on C_{PZ}. As a result, the remaining charge on C_{PZ} can be transferred to the next stage without being canceled and wasted by the negative-polarity I_{PZ}.

The switched capacitor converter plays two roles: impedance matching for maximum energy transfer and down-conversion of the output voltage V_{out}. During charge, the capacitors are stacked in series making its overall capacitance close to $3{\cdot}C_{PZ}$ to maximize the transferred energy into the storage. However, it results

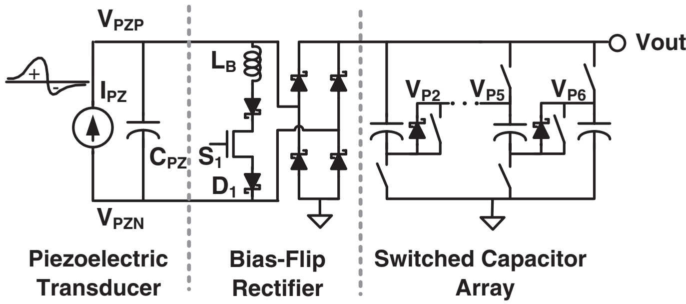


Fig. 12 A piezoelectric energy harvester using a bias-flip rectifier and switched capacitor array

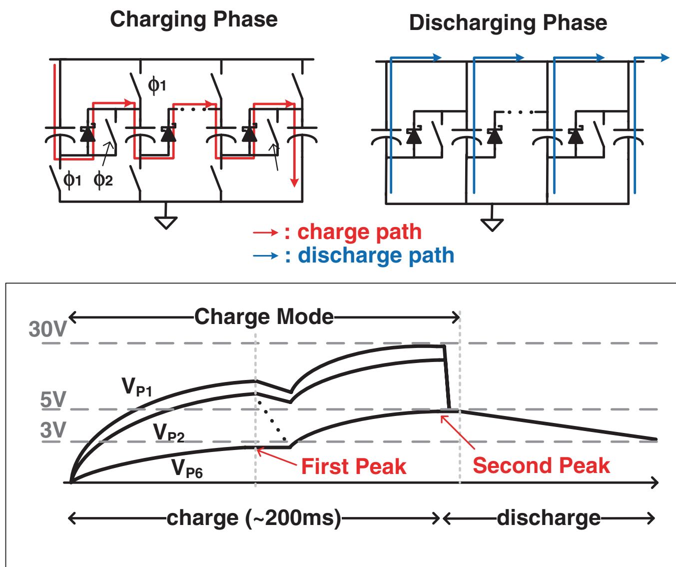


Fig. 13 Operation of the piezoelectric energy harvester in Fig. 12

in very high voltage stored on the capacitor array which is difficult to handle or convert down to a normal range with high efficiency. By switching the capacitors into a parallel configuration during discharge, the output voltage can be down-converted by $1/N$, where N is the number of capacitors.

Figure 13 illustrates the operation of this piezoelectric energy harvester employing the described bias-flip rectifier and switched capacitor converter. When a piezoelectric push-button is pressed, it generates a single, bipolar AC pulse. During the first half-period while the piezoelectric current I_{PZ} is positive, the current builds up a voltage V_{out} on the storage capacitors connected in series. When V_{out} reaches a peak, detected by a comparator that monitors the difference between V_{out} and its RC-filtered version, the switch S_1 in the bias-flip rectifier is enabled,

flipping the polarity of the voltage stored in the internal capacitor C_{PZ}. And charging continues into the second half-period. When the charging phase is completed, the switches in the switched capacitor array turn on, switching the capacitors from series to parallel configuration. The output voltage V_{out} is then down-converted by $1/N$. The described bias-flip rectifier and switched capacitor converter can greatly improve the efficiency of the piezoelectric energy harvester.

4 Summary

This chapter illustrated the design of energy harvester circuits interfacing two energy generators: thermoelectric generator and piezoelectric generator. For both examples, it is important to understand their key characteristics in order to design circuits that can collect the maximum energy. For instance, a thermoelectric generator is essentially a DC voltage source producing very low voltage in mV-range with negligible output resistance. On the other hand, a piezoelectric generator is an AC current source with a capacitive output impedance, of which open-circuit voltage can be as high as a few tens or hundreds of volts.

For each energy generator, this chapter has derived the conditions that keep its corresponding energy harvester circuit optimal. For instance, the design criteria for a blocking oscillator to harvest energy from the lowest possible input voltage are derived. Also, the maximum energy transfer condition for a full-bridge rectifier-based piezoelectric energy harvester is derived. These conditions also revealed that it is not always possible to achieve the best energy harvesting efficiency with the basic blocking oscillator or full-bridge rectifier.

To address these limitations, this chapter has also described some circuit techniques. For thermoelectric generators, the dual-path rectifier can double the output voltage with only a small degradation in efficiency. The switched capacitor array can maintain high efficiency by keeping the output voltage of the energy harvester circuit within a narrow range. For piezoelectric generators, the bias-flip rectifier and switched capacitor array both recover the energy loss due to the internal capacitance of the piezoelectric generator.

Acknowledgments This work is supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT & Future Planning as the Global Frontier Project.

References

1. Damaschke JM (1997) Design of a low-input voltage converter for thermoelectric generator. IEEE Trans Ind Appl 33(5):1203–1207
2. Im J-P et al (2012) A 40 mV transformer-reuse self-startup boost converter with MPPT control for thermoelectric energy harvesting. In: IEEE Int'l solid-state circuits conference (ISSCC) digital technology papers, pp 104–106, Feb 2012

3. Ramadass YK et al (2011) A battery-less thermoelectric energy harvesting interface circuit with 35 mV startup voltage. IEEE J Solid-State Circuits 46(1):333–341
4. Bandyopadhyay S et al (2012) Platform architecture for solar, thermal, and vibration energy combining with MPPT and single inductor. IEEE J Solid-State Circuits 47(9):2199–2215
5. Lim W et al (2013) A 5 V 33-kHz, 0.7-μW pulse generation circuit for ultra-low-power boost charging energy harvesters. IEEE Asian Solid-State Circuits Conference (A-SSCC), pp 449–452, Nov 2013
6. Paradiso JA, Feldmeier M (2001) A compact, wireless, self-powered pushbutton controller. In: Proceedings of the Ubicomp 2001: Ubiquitous Computing, Oct 2001
7. Tan YK et al (2006) Energy harvesting using piezoelectric igniter for self-powered radio frequency wireless sensors. In: Proceedings of IEEE Int'l conference on industrial technology, Dec 2006
8. Ottman G et al (2002) Adaptive piezoelectric energy harvesting circuit for wireless remote power supply. IEEE Trans Power Electron 17(5):669–676
9. Khaligh A et al (2010) Kinetic energy harvesting using piezoelectric and electromagnetic technologies—state of the art. IEEE Trans Industr Electron 57(3):850–860
10. Yang et al (2014) Maximum energy transfer condition for piezoelectric energy harvesters with single pulsed vibration inputs. IET Electron Lett 50(8):629–631
11. Yang Y, Tang L (2009) Equivalent circuit modeling of piezoelectric energy harvesters. J Intell Mater Syst Struct 20(18):2223–2235
12. Ramadass YK, Chandrakasan AP (2009) An efficient piezoelectric energy harvesting interface circuit using a bias-flip rectifier and shared inductor. IEEE J of Solid-State Circuits 45(1):189–204

Part III
Materials for Low Energy

Preparation of Porous Graphene-Based Nanomaterials for Electrochemical Energy Storage Devices

Yuanzhe Piao

Abstract Graphene-based nanostructures exhibit good mechanical strength, high porosity, outstanding electrical conductivity, and excellent thermal and chemical stability, which in addition to its low cost, versatile functionalization chemistry, and relative ease of large-scale preparation make it ideally suited to serve as a key component for the development of new electrode materials. Recently, a wide variety of methods have been developed for the formation of porous graphene architectures to further improve the performances. Porous graphene provides abundant pathways for rapid ion diffusion and high accessible surface area. In this chapter, the recent continued breakthroughs in the preparation of porous graphene-based nanoarchitectures as well as their applications as electrode materials for electrochemical energy storage devices are introduced.

Keywords Graphene · Porous · Electrochemistry · Electrochemical energy storage · Nanostructures · Lithium-ion rechargeable batteries · Supercapacitors

1 Introduction

Graphene is an atomic single layer of honeycomb carbon lattice. Recently, graphene and graphene-based nanomaterials have attracted increased attention because of their unique properties and great potential for numerous applications [3, 14, 21, 33, 36–38, 51, 63]. In particular, research on graphene-based nanomaterials for electrochemical energy storage has progressed rapidly during recent years due to the increasing demand for the development of these storage devices with improved performance including high energy, power density, and excellent cycle stability,

Y. Piao (✉)
Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea
e-mail: parkat9@snu.ac.kr

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_8

while maintaining low production cost [15, 43, 44, 71, 76, 83, 92]. Among various carbon materials, graphene has awakened a tremendous interest because of its superior electronic conductivity, large theoretical specific surface area, and physicochemical stability [22, 30, 52]. These interesting properties make graphene and graphene-based nanomaterials promising electrode materials for various electrochemical energy storage devices (e.g., supercapacitors and batteries), which have a wide range of applications from microelectromechanical systems to portable electronic devices, and to electric vehicles.

Graphene-based nanomaterial from graphite oxide is being manufactured in large scale at relatively low cost [58]. During the past decade, various graphene-based nanomaterials (Figs. 1, 2, and 3) have been investigated as potential electrode materials with high specific capacity and long cycle stability [26, 29, 31, 40–42, 46, 48, 53, 62, 82]. However, in spite of the previous efforts to prepare better electrochemical energy storage devices by using graphene, market demand for higher performances of these devices are ever increasing. For lithium-ion batteries, since lithium ions cannot pass through the carbon atomic arrays in 2D sheets of graphene, therefore, lithium ions have to tortuously detour to reach the electrolyte. This results in a longer diffusion distance and slow charge–discharge rate in lithium-ion battery systems [35, 65]. Supercapacitors, also called ultracapacitors or electrochemical capacitors, store electrical charge on high-surface-area conducting materials. Their widespread use is limited by their low energy storage density and

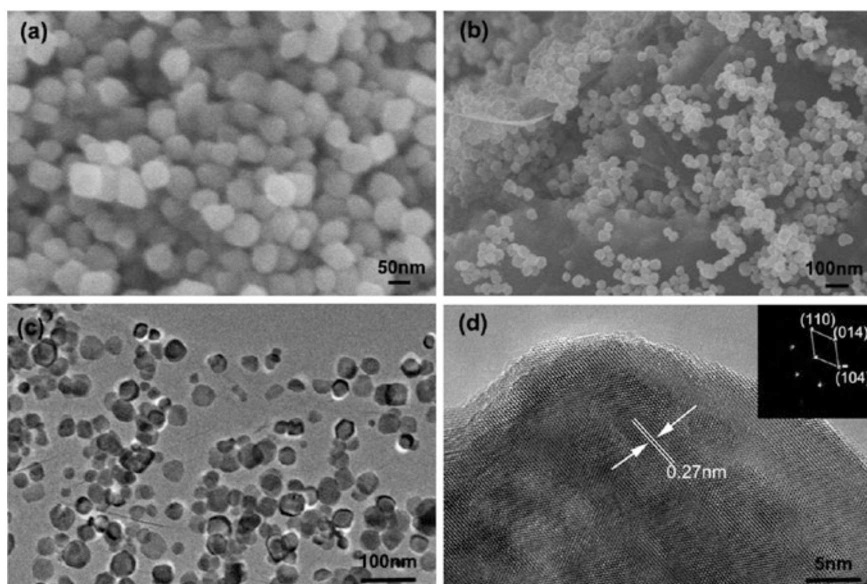


Fig. 1 **a** SEM image of Fe_2O_3; **b** SEM image of Fe_2O_3/graphene; **c** TEM image of Fe_2O_3/graphene; **d** TEM image of Fe_2O_3/graphene in high resolution. Reproduced from Ref. [82] with permission from Elsevier

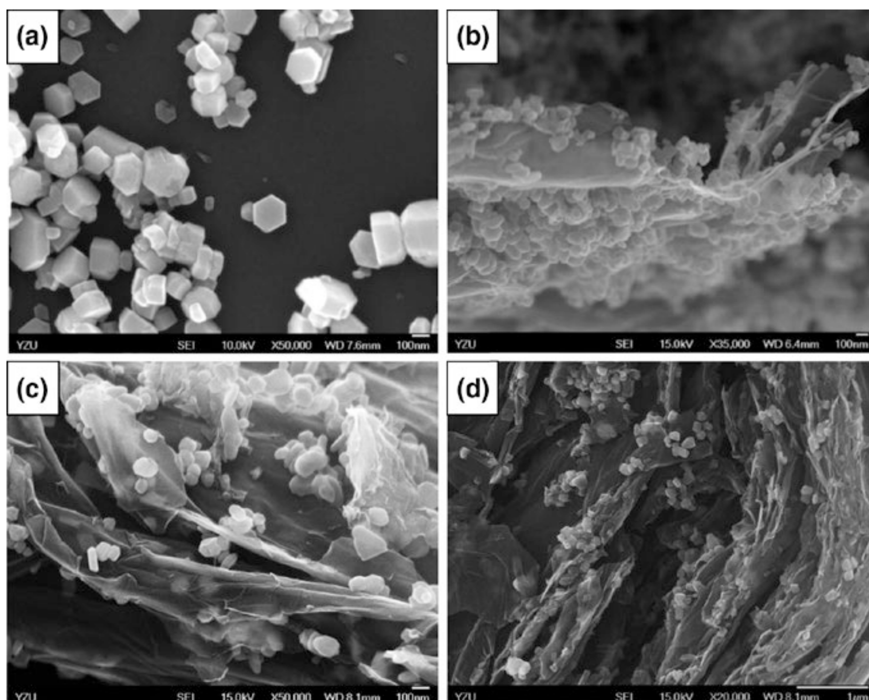


Fig. 2 FE-SEM photographs of **a** ZnO nanocrystals and different types of ZnO@GN hybrids, **b** ZnO@GN-1, **c** ZnO@GN-2, and **d** ZnO@GN-3. Reproduced from Ref. [26] with permission from Elsevier

relatively high effective series resistance. As previously reported, the theoretical specific surface area of a single graphene sheet is extremely high (~2600 $m^2\ g^{-1}$) [64]. However, the real accessible surface areas of graphene-based materials are far below this value, mainly due to the strong aggregation tendency of graphene sheets (Fig. 4).

To surmount these challenges, graphene sheets were further treated into porous structures which mean abundant pathways for rapid lithium-ion diffusion [79] and to achieve higher accessible surface area [90]. Due to its unique structural and electronic characteristics, a porous graphene opens up new opportunities for the development of electrode materials with novel nanoarchitectures in electrochemical energy storage devices. In this chapter, recent advances and novel strategies in the preparation of porous graphene architectures as well as their applications as electrode materials for electrochemical energy storage devices are introduced. Finally, the future prospects in the development of graphene-based nanocomposite materials with improved energy storage performances are discussed.

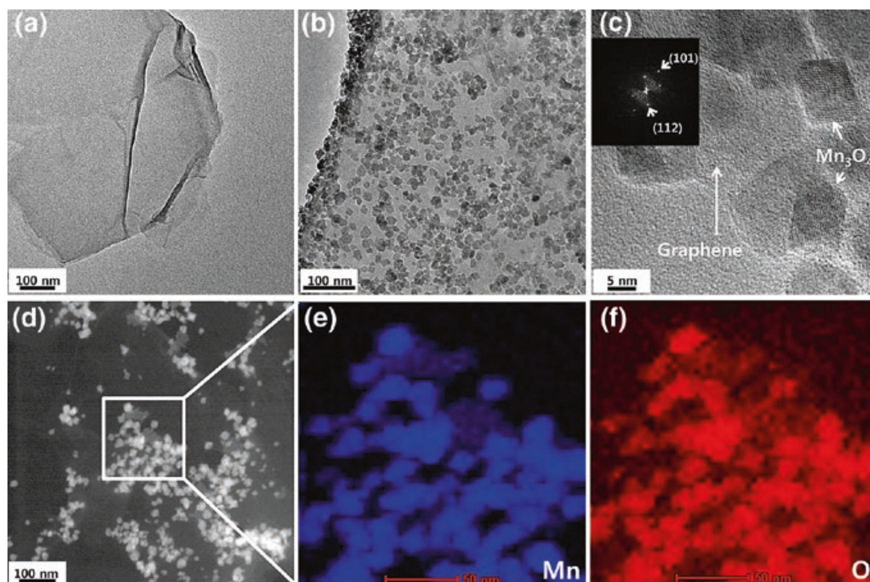


Fig. 3 TEM images of **a** GO before the one-step in-situ transformation reaction, and **b** Mn_3O_4/GNS composite after the one-step in-situ transformation reaction. **c** HRTEM image of individual Mn_3O_4 nanoparticle in Mn_3O_4/GNS composite; the inset shows the electron diffraction pattern of Mn_3O_4 nanoparticle in Mn_3O_4/GNS composite. **d** HAADF-STEM image of Mn_3O_4/GNS composite and EDS element mapping results for **e** Mn and **f** O species of Mn_3O_4/GNS composite. Reproduced from Ref. [48] with permission from Elsevier

2 Strategies to Buildup Porous Graphene

A wide variety of strategies have been developed for the formation of porous graphene architectures as outlined in Table 1. These can be roughly divided into two categories. One is the in-plane generation of defective pores into the graphene sheets and the other is the out-of-plane generation of 3D graphene-based porous superstructures (Scheme 1) (Figs. 5 and 6).

2.1 Generation of Defective Pores into the Graphene Sheets

When etching of graphene is performed under appropriate acid/oxidizer solution, carbon erosion will occur. Accordingly, porous graphene is produced when graphene oxide dispersed in water is treated with acid/oxidizer solution under sonication [84] or microwave irradiation [18]. HNO_3 and $KMnO_4$ are typical examples of acid and oxidizer used.

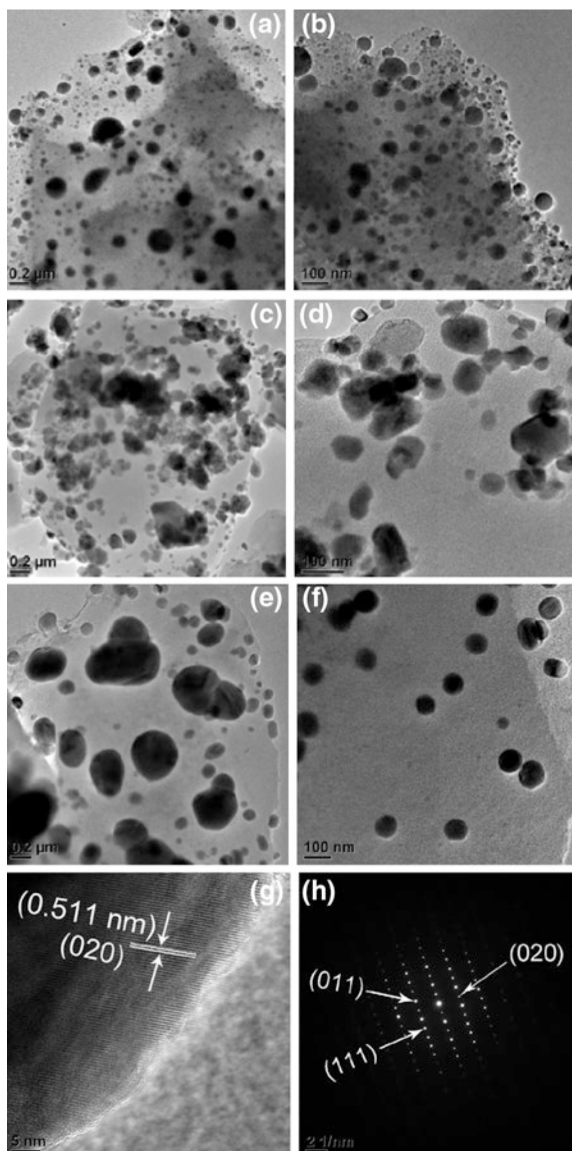


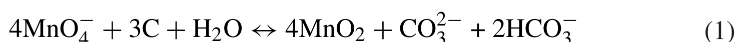
Fig. 4 TEM images of **a** and **b** LFP/CA-1, **c** and **d** LFP/CA-2, and **e** and **f** LFP/CA-3. And **g** HRTEM image and **h** SAED pattern of the $LiFePO_4$ particle in LFP/CA-3. Reproduced from Ref. [79] with permission from Elsevier

Fan et al. [18] investigated the preparation of porous graphene using $KMnO_4$ as oxidizer under microwave irradiation (Fig. 7). The obtained porous graphene revealed a pore size of approximately 3 nm and a specific surface area of 1374 $m^2\ g^{-1}$.

Table 1 Typical strategies for the formation of porous graphene

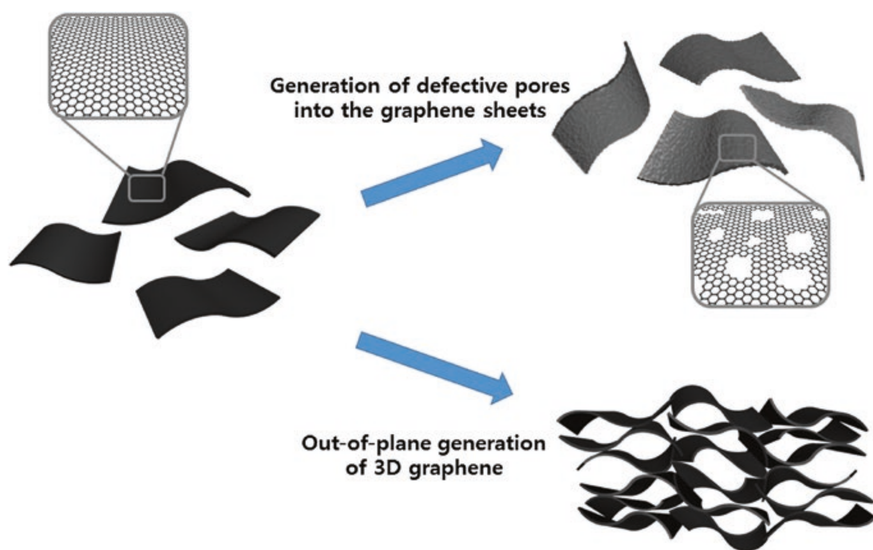
Method	Detail	Pore	Pore size	Specific surface area	Example
Acid/oxidizer	HNO_3, microwave	meso/macro	7–600 nm	15–25 m^2/g	[88]
	HNO_3, microwave	macro	100–1000 nm	784 m^2/g	[84]
	$KMnO_4$, microwave	meso	3 nm	1374 m^2/g	[18]
	$KMnO_4$, HCl, H_2O_2	meso	–	81 m^2/g	[11]
Base	KOH	micro/meso	0.6–5 nm	3100 m^2/g	[90]
	KOH	micro/meso	~5 nm	2400 m^2/g	[86]
	KOH	micro/meso	1–10 nm	3523 m^2/g	[87]
Hard template	Silica	meso	32.5 nm	851 m^2/g	[27]
	Silica	macro	140 nm	412 m^2/g	[2]
	PMMA	macro	107.3 nm	128.2 m^2/g	[9]
	SPS, PVP, Ni foam	micro/meso/ macro	–	70 m^2/g	[68]
	PS	macro	2000 nm	–	[13]
	Ice template	macro	–	–	[66]
	Ice template	macro	–	–	[17]
Deposition	Chemical vapor deposition	macro	~100 μm	670 m^2/g	[56]
	Chemical vapor deposition	macro	–	–	[8]
	Chemical vapor deposition	macro	–	1654 m^2/g	[50]
	Electrochemical deposition	macro	100–1000 nm	1000 m^2/g	[10]
Hydrothermal	Hydrothermal	meso	20 nm	–	[24]
	Hydrothermal	meso	9 nm	281 m^2/g	[20]
	Hydrothermal	macro	~1000 nm	–	[67]
	Hydrothermal	macro	–	–	[74]
	Hydrothermal	macro	–	–	[1]
Others	Polymer, thermal decompose	micro/meso	30 nm	1720 m^2/g	[77]
	Self-assembly	macro	1000 nm	–	[39]

The following equation shows the reaction of carbon with $KMnO_4$:



$$4MnO_4^- + 3C + H_2O \leftrightarrow 4MnO_2 + CO_3^{2-} + 2HCO_3^- \quad (1)$$

Porous graphene could also be produced by using a strong base. Romanos et al. [57] presented nanospace engineering of KOH-activated carbon. It is reported that



Scheme 1 Schematic illustration of in-plane generation of defective pores into the graphene sheets and out-of-plane generation of 3D graphene-based porous superstructures

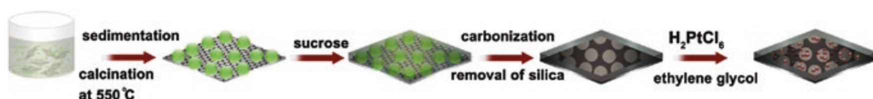


Fig. 5 Preparation pathway of the MC–GR and Pt/MC–GR composites. Reproduced from Ref. [2] with permission from Elsevier

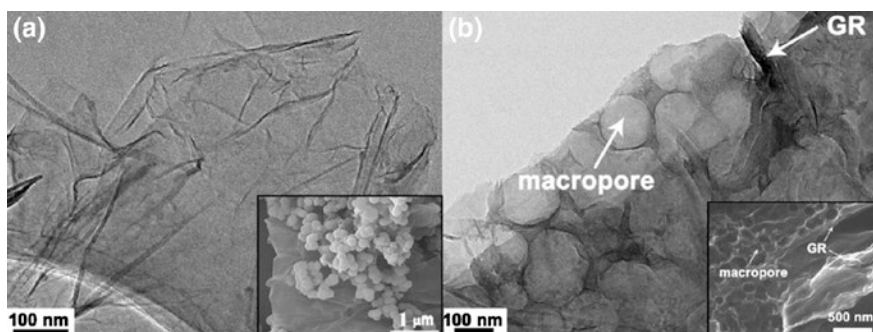


Fig. 6 **a** TEM image of GR. Inset shows SEM image of SiO_2-GO after calcination. **b** TEM image of MC–GR. Inset shows SEM image of MC–GR. Reproduced from Ref. [2] with permission from Elsevier

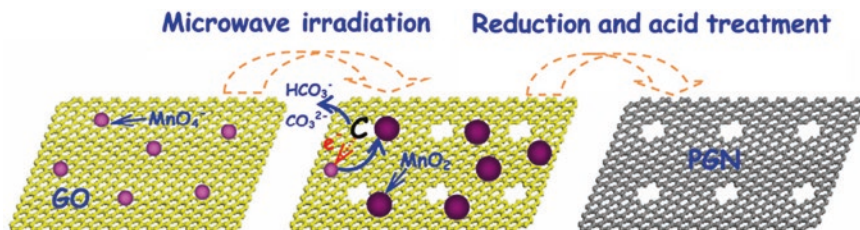


Fig. 7 Illustration of the formation of porous graphene material with pores on the surface of sheet. Reproduced from Ref. [18] with permission from Elsevier

high specific surface areas, porosities, subnanometer (<1nm), and suprananometer (1–5 nm) pore volumes could be quantitatively controlled by a combination of KOH concentration and activation temperature. Recently, the chemical process was used to prepare chemically activated graphene. Typical examples of these bases are KOH and NaOH. After chemical activation, the specific surface area of the porous graphene is increased to become closer to the theoretical value [86, 90].

2.2 *Out-of-Plane Generation of 3D Graphene-Based Porous Superstructures*

Besides the in-plane generation of defective pores into the graphene sheets, 3D graphene-based out-of-plane porous superstructures could be built up. Using the hard template approaches, graphene layers were deposited on inorganic/organic particles larger than 50 nm or in situ grown on metallic porous frameworks followed by the elimination of template that can result into graphene-based materials with 3D porous structures.

Using uniform polymethyl methacrylate (PMMA) latex spheres as hard templates, Chen et al. [9] prepared a controllable 3D macroporous bubble graphene film with tailorable microstructure. Zhao et al. [27] developed a novel hydrophobic interaction-driven hard templating approach for the rational designed preparation of nanoporous graphene foams with controlled pore size, high surface area, and ultra-large pore volume. Monodisperse silica particles were used as the templates to prepare nanoporous graphene foams. The generated graphene foams show the highest total pore volume value in all the reported porous graphene materials. Additionally, they demonstrated that metal oxide nanoparticles can be easily decorated on the pore walls, due to the ultra-large open-porous feature and the homogeneous hydrophobic surface nature. Huh et al. [13] built a 3D macroporous structure that consists of chemically modified graphene by using polystyrene particles as a sacrificial template. Furthermore, for further capacitance boost, a thin layer of MnO_2 was additionally deposited onto the embossed chemically modified graphene. The porous graphene nanostructure shows a large surface area facilitates fast ionic transport

within the electrode while preserving decent electronic conductivity and thus endows the composite electrodes with excellent electrochemical properties. Using sulfonated polystyrene (SPS) sphere as hard template, Zhang et al. [68] also prepared porous graphene electrode by an in situ constructing strategy.

Chemical vapor deposition and electrochemical deposition were used to prepare graphene-based materials with 3D porous structures using Ni foam, porous MgO, etc., as templates. Zhang et al. [4] prepared a novel 3D porous graphene networks by the scalable ethanol-chemical vapor deposition method. They demonstrated that the 3D graphene network can be used as a good platform to construct graphene/metal oxide composites for supercapacitor applications.

3D porous graphene-based composite materials were prepared by electrochemical deposition [10]. 3D graphene porous material is prepared electrochemically by reducing a concentrated graphene oxide dispersion. Subsequently, the second component is electrochemically deposited onto this 3D matrix, yielding graphene-based 3D porous composite material. The prepared graphene-based composite materials have a conductive graphene network as the matrix, onto which the second component is homogeneously coated.

Sun et al. [67] reported a hydrothermal approach to prepare nitrogen-doped graphene in various forms, including a stable dispersion, a hydrogel and an aerogel of nitrogen-doped graphene. The stable dispersion mainly consists of single-sheet graphene and the hydrogel is physically cross-linked to be quite strong.

2.3 Other Methods

Ogale et al. reported a one-step, catalyst-free process for the preparation of single-layer-graphene-assembled porous carbon by a polymer pyrolysis route [77]. The surface area of the single-layer-graphene-assembled porous carbon was found to be 1720 $m^2\ g^{-1}$. The nanomaterial was tested as a supercapacitor and showed a high capacitance value of ca. 154 $F\ g^{-1}$ in an aqueous electrolyte in a typical electrochemical cell.

Zhu et al. reported a simple approach to transform the CVD graphene films through self-assembly into porous and continuous fibers with tunable diameter, pore distribution, and high electrical conductivity [39]. Graphene is first self-assembled from a 2D film to a 1D fiber-like structure in an organic solvent and then dried to give the porous and crumpled structure.

3 Applications in Lithium-Ion Rechargeable Batteries

Due to their attractive advantages over other types of batteries, lithium-ion batteries have been widely used as power sources for various portable electronic devices. More recently, they have attracted growing attention as power supplies

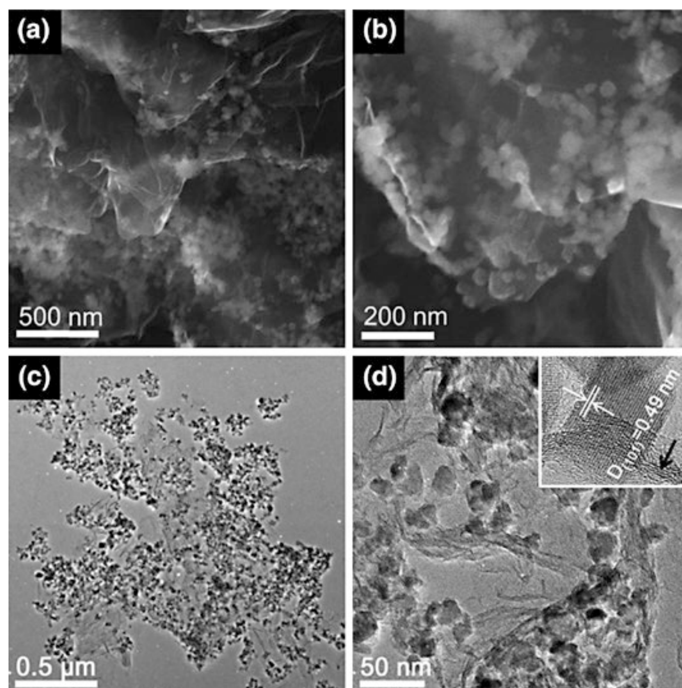


Fig. 8 **a**, **b** SEM images and **c**, **d** TEM images of M-NG composite (inset: HRTEM image of Mn_3O_4 nanoparticle on the graphene sheet). Reproduced from Ref. [54] with permission from Elsevier

for electric vehicles and hybrid electric vehicles. Intensive research has been performed to develop new electrode materials with improved performance for lithium-ion batteries. Accordingly, a great deal of effort has been made to find alternative electrode materials with improved electrochemical performance for lithium-ion batteries. To improve lithium storage capacity, the most promising carbon is disordered partially graphitic carbon from both a technological and scientific point of view, since defects provide large excess capacities [32]. Among the various novel nanostructured electrode candidate materials, graphene-based electrode materials (Fig. 8) are of particular interest due to their high surface area and good conductivity [54, 85].

3.1 Porous Graphene-Based Anode Materials

Porous graphene nanostructure could provide more space to accommodate the volume change of the active materials during the charge–discharge to enhance the electrochemical stability of the electrodes. Accordingly, porous graphene-based

materials hold promise as novel electrode materials to further improve the performance of lithium-ion batteries.

Feng et al. presented a bottom-up approach to the large-scale production of 2D sandwich like graphene-based mesoporous carbon [80]. Their high surface area, thinness, and numerous mesopores are favorable for the accessibility of the electrolyte, rapid diffusion of lithium ions, and host uptake. Additionally, the graphene layers can act as mini-current collectors dispersed in the electrode, which facilitates the fast transport of electrons during the charge–discharge cycling due to its high electrical conductivity. When the porous nanocomposite material is used as an anode material for lithium-ion batteries, the nanostructured electrode material exhibits a first discharge capacity of 915 mA h g^{-1}, which rapidly stabilizes and remains at 770 mA h g^{-1} even after 30 cycles, when cycled at a rate of C/5.

Graphene-based 3D macroporous materials are favorable electrode candidates for lithium-ion batteries. Yan et al. reported a simple method for the preparation of 3D graphene/Fe_3O_4 architectures by a mild chemical reduction of graphene oxide in the presence of Fe_3O_4 nanoparticles [7]. The obtained superparamagnetic, porous, and lightweight material shows good electrochemical performance as anode material in lithium-ion battery.

Using porous MgO sheets as a template, Fan et al. reported a simple CVD approach for scalable preparation of porous graphene materials (Figs. 9, and 10) [19]. The resulting porous graphene networks exhibit a high reversible capacity of 1723 mA h g^{-1}, and excellent high rate capability and cycling stability for Li-ion batteries.

A 3D porous architecture of Si/graphene nanocomposite was rationally designed and constructed through an in situ magnesiothermic reduction of SiO_2/graphene oxide composites series and spray-drying with additional graphene [73]. The porous nanoarchitectured composite has superior electrochemical stability and the 3D graphene network shows enhanced electrical conductivity as well as improves rate performance. Furthermore, the 3D nanoarchitechture can be cycled at extremely high Li^+ extraction rates.

3D graphene foams cross-linked with Fe_3O_4 nanospheres were prepared by hydrothermal treatment [70]. The Fe_3O_4 nanospheres are wrapped by graphene sheets and further confined within continuous graphene networks. Such hierarchical Fe_3O_4/graphene hybrids provide double protection against the volume changes of Fe_3O_4 nanospheres during electrochemical processes. The graphene shells suppress the aggregation of Fe_3O_4 nanospheres and buffer the volume expansion, while the interconnected 3D graphene networks act to reinforce the core–shell structure of Fe_3O_4@graphene shell (Fe_3O_4@GS) and thus enhance the electrical conductivity of the overall electrode. As a result, 3D graphene foams cross-linked with Fe_3O_4 nanospheres (Fe_3O_4 NSs) encapsulated with graphene (Fe_3O_4@GS/GF) delivers a high reversible capacity of 1059 mA h g^{-1} over 150 cycles, and excellent rate capability, thus exhibiting great potential as an anode material for lithium storage.

Fan et al. reported a bottom-up strategy assisted by atomic layer deposition to graft bicontinuous mesoporous nanostructure Fe_3O_4 onto 3D graphene foams and

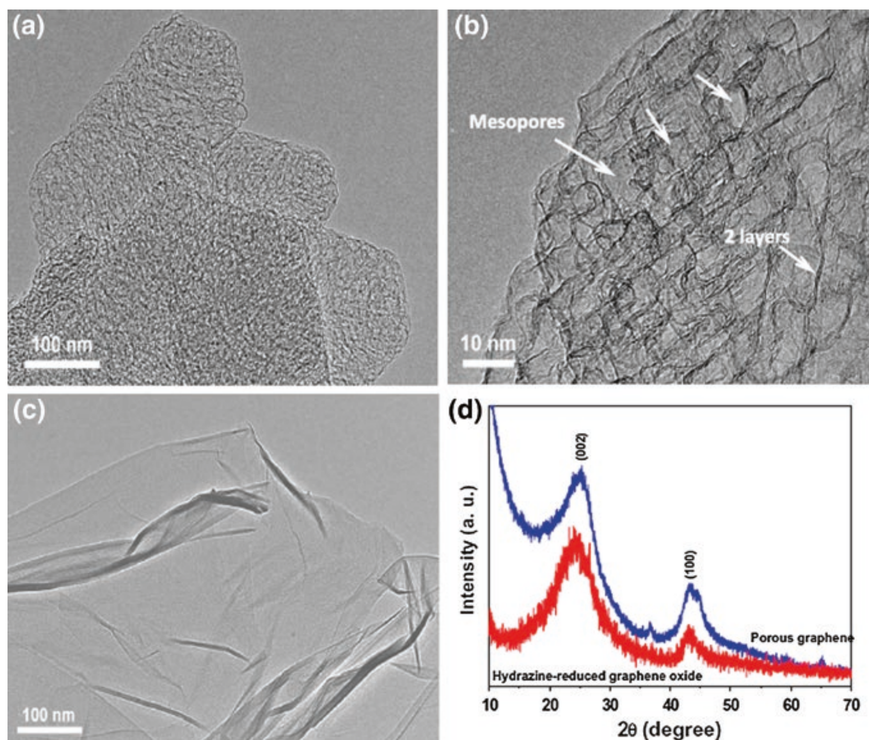


Fig. 9 TEM images of porous graphene (**a**) and (**b**), and hydrazine-reduced graphene oxide (**c**). XRD patterns of porous graphene and hydrazine-reduced graphene oxide (**d**). Reproduced from Ref. [19] with permission from Elsevier

directly use the composite as the lithium-ion battery anode [45]. This electrode exhibits high reversible capacity and fast charging and discharging capability. A high capacity of 785 mA h g^{-1} is achieved at 1 C rate and is maintained this high capacity up to 500 cycles. Moreover, the rate of up to 60 C is also demonstrated, rendering a fast discharge potential. For the first step, graphene foam was grown on Ni foam by CVD and then Ni was etched away by a mixture of $FeCl_3$ and HCl solution. In the second step, a layer of ZnO was coated onto the graphene foam by atomic layer deposition.

A facile and general method was reported to prepare ordered porous binder-free 3D porous graphene–metal oxide@carbon electrodes at a large scale [91]. Viscous precursor paint was prepared by mixing graphene oxide slurry, polystyrene aqueous solution and metal salt. The ordered porous binder-free electrodes were obtained after heat treatment of the paint at 400 °C under Ar for 60 min. The overall framework is macroporous structure and made of metal oxides or a mixture of graphene and metal oxides. There are secondary pores in the walls of the porous electrode with size in the range of 5–10 nm. The macropores are derived from the

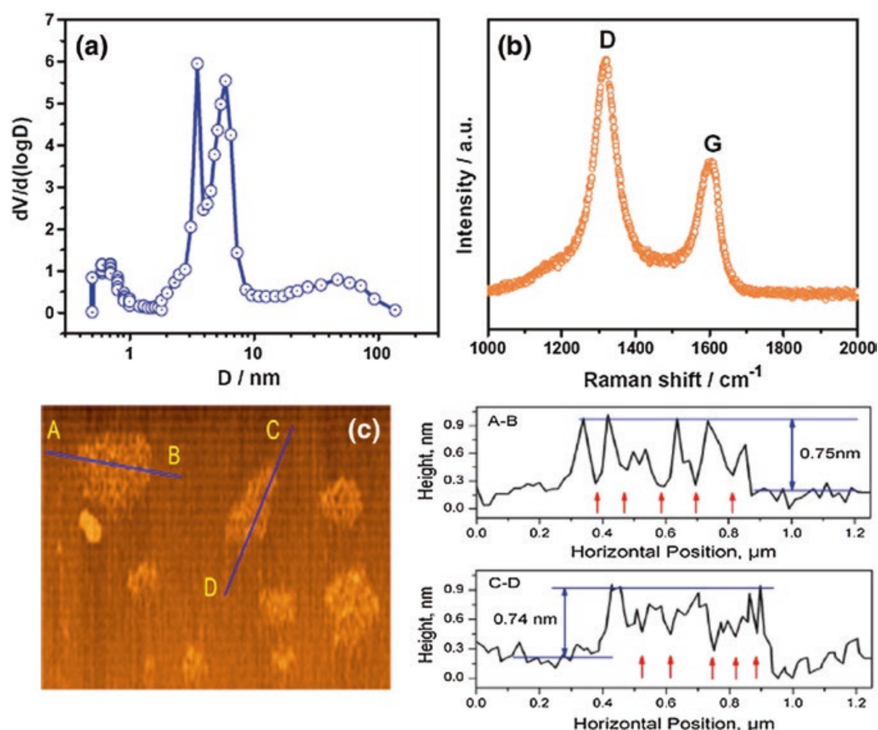


Fig. 10 **a** Pore size distribution and **b** Raman spectra of the porous graphene. **c** AFM image of the porous graphene, the arrows indicate the existence of pores in graphene sheet. Reproduced from Ref. [19] with permission from Elsevier

duplication of sacrificing polystyrene spheres, while the mesopores are generated from the gas release during decomposition of precursors. The preparation process allows the adjustment of the selected components, the amount of graphene added, the thickness of the electrodes. Such ordered porous binder-free electrodes demonstrated superior Li storage properties. For example, graphene-Fe_3O_4@C binder-free electrode depicts high capacities of 1123.8 and 505 mAh g^{-1} at current densities of 0.5 and 10 A g^{-1}, respectively. It shows that the surface Li storage mechanism contributes significantly to the total capacities in such 3D porous binder-free electrodes.

A novel composite, MoS_2-coated 3D graphene network, is synthesized by a facile CVD method [5]. The 3D graphene network serves as a template for the deposition of MoS_2 and provides good electrical contact between the current collector and deposited MoS_2. As proof of concept, the nanocomposite shows excellent electrochemical performance as an anode material for lithium-ion batteries, which exhibits reversible capacities of 877 and 665 mA h g^{-1} during the 50th cycle at current densities of 100 and 500 mA g^{-1}, respectively, indicating its

good cycling performance. Furthermore, the nanocomposite also shows excellent high-current-density performance.

Ultrahigh rate capabilities of transition metal oxide-based electrodes were derived from the design of ordered hierarchically porous 3D electrodes with entrapped active nanoparticle configuration [28]. In contrast to previous reports on hierarchically porous electrodes from irregular self-assembly or post-incorporation of active nanoparticles, the strategy relies on in situ formation and entrapment of active nanoparticles inside the simultaneously formed ordered hierarchically 3D porous carbon, in which the periodic macroporous-mesoporous carbon was directly integrated with the open-porous Ni foam current collector without organic binder, and the electrode active nanoparticles were spatially entrapped inside the periodic porous carbon. Based on the unique electrode configuration, the as-prepared ordered hierarchically porous 3D electrodes show extraordinary rate capabilities.

Zhang et al. developed a simple method for the preparation of metal-oxide coated 3D graphene composites through a facile two-step annealing process [6]. The metal–organic frameworks that served as the precursors of the metal oxides were first synthesized on the 3D graphene networks. The desired nanocomposites were then obtained by a two-step annealing process. The method is expected to be used for synthesis of other metal oxide/graphene composites with 3D structures.

3.2 Porous Graphene-Based Cathode Materials

Porous graphene-based cathode materials were also studied. Yang et al. reported a composite of chemically activated carbon and $LiFePO_4$ as a cathode active material. KOH activation was conducted to construct a 3D structure allowing for diffusion of lithium ions. The porous structure of chemically activated carbon is advantageous to lithium-ion diffusion due to its high rate capability [79]. A composite of chemically activated porous graphene and $LiFePO_4$ was developed to improve the speed of charging–discharging and the cycling stability of lithium-ion batteries using $LiFePO_4$ as a cathode material [23]. Chemically activated porous graphene was synthesized using KOH. Electrochemical properties have also been investigated after assembling coin cells with the porous graphene/$LiFePO_4$ composite as an active material. The composite electrode exhibited better electrochemical properties than the conventional graphene/$LiFePO_4$ composite as well as bare $LiFePO_4$, including exceptional speed of charging–discharging and excellent cycle stability. The porous graphene in the electrode composite provides abundant porous channels for the diffusion of lithium ions. Moreover, it acts as a conducting network for easy charge transfer and as a divider, preventing the aggregation of $LiFePO_4$ particles.

4 Applications in Supercapacitors

Electrochemical supercapacitors store energy using either ion adsorption or fast surface redox reactions [60]. Electrochemical capacitors are also promising energy storage devices due to the advantages of short charging times, a long cycle, and high power density [47, 49, 61]. However, current commercial electrochemical supercapacitors have much lower energy density than lithium-ion batteries. Design of a desirable, low-cost electrode material with a longer cycling lifetime and higher energy density is imperative for electrochemical capacitor. The performance of electrode materials for supercapacitors is dependent on the accessible specific surface area and the pore structure. The control over structure and morphology of carbon electrode materials is therefore an effective strategy to render them high surface area and efficient paths for ion diffusion. Accordingly, porous graphene-based materials have been proved favorable electrode candidates for supercapacitors due to their open-porous structure that allows electrolytes access more easily to the surface of frameworks.

4.1 Activation of Graphene for Supercapacitors

Defective pores could be generated on graphene by chemical methods to prepare activated graphene. Zhu et al. synthesized chemically activated graphene (CA-graphene) with a 3D morphology via KOH activation for application to supercapacitors. The extremely high energy and power density for supercapacitors were possible due to the large surface area from the abundant pore systems [33].

Hierarchical porous carbons are promising electrode materials in high-power supercapacitors. Kim et al. demonstrate the fabrication of highly porous graphene-derived carbons with hierarchical pore structures in which mesopores are integrated into macroporous scaffolds [34]. The macropores were introduced by assembling graphene-based hollow spheres, and the mesopores were derived from the chemical activation with KOH. The unique 3D pore structures in the graphene-based carbons give rise to a BET surface area value of up to 3290 $m^2\ g^{-1}$ and provide an efficient pathway for electrolyte ions to diffuse into the interior surfaces of electrode particles. These carbons exhibit both high gravimetric (174 $F\ g^{-1}$) and volumetric ($\sim$100 $F\ cm^{-3}$) specific capacitance in an ionic liquid electrolyte in acetonitrile. The energy density and power density of the cell assembled with this carbon electrode are also high, with gravimetric values of 74 $Wh\ kg^{-1}$ and 338 $kW\ kg^{-1}$ and volumetric values of 44 $Wh\ L^{-1}$ and 199 $kW\ L^{-1}$, respectively. The high supercapacitor performance achieved with these graphene-based carbons is attributed to their unique pore structure and makes them potentially promising for various energy storage devices.

Zheng et al. synthesized porous graphene/activated carbon composite by hydrothermal carbonization and subsequent two-step chemical activation with KOH

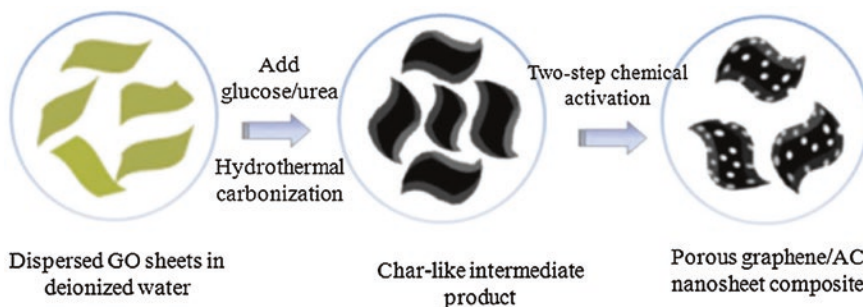


Fig. 11 Schematic illustration showing the experimental steps of preparing porous graphene/AC nanosheet composite. Reproduced from Ref. [89] with permission from Elsevier

(Fig. 11) [89]. The composite has a relatively high packing density and large specific surface area of 2106 $m^2\ g^{-1}$, as well as containing plenty of mesopores (Fig. 12). As supercapacitor electrode material, it exhibits specific capacitance up to 210 F g^{-1} in an aqueous electrolyte and 103 F g^{-1} in organic electrolyte, respectively. The specific capacitance decreases by only 5.3 % after 5000 cycles. In this composite material, a layer of porous activated carbon coats on graphene improves dispersion of graphene sheets and increases its packing density. The graphene integrated into activated carbon matrix also increases conductivity. Additionally, the nanosheet-like electrode material has a short diffusion pathway, which facilitates rapid transport of the electrolyte ions. Three-dimensional graphene-based frameworks are also fabricated by hydrolysis of TEOS with graphene aerogel as support and CTAB as soft template [72]. The resulting hierarchical macro- and mesoporous structures exhibit narrow mesopore size distribution (2–3.5 nm), high surface area, and low mass density. Benefiting from the integration of meso- and macroporous structures, the material manifests outstanding specific capacitance (226 F g^{-1}), high rate capability, and excellent cycling stability when it is applied in electrochemical capacitors.

4.2 3D Graphene-Based Porous Materials for Supercapacitors

Many research works have been published on 3D graphene-based porous nanostructures for supercapacitors. Here, some selected studies on this research field are reviewed.

Zhang et al. present a simple, green, and efficient approach using two standard and simple industry steps to make 3D graphene-based porous materials at bulk scale, with ultrahigh specific surface area (3523 $m^2\ g^{-1}$) and excellent bulk conductivity [87]. The good properties of these materials are demonstrated by their

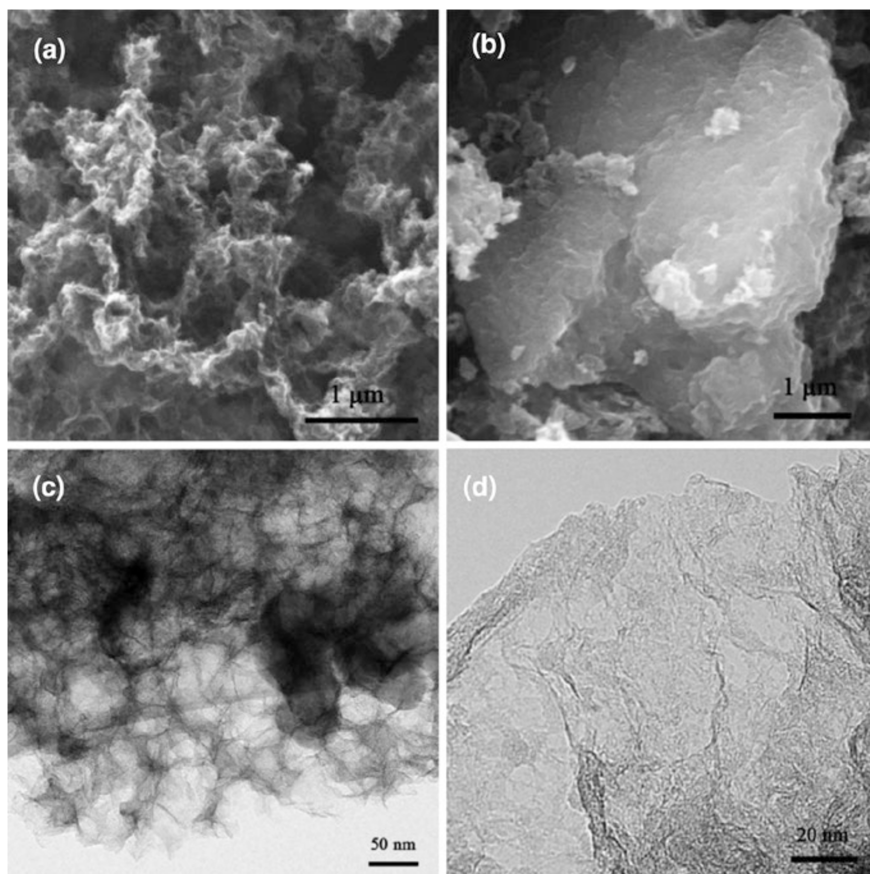


Fig. 12 **a** SEM image of char-like intermediate product, **b** SEM image of graphene/AC nanosheet composite, **c**, **d** TEM images of graphene/AC nanosheet composite. Reproduced from Ref. [89] with permission from Elsevier

superior supercapacitor performance in ionic liquid with specific capacitance and energy density of 231 F g^{-1} and 98 Wh kg^{-1}, respectively.

A self-assembled macrostructured graphene architecture was prepared by a convenient one-step hydrothermal method [74]. The self-assembled graphene hydrogel is electrically conductive, mechanically strong, and thermally stable and exhibits a high specific capacitance. The self-assembled graphene hydrogel as a 3D supercapacitor electrode material exhibits high specific capacitance (175 F g^{-1}) in an aqueous electrolyte.

Freestanding, lightweight, ultrathin, highly conductive, and flexible 3D graphene networks, loaded with MnO_2 by electrodeposition, were prepared as the electrodes of a flexible supercapacitor [25]. The 3D graphene networks showed an ideal supporter for active materials and permitted a large MnO_2 mass loading of

9.8 mg cm^{-2}, leading to a high area capacitance of 1.42 F cm^{-2} at a scan rate of 2 mV s^{-1}. The MnO_2 content with respect to the entire electrode was further optimized and a maximum specific capacitance of 130 F g^{-1} was achieved. The excellent electrochemical performance of a symmetrical supercapacitor consisting of a sandwich structure of two pieces of 3D graphene/MnO_2 composite network separated by a membrane and encapsulated in polyethylene terephthalate membranes was explored.

A hybrid structure of ZnO on 3D graphene foam has been synthesized by CVD growth of graphene followed by a facial in situ precipitation of ZnO nanorods under hydrothermal conditions [16]. The results show that the ZnO nanorods have high crystallinity and cluster uniformly on graphene skeleton to form flower-like nanostructures. It is found that the graphene/ZnO hybrids display superior capacitive performance with high specific capacitance of ~400 F g^{-1} as well as excellent cycle life, making them suitable for high-performance energy storage applications.

3D graphene architectures in the macroworld can in principle maintain all the extraordinary nanoscale properties of individual graphene flakes. However, current 3D graphene products suffer from poor electrical conductivity, low surface area, and insufficient mechanical strength/elasticity; the interconnected self-supported reproducible 3D graphenes remain unavailable. A sugar-blowing approach based on a polymeric predecessor to synthesize a 3D graphene bubble network was reported [69]. The bubble network consists of mono- or few-layered graphitic membranes that are tightly glued, rigidly fixed, and spatially scaffolded by micrometer-scale graphitic struts. Such a topological configuration provides intimate structural interconnectivities, freeway for electron/phonon transports, huge accessible surface area, as well as robust mechanical properties. The graphene network thus overcomes the drawbacks of presently available 3D graphene products and opens up a wide horizon for diverse practical uses, for example, high-power high-energy electrochemical capacitors, as highlighted in this work.

Fan et al. demonstrated the fabrication of functionalized graphene nanosheets via low temperature thermal treatment of graphite oxide with a slow heating rate using $Mg(OH)_2$ nanosheets as template [78]. Because of its dented sheet with high surface area, a certain amount of oxygen-containing groups, and low pore volume, the as-obtained graphene delivers both ultrahigh specific gravimetric and volumetric capacitances of 456 F g^{-1} and 470 F cm^{-3}, almost 3.7 times and 3.3 times higher than hydrazine-reduced graphene, respectively. The assembled supercapacitor exhibits an ultrahigh volumetric energy density of 27.2 Wh L^{-1}, which is among the highest values for carbon materials in aqueous electrolytes, as well as excellent cycling stability with 134 % of its initial capacitance after 10,000 cycles.

Mitlin et al. employed a microwave synthesis process of cobalt phthalocyanine molecules templated by acid-functionalized multiwalled carbon nanotubes to create 3D sponge-like graphene nanoarchitectures suited for ionic liquid-based electrochemical capacitor electrodes that operate at very high scan rates [75]. The 3D nanoarchitectures are able to deliver an energy density of 7.1 Wh kg^{-1} even at an extra high power density of 48 kW kg^{-1}. In addition, the ionic liquid supercapacitor based on this material works very well at room temperature due to its fully

opened structures, which is ideal for the high-power energy application requiring more tolerance to temperature variation. Moreover, the structures are stable in both ionic liquids and 1 M H_2SO_4, retaining 90 and 98 % capacitance after 10,000 cycles, respectively.

Porous yet densely packed carbon electrodes with high ion-accessible surface area and low ion-transport resistance are formed by capillary compression of adaptive graphene gel films in the presence of a nonvolatile liquid electrolyte [81]. This simple soft approach enables subnanometer-scale integration of graphene sheets with electrolytes to form highly compact carbon electrodes with a continuous ion-transport network. Electrochemical capacitors based on the resulting films can obtain volumetric energy densities approaching 60 Wh L^{-1}.

4.3 Flexible Supercapacitors Using 3D Graphene-Based Porous Materials

There has been much research interest in the development of flexible supercapacitors over the past few years due to their high mechanical compliance. Li et al. reported that the combination of graphene chemistry with ice physics can lead to the formation of ultralight and superelastic graphene-based cellular monoliths [55]. Chi et al. reported the preparation of freestanding paper-like electrode materials have trigged significant research interest for their practical application in flexible and lightweight energy storage devices [12]. The utilization of 3D porous graphene scaffold to load nanostructured polyaniline dramatically enhances the electrical conductivity, the specific capacitance, and the cycle stability of the graphene–polyaniline nanocomposite. Shao et al. demonstrated a simple method for preparing high-performance flexible asymmetric supercapacitors based on 3D porous graphene/MnO_2 nanorod and graphene/Ag hybrid thin-film electrodes [59]. These graphene hybrid films, which accelerate ion and electron transport by providing lower ion-transport resistances and shorter diffusion-distances, exhibit high specific capacitances and power performances, and excellent mechanical flexibility. These results suggest that such asymmetric graphene/MnO_2 nanorod and graphene/Ag hybrid thin-film architectures are promising for next-generation high-performance flexible supercapacitors.

5 Future Perspectives

To overcome the limitations of conventional materials, numerous novel nanocomposite materials have been prepared by combining different types of nanomaterials with porous graphene for various electrochemical energy storage devices. The major challenges for the preparation of porous graphene-based nanocomposite materials include the preparation of nanostructures with precisely controlled of

complex and hierarchical pore morphology as well as the designed fabrication of nanocomposite materials with novel compositions.

Porous graphene, as a new platform for nanocomposite materials, provides new possibilities to the nanodevices. Although there has been rapid growth in the development of porous graphene-based nanocomposite materials as promising candidate electrode materials for electrochemical energy storage application in the past few years, as evidenced by the sharp increase in research work published in this area, there remain several challenges needed to be overcome. In order for them to be used in real energy storage devices in the future, several issues including the production cost, scalable synthesis, and long-term mechanical stability of the nanocomposite materials need to be addressed. To study the performance of a porous material for electrochemical energy storage, more reliable parameter, such as the volumetric energy and power density against the whole electrochemical energy devices, should be measured. Nevertheless, we envisioned that numerous well-designed porous graphene-based novel nanocomposite materials with improved energy storage performances will be developed by using new preparation methods and novel compositions. These nanocomposite materials will provide many new chances for efficient electrochemical energy storage devices in the future.

Acknowledgments This work was supported by the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT and Future Planning, Republic of Korea, as a Global Frontier Project.

References

1. Bi H, Yin K, Xie X, Zhou Y, Wan N, Xu F, Banhart F, Sun L, Ruoff RS (2012) Low temperature casting of graphene with high compressive strength. Adv Mater 24:5124
2. Bo X, Guo L (2013) Simple synthesis of macroporous carbon–graphene composites and their use as a support for Pt electrocatalysts. Electrochim Acta 90:283
3. Bolotin KI, Ghahari F, Shulman MD, Stormer HL, Kim P (2009) Observation of the fractional quantum hall effect in graphene. Nature 462:196
4. Cao X, Shi Y, Shi W, Lu G, Huang X, Yan Q, Zhang Q, Zhang H (2011) Preparation of novel 3D graphene networks for supercapacitor applications. Small 7:3163
5. Cao X, Shi Y, Shi W, Rui X, Yan Q, Kong J, Zhang H (2013) Preparation of MoS_2-coated three-dimensional graphene networks for high-performance anode material in lithium-ion batteries. Small 9:3433
6. Cao X, Zheng B, Rui X, Shi W, Yan Q, Zhang H (2014) Metal oxide-coated three-dimensional graphene prepared by the use of metal-organic frameworks as precursors. Angew Chem Int Ed 53:1404
7. Chen W, Li S, Chen C, Yan L (2011) Self-assembly and embedding of nanoparticles by in situ reduced graphene for preparation of a 3D graphene/nanoparticle aerogel. Adv Mater 23:5679
8. Chen Z, Ren W, Gao L, Liu B, Pei S, Cheng HM (2011) Three-dimensional flexible and conductive interconnected graphene networks grown by chemical vapour deposition. Nat Mater 10:424
9. Chen CM, Zhang Q, Huang C-H, Zhao X-C, Zhang B-S, Kong Q-Q, Wang M-Z, Yang Y-G, Cai R, Su DS (2012) Macroporous 'bubble' graphene film via template-directed ordered-assembly for high rate supercapacitors. Chem Comm 48:7149

10. Chen K, Chen L, Chen Y, Bai H, Li L (2012) Three-dimensional porous graphene-based composite materials: electrochemical synthesis and application. J Mater Chem 22:20968
11. Chen S, Duan J, Tang Y, Qiao SZ (2013) Hybrid hydrogels of porous graphene and nickel hydroxide as advanced supercapacitor materials. Chem Eur J 19:7118
12. Chi K, Zhang Z, Xi J, Huang Y, Xiao F, Wang S, Liu Y (2014) Freestanding graphene paper supported three-dimensional porous graphene—polyaniline nanocomposite synthesized by inkjet printing and in flexible all-solid-state supercapacitor. ACS Appl Mater Interfaces 6:16312
13. Choi BG, Yang M, Hong WH, Choi JW, Huh YS (2012) 3D Macroporous graphene frameworks for supercapacitors with high energy and power densities. ACS Nano 6:4020
14. Dai L (2012) Functionalization of graphene for efficient energy conversion and storage. Acc Chem Res 46:31
15. Deng W, Ji X, Gómez-Mingot M, Lu F, Chen Q, Banks CE (2012) Graphene electrochemical supercapacitors: the influence of oxygen functional groups. Chem Commun 48:2770
16. Dong X, Cao Y, Wang J, Chan-Park MB, Wang L, Huang W, Chen P (2012) Hybrid structure of zinc oxide nanorods and three dimensional graphene foam for supercapacitor and electrochemical sensor applications. RSC Adv 2:4364
17. Estevez L, Kelarakis A, Gong Q, Da'as EH, Giannelis EP (2011) Multifunctional graphene/platinum/nafion hybrids via ice templating. J Am Chem Soc 133:6122
18. Fan Z, Zhao Q, Li T, Yan J, Ren Y, Feng J, Wei T (2012) Easy synthesis of porous graphene nanosheets and their use in supercapacitors. Carbon 50:1699
19. Fan Z, Yan J, Ning G, Wei T, Zhi L, Wei F (2013) Porous graphene networks as high performance anode materials for lithium ion batteries. Carbon 60:538
20. Fang Y, Lv Y, Che R, Wu H, Zhang X, Gu D, Zheng G, Zhao D (2013) Two-dimensional mesoporous carbon nanosheets and their derived graphene nanosheets: synthesis and efficient lithium ion storage. J Am Chem Soc 135:1524
21. Geim AK (2009) Graphene: status and prospects. Science 324:1530
22. Geim AK, Novoselov KS (2007) The rise of graphene. Nat Mater 6:183
23. Ha J, Park S-K, Yu S-H, Jin A, Jang B, Bong S, Kim I, Sung Y-E, Piao Y (2013) A chemically activated graphene-encapsulated $LiFePO_4$ composite for high-performance lithium ion batteries. Nanoscale 5:8647
24. Han TH, Huang YK, Tan ATL, Dravid VP, Huang J (2011) Steam etched porous graphene oxide network for chemical sensing. J Am Chem Soc 133:15264
25. He Y, Chen W, Li X, Zhang Z, Fu J, Zhao C, Xie E (2013) Freestanding three-dimensional graphene/MnO_2 composite networks as ultralight and flexible supercapacitor electrodes. ACS Nano 7:174
26. Hsieh C-T, Lin C-Y, Chen Y-F, Lin J-S (2013) Synthesis of ZnO@graphene composites as anode materials for lithium ion batteries. Electrochim Acta 111:359
27. Huang X, Qian K, Yang J, Zhang J, Li L, Yu C, Zhao D (2012) Functional nanoporous graphene foams with controlled pore sizes. Adv Mater 24:4419
28. Huang X, Yu H, Chen J, Lu Z, Yazami R, Hng HH (2014) Ultrahigh rate capabilities of lithium-ion batteries from 3d ordered hierarchically porous electrodes with entrapped active nanoparticles configuration. Adv Mater 26:1296
29. Jang B, Park M, Chae OB, Park S, Kim Y, Oh SM, Piao Y, Hyeon T (2012) Direct synthesis of self-assembled ferrite/carbon hybrid nanosheets for high performance lithium-ion battery anodes. J Am Chem Soc 134:15010
30. Jang B, Choi E, Piao Y (2013) Preparation of well-dispersed Pt nanoparticles on solvothermal graphene and their enhanced electrochemical properties. Mater Res Bull 48:834
31. Jang B, Chae OB, Park S-K, Ha J, Oh SM, Na HB, Piao Y (2013) Solventless synthesis of an iron-oxide/graphene nanocomposite and its application as an anode in high-rate Li-ion batteries. J Mater Chem A 1:15442
32. Kaskhedikar NA, Maier J (2009) Lithium storage in carbon nanostructures. Adv Mater 21:2664

33. Kim KS, Zhao Y, Jang H, Lee SY, Kim JM, Kim KS, Ahn J-H, Kim P, Choi J-Y, Hong BH (2009) Large-scale pattern growth of graphene films for stretchable transparent electrodes. Nature 457:706
34. Kim T, Jung G, Yoo S, Suh KS, Ruoff RS (2013) Activated graphene-based carbons as supercapacitor electrodes with macro- and mesopores. ACS Nano 7:6899
35. Kucinskis G, Bajars G, Kleperis J (2013) Graphene in lithium ion battery cathode materials: a review. J Power Sources 240:66
36. Li D, Kaner RB (2008) Graphene-based materials. Science 320:1170
37. Li X, Wang X, Zhang L, Lee S, Dai H (2008) Chemically derived, ultrasmooth graphene nanoribbon semiconductors. Science 319:1229
38. Li X, Cai W, An J, Kim S, Nah J, Yang D, Piner R, Velamakanni A, Jung I, Tutuc E, Banerjee SK, Colombo L, Ruoff RS (2009) Large-area synthesis of high-quality and uniform graphene films on copper foils. Science 324:1312
39. Li X, Zhao T, Wang K, Yang Y, Wei J, Kang F, Wu D, Zhu H (2011) Directly drawing self-assembled, porous, and monolithic graphene fiber from chemical vapor deposition grown graphene film and its electrochemical properties. Langmuir 27:12164
40. Li L, Guo Z, Du A, Liu H (2012) Rapid microwave-assisted synthesis of Mn_3O_4-graphene nanocomposite and its lithium storage properties. J Mater Chem 22:3600
41. Li L, Seng KH, Chen Z, Liu H, Nevirkovets IP, Guo Z (2013) Synthesis of Mn_3O_4-anchored graphene sheet nanocomposites via a facile, fast microwave hydrothermal method and their supercapacitive behavior. Electrochim Acta 87:801
42. Liang MH, Zhi LJ (2009) Graphene-based electrode materials for rechargeable lithium batteries. J Mater Chem 19:5871
43. Liu C, Yu Z, Neff D, Zhamu A, Jang BZ (2010) Graphene-based supercapacitor with an ultrahigh energy density. Nano Lett 10:4863
44. Luo B, Liu S, Zhi L (2012) Chemical approaches toward graphene-based nanomaterials and their applications in energy-related areas. Small 8:630
45. Luo J, Liu J, Zeng Z, Ng CF, Ma L, Zhang H, Lin J, Shen Z, Fan HJ (2013) Three-dimensional graphene foam supported Fe_3O_4 lithium battery anodes with long cycle life and high rate capability. Nano Lett 13:6136
46. Mai YJ, Wang XL, Xiang JY, Qiao YQ, Zhang D, Gu CD, Tu JP (2011) CuO/graphene composite as anode materials for lithium-ion batteries. Electrochim Acta 56:2306
47. Miller JR, Simon P (2008) Electrochemical capacitors for energy management. Science 321:651
48. Nam I, Kim ND, Kim G-P, Park J, Yi J (2013) One step preparation of Mn_3O_4/graphene composites for use as an anode in Li ion batteries. J Power Sources 244:56–62
49. Naoi K, Naoi W, Aoyagi S, Miyamoto J, Kamino T (1075) New generation “nanohybrid supercapacitor”. Acc Chem Res 2012:46
50. Ning G, Fan Z, Wang G, Gao J, Qian W, Wei F (2011) Gram-scale synthesis of nanomesh graphene with high surface area and its application in supercapacitor electrodes. Chem Comm 47:5976
51. Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA (2004) Electric field effect in atomically thin carbon films. Science 306:666
52. Novoselov KS, Geim AK, Morozov SV, Jiang D, Katsnelson MI, Grigorieva IV, Dubonos SV, Firsov AA (2005) Two-dimensional gas of massless Dirac fermions in graphene. Nature 438:197
53. Park S-K, Yu S-H, Pinna N, Woo S, Jang B, Chung YH, Cho YH, Sung Y-E, Piao Y (2012) A facile hydrazine-assisted hydrothermal method for the deposition of monodisperse SnO_2 nanoparticles onto graphene for lithium ion batteries. J Mater Chem 22:2520
54. Park S-K, Jin A, Yu S-H, Ha J, Jang B, Bong S, Woo S, Sung Y-E, Piao Y (2014) In situ hydrothermal synthesis of Mn_3O_4 nanoparticles onnitrogen-doped graphene as high-performance anode materials for lithium ion batteries. Electrochim Acta 120:452
55. Qiu L, Liu JZ, Chang SLY, Wu Y, Li D (2012) Biomimetic superelastic graphene-based cellular monoliths. Nat Commun 3:1241

56. Qiu H, Dong X, Sana B, Peng T, Paramelle D, Chen P, Lim S, Appl ACS (2013) Mater Inter 5:782
57. Romanos J, Beckner M, Rash T, Firlej L, Kuchta B, Yu P, Suppes G, Wexler C, Pfeifer P (2012) Nanospace engineering of KOH activated carbon. Nanotechnology 23:015401
58. Segal M (2009) Selling graphene by the ton. Nat Nanotechnol 4:612
59. Shao Y, Wang H, Zhang Q, Li Y (2013) High-performance flexible asymmetric supercapacitors based on 3D porous graphene/MnO_2 nanorod and graphene/Ag hybrid thin-film electrodes. J Mater Chem C 1:1245
60. Simon P, Gogotsi Y (2008) Materials for electrochemical capacitors. Nat Mater 7:845
61. Simon P, Gogotsi Y, Dunn B (2014) Where do batteries end and supercapacitors begin. Science 343:1210
62. Song Z, Zhang Y, Liu W, Zhang S, Liu G, Chen H, Qiu J (2013) Hydrothermal synthesis and electrochemical performance of Co_3O_4/reduced graphene oxide nanosheet composites for supercapacitors. Electrochim Acta 112:120
63. Stankovich S, Dikin DA, Dommett GHB, Kohlhaas KM, Zimney EJ, Stach EA, Piner RD, Nguyen ST, Ruoff RS (2006) Graphene-based composite materials. Nature 442:282
64. Stoller MD, Park S, Zhu Y, An J, Ruoff RS (2008) Graphene-based ultracapacitors. Nano Lett 8:3498
65. Su F-Y, He Y-B, Li B, Chen X-C, You C-H, Wei W, Lv W, Yang Q-H, Kang F (2012) Could graphene construct an effective conducting network in a high-power lithium ion battery. Nano Energy 1:429
66. Vickery JL, Patil AJ, Mann S (2009) Fabrication of graphene-polymer nanocomposites with higher-order three-dimensional architectures. Adv Mater 21:2180
67. Wang G, Jia L-T, Zhu Y, Hou B, Li D-B, Sun Y-H (2012) Novel preparation of nitrogen-doped graphene in various forms with aqueous ammonia under mild conditions. RSC Adv 2:11249
68. Wang Z-L, Xu D, Wang H-G, Wu Z, Zhang X-B (2013) In situ fabrication of porous graphene electrodes for high-performance energy storage. ACS Nano 7:2422
69. Wang X, Zhang Y, Zhi C, Wang X, Tang D, Xu Y, Weng Q, Jiang X, Mitome M, Golberg D, Bando Y (2013) Three-dimensional strutted graphene grown by substrate-free sugar blowing for high-power-density supercapacitors. Nat Commun 4:2905
70. Wei W, Yang S, Zhou H, Lieberwirth I, Feng X, Müllen K (2013) 3D Graphene foams cross-linked with pre-encapsulated Fe_3O_4 nanospheres for enhanced lithium storage. Adv Mater 25:2909
71. Wu Z-S, Ren W, Wang D-W, Li F, Liu B, Cheng H-M (2010) High-energy MnO_2 nanowire/graphene and graphene asymmetric electrochemical capacitors. ACS Nano 4:5835
72. Wu ZS, Sun Y, Tan YZ, Yang S, Feng X, Müllen K (2012) Three-dimensional graphene-based macro- and mesoporous frameworks for high-performance electrochemical capacitive energy storage. J Am Chem Soc 134:19532
73. Xin X, Zhou X, Wang F, Yao X, Xu X, Zhu Y, Liu Z (2012) A 3D porous architecture of Si/graphene nanocomposite as high-performance anode materials for Li-ion batteries. J Mater Chem 22:7724
74. Xu Y, Sheng K, Li C, Shi G (2010) Self-assembled graphene hydrogel via a one-step hydrothermal process. ACS Nano 4:4324
75. Xu Z, Li Z, Holt CMB, Tan X, Wang H, Amirkhiz BS, Stephenson T, Mitlin D (2012) electrochemical supercapacitor electrodes from sponge-like graphene nanoarchitectures with ultrahigh power density. J Phys Chem Lett 3:2928
76. Xu C, Xu B, Gu Y, Xiong Z, Sun J, Zhao XS (2013) Graphene-based electrodes for electrochemical energy Storage. Energy Environ Sci 6:1388
77. Yadav P, Banerjee A, Unni S, Jog J, Kurungot S, Ogale S (2012) A 3D hexaporous carbon assembled from single-layer graphene as high performance supercapacitor. Chemsuschem 5:2159
78. Yan J, Wang Q, Wei T, Jiang L, Zhang M, Jing X, Fan Z (2014) Template-assisted low temperature synthesis of functionalized graphene for ultrahigh volumetric performance supercapacitors. ACS Nano 8:4720

79. Yang M, Gao Q (2011) $LiFePO_4$/C composite cathode material with a continuous porous carbon network for high power lithium-ion battery. J Alloys Compd 509:3690
80. Yang SB, Feng XL, Wang L, Tang K, Maier J, Müllen K (2010) Graphene-based nanosheets with a sandwich structure. Angew Chem Int Ed 49:4795
81. Yang X, Cheng C, Wang Y, Qiu L, Li D (2013) Liquid-mediated dense integration of graphene materials for compact capacitive energy storage. Science 341:534
82. Ye J, Zhang J, Wang F, Su Q, Du G (2013) One-pot synthesis of Fe_2O_3/graphene and its lithium-storage performance. Electrochim Acta 113:212–217
83. Yoo E, Kim J, Hosono E, Zhou H, Kudo T, Honma I (2008) Large reversible Li storage of graphene nanosheet families for use in rechargeable lithium ion batteries. Nano Lett 8:2277
84. Yu D, Wei L, Jiang W, Wang H, Sun B, Zhang Q, Goh K, Si R, Chen Y (2013) Nitrogen doped holey graphene as an efficient metal-free multifunctional electrochemical catalyst for hydrazine oxidation and oxygen reduction. Nanoscale 5:3457
85. Yu S-H, Conte DE, Baek S, Lee D-C, Park S-K, Lee KJ, Piao Y, Sung Y-E, Pinna N (2013) Structure-properties relationship in iron oxide-reduced graphene oxide nanostructures for Li-ion batteries. Adv Funct Mater 23:4293
86. Zhang LL, Zhao X, Stoller MD, Zhu Y, Ji H, Murali S, Wu Y, Perales S, Clevenger B, Ruoff RS (1806) Highly conductive and porous activated reduced graphene oxide films for high-power supercapacitors. Nano Lett 2012:12
87. Zhang L, Zhang F, Yang X, Long G, Wu Y, Zhang T, Leng K, Huang Y, Ma Y, Yu A, Chen Y (2013) Porous 3D graphene-based bulk materials with exceptional high surface area and excellent conductivity for supercapacitors. Sci Rep 3:1408
88. Zhao X, Hayner CM, Kung MC, Kung HH (2011) Flexible holey graphene paper electrodes with enhanced rate capability for energy storage applications. ACS Nano 5:8739
89. Zheng C, Zhou X, Cao H, Wang G, Liu Z (2014) Synthesis of porous graphene/activated carbon composite with high packing density and large specific surface area for supercapacitor electrode material. J Power Sources 258:290
90. Zhu Y, Murali S, Stoller MD, Ganesh KJ, Cai W, Ferreira PJ, Pirkle A, Wallace RM, Cychosz KA, Thommes M, Su D, Stach EA, Ruoff RS (2011) Carbon-based supercapacitors produced by activation of graphene. Science 332:1537
91. Zhu J, Yang D, Rui X, Sim D, Yu H, Hng HH, Hoster HE, Ajayan PM, Yan Q (2013) Facile preparation of ordered porous graphene-metal oxide@C binder-free electrodes with high Li storage performance. Small 9:3390
92. Zhu J, Yang D, Yin Z, Yan Q, Zhang H (2014) Graphene and graphene-based materials for energy storage applications. Small 10:3480

Graphene and Two-Dimensional Transition Metal Dichalcogenide Materials for Energy-Related Applications

Gyeong Sook Bang and Sung-Yool Choi

Abstract The energy problem is one of the most challenging issues in the twenty-first century. Many energy applications for portable electronics, electric vehicles, spacecraft, and renewable energy are under extensive investigation worldwide. New alternative energy with renewable energy devices are competitive with fossil fuels. To develop the advanced energy storage and harvesting/conversion system, renewable energy nanomaterials are in high demand. Two-dimensional nanomaterials composed of graphene and two-dimensional transition-metal chalcogenides (2D-TMDs) have attracted a great deal of interest due to their unique properties. From the prospect of energy applications, graphene and 2D-TMD nanosheets have many interesting properties, such as large surface area, atomically thin sheet with high flexibility, and a wide range of electrical conductivity. Graphene has proved to be a good material for nanoscale devices used in energy harvesting/conversion and storage applications. Recently, 2D-TMDs are also attracting significant attention in many energy-related applications. In this chapter, we focus on the recent advances in graphene (including graphene oxide, GO) and 2D-TMD nanosheets research for energy devices: electrodes in solar cell, electrocatalysts or photocatalysts for fuel cell, electrodes in Li-ion battery, and electrodes for supercapacitors.

Keywords Graphene · Transition-metal dichalcogenides (TMDs) · Nanosheets · Solar cell · Fuel cell · Li-ion battery · Supercapacitor

G.S. Bang · S.-Y. Choi (✉)
School of Electrical Engineering, Graphene Research Center,
Korea Advanced Institute of Science and Technology, Daejeon, South Korea
e-mail: sungyool.choi@kaist.ac.kr

© Springer Science+Business Media Dordrecht 2016

C.-M. Kyung (ed.), *Nano Devices and Circuit Techniques for Low-Energy Applications and Energy Harvesting*, KAIST Research Series,
DOI 10.1007/978-94-017-9990-4_9

1 Introduction

The rapid development of better energy-related devices has generated some of the most important challenges in the twenty-first century. Extensive studies around the world have been focused on energy products such as portable electronics, electric vehicles, spacecraft, and renewable energy. Low cost, high efficiency, sustainable-energy devices are in high demand. As new alternative energy devices have become more competitive with fossil fuels, renewable energy nanomaterials for energy harvesting and energy storage are in high demand.

Two-dimensional (2D) nanomaterials composed of graphene and nanostructured transition-metal dichalcogenides (TMDs) have attracted a great deal of interest due to their unique properties. These make them good candidates for potential applications in electronic devices such as supercapacitors, light-emitting diodes, solar cells, fuel cells, piezoelectric nano-generators, and lithium-ion batteries.

When it comes to energy-related applications, graphene and 2D-TMD single-layers have many advantages: (1) atomic thickness for quantum confinement of charge carriers, (2) electrical properties that vary from those of metal to semiconductor to insulator, (3) good cycle stability and flexibility, (4) high catalytic activity, and (5) short path-lengths for electron transport. Moreover, 3D structures incorporating graphene and 2D-TMDs are more attractive electrode materials in applications related to renewable energy. In particular, their properties related to the versatile electrical conductivity and their huge surface area are useful in enhancing the performance of energy devices.

Since its discovery in 2004 [1] graphene has demonstrated its role in enabling rapid advances in energy technologies for high-performance energy devices. Moreover, 2D-graphene has proven a good material for nanoscale devices used in energy harvesting, energy conversion, and storage applications. Recently, layered TMD materials (e.g., MoS_2, WS_2, $MoSe_2$, and WSe_2) are attracting significant attention in many energy-related applications. Their crystal structure varies from hexagonal (MoS_2, WS_2) to ortho-rhombic ($MoTe_2$, WTe_2). Nanosheets of 2D-TMDs possess remarkably different electronic properties than those of bulk TMDs, which can exhibit either semiconductor or metallic behavior depending on their elemental composition and method of synthesis. The optical properties of 2D-TMD nanosheets make them potentially useful for energy harvesting and conversion materials in solar cells, photoelectrochemical cells, and photo-fuel cells. Their graphene-like layer can also be used for energy storage devices such as Li-ion batteries and supercapacitors.

Here, we focus on the recent advances on graphene (including graphene oxide, GO) and 2D-TMD nanosheets for use in energy devices: specifically, electrodes in solar cells, electrocatalysts and photocatalysts for fuel cell, electrodes in Li-ion batteries, and electrodes for supercapacitors.

2 Electronic Structure and Synthesis of 2D Materials

2.1 Electronic Structure

Graphene This substance can be made to assume the form of single-atom-thick, 2D single-layer of sp^2 hybridized carbon atoms in a honeycomb crystal lattice. In this form, it has unique structure and properties. These include rich electronic states (excellent conductivity of 10^6 S cm^{-1} and thermal conductivity of 5000 W m^{-1} K^{-1}), good mechanical properties (Young's modulus of 1.0 TPa), large surface area (2630 m^2 g^{-1}), optical absorbance of ~2.3 % for visible light, high transparency, high flexibility, and easy modification using organic and inorganic molecules. Detailed descriptions of the overall properties of graphene have been published elsewhere [2–7].

Because it possesses these superior properties, graphene is still being studied intensely as an attractive candidate for many energy-related applications [8–21].

2D-TMDs Atomically thin 2D-TMDs (6.5 Å thick) have received much attention in recent years, because they are naturally abundant, have unique properties and form semiconductors with various bandgaps [6, 22]. Layered TMDs are materials with the formula MX_2, where M is a transition metal from Group-4 to Group-7 and X is a chalcogen of S, Se, or Te. Their structures are formed of covalently bonded X-M-X single-layers that interact by van der Waals forces (Fig. 1a). Each single layer consists of two X atom layers and a layer of metal atom sandwiched between two layers of chalcogens. Transition metal atoms provide four electrons to fill the bonding states of MX_2, oxidation states of +4 for metal atoms, and oxidation states of −2 for chalcogen atoms. The d-electron count of the transition metal relates to the MX_2 phase. The metal atoms of MX_2 can have either

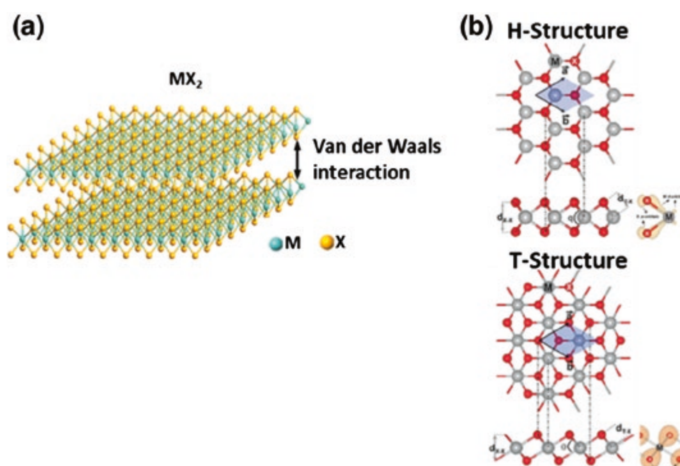


Fig. 1 **a** Layered MX_2 structure. **b** Atomic structure for MX_2 in the H and T structure. (Reproduced with permission from Ataca et al. 2012 [24]. Copyright © 2012 American Chemical Society)

trigonal prismatic or octahedral coordination. Bulk MX_2 exhibits three different stacking-types: 1T (trigonal), 2H (hexagonal), and 3R (rhombohedral) [6, 23]. Group-6 bulk MX_2 is mostly of 2H structure, which is more stable than the 1T phase. Single-layer MX_2 is of only two types: 1H and 1T (Fig. 1b) [24]. Each M atom has six X atoms, and each X atom has three M atoms forming hybridization of the M d-orbital and X p-orbital. The change in symmetry induces changes in the electronic properties from semiconducting to metallic [25, 26]. Thus, MX_2 nanosheets have electronic properties from semiconducting to metallic, depending on their geometry, composition, thickness, and electron density. The phase transition from 2H to 1T can be induced by specific procedures.

The bandgap of MX_2 varies with a number of layers. Calculations using density functional theory (DFT) have shown that the MoX_2 and WX_2 materials can change from indirect-to-direct bandgaps, depending on the number of layers (Fig. 2) [6, 27–30]. Single-layer MX_2 is a direct bandgap semiconductor with bandgap energy of 1.2−2.1 eV, whereas bulk MX_2 is an indirect bandgap semiconductor with bandgap energy of 1.0−1.4 eV. The electronic properties of MX_2 are summarized in Table 1.

Contrary to zero-bandgap graphene, single-layer 2D MX_2 sheets are semiconductors with a small direct-bandgap, which is a useful material for a wide range of applications (e.g., field effect transistors, energy harvesting/conversion devices, optoelectronics, and sensors). Transistors of single-layer MoS_2 using HfO_2, exhibited a mobility of 200 cm^2 V^{-1} s^{-1} [32]. Other properties shown are Young's

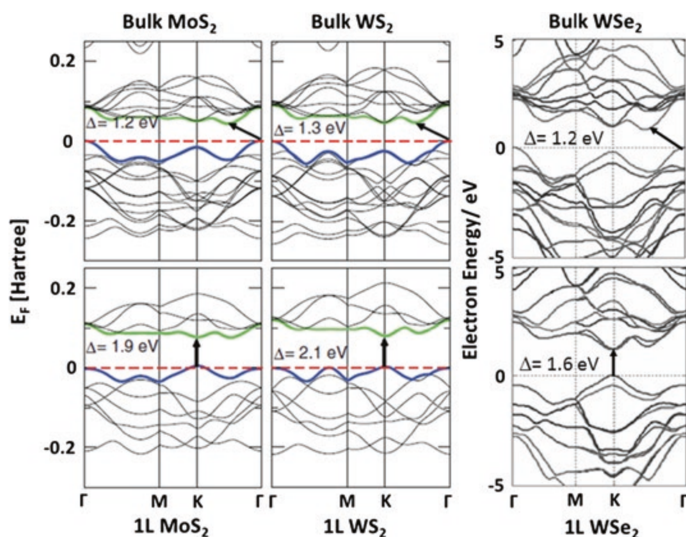


Fig. 2 Electronic band structures of bulk and single-layer MX_2. The *black arrows* are their band gaps. (Reproduced with permission from: (MoS_2, WS_2) Kuc et al. 2011 [28]. Copyright © 2011 American Physical Society. (WSe_2) Huang et al. 2014 [30]. Copyright © 2014 Royal Society of Chemistry)

Table 1 Electronic properties of MX_2 materials [6, 28, 30, 31]

Metal	X, chalcogen		
	S	Se	Te
Mo	Semiconductor	Semiconductor	Semiconductor
	Bulk: 1.2 eV	Bulk: 1.1 eV	Bulk: 1.0 eV
	1 L: 1.8–1.9 eV	1 L: 1.5–1.6 eV	1 L: 1.2 eV
W	Semiconductor	Semiconductor	Semiconductor
	Bulk: 1.4 eV	Bulk: 1.2 eV	Bulk:
	1 L: 1.9–2.1 eV	1 L: 1.6–1.7 eV	1 L: 1.2 eV

modulus of 270 ± 100 GPa for MoS_2 single-layer [33] and in-plane resistivity of 3–40 kΩ/sq [34] and 14–28 kΩ/sq [35] for micromechanical exfoliated MoS_2 and CVD-grown MoS_2, respectively. The direct-bandgap transition in single-layer MX_2 provide enhanced photoluminescence (PL) by quantum confinement effects. Moreover, MoS_2 single-layer exhibits optical-valley polarization [36, 37] and lesser stiffness than graphene [38]. The electronic band structure of 2D-MX_2 not only depends on the number of layers but also on the strain [27].

In summary, both graphene and 2D-MX_2 nanosheets have a great number of superior properties that make them attractive candidates for use in energy harvesting, energy conversion, and storage applications.

2.2 Synthesis Methods

Graphene and 2D-MX_2 can be obtained from a variety of methods including micromechanical exfoliation (also called scotch-tape method), chemical vapor deposition (CVD), and liquid-based exfoliation. Each preparation method results in products with particular structures and properties. Graphene and 2D-MX_2 processed using liquid-based-exfoliation and CVD methods have often been used in energy harvesting, conversion, and storage applications due to their large surface area and ease of mass production. [7, 39–48].

2.2.1 Micromechanical Exfoliation

Graphene become an important material after discovery through micromechanical exfoliation method by Novoselov et al. (2004) [1]. Micromechanical exfoliation is the best known method for obtaining high quality graphene. This method produces graphene from bulk crystals of layered graphite by repeated exfoliation using scotch tape [1, 49] (Fig. 3). Like graphene, a 2D-MX_2 single-layer can be exfoliated from bulk MX_2 crystal using scotch tape [50]. Other layered materials can also be exfoliated into single-layer flakes by micromechanical exfoliation [51, 52]. This method does not control flake thickness and size; however, and is

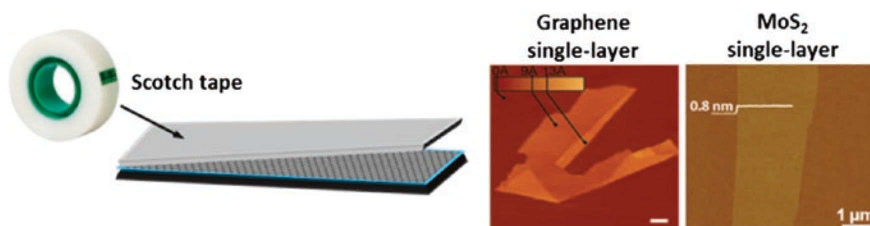


Fig. 3 Micromechanical exfoliation. AFM images of mechanically exfoliated single-layer graphene and MoS_2. (Reproduced with permission from: (graphene) Novoselov et al. 2005 [49]. Copyright © 2005 American Chemical Society). (MoS_2) (Reproduced with permission from Li et al. 2012 [50]. Copyright © 2012 John Wiley and Sons)

not appropriate for mass production, or for production of large 2D-MX_2 single-layers. For energy harvesting/conversion and storage devices, 2D nanosheet-based devices require nanosheets of large surface area for solar cells, and large amounts of nanosheets for energy storage applications. Thus, this mechanical exfoliation method is used primarily for fundamental studies of new physics and new devices [32, 33, 40, 53–61].

2.2.2 Chemical Vapor Deposition

Among methods for graphene synthesis, CVD is widely used to produce uniform single-layer of large area [44, 62]. This is also the most appropriate method to synthesize 2D-MX_2 single-layer with a vertically hybrid structure. This method also makes it possible to introduce dopants to 2D-MX_2 to control bandgap [63].

A typical, low-pressure CVD system consists of a furnace with a quartz tube, for heat-vaporization of chalcogenides, and an outflux cold trap. The CVD synthesis of MoS_2 has involved precursors of various phases, such as MoO_3 and S powder vaporized and co-deposited onto substrate [64–66], sulfurization of pre-deposited Mo precursors (e.g., MoO_2 film [67, 68], MoO_3 film [68], Mo [35], aromatic molecules [69]), and decomposition of ammonium thiomolybdate ($(NH_4)_2MoS_4$) film under S gas [70].

Regarding the first case, CVD methods have been reported for creating MoS_2 single-layer on SiO_2/Si substrates by Lee et al. [64]. They reported that large-area MoS_2 sheets with 1–3 layers were directly synthesized on SiO_2/Si substrate by heating MoO_3 and S powder (Fig. 4). The synthesized MoS_2 showed high crystallinity in a six-fold hexagonal lattice and n-type semiconductor properties. Zhan et al. demonstrated that large-area MoS_2 could be synthesized by sulfurization of pre-deposited Mo film by e-beam evaporation [35]. The resulting film showed polycrystalline MoS_2 due to suppressed Mo migration because the melting point of Mo (2610 °C) was higher than the growth temperature (~750 °C). Another similar method was reported by Wang et al. [67]. They demonstrated formation of a MoS_2 bilayer 1.5 nm thick using layer-by-layer sulfurization of MoO_2

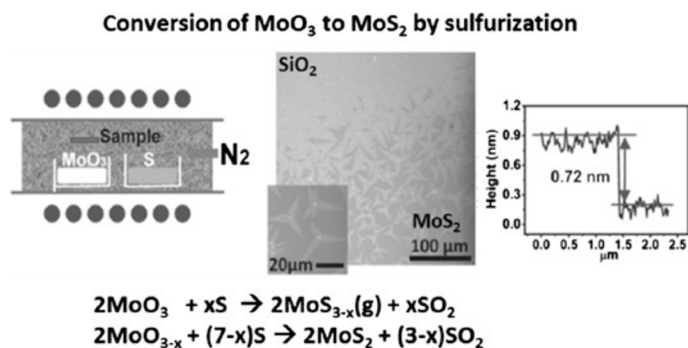


Fig. 4 Schematic illustration for CVD (*left*) and the optical image of MoS_2 on SiO_2/Si (*center*). The thickness of single-layer MoS_2 is 0.72 nm, which is measured by AFM (*right*). (Reproduced with permission from Lee et al. 2012 [64]. Copyright © 2012 John Wiley and Sons)

microplates at 850–950 °C. Using the weak adhesion between a MoS_2 layer and MoO_2 precursor film, MoS_2 flakes with domain size of 10 μm were obtained by separation from the MoS_2/MoO_2 film. MoS_2 synthesis using substrate dip-coated in $(NH_4)_2MoS_4$ solution and exposed to S gas was reported by Liu et al. [70]. These workers produced large-area MoS_2 film with uniform thickness in 2–3 layers. Sulfurization using pre-deposited metal precursor films is an effective method for preparation of large-area MX_2 single-layer.

Recently, Najmaei et al. reported forming triangular MoS_2 single-layer with large grain size and an edge-length of 10 μm using MoO_3 film on Si substrate [68] (Fig. 5a). Lee et al. demonstrated triangular MoS_2 sheets on various substrates using atmospheric pressure CVD (APCVD) with aromatic-molecule seeds (perylene-3,4,9,10-tetracarboxylic acid tetrapotassium salt, PTAS) [69] (Fig. 4b). In addition, there have been reports of demonstrations in which insulating single-crystals, such as sapphire, quartz, and mica, were used to form high quality 2D-MX_2 [71–76].

Formation of large-area MoS_2 single-layer was reported by van der Zande et al. [66]. They used solid MoO_3 and S precursors to produce large MoS_2 single-layer by APCVD growth. The resulting sheets of triangular MoS_2 had very long edge-length (~120 μm). Another large MoS_2 layer was grown on mica substrate, resulting in a continuous MoS_2 film of high uniformity, by coalescence of aligned MoS_2 domains [75].

Besides MoS_2, recently, our group reported CVD growth of large-area $MoSe_2$ single-layer with high quality and uniformity, on SiO_2/Si and sapphire substrate [73] (Fig. 6). A $MoSe_2$-graphene hetero-structure created by the CVD method was also demonstrated. It exhibited PL quenching by fast transfer of photogenerated charge carriers in the stacked heterostructure. Cong et al. reported triangular WS_2 single-layer with long edge-length (178 μm) formed in a modified quartz-tube furnace [77]. In this case, WO_3 and S powder were loaded into a small, inner quartz tube sealed on one side. By direct sulfurization of the WO_3, formation of

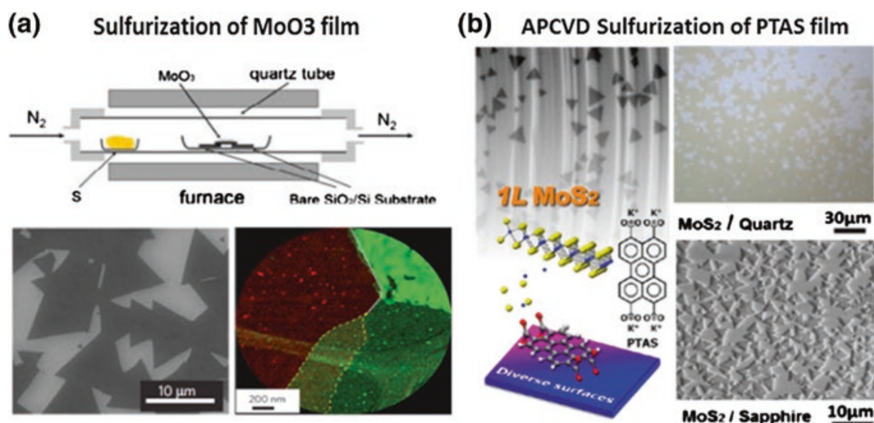


Fig. 5 **a** Schematic illustration for CVD (*upper*), SEM image (*bottom left*) and False-color dark-field TEM image of MoS_2 (*bottom right*) (Reproduced with permission from Najmaei et al. 2013 [68]. Copyright © 2013 Nature Publishing Group). **b** Chemical structure of PTAS (*bottom right*), schematic illustration of MoS_2 growth (*bottom left*) and SEM images of single-layer MoS_2 (*upper left*). Optical images of MoS_2 film on quartz (*upper right*) and sapphire (*bottom right*). (Reproduced with permission from Lee et al. 2013 [69]. Copyright © 2013 American Chemical Society)

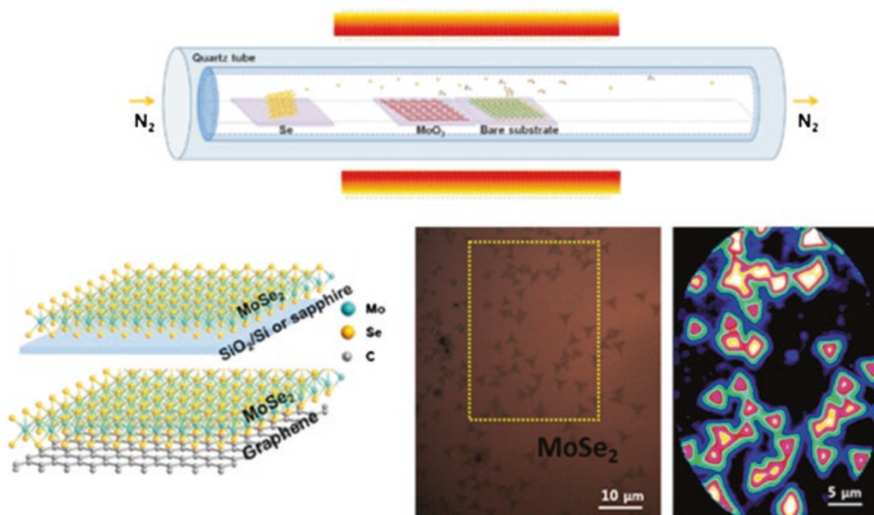


Fig. 6 Schematic illustration of CVD system (*upper*). The structure of $MoSe_2$ on SiO_2/Si (or sapphire) and graphene/SiO_2/Si substrate, respectively (*bottom left*). Optical image of $MoSe_2$ single-layer and PL intensity map for the yellow box shown in the left optical image (*bottom right*). (Reproduced with permission from Shim et al. 2014 [73]. Copyright © 2014 American Chemical Society)

large-area, WS_2 single-layer of high quality was demonstrated. First, triangular WS_2 sheets formed and then a few layer sheets grew at the apexes of the triangular sheets. Formation of large-area WSe_2 monolayer was also demonstrated using gas-phase reaction of WO_3 and Se powder in hydrogen [72].

The CVD method can produce MX_2 single-layer with large domain and high quality. In particular, well-aligned MX_2 domains may produce wafer-scale, single crystals by domain coalescence during CVD growth.

2.2.3 Liquid-Based Exfoliation

Graphene and 2D-MX_2 can also be obtained by the liquid-based exfoliation method. This method is the most economic for mass production. Therefore, it has been considered a useful approach for obtaining large quantities of 2D nanosheets from bulk powder. Liquid-based exfoliation is a suitable method for applications that require large amounts, such as electrochemical energy storage, flexible energy devices, composite materials, inkjet printing, thin films, and spray coating.

Graphene Nanosheets

Liquid-based exfoliation involves sonication of graphite, intercalated graphite, or graphite oxide powder in suitable solvents [43, 78–80].

Graphite oxide Liquid-based exfoliation of graphite oxide is widely used for graphene production [41, 43, 81–89]. This method involves a two-step process: intercalation with strong oxidizing agents and expansion of graphite layers by sonication. These chemical processes introduce functional groups (i.e., epoxide, –OH, –COOH, and –COH groups) in the basal plane and at the edges [80] (Fig. 7). Thus, GO by oxidative exfoliation has a defective structure and insulating properties. The reduction of GO by thermal [89, 90], electrochemical [86, 88, 91], or chemical treatment [87, 90] can transform the insulating GO to conductive graphene, which is then referred to as reduced GO (rGO).

Direct exfoliation of graphite Direct liquid exfoliation of natural graphite can produce single- or few-layer graphene by ultrasonication in organic solvents (e.g., N-methyl-2-pyrrolidone (NMP), N,N-dimethylformamide (DMF), and benzyl benzoate) with suitable surface energy (40–50 mJ m^{-2}) [92]. This method gives the highest yield in NMP and high quality graphene with low defects. Although this type of exfoliation can be used directly to obtain conductive graphene sheets with low defects, this method has some disadvantages (e.g., very low yield, toxic solvents, and reaggregation).

Intercalated graphite Intercalated graphite formed by insertion of intercalant species between graphite layers can be used to expand the graphite-interlayer distance by rapid increase in the vapor pressure of inserted materials, during microwave or thermal treatment. This form of exfoliation provides a high-yield method

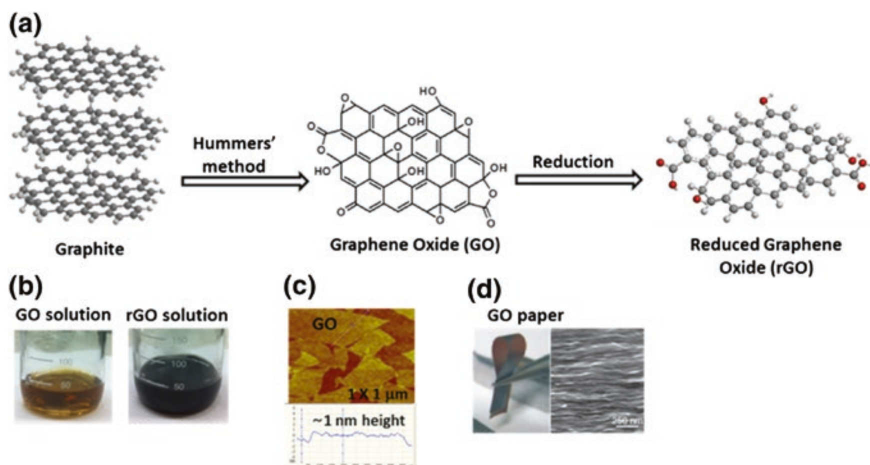


Fig. 7 **a** Scheme for GO synthesis and the structure of GO and rGO (*gray ball*: carbon, *red ball*: oxygen). **b** Optical image of GO and rGO solution, and **c** AFM image of GO sheets on SiO_2/Si with 1 nm height. **d** Optical image of GO paper (*left*) and SEM image (*right*) of cross-section of the GO paper. (Reproduced with permission from Dikin et al. 2007 [85]. Copyright © 2007 Nature Publishing Group)

of producing graphene with high quality and large lateral size. However, many methods for creating intercalated-graphite involve oxidation in ambient air and dangerous chemical reactions.

2D-TMD Nanosheets

With a layered structure similar to graphene, MX_2 can also be made to produce single-layer or few-layer MX_2 by liquid-based exfoliation. There are two pertinent types of liquid-based exfoliation: direct exfoliation by sonication in appropriate solvents and chemical exfoliation of Li-intercalated MX_2 in water.

Liquid-based direct exfoliation Direct exfoliation involves ultrasonication of bulk MX_2 powder in organic solvents and in aqueous surfactant solutions [93–97]. The efficiency of direct exfoliation depends on solvent, surfactant, and sonication time. Direct liquid exfoliation using organic solvents can produce single- or few-layer 2D-MX_2 through ultrasonication. Nanosheets of 2D-MX_2 also showed the highest yield in NMP as graphene. Coleman et al. demonstrated this form of direct exfoliation and reported surface energy of 65–75 mJ m^{-2} for MoS_2, WS_2, and $MoSe_2$ [94, 96]. NMP has surface energy similar to MX_2. In the case of direct exfoliation using surfactants, the surfactants prevent the restacking of MX_2 nanosheets. However, these direct exfoliations exhibited very low yield (2–10 %), which also required lengthy sonication time to improve dispersion. Our group, has recently reported direct exfoliation of MoS_2 with alkali-metal-hydroxide

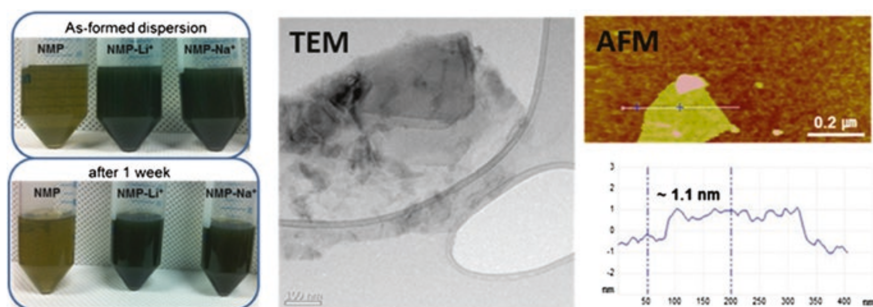


Fig. 8 Optical image of MoS_2 dispersion with alkali-metal assistant in DMF (*left*). TEM image and AFM image of the exfoliated MoS_2 single-layer. (Reproduced with permission from Bang et al. 2014 [93]. Copyright © 2014 American Chemical Society)

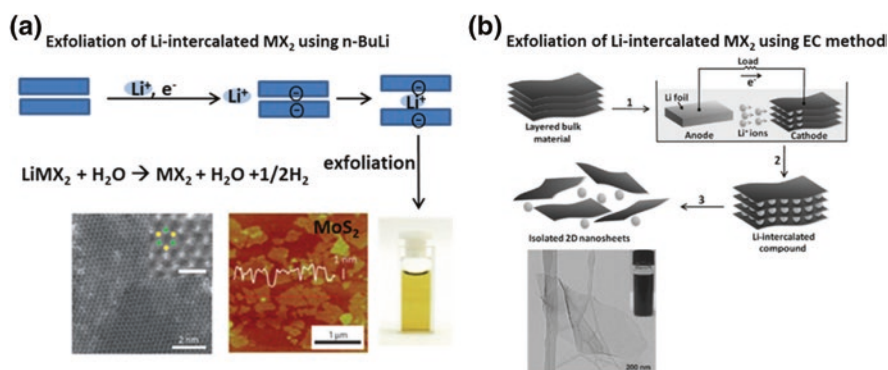


Fig. 9 **a** Schematic illustration for Li intercalation into MX_2 layers (*upper*) and optical image of the exfoliated MoS_2 dispersion in water. TEM image of MoS_2 single-layer and AFM image the exfoliated MoS_2 sheets (*bottom*). (Reproduced with permission from Eda et al. 2011 [26]. Copyright © 2011 American Chemical Society). **b** Schematic illustration for electrochemical lithiation (2) and exfoliation (3) of the Li-intercalated MX_2 (*upper*). TEM image of MoS_2 nanosheet. Inset: optical image of the exfoliated MoS_2 dispersion (*bottom*). (Reproduced with permission from Zeng et al. 2011 [98]. Copyright © 2011 John Wiley and Sons)

assistance to improve the direct-exfoliation efficiency [93] (Fig. 8). Our method exhibited high yield (65 %) and good dispersibility.

Liquid-based chemical exfoliation Liquid-based chemical exfoliation of Li-intercalated MX_2 in water gives a high yield of single-layer MX_2 (nearly 100 %). Bulk MX_2 is intercalated with Li using n-butyllithium (n-BuLi) [26] (Fig. 9a) or electrochemical lithiation [98, 99] (Fig. 9b), forming Li_xMX_2. Electrochemical lithiation can be controlled by adjusting the amount of Li ions inserted [99]. The Li_xMX_2 materials can easily be exfoliated using brief ultrasonication in water. The resulting 2D-MX_2 flakes showed single-layer thickness of 1–1.2 nm and metallic 1T phase. The initial 2H phase was induced to form the metallic 1T phase by

charge transfer from Li to MX_2. The induced 1T-phase can recover to the initial 2H phase upon annealing at temperatures of 200–300 °C [26]. This method is one of the most effective methods for mass production of metallic 2D-MX_2 single-layer.

3 Energy Harvesting and Energy Conversion Applications

3.1 Solar Cells

Solar energy is abundant and sustainable and has attracted interest for many years. Solar cells (dye-sensitized, polymer, organic, hetero-junction, and quantum-dot sensitized) are promising devices for conversion of solar energy into electricity. Graphene has been extensively investigated in relation to solar cell applications because of its high optical transparency, good mechanical flexibility, high thermal stability, and high electrical conductivity. For these reasons, it is now a candidate to replace indium tin oxide (ITO) in photovoltaics and opto-electronics. Graphene-analogs, such as 2D-MX_2 materials, have also potential as photovoltaic and opto-electronic materials. There have been reports of their use for transparent anodes or cathodes, catalytic counter electrodes, and the active layer for energy harvesting and conversion. In particular, heterostructures, such as those formed by combining 2D-MX_2 and graphene, have recently attracted attention as a new structure for light-harvesting applications. Heterostructures can induce significant photon absorption and photocurrent.

3.1.1 Graphene Nanosheets for Solar Cells

Graphene and graphene-based materials are attractive for solar energy harvesting and conversion. They have been investigated extensively for solar cell applications. As mentioned in Sect. 2.2, graphene has unique properties such as high carrier mobility, high optical transparency, very large surface area, high chemical and thermal stability, and high mechanical flexibility. It can be used in diverse ways as transparent anodes and cathodes, electron and hole-transport materials, catalytic counter electrodes, and active layers in solar cells.

Graphene as photo-anode In dye-sensitized solar cells (DSSCs), graphene can be used as photo-anode. Transparent indium tin oxide (ITO) and fluorine-doped tin oxide (FTO) are conventional electrodes used in DSSCs. However, indium is a rare-earth metal, mechanically brittle, and unstable under high temperature processing. Graphene, because of its high electrical conduction and optical transparency, may replace ITO in solar cells. Coleman et al. analyzed the transmittance and sheet resistance of graphene-based transparent electrodes from published data [100] (Fig. 10a). They mentioned that the conductivity of graphene could be increased by increasing the concentration of carriers. Bae et al. reported creation

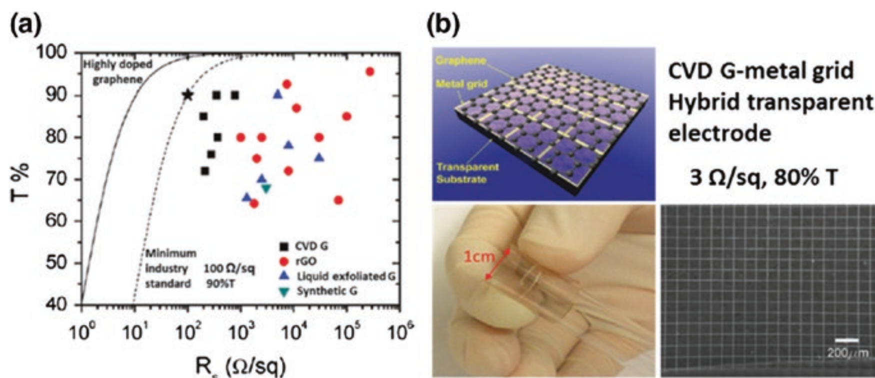


Fig. 10 **a** Plot for transmittance and sheet resistant from data in the published papers. (Reproduced with permission from De and Coleman 2010 [100]. Copyright © 2010 American Chemical Society). **b** Schematic illustration of hybrid film of CVD-graphene and metal grid as transparent electrode (*upper*). Optical image of the bendable transparent electrode (*bottom left*) and SEM image (*bottom right*) of CVD graphene-metal grid on PET. (Reproduced with permission from Zhu et al. 2011 [102]. Copyright © 2011 American Chemical Society)

of large-area CVD-graphene film (4-layered) with HNO_3 doping by the roll-to-roll method [101]. This acid-doped graphene film exhibited low sheet resistance (30 Ω/sq) with 90 % transmittance, which is superior to ITO electrodes. In addition to these, hybrid films of CVD-graphene and metal grid showed the best performance (3 Ω/sq with 80 % transmittance), in the form of bendable, transparent electrodes [102] (Fig. 10b).

Wang et al. reported rGO film used as a transparent conducting electrode in DSSCs [103]. The rGO film exhibited conductivity of 550 S cm^{-1} with 70 % transmittance (1000–3000 nm). However, DSSCs using rGO film (device structure: rGO/TiO_2/dye/spiro-OMeTAD/Au) showed very low power-conversion efficiency (PCE) of 0.26 %. Li et al. reported the use of GO film as the hole-transport layer in organic solar cells (OSCs) [104]. The OSC using GO film (device structure: ITO/GO(2 nm)/P3HT:PCBM/Al) had PCE of 3.5 ± 0.3 %. DeArco et al. reported flexible CVD-graphene films on polyethylene terephthalate (PET) film [105]. The CVD-graphene on transparent PET film showed sheet resistance of 230 Ω/sq with 72 % transparency. Organic solar cells using this CVD-graphene as flexible transparent anode exhibited 1.27 % PCE and good performance under bending of up to 138° (Fig. 11). For use in high performance organic solar cells, Hsu et al. reported a CVD graphene-tetracyanoquinodimethane (TCNQ) multilayer anode with the resistance of 182 Ω/sq and 88 % transmittance [106]. The OSC using graphene-TCNQ (device structure: Graphene-TCNQ/PEDOT:PSS/P3HT:PCBM/Ca/Al) showed 2.58 % PCE.

CVD-graphene and rGO can be used as diverse roles in solar cells. CVD-graphene is more attractive as transparent conducting electrode due to their high electrical conductivity and relative inertness compared to rGO.

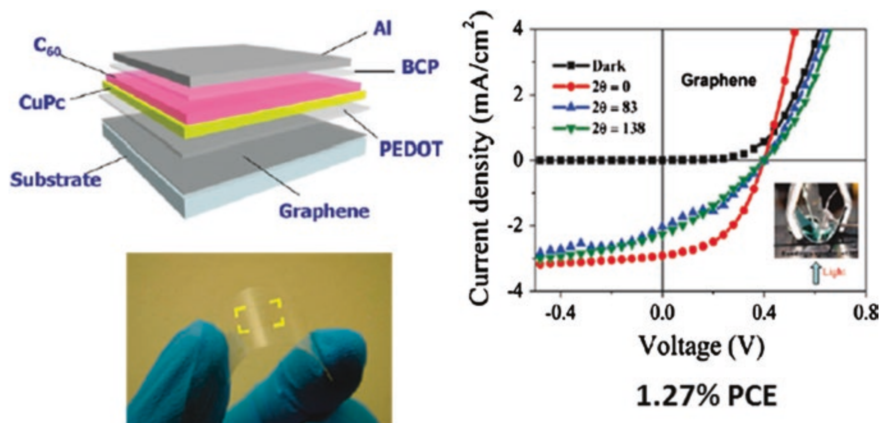


Fig. 11 Schematic illustration of the organic solar cell with CVD graphene as anodic electrode (*upper*). Current-voltage characteristic curves for the photovoltaic device under A.M. 1.5 illumination at 100 mW cm^{-2} for different bending angles (*bottom*). (Reproduced with permission from De Arco et al. 2010 [105]. Copyright © 2010 American Chemical Society)

Graphene as catalytic counter electrodes Graphene is also a promising material for use as catalytic counter electrodes (CE) in DSSCs due to its low cost, high surface area, high conductivity, and electrocatalytic properties. Many works have reported using graphene and graphene composites as catalytic CEs in Pt-free DSSCs. Recently, Gong et al. demonstrated that rGO embedded in a polypyrrol (ppy) matrix, is a good candidate conductive and catalytic CE in DSSCs [107]. Certain DSSCs (device structure FTO/TiO_2/dye/I_3, I^- mediated electrolyte/rGO-ppy) exhibited a high PCE of 8.14 %, which is comparable to a Pt counter electrode (8.34 % PCE). This rGO-ppy electrode can be considered a promising CE for Pt-free DSSCs.

Graphene Quantum Dots (GQDs) GQDs are graphene nanoparticles with lateral size less than 100 nm. They have excellent properties that include high optical absorptivity [108–110]. Yan et al. first reported GQD-sensitized DSSC [111]. Zhang et al. demonstrated that graphene synthesized by a hydrothermal method could act as a photosensitizer [112]. Recently, Williams et al. suggested that photoexcited GQDs inject electrons into TiO_2 within 15 fs [113].

Metallic graphene with a semiconductor can form Schottky-junction layer in DSSCs. Miao et al. reported an enhanced Schottky-junction solar cell formed using chemical-doped graphene/n-Si [114]. The graphene had been doped with bis(trifluoromethanesulfonyl)-amide (TFSA). The TFSA-doped graphene showed low sheet resistance and increased work function. This solar cell exhibited a high PCE (8.6 %). Song et al. reported the rGO-TiO_2 Schottky-junction in a DSSC (device structure: FTO/graphene-TiO_2/dye/I_3, I^- mediated electrolyte/Pt). DSSC using rGO-TiO_2 showed an improved PCE (6.06 %) compared to pure TiO_2 without rGO [115]. Liu et al. reported GO (work function of 4.6–4.8 eV) used for hole

transport and Cs-doped GO (work function of 3.9–4.1 eV) for electron transport in polymer solar cells [116].

Thus, graphene and graphene derivatives could be widely used to improve the performance of solar cells. These materials have high electrical conductivity and optical transparency, and are promising candidates to replace ITO in solar cells.

3.1.2 Nanosheets of 2D-TMDs for Solar Cells

Nanosheets of 2D-MX_2 have recently been receiving a lot of interest due to their role as semiconductors with a small direct bandgap (1–2 eV) and n or p-type carriers depending on their layer thickness and elemental compositions [22]. They can absorb significant portions of the solar spectrum, and may be considered efficient nanomaterials for use in solar energy conversion. These 2D-MX_2 sheets have higher sunlight absorption (5–10 % in visible range at 6.5 Å thickness) than the commonly used absorbers, Si and GaAs [117]. Moreover, the 2D-MX_2 single-layer is flexible, transparent, and ultrathin. Thus, 2D-MX_2 single-layer can be used as an efficient photovoltaic material.

2D-TMDs for cathode catalysts In DSSC, platinum (Pt) is widely used as a cathode material due to its excellent electrocatalytic properties. However, it has several disadvantages in DSSC applications (i.e., high price and limited reserves). To replace expensive Pt electrodes, 2D-MX_2-based materials have been studied for use as cathode catalysts in DSSC. When DSSC was chemically synthesized using MoS_2 and WS_2 it exhibited power conversion efficiency (PCE) of 7.59 and 7.73 %, respectively [118]. Recently, Patil et al. prepared MoS_2 from $MoCl_5$ and thioacetamide at low temperature (70 °C) and under wet conditions [119]. This DSSC exhibited 7.01 % PCE, which is comparable to DSSC with a Pt counter electrode (7.31 %). Chen et al. reported the few-layer $MoSe_2$ fabricated by surface selenization of Mo-coated glass in a CVD system [120]. The few-layer $MoSe_2$ DSSC showed a higher PEC (9 %) than a counter electrode based on Pt nanoparticles on FTO glass. MoS_2 and graphene composite as a counter electrode in DSSC showed a PEC of about 6 % [121, 122]. The 2D-MX_2-based CE exhibited good performance, and are very useful for replacement of Pt in DSSCs.

2D-TMDs for polymer solar cells 2D-MX_2-based materials have been studied for applications in polymer solar cells. Yu et al. reported a MoS_2-Au Schottky-junction solar cell with 1.8 % PCE [123]. In the same group, a MoS_2-TiO_2 composite structure with a P3HT active layer (device structure: ITO/TiO_2/MoS_2/P3HT/Au) exhibited 1.3 % PCE under 100 mW cm^{-2} illumination [124]. Recently, Yun et al. reported creation of a polymer solar cell using p- and n-doped MoS_2 film for modulation of the work function of the interfacial layer. A P-doped MoS_2 hole-transport layer showed enhanced performance (3.4 % PCE) in a polymer solar cell [125]. Gu et al. prepared a 2D-MoS_2 single-layer (AFM image in Fig. 13) using liquid-based exfoliation from Li-intercalated MoS_2 [126]. A device using liquid-exfoliated 2D-MoS_2 nanosheets as a hole-extraction layer exhibited 4.03 % PCE for the P3HT:PC_{61}BM active layer and 8.11 % PCE for

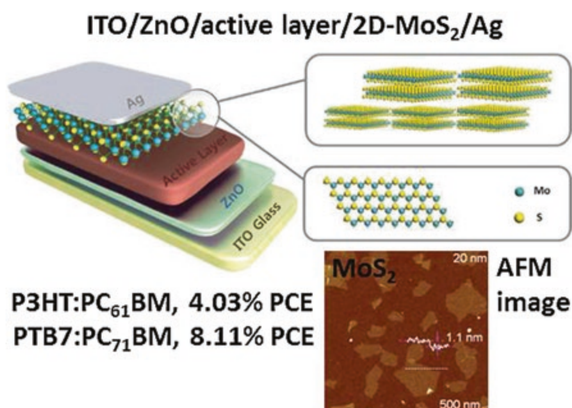


Fig. 12 Schematic illustration of inverted-type OSC with MoS_2 thin layer (*upper*) and AFM image of MoS_2 nanosheets on APTES modified SiO_2/Si (*bottom*). (Reproduced with permission from Gu et al. 2013 [126]. Copyright © 2013 John Wiley and Sons)

the PTB7:PC_{71}BM active layer, respectively (Fig. 12). Niu et al. reported single and few-layer 2D-TMDs nanosheets obtained by salt-assisted liquid exfoliation [127]. Some OSCs using the 2D-MoS_2 nanosheets as hole-transport layers and P3HT:PC_{61}BM as the photoactive layer (device structure: ITO/2D-MoS_2/P3HT:PC_{61}BM/Al) showed 1.81 % PCE under AM 1.5 illumination. Recent OSCs using MoO_3/MoS_2 exhibited a PCE of 6.9 % and high air-stability of 5.5 % PCE after 16 days [128]. These results also demonstrate that 2D-MoS_2-based nanosheets are a promising hole-transport material for high performance solar cells. To improve light harvesting in OSCs, plasmonic OSCs using MoS_2-Au NP composite (MoS_2@Au) as the hole-transport layer exhibited enhanced short-circuit photocurrent density, and a PCE of 7.25 % [129].

Heterostructure of 2D-TMDs Van der Waals heterostructures of semiconducting MX_2 have recently attracted attention as a new structure for light-harvesting applications [117, 130–135]. Britnell et al. prepared vertical heterostructures of MX_2-graphene using 2D-MX_2 as good photoactive materials and graphene as a good transparent electrode [135]. They demonstrated effective photovoltaic devices with photosensitivity above 0.1 A W^{-1} using vertical MX_2-graphene heterostructures. From DFT calculation, Bernardi et al. predicted the performance of heterojunction solar cells with a Schottky barrier (MoS_2-graphene) and bilayer excitonic MX_2 (MoS_2-WS_2) [117]. They estimated 0.4–1.5 % PCE for a bilayer of MoS_2-WS_2 1.2 nm thick, and 0.1–1.0 % PCE for a bilayer of MoS_2-graphene 0.9 nm thick. Recently, Lopez-Sanchez et al. reported a diode based on a p–n heterojunction of single-layer MoS_2 and p-type silicon [132]. This heterojunction diode of MoS_2-Si operates as a photovoltaic device, which converts incident light into electrical power with an external quantum efficiency (EQE) of 4.4 %. Lee et al. fabricated at sandwiched p–n heterojunction of n-MoS_2 and p-WSe_2 single-layer between graphene layers, and measured the external quantum efficiency [134] (Fig. 13). The reported EQE of the p-WSe_2/n-MoS_2 heterojunctions of different thicknesses were 2.4, 12, and 34 % for single layer, bilayer, and multilayer, respectively. In addition, Hong et al. reported the experimental observation of ultrafast hole-transfer within 50 fs in the stacked heterostructure of MoS_2

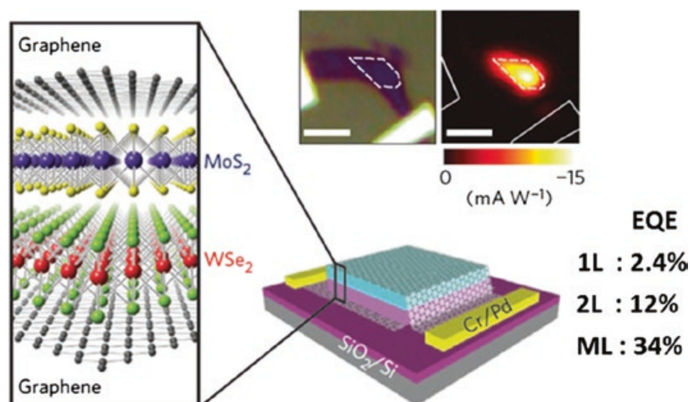


Fig. 13 Schematic illustration of Graphene/n-MoS_2/p-WSe_2/Graphene. (Reproduced with permission from Lee et al. 2014 [134]. Copyright © 2014 Nature Publishing Group)

and WS_2 [133]. Heterostructures of semiconducting 2D-MX_2 can enhance light–matter interaction, which induces major photon absorption and production of photocurrent.

Semiconductors of 2D-MX_2 are materials that separate electrons and holes for energy conversion. The physical properties of heterostructures using them can be controlled to induce fast charge-separation. Therefore, 2D MX_2-based devices are promising for light-harvesting and conversion.

3.2 Fuel Cells

Fuel cells are clean, sustainable energy conversion devices that convert chemical energy of a fuel directly into electricity. These technologies are approaching commercialization in small portable power sources [136]. The key parts of a fuel cell are electrodes for fuel oxidation (anode) and oxygen reduction (cathode) (Fig. 14). Expensive noble metals are well-known active catalysts for both anodes and

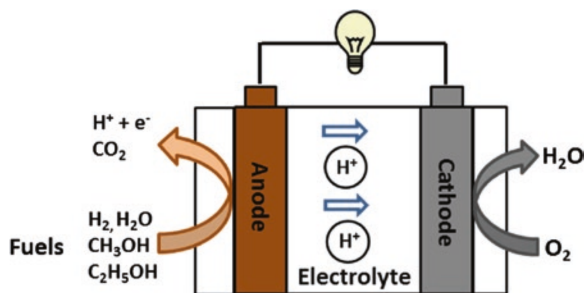


Fig. 14 The structure of fuel cell

cathodes in fuel cells. Among them, platinum-nanoparticles are regarded as the best catalyst, and they have the highest electrocatalytic activity for oxygen reduction at the cathode. However, Pt has several disadvantages in practical applications of fuel cells. These include its high price, limited (global) reserves, and deactivation by self-poisoning from CO adsorption. Therefore, it is essential to develop low-cost non-platinum catalysts. The activities of catalysts depend on their surface properties. Good catalysts for high performance and commercialization require low cost, high surface area, maximum contact area, good electronic conductivity, and high electrochemical stability in acidic and alkaline electrolytes. Graphene has received a great deal of attention as a good catalyst due to its high surface area, good chemical and environmental stability, and unique structures. Graphene-analogous 2D-MX_2 has also begun receiving considerable interest recently.

3.2.1 Graphene Nanosheets for Fuel Cells

Oxygen reduction reaction (ORR) at the cathode The oxygen reduction reaction determines the overall performance of a fuel cell. Pt nanoparticles are the best catalysts and are still used in practical applications due to their high catalytic activity and good stability. Graphene-based Pt nanoparticles have been also studied for ORR [13, 137–142]. However, Pt has the disadvantage of reserves that are too limited, and prices that are too high for the commercialization of fuel cells. Therefore, development of Pt-free catalysts is important for commercialization of fuel cells. Many researchers have addressed the electrocatalytic activity of metal-free graphene-based nanomaterials. These graphene-based nanomaterials exhibited enhanced catalytic activity and durability when used for ORR in fuel cells. Graphene appears to be a promising candidate as catalyst support than commercial Pt-C due to high surface area, high electrical conductivity, 2D nanosheets with atomic layer, basal-plane structure of sp^2-hybridized carbon, and good stability [13, 140–149].

As metal-free alternatives, metal-free graphene-based nanomaterials are important for commercialization of fuel cells. As a result, metal-free graphene-based nanomaterials have been extensively developed for ORR [13, 140–142, 150–161]. Heteroatom-doped graphene is one of the metal-free nanocatalysts. Heteroatom doping can induce charge redistribution in graphene and create active sites for oxygen adsorption. Heteroatom-doped graphene induces intramolecular charge transfer between graphene and dopants, and exhibits high catalytic activity and stability. Qu et al. reported creating N-doped graphene by growth of CVD graphene with ammonia [140]. The resulting N-doped graphene has shown higher catalytic activity and stability than commercial Pt/C. Yang et al. reported an S-doped rGO with good catalytic activity [141]. Li et al. synthesized N-doped graphene quantum dots (N-GQDs) with oxygen rich functional groups [142]. Zhang et al. reported amine-functionalized rGO with good performance for ORR [13]. These heteroatom-doped graphenes provided advantages as ORR catalysts. Although it is still not exactly clear about the active sites, theoretical and

experimental studies suggest that heteroatom-based graphene or its composites may provide metal-free catalysts with high activity and low cost.

Fuel oxidation at the anode Graphene-based noble metals have been attractive as anode material for use in direct methanol fuel cells (DMFCs). Methanol is a fuel appropriate for fuel cells, and the methanol oxidation process includes methanol adsorption and subsequent dissociation into adsorbed intermediates [162]. As mentioned earlier, CO-poisoning can greatly reduce the catalytic activity of Pt NPs [163]. In the case of the rGO-Pt NP catalysts, the oxygen-containing functional group of rGO or graphene can improve the electrocatalytic activity by removing the adsorbed CO from Pt sites [164]. Actually, the rGO-Pt nanoparticles exhibited higher catalytic performance than commercial Pt/C catalyst for the methanol oxidation reaction [143–148, 163–165]. Li et al. reported high catalytic activity for methanol oxidation using rGO-Pt nanoparticle (rGO-PtNPs) electrodes [145]. The graphene-based alloy metal NPs (e.g., PtRu-graphene, PtNi-graphene, and PtPd-rGO, PtFe-graphene) also displayed high electrocatalytic activities for methanol oxidation. Zhao et al. used one metal-free graphene-based nanomaterial (rGO-PPy-PdNPs) as anode [164]. In addition to methanol, the catalytic activity of graphene-based metal NPs catalysts has also been studied for use in oxidation reaction of ethanol [166, 167], formic acid [167–169], and hydrogen [163].

Graphene-based nanomaterials can be used as catalyst at both anodes and cathodes in fuel cells. Further study is necessary to determine the exact catalytic mechanisms needed to develop graphene-based catalysts with high activity and durability.

3.2.2 Nanosheets of 2D-TMDs for Fuel Cells

Hydrogen is a clean fuel that leaves only water behind when used, and is one of the promising new energy sources. To produce hydrogen in a fuel cell effectively, an electrocatalyst is required. As already known, Pt is the best known electrocatalyst for the hydrogen evolution reaction (HER), though it has the disadvantages mentioned previously. Layered MX_2 products are made from abundant materials, and 2D-MX_2 nanosheets have received increasing attention as HER catalysts for fuel cells [170–172].

2D-TMDs for HER catalysts As seen in Sect. 2.1.2 (CVD), 2D-MX_2 nanosheets exhibit the shape of triangle or truncated triangles. Their structure has two different edge sites: X and M [173] (Fig. 15a). The elements Co and Ni incorporated into MoS_2 induce morphology change resulting in truncated triangles with prominently exposed S edge [174]. The ratio of the length of the basal plane to that of the edge sites changes with the size of the nanosheet. These edge sites are related to HER activity. In density-functional calculations, the free energy for hydrogen adsorption on the MoS_2 edge was shown to be similar to the activity of biological catalysts [170] (Fig. 15b).

Jaramillo et al. [172] experimentally demonstrated that MoS_2 edges are catalytically active sites for HER. Since then, Kibsgaard et al. synthesized double-gyroid

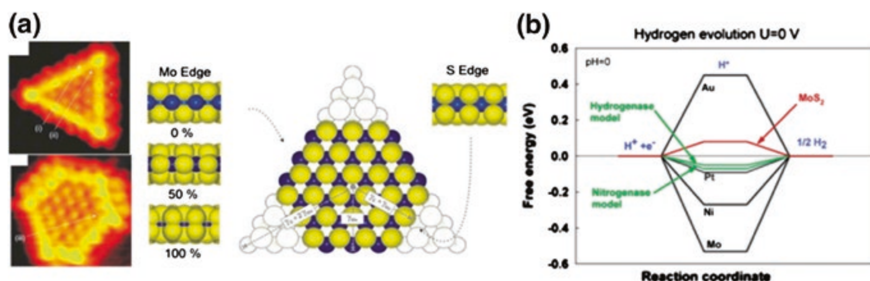


Fig. 15 **a** STM images of MoS_2 single-layer (*left*) and white lines for the scan orientation. *Top view* of atomic ball model of the truncated MoS_2 (*right*, Mo atoms: *blue*, S atoms: *yellow*). (Reproduced with permission from Lauritsen et al. [173]. Copyright © 2004 Elsevier). **b** Free energy diagram for hydrogen evolution by density functional calculations. (Reproduced with permission from Hinnemann et al. 2005 [170]. Copyright © 2005 American Chemical Society)

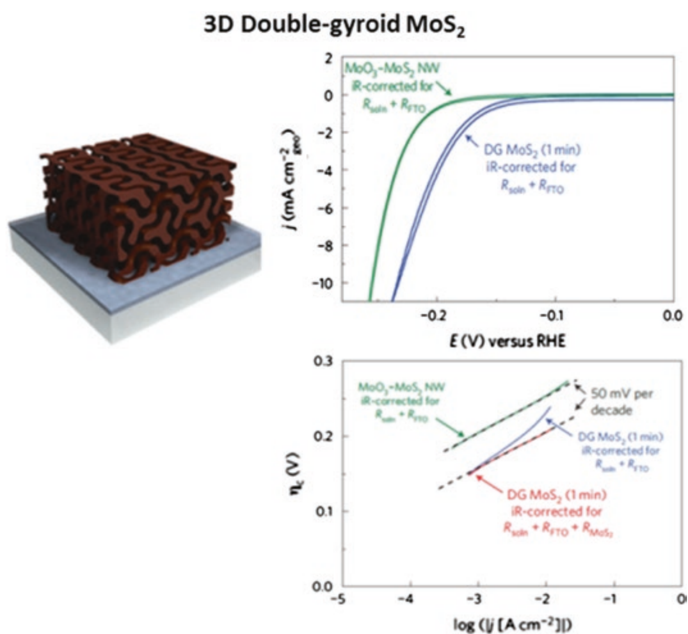


Fig. 16 The mesoporous structure of double-gyroid MoS_2 (*left*). CVs of the double-gyroid MoS_2 electrode at 5 mV s^{-1} (*right upper*) and tafel plot (*right bottom*). (Reproduced with permission from Kibsgaard et al. 2012 [171]. Copyright © 2012 Nature Publishing Group)

MoS_2 to largely expose the edge sites [171] (Fig. 16). This double-gyroid MoS_2 exhibited high activity, with a Tafel slope of 50 mV $decade^{-1}$. An amorphous MoS_2 with many defects also exhibited catalytically activity, with an average Tafel slope of 50 mV $decade^{-1}$ [175–177].

As mentioned in Sect. 2.1 (Electronic structure), the symmetry of single-layer 2D-MX_2 has trigonal prismatic (2H phase) or octahedral (1T phase) coordination for transition-metal atoms [22]. Materials of 2D-MX_2 with 2H phase are mainly semiconducting, whereas 1T phase is metallic. Moreover, 2D-MX_2 nanosheets chemically exfoliated with Li intercalation exhibit metallic properties after phase change from 2H to 1T. An MX_2 single-layer in 1T phase was much more active for HER than in 2H phase [99, 178, 179]. The higher activity can be attributed to the presence of the metallic 1T phase. In particular, strained single-layer 1T-WS_2 produced by chemical exfoliation with Li intercalation exhibited excellent catalytic activity for HER [179]. Yi et al. demonstrated the correlation between MoS_2 properties and HER activity by electrochemical tuning of Li intercalation [99]. They also reported that 1T-MoS_2 have high catalytic activity for HER. In other work, MX_2-graphene hybrid types were used as catalysts for HER [180, 181]. Li et al. reported use of a MoS_2-rGO hybrid to improve the conducting network and edge sites, and this hybrid exhibited excellent electrocatalytic activity with high current, low over-potential of −0.1 V and a low Tafel slope of 41 mV decade^{-1} [181]. Liao et al. synthesized MoS_2 NPs on mesoporous graphene sheet (MoS_2 NPs-MGF) with high surface area and conductive skeleton, and this product showed high electro-catalytic activity with rapid electron transfer and a low Tafel slope of 42 mV decade^{-1} [180]. A WS_2-rGO hybrid exhibited electrocatalytic activity with a Tafel slope of 58 mV decade^{-1}. There are also studies in which noble-metal NPs (Pt and Au)-MoS_2 nanosheets [182, 183] were used for HER and Pd-MoS_2 [184] for methanol oxidation. The hybrid materials showed higher HER activity than pure 2D-MoS_2 or WS_2. High HER performance was shown using 2D-MX_2 or 2D MX_2-G hybrid types (amorphous sheets, defective nanosheets, porous structure, and metallic 1T phase) with highly active sites. A summary of MX_2-based HER catalysts is provided in Table 2.

Table 2 MX_2-based HER catalysts

Material	Tafel slope (mV/decade)	References
MoS_2 nanoparticles	55–60	[172]
Double-gyroid MoS_2	50	[171]
Amorphous MoS_2	40	[177]
Defect-rich MoS_2	50	[176]
1T-MoS_2 nanosheets	40–44	[99, 185, 186]
1T-WS_2 nanosheets	60	[179]
MoS_2-rGO	41	[181]
MoS_2 on MGF	42.8	[180]
WS_2-rGO	58	[187]
Pt on single-layer MoS_2	40	[183]
Pt-2H MoS_2	110	[185]
Pt-1T MoS_2	43	

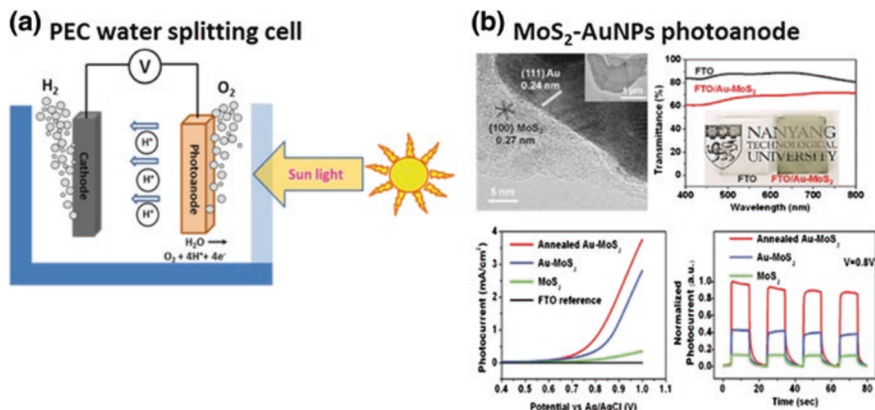


Fig. 17 **a** The structure of PEC water splitting cell. **b** TEM image of Au NPs on MoS_2 (*upper left*) and transmittance versus wavelength for MoS_2-Au NPs on FTO (*upper right*). LSVs for PEC with MoS_2-Au NPs photoanode in 0.1 M KH_2PO_4 under visible light of 350 mW cm^{-2} (*bottom left*) and normalized amperometric I-t cycles at 0.8 V for PEC on MoS_2-Au NPs photoanode (*bottom right*). (Reproduced with permission from Yin et al. 2014 [188]. Copyright © 2014 John Wiley and Sons)

2D-TMDs for water-splitting catalysts Semiconductor-based photocatalysts can be used for hydrogen production from photoelectrochemical (PEC) water splitting powered by solar energy (Fig. 17a). Single-layer 1H-MoS_2 is a photoactive semiconductor with a direct bandgap of 1.8 eV, which exhibits quantum confinement and excellent catalytic activity. Bandgaps of the 2D MX_2 single-layer can be matched with the visible region of the solar spectrum (1−2 eV). Therefore, the 2D MX_2 single-layer can absorb more sunlight, which can lead to high efficiency. Very recently, the Au-MoS_2 composite photo-anode exhibited enhanced photocatalytic water splitting under visible light [188] (Fig. 17b).

To achieve high efficiency of solar energy conversion, the development of active semiconductors under visible-or-longer wavelength regions will become important. Nano materials of 2D MX_2 are potential candidates for renewable hydrogen production. However, many challenges remain regarding solar hydrogen fuel.

4 Energy Storage Applications

Batteries and supercapacitors are important electrochemical energy storage devices, and have been extensively developed for a wide range of applications. The use of these energy storage devices in many energy-related products demands high energy storage capability, power delivery capability, and cycle stability. The rapid advance of these technologies depends on the development of better electrode materials. Nanostructured materials offer excellent energy storage, long lifecycles, and high rate capability.

Atomically thin graphene and 2D-MX_2 nanosheets have also attracted attention for use as electrode materials and electrode assistants for developing energy storage devices.

4.1 Lithium-Ion Batteries

Li-ion batteries (LIBs) have the significant advantages of low weight and higher energy storage. They are at the heart of the most promising energy storage systems for portable electronic devices and future electric vehicles. An LIB system consists of three parts: anode, cathode, and electrolyte. Commercial batteries utilize graphite as anode and lithium cobalt oxide ($LiCoO_2$) as cathode. The electrode materials are closely related to battery performance. Fast insertion and extraction of Li ions can be sustained using 2D nanosheets of materials such as graphene and MX_2. They are ultrathin, flexible, stretchable, and have high surface area, which will be useful for development of future portable, flexible devices.

4.1.1 Graphene Nanosheets for LIBs

Graphite is a common anode-electrode material in LIBs, but exhibits low Li storage capacity. The theoretical capacity of graphite is 372 mAh g^{-1} for LiC_6 [189]. Great efforts have been great made to overcome the capacity limitations of graphite for use in advanced LIBs. Graphene has received significant interest as an electrode material due to its high specific surface area, good chemical and thermal stability, wide potential window, high electrical conductivity, and excellent mechanical flexibility. The theoretical surface area of graphene is 2630 m^2 g^{-1} [7]. This is much higher than that of graphite (~ 10 m^2 g^{-1}) and CNT (~400 m^2 g^{-1}). Thus, atomically thin graphene can be considered an electrode material with good power capability for electrochemical energy storage. In particular, chemically prepared rGO offers a large number of porous sites, good conductivity, and increased interlayer spacing. Furthermore, they can be made to form layered structures with large interlayer space. The specific capacity of rGO sheets, with specific surface area of 492.5 m^2 g^{-1}, was 1264 mAh g^{-1} at 100 mA g^{-1}, a value higher than that of graphite [190]. However, the rGO electrode showed limited rate capability with capacity fluctuation due to induction of instability during lithiation and de-lithiation. To improve the rate capability, N or B-doped graphene has been used. It showed high rate capability and high specific capacity (1043 mAh g^{-1} for N-doped graphene and 1540 mAh g^{-1} for B-doped graphene) [19]. The high performance of these doped-graphene electrodes is due to fast Li-ion diffusion and electron transport supported by heteroatomic defects, increased interlayer distance between graphene sheets, improved electrical conductivity, and thermal stability. Graphene sponges could increase porosity, resulting in improved capacity. However, porosity also reduces the volumetric capacity and leads to a

large amount of electrolyte insertion, resulting in increased irreversible capacity. The first-cycle capacity of rGO sponge was 1059 mAh g^{-1} at 50 mA g^{-1}, but the capacity after a few cycles was only 400 mAh g^{-1} at 50 mA g^{-1} [191]. In the first cycle, mesoporous graphene had a specific capacity of 3535 mAh g^{-1} at 100 mA g^{-1}, but only 1040 mAh g^{-1} in the second cycle [192]. Graphene itself used as electrode material show low rate capability, poor cycle stability, and much lower capacity than a silicon-based anode (4200 mAh g^{-1}).

For high performance LIBs, graphene-based hybrid composites have been widely explored as electrode materials. These include graphene-metal oxide composites, graphene-2D MX_2 composites, graphene-CNT composites, and graphene-Si nanoparticle composites. In graphene-based composites, graphene can act as a highly conductive layer and as a mechanical support layer. The hybrid composites may reduce restacking of graphene layers and maintain a highly active surface area. Thus, Li storage capacity and the cycling performance of graphene-based hybrid composites can be enhanced. For example, the specific capacity of rGO-Fe_2O_3 composite in the 1st and 50th cycles exhibited 1693 and 1027 mAh g^{-1} at 100 mA g^{-1}, respectively [193]. Honeycomb film of 3D rGO-dimethyldioctadecylammonium (rGO-DODA) showed a large specific capacity of about 3025 mAh g^{-1} in the first cycle and a reversible capacity of 1612 mAh g^{-1} at 50 mA g^{-1} [194]. A graphene-silicon hybrid structure may significantly improve the energy density. The rGO-Si nanoparticle composite (SiNP@rGO) exhibited specific capacity of 2920 mAh g^{-1} in the first cycle and capacity of over 1205 mAh g^{-1} after 150 cycles with high cycling stability [195]. Sandwiching Si nanowires in rGO creates 3D porous structure (SiNW@G@rGO) produced a reversible capacity of 1600 mAh g^{-1} at 2100 mA g^{-1} [196]. Porous 3D graphene networks connected with Sn nanoparticles encapsulated within graphene shells (Sn@G-PGNW)were used as LIB anode [9]. Its capacity was 1022 mAh g^{-1} at 0.2 C and it exhibited very long-term cycle stability with capacity of 96.3 % after 1000 cycles. Graphene-based hybrid structures have shown better performance than bare graphene. Generally, graphene-Si composites showed high capacity (more than 2000 mAh g^{-1}). Various graphene-based anode materials useful for LIBs have been listed in Table 3.

One-atom-thick graphene was also studied for use in flexible LIBs, due to its high surface area, excellent flexibility, high conductivity, and short ion diffusion length. Graphene paper is highly conductive and mechanically strong with a Young's modulus of 41.8 GPa and a tensile strength of 293.3 MPa [197]. Some groups have demonstrated it using conductive rGO paper or CVD graphene. This graphene paper exhibited a capacity of 822 mAh g^{-1} at 50 mA g^{-1} [198]. The battery using CVD graphene can be bent and showed energy density of 10 Wh L^{-1} at 50 W L^{-1} with good cycle stability over 100 cycles [12]. These flexible batteries using graphene have shown good flexibility, high capacity, high rate, and long cycle performance even under conditions of repeated bending.

Graphene also serves as cathode material due to its high capacity and long cycle stability. $LiFePO_4$-CVD graphene sponge as cathode material exhibited a reversible capacity of 120 mAh g^{-1} at 10 °C without capacity loss after 500 cycles [199]. Composite sponge of rGO-VO_2 ribbons 10 nm thick showed a high reversible

Table 3 Graphene-based anode materials for LIBs

Material	Surface area ($m^2\ g^{-1}$)	Capacity ($mAh\ g^{-1}$)	Cycle stability ($mAh\ g^{-1}$)	References
rGO	492.5	1264 at 0.1 A/g	848 after 40 cycles	[190]
N-doped rGO	290	1043 at 0.05 A/g	872 after 30 cycles	[19]
B-doped rGO	256	1549 at 0.05 A/g	1227 after 30 cycles	[19]
rGO sponge		1059 at 0.05 A/g	82 % capacity retention after 100 cycles	[191]
Mesoporous graphene	281	3535 at 0.1 A/g	1040 at the 2nd cycle	[192]
			833 after 60 cycles	
rGO-Fe_2O_3		1693 at 0.1 A/g	1027 after 50 cycles	[193]
rGO-DODA honey-comb film		3025 at 0.05 A/g	1612: reversible capacity, 1150 after 50 cycles	[194]
SiNP@rGO		2920 at 0.1 A/g	1720: reversible capacity, 1205 after 150 cycles	[195]
SiNP@G@rGO		1600 at 2.1 A/g	80 % capacity retention after 100 cycles	[196]
Sn@G-PGNW		1022 at 0.2 C	96.3 % capacity retention after 1000 cycles	[9]

specific capacity of 415 mAh g^{-1} at 1 C and cycle stability of 10 % capacity loss for over 1000 cycles [200].

Graphene-sponge-type electrodes could be used as anodes or cathodes in LIBs, with improved performance. Graphene-based hybrid materials that are thin, flexible, and stretchable and have high surface area will be applied in future electronic devices. Despite many achievements, however, new structures of graphene-based electrodes are still needed for higher performance energy-storage devices.

4.1.2 Nanosheets of 2D-TMDs for LIBs

Layered bulk-MX_2 has been explored for use as LIB-anode material due to its potential for improved Li-insertion and extraction. For example, MoS_2 has interlayer spacing of 0.615 nm, larger than that of graphite (0.335 nm), which may easily diffuse Li ions. MoS_2 nanosheets have recently received great attention because of their high theoretical specific capacity (669 mAh g^{-1}), and because of a voltage for Li-ion insertion that is higher than that of graphite, which is a good anode material [201, 202]. The theoretical capacity of MoS_2 is two times higher than that of graphite. However, the electric conductivity of MoS_2 is very low, resulting in poor rate performance. One effective method for enhancing the conductivity is combining graphene, CNT, or conducting polymers with poorly conductive MoS_2 nanosheets. There are many reports of MoS_2 composites used as LIB anodes. The capacity of the MoS_2-rGO composite as anode was 1100 mAh g^{-1} at 100 mA g^{-1} and showed good rate capability [22]. The capacity of MoS_2 (66.7 %)-polyaniline nanowires

Table 4 MX_2-based anode materials for LIBs

Material	Surface area ($m^2\ g^{-1}$)	Capacity ($mAh\ g^{-1}$)	Cycle stability ($mAh\ g^{-1}$)	References
Bulk MoS_2	4.89	800 at 0.05 A/g	226 after 50 cycles	[202]
Restacked MoS_2	9.83	800 at 0.05 A/g	750 after 50 cycles	
MoS_2 nanosheets		1062 at 1 C	907 after 50 cycles	[206]
WS_2 nanosheets		886 at 1 A/g	318.6 after 500 cycles	[205]
MoS_2-GNS(1:2)		2200 at 0.1 A/g	1290 after 50 cycles	[207]
MoS_2-rGO		1100 at 0.1 A/g		[22]
95 % MoS_2-PEO		1131 at 0.05 A/g	~900 after 50 cycles	[208]
66.7 % MoS_2-PANI NW		1063.9 at 0.1 A/g	90.2 % capacity retention after 50 cycles	[209]
Flower-like C@MoS_2	31	1419 at 0.1 A/g	80 % capacity retention after 50 cycles	[203]
MoS_2-MWCNT		1549 at 0.05 A/g	98.6 % capacity retention after 10 cycles	[210]
90 % MoS_2-G nanocable	20	1150 at 0.5 A/g reversible capacity	100 % capacity retention after 160 cycles	[204]
WS_2NT-G		996 at 0.1 A/g	500.2 after 50 cycles	[205]

(66.7 % MoS_2-PANI NW) was 952.6 mAh g^{-1} at 100 mA g^{-1}. Composites of MoS_2 (C@MoS_2) with a coating of flower-like carbon, synthesized with D-glucose and MoO_3, exhibited high reversible specific capacity (1419 mAh g^{-1} at 0.1 A g^{-1}) and good rate performance [203]. MoS_2(90 %)-graphene with a nano-cable structure exhibited a specific capacity of about 1150 mAh g^{-1} at 0.5 A g^{-1} and a long cycle life (100 % capacity retention after 160 cycles) [204]. Nano MoS_2-based composites as anode material exhibited a significant improvement in cycle performance and rate capability. WS_2 could also be a candidate material for anodes. A hybrid of 3D WS_2 nanotubes-graphene (WS_2NT-G) exhibited improved cycling stability and rate capability without additional materials. Its initial capacity was 886.1 mAh g^{-1} at 1 A g^{-1} and 318.6 mAh g^{-1} after 500 cycles [205]. Anode materials for LIBs made of 2D-MX_2 are listed in Table 4.

MoS_2 and MoS_2-graphene composites are also attractive material for novel Na-ion batteries [93, 211, 212]. Sodium-ion batteries (SIBs) have an advantage in large-scale applications that require a large amount, due to the low cost and abundance of Na. Our group worked with MoS_2 nanosheet and MoS_2-rGO composite and these materials showed high initial capacities of 254 mAh g^{-1} (MoS_2 nanosheets) and 376 mAh g^{-1} (MoS_2-rGO composite) [93]. Recently, anodes of single-layer MoS_2-carbon-nanofiber composite exhibited the best rate performance and cycling stability for Na storage in MoS_2. It achieved a specific capacity of 854 mAh g^{-1} at 0.1 A g^{-1} after 1000 cycles [213].

These results demonstrate the advantages of MX_2 nanosheets, graphene, and their hybrid composites as electrode materials for LIBs or SIBs.

4.2 Supercapacitors

Supercapacitors, another energy storage device, have advantages over batteries in their high power density and excellent cycle ability. There are two types, according to the energy storage mechanism used [214, 215]. One type is called the electrical double layer capacitors (EDLCs). They store charge in electric double layer formed at the interface between an activated electrode and an electrolyte. The others are called pseudocapacitors, and the charge storage of pseudocapacitors depends on fast faradaic redox reactions.

EDLC depends on the charge in the double layer of the electrodes used. The capacitance is given by $C = \varepsilon A/4\pi t$, where ε is the dielectric constant of the electrical double-layer region, A the electrode surface area, and t the thickness of electrical double layer. To achieve high capacitance, large specific surface area and thin double layers are necessary. Double-layer charge storage is a surface process and the surface properties of the electrode greatly influence the capacitance. Therefore, the performance of EDLCs is determined by the choice of electrode material, in relation to large surface area and high electrical conductivity. Recently, 2D layered materials (i.e., graphene, rGO, MX_2 and their composites) have been shown to be efficient, promising materials for high-performance supercapacitor electrodes due to their large surface area and large in-plane conductivity [10, 16, 216–220].

Generally, activated carbon materials have been used as electrodes for EDLCs, while transition metal chalcogenides have been investigated for use in pseudocapacitors. Thus, atomically thin graphene and 2D-MX_2 nanosheets are attractive for use in supercapacitors.

4.2.1 Graphene Nanosheets for Supercapacitors

The capacitance of single-layer graphene was reported to be 21 mF cm^{-2}. Theoretical gravimetric capacitance of graphene materials is about 550 F g^{-1}, which is the highest capacitance value among all carbon-based electrodes [215, 221]. However, the reported capacitance of graphene-based materials is still below the theoretical value. For example, the specific capacitance of reduced graphite oxide was 135 F g^{-1} in aqueous KOH electrolyte and 99 F g^{-1} in organic electrolyte [18]. The specific capacitances of other graphene materials with different treatments were also low values (120 F g^{-1} in an organic electrolyte for a reduced graphene oxide (rGO) electrode by thermal heating [222], 282 F g^{-1} at 1 A g^{-1} for an N-doped rGO electrode) [218]. To enhance storage capacity, several groups have reported results from work on various supercapacitors using graphene-hybrid composites (e.g., graphene-CNTs composites and graphene-conductive polymer composites). There have also been many efforts to enhance capacity using porous graphene. KOH-activated rGO had large surface area (3100 m^2 g^{-1}) and specific capacitance of 165 F g^{-1} at 1.4 A g^{-1} [219]. The porous graphene grown on a

porous MgO layer gave a specific capacitance of 255 F g^{-1} at 10 mV s^{-1} in 6 M KOH aqueous solution [223]. Vertically oriented graphene showed high power density (112.6 kW kg^{-1} at 600 A g^{-1}) [10]. Restacking inhibited rGO using melamine resin exhibited high specific capacitance of 210 F g^{-1} at 0.5 A g^{-1} due to its high specific surface area (~1040 m^2 g^{-1}) and large macro pore distributions [224]. 3D macroporous embossed graphene frameworks showed both high energy (44 Wh kg^{-1}) and power densities (25 kW kg^{-1}) [225]. Recently, a supercapacitor with holey-graphene-framework electrode yielded a high capacitance (298 F g^{-1} at 1 A g^{-1}) [16].

Portable electronic devices require on-chip energy storage. Micro- or nanodevices have advantages such as small thermal time constants, high sensitivity, and integrated circuit fabrication. Micro-supercapacitor devices formed by the patterning of graphite oxide thin film exhibited good energy storage capacity and excellent cycle stability [226]. However, they also showed large internal resistance (6.5 kΩ) and poor frequency response. Recently, more than 100 micro-supercapacitors exhibited high power (200 W cm^{-3}), excellent frequency response, and were highly bendable [15]. A graphene supercapacitor using plain-woven fabric composites showed a specific capacitance of 8 mF cm^{-2} (267 F g^{-1}) [227]. This electrode had excellent flexibility, an electrode about 1–7 nm thick, and device thickness of less than 1 mm. Thus, this flexible electrode could be useful for energy storage devices in portable and wearable electronics. Although many improved electrodes for EDLC have been demonstrated, their capacitance is still not sufficient for high-performance energy storage devices.

Pseudocapacitors (another type of supercapacitor), have large specific capacitance and hybrid electrodes made of redox-active materials and highly conductive graphene-based materials. Transition metal oxides (RuO_2 and MnO_2) are widely used as pseudocapacitor electrode materials, and they use fast and reversible redox reactions for charge storage [228–230]. Graphene-38w% RuO_2 composites exhibited high specific capacitance of 570 F g^{-1} and excellent cycle stability of 97.9 % capacitance retention after 1000 cycles [229]. Graphene-MnO_2 nanoparticle composites with 3D-porous structure showed a specific capacitance of 389 F g^{-1} at 1.0 A g^{-1}, energy density of 44 Wh kg^{-1}, and power density of 25 kW kg^{-1} [225]. Micro-supercapacitor of rGO-polyaniline films showed electrochemical capacitance of 970 F g^{-1} at 2.5 A g^{-1} [230]. Cobalt oxide nanowires on 3D-graphene possessed high specific capacitance (1100 F g^{-1} at 10 A g^{-1}) [228]. Thus, hybrid graphene combined with other active materials give much higher pseudo-capacitance. Details of graphene-based supercapacitors are summarized in Table 5.

4.2.2 Nanosheets of 2D-TMDs for Supercapacitors

Nanosheets of 2D-MoS_2 have graphene-like morphology including a basal plane and an edge plane similar to graphene, which provide a large surface area. These 2D-MoS_2 nanosheets can be stacked using van der Waals interaction. The Mo atoms possess a range of oxidation states from Mo^{2+} to Mo^{6+} and show promising

Table 5 Graphene-based supercapacitors

Material	Capacitance	Electrolyte	References
EDLC			
Reduced graphite oxide	135 F/g	5.5 M KOH	[18]
	99 F/g	1 M TEA BF_4/AN	
rGO by thermal treatment	122 F/g at 1 A/g	1 M TEA BF_4/PC	[222]
N-doped rGO	282 F/g at 1 A/g	1 M TEA BF_4/AN	[218]
Microwave exfoliated porous rGO	165 F/g at 1.4 A/g	1 M BMIM BF_4/AN	[219]
Porous graphene grown on porous MgO template	255 F/g at 10 mV/s	6 M KOH	[223]
Vertically oriented graphene	156 F/g at 100 A/g	6 M KOH	[10]
3D Holey graphene	298 F/g at 1 A/g	6 M KOH	[16]
Pseudocapacitor			
rGO-RuO_2 composite	570 F/g at 1 A/g	1 M H_2SO_4	[229]
rGO-MnO_2 composite	389 F/g at 1 A/g	1 M Na_2SO_4	[225]
rGO-PANI	970 F/g at 2.5 A/g	1 M Na_2SO_4	[230]
Co_3O_4 NW on 3D graphene	1100 F/g at 10 A/g	2 M KOH	[228]

behavior as pseudocapacitors. The theoretical gravimetric capacitance of MoS_2 is about 1000 F g^{-1} [207]. Their performance in supercapacitors is comparable to that of CNT array electrodes. Liquid-based exfoliated 2D-MoS_2 nanosheets could be used as electrode materials for high-performance micro-supercapacitors. Micro-supercapacitors with finger-like electrodes using 2D-MoS_2 films exhibited area-specific capacitance of 8 mF cm^{-2} with excellent cycle stability [220]. Composites of MoS_2-polypyrrole exhibited high specific capacitance (553. 7 F g^{-1} at 1 A g^{-1}) and high cycle stability [217]. Hybrid electrodes of 2D WS_2-rGO also exhibited high specific capacitance (350 F g^{-1}) [216]. Edge-oriented MoS_2-nanowall films made excellent supercapacitors [231]. Supercapacitors using 2D MX_2 nanosheet-graphene composites showed enhanced specific capacitance and excellent cycle stability.

Graphene and 2D-MX_2 nanosheets could provide electrode materials for supercapacitors to be used in portable, flexible, transparent microelectronic devices. In spite of the substantial research already done, there is still a pressing need to develop higher quality electrode materials with higher power, higher energy density, and lower cost for supercapacitor applications.

5 Summary

Today, some of the fastest growing technologies are related to electronic devices in communication, health care, and environmental monitoring. Nanomaterials that might be used to enhance energy harvesting, energy conversion, and energy

storage devices are in great demand. High-quality electrode materials are the main driving force for energy-related devices involving high power and high energy density at lower cost.

Graphene and 2D-MX_2 nanosheets are very attractive for energy harvesting, energy conversion and storage applications, due to their superior electrical, optical, and mechanical properties. Graphene, in particular, shown to be an ideal material for use in many of the nanoscale devices used in energy harvesting/conversion and storage applications, is still being studied for these and other energy-related purposes. Recently, 2D-MX_2 nanomaterials are also attracting significant attention in many energy-related applications. These 2D-MX_2 materials exhibit controllable bandgap properties and MX_2 single-layer is a direct bandgap semiconductor with bandgap energy of 1.2–2.1 eV in its elemental composition. The optical properties of 2D-MX_2 nanosheets have potential as energy harvesting and energy conversion materials in solar cells, photoelectrochemical cells, and photo-fuel cells. Moreover, the heterostructure of semiconducting 2D-MX_2 can be used to control its physical and optical properties, which include large photon absorption, high photocurrent production, and fast charge separation. Thus, atomically thin layers of 2D-MX_2 and graphene, with good flexibility and high transparency, are promising materials for use in future devices. In particular, 3D assembly technology will provide intrinsic advantages necessary for high efficiency in practical applications.

Acknowledgments This work was supported by the Nano-Material Technology Development Program (2012M3A7B4049807), Global Frontier Project (CASE-2011-0031640), and LG Display Co., Ltd.

References

1. Novoselov KS et al (2004) Electric field effect in atomically thin carbon films. Science 306:666–669
2. Georgakilas V et al (2012) Functionalization of graphene: covalent and non-covalent approaches, derivatives and applications. Chem Rev 112:6156–6214
3. Molitor F et al (2011) Electronic properties of graphene nanostructures. J Phys: Condens Matter 23(243201):243215
4. Nair RR et al (2008) Fine structure constant defines visual transparency of graphene. Science 320:1308
5. Sarma SD et al (2011) Electronic transport in two-dimensional graphene. Rev Mod Phys 83:407–470
6. Wang QH et al (2012) Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. Nat Nanotech 7:699–712
7. Zhu Y et al (2010) Graphene and graphene oxide: synthesis, properties, and applications. Adv Mater 22:3906–3924
8. Kim K et al (2014) Ultrathin organic solar cells with graphene doped by ferroelectric polarization. ACS Appl Mater Interfaces 6:3299–3304
9. Qin Jian et al (2014) Graphene networks anchored with sn@graphene as lithium ion battery anode. ACS Nano 8:1728–1738
10. Bo Z et al (2013) Vertically oriented graphene bridging active-layer/current-collector interface for ultrahigh rate supercapacitors. Adv Mater 25:5799–5806

11. Liu Z et al (2013) Package-Free flexible organic solar cells with graphene top electrodes. Adv Mater 25:4296–4301
12. Wei D et al (2013) Ultrathin rechargeable all-solid-state batteries based on monolayer graphene. J Mater Chem A 1:3177–3181
13. Zhang C et al (2013) Synthesis of amino-functionalized graphene as metal-free catalyst and exploration of the roles of various nitrogen states in oxygen reduction reaction. Nano Energy 2:88–97
14. Wang JT-W et al (2014) Low-temperature processed electron collection layers of graphene/tio_2 nanocomposites in thin film perovskite solar cells. Nano Lett 14:724–730
15. El-Kady MF, Kaner RB (2013) Scalable fabrication of high-power graphene micro-supercapacitors for flexible and on-chip energy storage. Nat Commun 4:1475–1484
16. Xu Y et al (2014) Holey graphene frameworks for highly efficient capacitive energy storage. Nat Commun 5:4554–4562
17. Xue Y et al (2014) Controllable synthesis of doped graphene and its applications. Small 10:2975–2991
18. Stoller MD et al (2008) Graphene-based ultracapacitors. Nano Lett 8:3498–3502
19. Wu Z-S et al (2011) Doped graphene sheets as anode materials with superhigh rate and large capacity for lithium ion batteries. ACS Nano 5:5463–5471
20. Liu C et al (2010) Graphene-based supercapacitor with an ultrahigh energy density. Nano Lett 10:4863–4868
21. Wang G et al (2009) Graphene nanosheets for enhanced lithium storage in lithium ion batteries. Carbon 47:2049–2053
22. Chhowalla M et al (2013) The chemistry of two-dimensional layered transition metal dichalcogenide nanosheets. Nat Chem 5:263–275
23. Enyashin A et al (2007) Nanosized allotropes of molybdenum disulfide. Eur Phys J Special Topics 149:103–125
24. Ataca C et al (2012) Stable, single-layer MX_2 transition-metal oxides and dichalcogenides in a honeycomb-like structure. J Phys Chem 116:8983–8999
25. Putungan DB, Kuo J-L (2014) Structural and electronic properties of monolayer 1T-MoS_2 phase, and its interaction with water adsorbed on perfect, single S-vacated and MoS_2-unit-vacated surface: density functional theory calculations. Integrated Ferroelectrics 156:93–101
26. Eda G et al (2011) Photoluminescence from chemically exfoliated MoS_2. Nano Lett 11:5111–5116
27. Yun WS et al (2012) Thickness and strain effects on electronic structures of transition metal dichalcogenides: 2H-MX_2 semiconductors (M = Mo, W; X = S, Se, Te). Phys Rev B 85:033305
28. Kuc A et al (2011) Influence of quantum confinement on the electronic structure of the transition metal sulfide TS_2. Phys Rev B 83:245213
29. Kadantsev ES, Hawrylakb P (2012) Electronic structure of a single MoS_2 monolayer. Solid State Comm 152:909–913
30. Huang W et al (2014) Theoretical study of thermoelectric properties of few-layer MoS_2 and WSe_2. Phys Chem Chem Phys 16:10866–10874
31. Kumara A, Ahluwalia PK (2012) Electronic structure of transition metal dichalcogenides monolayers 1H-MX_2 (M = Mo, W; X = S, Se, Te) from ab-initio theory: new direct band gap semiconductors. Eur Phys J B 85:186–193
32. Radisavljevic B et al (2011) Single-layer MoS_2 transistors. Nat Nanotech 6:147–150
33. Bertolazzi S et al (2011) Stretching and breaking of ultrathin MoS_2. ACS Nano 5:9703–9709
34. Ayari A et al (2007) Realization and electrical characterization of ultrathin crystals of layered transition-metal dichalcogenides. J Appl Phys 101:014057
35. Zhan Y et al (2012) Large-area vapor-phase growth and characterization of MoS_2 atomic layers on a SiO_2 substrate. Small 8:966–971
36. Zeng H et al (2012) Valley polarization in MoS_2 monolayers by optical pumping. Nat Nanotech 7:490–493

37. Mak KF et al (2012) Control of valley polarization in monolayer MoS_2 by optical helicity. Nat Nanotech 7:494–498
38. Yue Q et al (2012) Mechanical and electronic properties of monolayer MoS_2 under elastic strain. Phys Lett A 376:1166–1170
39. Zhou M et al (2014) Production of graphene by liquid-phase exfoliation of intercalated graphite. Int J Electrochem Sci 9:810–820
40. Bonaccorso F et al (2012) Production and processing of graphene and 2d crystals. Mater Today 15:564–589
41. Allen MJ et al (2010) Honeycomb carbon: a review of graphene. Chem Rev 110:132–145
42. Kumar P et al (2013) Graphene: synthesis, properties and application in transparent electronic devices. Rev Adv Sci Eng 2:1–21
43. Park S, Ruoff RS (2009) Chemical methods for the production of graphenes. Nat Nanotech 4:217–224
44. Kim KS et al (2009) Large-scale pattern growth of graphene films for stretchable transparent electrodes. Nature 457:706–710
45. Shen Y, Lua AC (2013) A facile method for the large-scale continuous synthesis of graphene sheets using a novel catalyst. Sci Rep 3:3037–3042
46. Zhou H et al (2013) Chemical vapour deposition growth of large single crystals of monolayer and bilayer graphene. Nat Commun 4:2096–2103
47. Lee J-H et al (2014) Wafer-scale growth of single-crystal monolayer graphene on reusable hydrogen-terminated germanium. Science 344(6181):286–289
48. Muñoz R, Gómez-Aleixandre C (2013) Review of CVD synthesis of graphene. Chem Vap Depos 19:297–322
49. Novoselov KS et al (2005) Two-dimensional atomic crystals. Proc Natl Acad Sci USA 102:10451–10453
50. Li H et al (2012) Fabrication of single- and multilayer MoS_2 film-based field-effect transistors for sensing no at room temperature. Small 8:63–67
51. Osada M, Sasaki T (2012) Two-dimensional dielectric nanosheets: novel nanoelectronics from nanocrystal building blocks. Adv Mater 24:210–228
52. Kalantar-zadeh K et al (2010) Synthesis of atomically thin WO_3 sheets from hydrated tungsten trioxide. Chem Mater 22:5660–5666
53. Late DJ et al (2012) Hysteresis in single-layer MoS_2 field effect transistors. ACS Nano 6:5635–5641
54. Lee C et al (2010) Anomalous lattice vibrations of single- and few-layer MoS_2. ACS Nano 4:2695–2700
55. McDonnell S et al (2013) HfO_2 on MoS_2 by atomic layer deposition: adsorption mechanisms and thickness scalability. ACS Nano 7:10354–10361
56. Radisavljevic B et al (2011) Integrated circuits and logic operations based on single-layer MoS_2. ACS Nano 5:9934–9938
57. Bao W et al (2013) High mobility ambipolar MoS_2 field-effect transistors: substrate and dielectric effects. Appl Phys Lett 102:042104
58. Splendiani A et al (2010) Emerging photoluminescence in monolayer MoS_2. Nano Lett 10:1271–1275
59. Sun Y et al (2014) Probing local strain at MX_2—metal boundaries with surface plasmon-enhanced raman scattering. Nano Lett 14:5329–5334
60. Zhang Y et al (2012) Ambipolar MoS_2 thin flake transistors. Nano Lett 12:1136–1140
61. Mak KF et al (2010) Atomically thin MoS_2: a new direct-gap semiconductor. Phys Rev Lett 105:136805
62. Zhang Y et al (2013) Review of chemical vapor deposition of graphene and related applications. Acc Chem Res 46:2329–2339
63. Gong Y et al (2013) Stabilizing nanostructured solid oxide fuel cell cathode with atomic layer deposition. Nano Lett 13:4340–4345
64. Lee Y-H et al (2012) Synthesis of large-area mos_2 atomic layers with chemical vapor deposition. Adv Mater 24:2320–2325

65. Lin Y-C et al (2012) Wafer-scale MoS_2 thin layers prepared by MoO_3 sulfurization. Nanoscale 4:6637–6641
66. Zande AMvd et al (2013) Grains and grain boundaries in highly crystalline monolayer molybdenum disulphide. Nature Mater 12:554–561
67. Wang X et al (2013) Controlled synthesis of highly crystalline MoS_2 flakes by chemical vapor deposition. J Am Chem Soc 135:5304–5307
68. Najmaei S et al (2013) Vapour phase growth and grain boundary structure of molybdenum disulphide atomic layers. Nat Mater 12:754–759
69. Lee Y-H et al (2013) Synthesis and transfer of single-layer transition metal disulfides on diverse surfaces. Nano Lett 13:1852–1857
70. Liu K-K et al (2012) Growth of large-area and highly crystalline MoS_2 thin layers on insulating substrates. Nano Lett 12:1538–1544
71. Chang Y-H et al (2014) Monolayer $MoSe_2$ grown by chemical vapordeposition for fast photodetection. ACS Nano 8:8582–8590
72. Huang J-K et al (2014) Large-area synthesis of highly crystalline WSe_2 monolayers and device applications. ACS Nano 8:923–930
73. Shim GW et al (2014) Large-area single-layer $MoSe_2$ and Its van der waals heterostructures. ACS Nano 8:6655–6662
74. Zhang Y et al (2013) Controlled growth of high-quality monolayer WS_2 layers on sapphire and imaging its grain boundary. ACS Nano 7:8963–8971
75. Ji Q et al (2013) Epitaxial monolayer MoS_2 on mica with novel photoluminescence. Nano Lett 13:3870–3877
76. Yu Y et al (2013) Controlled scalable synthesis of uniform, high-quality monolayer and few-layer MoS_2 films. Sci Rep 3:1866–1871
77. Cong C et al (2014) Synthesis and optical properties of large-area single-crystalline 2d semiconductor WS_2 monolayer from chemical vapor deposition. Adv Optical Mater 2:131–136
78. Cai M et al (2012) Methods of graphite exfoliation. J Mater Chem 22:24992–25002
79. Park KH et al (2012) Exfoliation of non-oxidized graphene flakes for scalable conductive film. Nano Lett 12:2871–2876
80. Tung VC et al (2009) High-throughput solution processing of large-scale graphene. Nat Nanotech 4:25–29
81. Eda G, Chhowalla M (2010) Chemically derived graphene oxide: towards large-area thin-film electronics and optoelectronics. Adv Mater 22:2392–2415
82. Wei D, Liu Y (2010) Controllable synthesis of graphene and its applications. Adv Mater 22:3225–3241
83. Dreyer DR et al (2010) The chemistry of graphene oxide. Chem Soc Rev 39:228–240
84. Gao W et al (2009) New insights into the structure and reduction of graphite oxide. Nat Chem 1:403–408
85. Dikin DA et al (2007) Preparation and characterization of graphene oxide paper. Nat 448:457–460
86. Sundaram RS et al (2008) Electrochemical modification of graphene. Adv Mater 20:3050–3053
87. Stankovicha S et al (2007) Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphite oxide. Carbon 45:1558–1565
88. Zhou M et al (2009) Controlled synthesis of large-area and patterned electrochemically reduced graphene oxide films. Chem Eur J 15:6116–6120
89. McAllister MJ et al (2007) Single sheet functionalized graphene by oxidation and thermal expansion of graphite. Chem Mater 19:4396–4404
90. Yang D et al (2009) Chemical analysis of graphene oxide films after heat and chemical treatments by X-ray photoelectron and micro-raman spectroscopy. Carbon 47:145–152
91. Shao Y et al (2010) Facile and controllable electrochemical reduction of graphene oxide and its applications. J Mater Chem 20:743–748
92. Hernandez Y et al (2008) High-yield production of graphene by liquid-phase exfoliation of graphite. Nat Nanotech 3:563–568

93. Bang GS et al (2014) Effective liquid-phase exfoliation and sodium ion battery application of MoS_2 nanosheets. ACS Appl Mater Interfaces 6:7084–7089
94. Cunningham G et al (2012) Solvent exfoliation of transition metal dichalcogenides: dispersibility of exfoliated nanosheets varies only weakly between compounds. ACS Nano 6:3468–3480
95. Zhou K-G et al (2011) A mixed-solvent strategy for efficient exfoliation of inorganic graphene analogues. Angew Chem Int Ed 50:10839–10842
96. Coleman JN et al (2011) Two-dimensional nanosheets produced by liquid exfoliation of layered materials. Science 331:568–571
97. Smith RJ et al (2011) Large-scale exfoliation of inorganic layered compounds in aqueous surfactant solutions. Adv Mater 23:3944–3948
98. Zeng Z et al (2011) Single-layer semiconducting nanosheets: high-yield preparation and device fabrication. Angew Chem Int Ed 50:11093–11097
99. Wang H et al (2013) Electrochemical tuning of vertically aligned MoS_2 nanofilms and its application in improving hydrogen evolution reaction. Proc Natl Acad Sci USA 110:19701–19706
100. De S, Coleman JN (2010) Are there fundamental limitations on the sheet resistance and transmittance of thin graphene films? ACS Nano 4:2713–2720
101. Bae S et al (2010) Roll-to-roll production of 30-inch graphene films for transparent electrodes. Nat Nanotech 5:574–578
102. Zhu Y et al (2011) Rational design of hybrid graphene films for high-performance transparent electrodes. ACS Nano 5:6472–6479
103. Wang X et al (2008) Transparent, conductive graphene electrodes for dye-sensitized solar cells. Nano Lett 8:323–327
104. Li S-S et al (2010) Solution-processable graphene oxide as an efficient hole transport layer in polymer solar cells. ACS Nano 4:3169–3174
105. DeArco LG et al (2010) Continuous, highly flexible, and transparent graphene films by chemical vapor deposition for organic photovoltaics. ACS Nano 4:2865–2873
106. Hsu C-L et al (2012) Layer-by-layer graphene/TCNQ stacked films as conducting anodes for organic solar cells. ACS Nano 6:5031–5039
107. Gong F et al (2013) Enhanced charge transportation in a polypyrrole counter electrode via incorporation of reduced graphene oxide sheets for dye-sensitized solar cells. Phys Chem Chem Phys 15:546–552
108. Chen L et al (2013) Graphene quantum-dot-doped polypyrrole counter electrode for high-performance dye-sensitized solar cells. ACS Appl Mater Interfaces 5:2047–2052
109. Peng J et al (2012) Graphene quantum dots derived from carbon fibers. Nano Lett 12:844–849
110. Zhang Y et al (2013) Graphene quantum dots/gold electrode and its application in living cell H_2O_2 detection. Nanoscale 5:1816–1819
111. Yan X et al (2010) Large, solution-processable graphene quantum dots as light absorbers for photovoltaics. Nano Lett 10:1869–1873
112. Zhang Y et al (2012) Graphene transforms wide band gap zns to a visible light photocatalyst. the new role of graphene as a macromolecular photosensitizer. ACS Nano 6:9777–9789
113. Williams KJ et al (2013) Hot electron injection from graphene quantum dots to TiO_2. ACS Nano 7:1388–1394
114. Miao X et al (2012) High efficiency graphene solar cells by chemical doping. Nano Lett 12:2745–2750
115. Song J et al (2011) Enhancement of photogenerated electron transport in dye-sensitized solar cells with introduction of a reduced graphene oxide-TiO_2 junction. Chem Eur J 17:10832–10837
116. Liu J et al (2012) Hole and electron extraction layers based on graphene oxide derivatives for high-performance bulk heterojunction solar cells. Adv Mater 24:2228–2233
117. Bernardi M et al (2013) Extraordinary sunlight absorption and one nanometer thick photovoltaics using two-dimensional monolayer materials. Nano Lett 13:3664–3670

118. Wu M et al (2011) Economical and effective sulfide catalysts for dye-sensitized solar cells as counter electrodes. Phys Chem Chem Phys 13:19298–19301
119. Patil SA et al (2014) Highly efficient and stable DSSCs of wet-chemically synthesized MoS_2 counter electrode. Dalton Trans 43:5256–5259
120. Lee LTL et al (2014) Few-layer $MoSe_2$ possessing high catalytic activity towards iodide/triiodide redox shuttles. Sci Rep 4:4063–4069
121. Yue G et al (2012) A catalytic composite film of MoS_2/graphene flake as a counter electrode for Pt-free dye-sensitized solar cells. Electrochim Acta 85:162–168
122. Liu C-J et al (2012) Facile synthesis of MoS_2/graphene nanocomposite with high catalytic activity toward triiodide reduction in dye-sensitized solar cells. J Mater Chem 22:21057–21064
123. Shanmugam M et al (2012) Layered semiconductor molybdenum disulfide nanomembrane based Schottky-barrier solar cells. Nanoscale 4:7399–7405
124. Shanmugam M et al (2012) Molybdenum disulphide/titanium dioxide nanocomposite-poly 3-hexylthiophene bulk heterojunction solar cell. Appl Phys Lett 100:153901
125. Yun J-M et al (2013) Efficient work-function engineering of solutionprocessed MoS_2 thin-films for novel hole and electron transport layers leading to high-performance polymer solar cells. J Mater Chem C 1:3777–3783
126. Gu X et al (2013) A solution-processed hole extraction layer made from ultrathin MoS_2 nanosheets for effi cient organic solar cells. Adv Energy Mater 3:1262–1268
127. Niu L et al (2014) Salt-assisted high-throughput synthesis of single- and few-layer transition metal dichalcogenides and their application in organic. Solar Cells. small. doi:10.1002/smll.201401647
128. Yun J-M et al (2014) Exfoliated and partially oxidized MoS_2 nanosheets by one-pot reaction for efficient and stable organic solar cells. Small 10:2319–2324
129. Yang X et al (2014) Au nanoparticles on ultrathin MoS_2 sheets for plasmonic organic solar cells. J Mater Chem A 2:14798–14806
130. Geim AK, Grigorieva IV (2013) Van der Waals heterostructures. Nature 499:419–425
131. Haigh SJ et al (2012) Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices. Nat Mater 11:764–767
132. Lopez-Sanchez O et al (2014) Light generation and harvesting in a van der waals heterostructure. ACS Nano 8:3042–3048
133. Hong X et al (2014) Ultrafast charge transfer in atomically thin MoS_2/WS_2 heterostructures. Nat Nanotech 9:682–686
134. Lee C-H et al (2014) Atomically thin p-n junctions with van der Waals heterointerfaces. Nat Nanotech 9:676–681
135. Britnell L et al (2013) Strong light-matter interactions in heterostructures of atomically thin films. Science 340:1311–1314
136. Arico AS et al (2005) Nanostructured materials for advanced energy conversion and storage devices. Nat Mater 4:366–377
137. Shao Y et al (2010) Nitrogen-doped graphene and its electrochemical applications. J Mater Chem 20:7491–7496
138. Li Y et al (2012) Stabilization of high-performance oxygen reduction reaction pt electrocatalyst supported on reduced graphene oxide/carbon black composite. J Am Chem Soc 134:12326–12329
139. Wang C et al (2013) A unique platinum-graphene hybrid structure for high activity and durability in oxygen reduction reaction. Sci Rep 3:2580–2586
140. Qu L et al (2010) Nitrogen-doped graphene as efficient metal-free electrocatalyst for oxygen reduction in fuel cells. ACS Nano 4:1321–1326
141. Yang Z et al (2012) Sulfur-doped graphene as an efficient metal-free cathode catalyst for oxygen reduction. ACS Nano 6:205–211

142. Li Q et al (2012) Nitrogen-doped colloidal graphene quantum dots and their size-dependent electrocatalytic activity for the oxygen reduction reaction. J Am Chem Soc 134:18932–18935
143. Li Y et al (2010) Catalytic performance of Pt nanoparticles on reduced graphene oxide for methanol electro-oxidation. Carbon 48:1124–1130
144. Zhou Y-G et al (2010) A facile approach to the synthesis of highly electroactive Pt nanoparticles on graphene as an anode catalyst for direct methanol fuel cells. Chem Commun 46:5951–5953
145. Li Y et al (2009) Preparation and electrochemical performance for methanol oxidation of pt/graphene nanocomposites. Electrochem Commun 11:846–849
146. Kakaei K, Zhiani M (2013) A new method for manufacturing graphene and electrochemical characteristic of graphene-supported Pt nanoparticles in methanol oxidation. J Power Sour 225:356–363
147. Xin Y et al (2011) Preparation and characterization of Pt supported on graphene with enhanced electrocatalytic activity in fuel cell. J Power Sour 196:1012–1018
148. Maiyalagan T et al (2012) Electrodeposited Pt on three-dimensional interconnected graphene as a free-standing electrode for fuel cell application. J Mater Chem 22:5286–5290
149. Yoo E et al (2011) Sub-nano-Pt cluster supported on graphene nanosheets for CO tolerant catalysts in polymer electrolyte fuel cells. J Power Sour 196:110–115
150. Wang H et al (2012) Review on recent progress in nitrogen-doped graphene: synthesis, characterization, and its potential applications. ACS Catal 2:781–794
151. Sheng Z-H et al (2011) Catalyst-free synthesis of nitrogen-doped graphene via thermal annealing graphite oxide with melamine and its excellent electrocatalysis. ACS Nano 5:4350–4358
152. Wang S et al (2011) Polyelectrolyte-functionalized graphene as metal-free electrocatalysts for oxygen reduction. ACS Nano 5:6202–6209
153. Lin Z et al (2012) Facile synthesis of nitrogen-doped graphene via pyrolysis of graphene oxide and urea, and its electrocatalytic activity toward the oxygen-reduction reaction. Adv Energy Mater 2:884–888
154. Zhao Y et al (2012) A versatile, ultralight, nitrogen-doped graphene framework. Angew Chem Int Ed 51:11371–11375
155. Lee KR et al (2010) Electrochemical oxygen reduction on nitrogen doped graphene sheets in acid media. Electrochem Commun 12:1052–1055
156. Ahmed MS, Jeon S (2012) New functionalized graphene sheets for enhanced oxygen reduction as metal-free cathode electrocatalysts. J Power Sour 218:168–173
157. Wu J et al (2013) Electrocatalytic activity of nitrogen-doped graphene synthesized via a one-pot hydrothermal process towards oxygen reduction reaction. J Power Sour 227:185–190
158. Yang Z et al (2013) Recent progress in doped carbon nanomaterials as effective cathode catalysts for fuel cell oxygen reduction reaction. J Power Sour 236:238–249
159. Jafri RI et al (2010) Nitrogen doped graphene nanoplatelets as catalyst support for oxygen reduction reaction in proton exchange membrane fuel cell. J Mater Chem 20:7114–7117
160. He C et al (2013) A strategy for mass production of self-assembled nitrogen-doped graphene as catalytic materials. J Mater Chem A 1:1401–1406
161. Zhang L et al (2012) Effect of microstructure of nitrogen-doped graphene on oxygen reduction activity in fuel cells. Langmuir 28:7542–7550
162. Leger J-M (2001) Mechanistic aspects of methanol oxidation on platinum-based electrocatalysts. J Appl Electrochem 31:767–771
163. Yoo E et al (2011) Sub-nano-Pt cluster supported on graphene nanosheets for CO tolerant catalysts in polymer electrolyte fuel cells. J Power Sour 196:110–115
164. Zhao Y et al (2011) Enhanced electrocatalytic oxidation of methanol on Pd/polypyrrole-graphene in alkaline medium. Electrochim Acta 56:1967–1972

165. Yoo E et al (2009) Enhanced electrocatalytic activity of pt subnanoclusters on graphene nanosheet surface. Nano Lett 9:2255–2259
166. Yang X et al (2012) Bimetallic PtPd nanoparticles on Nafion–graphene film as catalyst for ethanol electro-oxidation. J Mater Chem 22:8057–8062
167. Chen X et al (2011) Synthesis of “Clean” and well-dispersive Pd nanoparticles with excellent electrocatalytic property on graphene oxide. J Am Chem Soc 133:3693–3695
168. Zhao H et al (2011) Fabrication of a palladium nanoparticle/graphene nanosheet hybrid via sacrifice of a copper template and its application in catalytic oxidation of formic acid. Chem Commun 47:2014–2016
169. Rao CV et al (2011) Graphene-supported ptau alloy nanoparticles: a highly efficient anode for direct formic acid fuel cells. J Phys Chem C 115:21963–21970
170. Hinnemann B et al (2005) Biomimetic hydrogen evolution: MoS_2 nanoparticles as catalyst for hydrogen evolution. J Am Chem Soc 127:5308–5309
171. Kibsgaard J et al (2012) Engineering the surface structure of MoS_2 to preferentially expose active edge sites for electrocatalysis. Nat Mater 11:963–969
172. Jaramillo TF et al (2007) Identification of active edge sites for electrochemical H_2 evolution from MoS_2 nanocatalysts. Science 317:100–103
173. Lauritsen JV et al (2004) Atomic-scale insight into structure and morphology changes of MoS_2 nanoclusters in hydrotreating catalysts. J Catal 221:510–522
174. Bonde J et al (2009) Hydrogen evolution on nano-particulate transition metal sulfides. Faraday Discuss 140:219–231
175. Benck JD et al (2012) Amorphous molybdenum sulfide catalysts for electrochemical hydrogen production: insights into the origin of their catalytic activity. ACS Catal 2:1916–1923
176. Xie J et al (2013) Defect-rich MoS_2 ultrathin nanosheets with additional active edge sites for enhanced electrocatalytic hydrogen evolution. Adv Mater 25:5807–5813
177. Merki D et al (2011) Amorphous molybdenum sulfide films as catalysts for electrochemical hydrogen production in water. Chem Sci 2:1262–1267
178. Wang H et al (2014) Electrochemical tuning of MoS_2 nanoparticles on three-dimensional substrate for efficient hydrogen evolution. ACS Nano 8:4940–4947
179. Voiry D et al (2013) Enhanced catalytic activity in strained chemically exfoliated WS_2 nanosheets for hydrogen evolution. Nat Mater 12:850–855
180. Liao L et al (2013) MoS_2 formed on mesoporous graphene as a highly active catalyst for hydrogen evolution. Adv Funct Mater 23:5326–5333
181. Li Y et al (2011) MoS_2 nanoparticles grown on graphene: an advanced catalyst for the hydrogen evolution reaction. J Am Chem Soc 133:7296–7299
182. Kim J et al (2013) Enhanced electrocatalytic properties of transition-metal dichalcogenides sheets by spontaneous gold nanoparticle decoration. J Phys Chem Lett 4:1227–1232
183. Huang X et al (2013) Solution-phase epitaxial growth of noble metal nanostructures on dispersible single-layer molybdenum disulfide nanosheets. Nat Commun 4:1444–1451
184. Yuwen L et al (2014) General synthesis of noble metal (Au, Ag, Pd, Pt) nanocrystal modified MoS_2 nanosheets and the enhanced catalytic activity of Pd-MoS_2 for methanol oxidation. Nanoscale 6:5762–5769
185. Lukowski MA et al (2013) Enhanced hydrogen evolution catalysis from chemically exfoliated metallic MoS_2 nanosheets. J Am Chem Soc 135:10274–10277
186. Voiry D et al (2013) Conducting MoS_2 nanosheets as catalysts for hydrogen evolution reaction. Nano Lett 13:6222–6227
187. Yang J et al (2013) Two-dimensional hybrid nanosheets of tungsten disulfide and reduced graphene oxide as catalysts for enhanced hydrogen evolution. Angew Chem Int Ed 52:13751–13754
188. Yin Z et al (2014) Au nanoparticle-modified MoS_2 nanosheet-based photoelectrochemical cells for water splitting. Small 10:3537–3543
189. Tarascon JM, Armand M (2001) Issues and challenges facing rechargeable lithium batteries. Nature 414:359–367

190. Lian P et al (2010) Large reversible capacity of high quality graphene sheets as an anode material for lithium-ion batteries. Electrochim Acta 55:3909–3914
191. Zhou X, Liu Z (2011) Graphene foam as an anode for high-rate Li-ion batteries. IOP Conf Series: Mater Sci Eng 18:062006
192. Fang Y et al (2013) Two-dimensional mesoporous carbon nanosheets and their derived graphene nanosheets: synthesis and efficient lithium ion storage. J Am Chem Soc 135:1524–1530
193. Zhu X et al (2011) Nanostructured reduced graphene oxide/Fe_2O_3 composite as a high-performance anode material for lithium ion batteries. ACS Nano 5:3333–3338
194. Yin S et al (2011) Assembly of graphene sheets into hierarchical structures for high-performance energy storage. ACS Nano 5:3831–3838
195. Zhou X et al (2012) Self-assembled nanocomposite of silicon nanoparticles encapsulated in graphene through electrostatic attraction for lithium-ion batteries. Adv Energy Mater 2:1086–1090
196. Wang B et al (2013) Adaptable siliconcarbon nanocables sandwiched between reduced graphene oxide sheets as lithium ion battery anodes. ACS Nano 7:1437–1445
197. Wang C et al (2009) Electrochemical properties of graphene paper electrodes used in lithium batteries. Chem Mater 21:2604–2606
198. Ning G et al (2013) Chemical vapor deposition derived flexible graphene paper and its application as high performance anodes for lithium rechargeable batteries. J Mater Chem A 1:408–414
199. Li N et al (2012) Flexible graphene-based lithium ion batteries with ultrafast charge and discharge rates. Proc Natl Acad Sci USA 109:17360–17365
200. Yang S et al (2013) Bottom-up approach toward single-crystalline VO_2-graphene ribbons as cathodes for ultrafast lithium storage. Nano Lett 13:1596–1601
201. J-w Seo et al (2007) Two-dimensional nanosheet crystals. Angew Chem Int Ed 46:8828–8831
202. Du G et al (2010) Superior stability and high capacity of restacked molybdenum disulfide as anode material for lithium ion batteries. Chem Commun 46:1106–1108
203. Hu S et al (2014) Preparation of carbon coated MoS_2 flower-like nanostructure with self-assembled nanosheets as high-performance lithium-ion battery anodes. J Mater Chem A 2:7862–7872
204. Kong D et al (2014) Rational design of MoS_2@graphene nanocables: towards high performance electrode materials for lithium ion batteries. Energy Environ Sci 7:3320–3325
205. Chen R et al (2014) Free-standing hierarchically sandwich-type tungsten disulfide nanotubes/graphene anode for lithium-ion batteries. Nano Lett 14:5899–5904
206. Hwang H et al (2011) MoS_2 nanoplates consisting of disordered graphene-like layers for high rate lithium battery anode materials. Nano Lett 11:4826–4830
207. Chang K, Chen W (2011) In situ synthesis of MoS_2/Graphene nanosheet composites with extraordinarily high electrochemical performance for lithium ion batteries. Chem Commun 47:4252–4254
208. Xiao J et al (2010) Exfoliated MoS_2 nanocomposite as an anode material for lithium ion batteries. Chem Mater 22:4522–4524
209. Yang L et al (2013) Hierarchical MoS_2/Polyaniline nanowires with excellent electrochemical performance for lithium-ion batteries. Adv Mater 25:1180–1184
210. Shi Y et al (2013) Self-assembly of hierarchical MoSx/CNT nanocomposites (2, x,3): towards high performance anode materials for lithium ion batteries. Sci Rep 3:2169–2176
211. David L et al (2014) MoS_2/Graphene composite paper for sodium-ion battery electrodes. ACS Nano 8:1759–1770
212. Park J et al (2013) Discharge mechanism of MoS_2 for sodium ion battery: Electrochemical measurements and characterization. Electrochim Acta 92:427–432
213. Zhu C et al (2014) Single-layered ultrasmall nanoplates of mos_2 embedded in carbon nanofibers with excellent electrochemical performance for lithium and sodium storage. Angew Chem Int Ed 53:2152–2156

214. Simon P, Gogotsi Y (2008) Materials for electrochemical capacitors. Nat Mater 7:845–854
215. Conway BE et al (1997) The role and utilization of pseudocapacitance for energy storage by supercapacitors. J Power Sour 66:1–14
216. Ratha S, Rout CS (2013) Supercapacitor electrodes based on layered tungsten disulfide-reduced graphene oxide hybrids synthesized by a facile hydrothermal method. ACS Appl Mater Interfaces 5:11427–11433
217. Ma G et al (2013) In situ intercalative polymerization of pyrrole in graphene analogue of MoS_2 as advanced electrode material in supercapacitor. J Power Sour 229:72–78
218. Jeong HM et al (2011) Nitrogen-doped graphene for high-performance ultracapacitors and the importance of nitrogen-doped sites at basal planes. Nano Lett 11:2472–2477
219. Zhu Y et al (2011) Carbon-based supercapacitors produced by activation of graphene. Science 332:1537–1541
220. Cao L et al (2013) Direct laser-patterned micro-supercapacitors from paintable MoS_2 films. Small 9:2905–2910
221. Xia J et al (2009) Measurement of the quantum capacitance of graphene. Nat Nanotech 4:505–509
222. Zhu Y et al (2010) Exfoliation of graphite oxide in propylene carbonate and thermal reduction of the resulting graphene oxide platelets. ACS Nano 4:1227–1233
223. Ning G et al (2011) Gram-scale synthesis of nanomesh graphene with high surface area and its application in supercapacitor electrodes. Chem Commun 47:5976–5978
224. Lee JH et al (2013) Restacking-inhibited 3D reduced graphene oxide for high performance supercapacitor electrodes. ACS Nano 7:9366–9374
225. Choi BG et al (2012) 3D macroporous graphene frameworks for supercapacitors with high energy and power densities. ACS Nano 6:4020–4028
226. Gao W et al (2011) Direct laser writing of micro-supercapacitors on hydrated graphite oxide films. Nat Nanotech 6:496–500
227. Zang X et al (2014) Highly flexible and adaptable, all-solid-state supercapacitors based on graphene woven-fabric film electrodes. Small 10:2583–2588
228. Dong X-C et al (2012) 3D graphene-cobalt oxide electrode for high-performance supercapacitor and enzymeless glucose detection. ACS Nano 6:3206–3213
229. Wu Z-S et al (2010) Anchoring hydrous RuO_2 on graphene sheets for high-performance electrochemical capacitors. Adv Funct Mater 20:3595–3602
230. Xue M et al (2012) Structure-based enhanced capacitance. in situ growth of highly ordered polyaniline nanorods on reduced graphene oxide patterns. Adv Funct Mater 22:1284–1290
231. Soon JM, Loh KP (2007) Electrochemical double-layer capacitance of MoS_2 nanowall films. Electrochem Solid-State Lett 10:A250–A254