博士論文

Operation Mechanisms and Intrinsic Performances of

Non-volatile Memory Devices

Based on Two-dimensional van der Waals Heterostructures

（二次元ファンデルワールスヘテロ構造からなる

不揮発性メモリデバイスの動作機構と本質的なデバイス性能に関する研究）

佐々木 太郎

Abstract

It is no doubt that our society has been strongly supported by the remarkable progress of Si-based electronic devices. On the contrary, the emergence of new kinds of applications such as neuromorphic computing, and the predicted arrival of geometrical and economical limit of the device scaling have required the alternative strategy for future fertile society. In this context, two-dimensional (2D) materials have been increasing their prominence as a candidate for next-generation electronic material. As well as its field-effect-transistor (FET) applications, its non-volatile memory (NVM) applications have been aggressively studied. With the help of dry transfer technique, NVM devices based on 2D van der Waals heterostructures (2D heterostructured NVM devices) have been realized, and considered as promising candidates of next-generation NVM, since its dangling-bond free interfaces have potential to realize the interface trapped charge free NVM devices. In 2D heterostructured NVM devices, typically, transition metal dichalcogenides (TMDCs) are often used as channel materials, while hexagonal boron nitride (*h*-BN) and graphite or graphene are used as a tunneling barrier and floating gate (FG), respectively. In addition, because of the increasing attention of 2D materials-based FETs, future material compatibility can also be expected for 2D heterostructured NVM device. However, even though a lot of progress has been achieved from graphene channel device with non-2D materials to 2D heterostructured device, it has not been sufficiently supported by the understandings of their operation mechanisms, preventing further improvement and control of their performances as desired. The lack of understandings also leads to the invalid evaluation of their performances, that is, *intrinsic* performances of them are still unclear. Therefore, the objective of this study is to reveal the comprehensive operation mechanisms and their intrinsic performances. The study is conducted in bottom-up approach as follows.

First of all, new measurement technique called *floating gate voltage (V_{FG}) measurement* is proposed, for unveiling the operation mechanisms behind standard

current–voltage (*I–V*) round sweep curves. Since V_{FG} measurement can be performed by just adding another electrode onto the FG, it is applicable to any 2D heterostructured NVM devices. Measured V_{FG} during *I–V* round sweep called *V_{FG} trajectory* can be explained by two regions, which are capacitive coupling region with tilted V_{FG} and feedback region with pinned V_{FG}. The principle has been confirmed by experimental results.

Next, for reasonable performance evaluation, the validity of memory window extraction method is investigated since there are two different extraction methods which strongly depend on the research field. While *I–V* round sweep is widely used in 2D research field, *I–V* single sweep after program and erase (P/E) operation is used in Si research field. Consequently, V_{FG} trajectory-based analysis has revealed that the memory window extracted by *I–V* round sweep is often overestimated. The criterion for the overestimation by *I–V* round sweep is also derived, and it is found that the criterion is easy to be satisfied. Therefore, as well as Si research field, *I–V* single sweep should be used for the extraction of memory window even in 2D research field.

Then, the V_{FG} measurement is applied for three kinds of 2D heterostructured NVM devices. Since bandgap of 2D channel material and metal/2D interface are considered as the keys of device operation, three different materials (MoS_2, WSe_2, $MoTe_2$) are used as the channels. On the other hand, all three devices have *h*-BN tunneling barrier and graphite FG. Interestingly, each device has inherent V_{FG} trajectory while their *I–V* round sweep curves are very similar as well as previous studies. By analyzing the trajectories, comprehensive understandings of their operation mechanism can be revealed, in terms of three tunneling current limiting paths. Moreover, the validity of understandings is confirmed by experimental controls of V_{FG} trajectory where the temperature, device structure, and sweeping rate are varied. Tunneling between source/drain (S/D) metal electrode and FG are experimentally proved for the first time, while previous studies have claimed that the 2D heterostructure was tunneling path.

Finally, intrinsic performances of 2D heterostructured NVM device is reasonably investigated as mainly focused on the P/E speed. Remarkably, 50 ns ultra-fast P/E

operation can be achieved by appropriately designed devices. To minimize the size of FG pad is the key for preventing the large tunneling current during fast P/E operation, which leads the *h*-BN breakdown. Although the superior 2D/2D interface seems to be the key, the controlled experiment revealed it is not the key. On the other hand, as compared with conventional SiO_2, *h*-BN possesses stronger breakdown strength under high-speed voltage pulse stress, suggesting that the larger tunneling current can be allowed for *h*-BN. Importantly, the ultra-fast nature of 2D heterostructured NVM device is in the range of storage class memory, which is now strongly desired. Retention and endurance characteristics are also investigated, and 10^5 s retention and 5×10^4 P/E cycles endurance can be achieved for the MoS_2 device with access region. Since the MoS_2 device has large FG pad for V_{FG} and tunneling current measurements, *h*-BN may be severely degraded due to large tunneling current as well as the speed test, suggesting that further improvement can be expected by the device with appropriately designed FG pad.

As the conclusion of this study, 2D heterostructured NVM device with graphite contact is proposed as a promising device structure. Although 2D/2D interface is not the key for the ultra-fast operation, it will be the key of great reliability.

Thesis supervisor: Kosuke Nagashio

Title: Professor of Materials Engineering

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Prof. Kosuke Nagashio for the exciting discussions and clear guidance throughout my doctoral course. The greatest fortunate in my doctoral course is that, Nagashio-sensei treated my curious result, V_{FG} *trajectory* as an interesting and meaningful one. Without this encouragement, the comprehensive understandings of operation mechanism had never been seen in the world. In addition, all of discussions with Nagashio-sensei are very exciting. Even if I was disappointed due to my poor results, Nagashio-sensei always gave me an alternative perspective which re-activates my motivation. These experiences must be a guideline of my future carrier.

I would like to express my gratitude to Prof. Ken Uchida, Prof. Koji Kita, Prof. Tomoki Machida, and Prof. Masaharu Kobayashi, for insightful comments on my doctoral dissertation committee. By their help, I can gain important perspectives and drastically improved the discussions in this thesis.

I would also like to thank Dr. Tomonori Nishimura for his help of my experiments and a lot of significant comments. Whenever I encountered the trouble in my experiments, Nishimura-san kindly helped me by not only trouble shootings, but also suggestions of alternatives. At the weekly meeting, Nishimura-san precisely pointed out the quite important things which I missed until finishing the presentation, which drastically improved my experiments and discussions.

I would also like to offer my thanks to Prof. Keiji Ueno (Saitama Univ.) for providing us high-quality TMDC crystals, Dr. Takashi Taniguchi and Dr. Kenji Watanabe (NIMS) for providing us high-quality *h*-BN crystals, and Dr. Yukinori Morita and Dr. Yongxun Liu (AIST) for meaningful suggestions at the early stage of this study. I would like to offer my special thanks to Masaaki Niwa, Senior Fellow (UTokyo), for his kind attention throughout my doctoral course.

In Nagashio lab., I would also like to thank Mrs. Kyoko Ogawa for her support of our daily life in the lab. I would also like to thank Drs. Nan Fang, Teerayut Uwanno, and Naoki Higashitarumizu for their fruitful help and suggestions in my experiments. I would also like to thank Dr. Haruki Uchiyama for giving me his new and interesting insights. I really thank Yih-Ren Chang, Supawan Ngamprapawat, Satoshi Toyoda, Kohei Maruyama, Keigo Nakamura, Hayami Kawamoto, Masaya Umeda, Yuichiro Sato, Masahiro Kobayashi, Shuhong Li, Roichi Kato, Wataru Nishiyama, Tomohiro Fukui, Ryuichi Nakajima, Hiryu Kozaki, and Ryo Nanae for significant discussions and fun memories. Thanks to all of the members, my doctoral course at the University of Tokyo become very enjoyable.

I would also like to thank the administrative staffs in the department of materials engineering for their support on my daily life in the department. I would like to extend my gratitude to my friends for their personal support, which makes my life fertile.

Finally, I want to give sincere thanks to my parents and family for their long support of my education. Thanks to the support, I can finally come this far.

Citations to Previously Published Works

Chapter 2, 3, and 4 in this thesis are based on the following publications:

(1) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Understanding the Memory Window Overestimation of 2D Materials Based Floating Gate Type Memory Devices by Measuring Floating Gate Voltage. *Small* **2020**, *16* (47), 2004907. (https://doi.org/10.1002/smll.202004907)

(2) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Material and Device Structure Designs for 2D Memory Devices Based on the Floating Gate Voltage Trajectory. *ACS Nano* **2021**, *15* (4), 6658–6668. (https://doi.org/10.1021/acsnano.0c10005)

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendix

List of Tables

Chapter 1
Introduction

1.1 Non-volatile Memory Technology

Non-volatile memory (NVM) is one of the most important technologies in our society. Today, people use a smartphone, personal computer (PC), tablet *etc.* in their daily life. These computing systems must employ a non-volatile memory to store the huge amount of data without power supply. Thanks to its non-volatility, we can save the data and shut down our PC whenever we want, and we do not have to worry about our smartphone running out of battery with data loss. In addition, the popularization of internet and great progress of hardware technologies have resulted in the data driven society. As shown in **Fig. 1-1**, International Data Corporation reported the exponential expansion of global data in 2018[1]. This means that, the importance of NVM technology will continue to increase in the future.

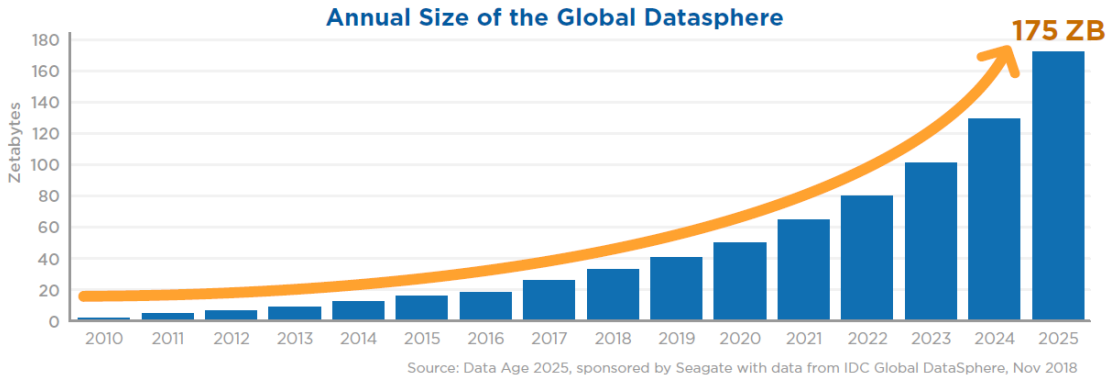


Fig. 1-1 Exponential expansion of global data[1].

At the beginning of information and communication technology (ICT) era, hard disk drive (HDD) had played an important role. Due to its huge capacity, HDD is often treated as a *storage* rather than a *memory*. In case of an HDD, ferromagnetic materials are significant to store the binary data "0" and "1" which correspond to poles of magnet. Magnetization of these materials can be electrically controlled[2], and has an immunity against an environment, resulting in non-volatility of HDD. In other words, *spin* of electron

is a key property for non-volatility.

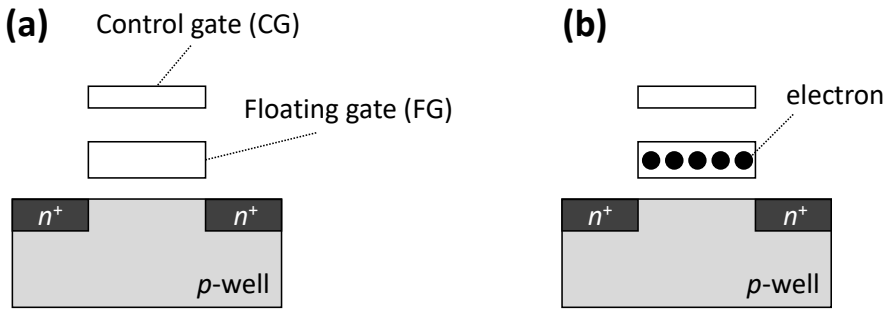


Fig. 1-2 Schematic of a flash memory cell with **(a)** empty FG and **(b)** filled FG.

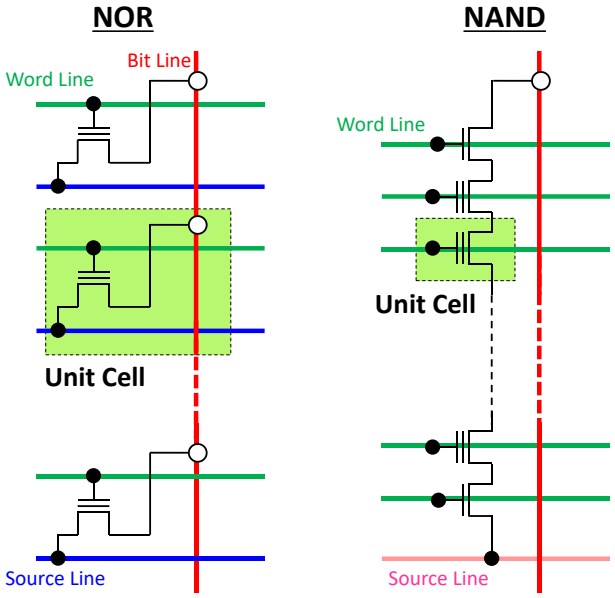


Fig. 1-3 NOR and NAND array configuration of flash memory cells[4].

In 1984, F. Masuoka et al. has developed epoch-making non-volatile memory device called *flash memory*[3]. Schematic of flash memory is shown in **Fig. 1-2**. Charge trapping layer called floating gate (FG) is inserted between a control gate (CG) and channel. By applying the large positive (negative) voltage to the CG, electrons can be injected to (ejected from) the FG. Depending on a state of FG, that is, whether filled by electrons or empty, the device has two threshold voltages, corresponding to binary "0" and "1", respectively. In addition, NAND type array configuration has realized the large-capacity semiconductor non-volatile memory, whereas NOR type array configuration has been used as an instruction storage memory for a processing unit. Both array configurations are

illustrated in **Fig. 1-3**[4]. Thanks to the large capacity of NAND flash memory, it has created new market such as USB drive and MP3 player instead of tape media[5]. Over the past 40 years, NAND flash memory technology has been drastically updated. The details of these improvement and challenges are presented in next section. For the flash memory technology, *charge* of electron is a key property for non-volatility.

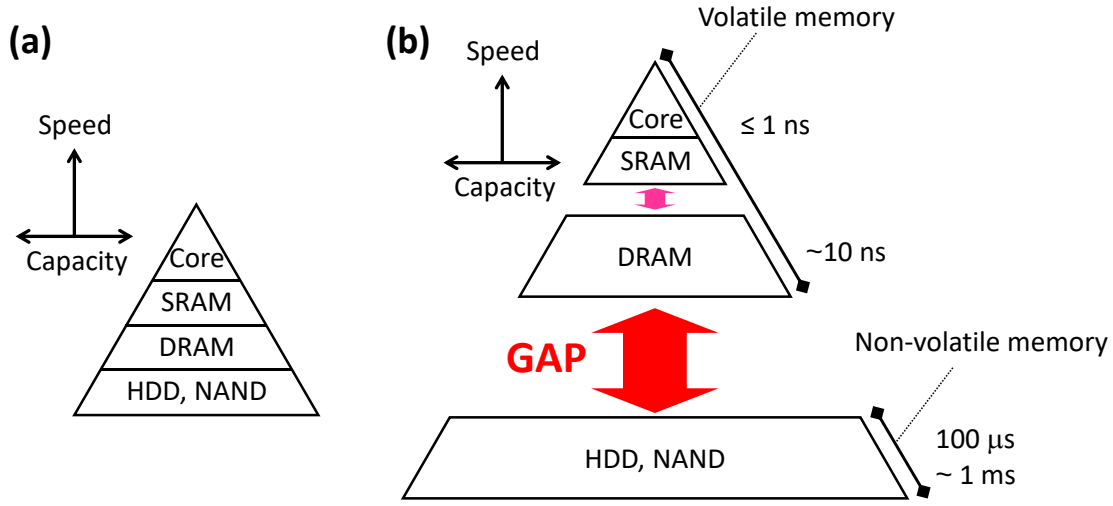


Fig. 1-4 Memory hierarchy in computing system at **(a)** early stage of ICT era, and **(b)** current.

Of course, volatile memories such as static random access memory (SRAM) and dynamic RAM (DRAM) have also played an important role in computing system. Since the volatile memories have lower capacity but higher speed than HDD and flash memory, these are used as a cache memory or working memory. At the early stage of ICT era, as illustrated in **Fig. 1-4**, non-volatile memories (NAND flash, HDD) and volatile memories (SRAM, DRAM) had well cooperated for an efficient computing. However, drastic improvements of each technology have resulted in the huge speed and capacity gap between memories. In addition, scaling of the metal-oxide-semiconductor field effect transistor (MOSFET) have led unexpected performance degradations called short channel effects (SCEs)[6]. One of SCEs is the increase of leakage current of MOSFET, which increases static power of SRAM and DRAM since MOSFET is a key component of them. For the purpose of filling the gap and achieving energy efficient computing, novel non-volatile memory technologies have been emerged recently. At this stage, not only *spin* and

charge of electron, but also any physics can be a key component of non-volatility. For example, phase change RAM (PCRAM) utilizes the *phase change* between amorphous phase and crystal phase of the material[7], and resistive RAM (ReRAM) often utilizes the *soft breakdown* of the metal-insulator-metal structure which can be controlled by applying voltage[8]. Especially, the ReRAM whose filament is formed by metal ion is often referred to as conductive bridge RAM (CBRAM). Whereas spin-transfer-torque magnetic RAM (STT-MRAM) utilizes the *spin* of electron as a key physics as well as HDD, it is manipulated in a different way. The new storage device called magnetic tunnel junction (MTJ) is employed, and spin-transfer-torque by spin polarized current changes the state of MTJ[9,10]. In addition, ferroelectric RAM utilize the *ferroelectricity* of the material[11]. Typical performances of them are summarized in **Table 1-1**[12]. Although some of emerging memories are in the market, flash memory still maintains its dominant position as solid-state drives (SSDs) or embedded storage in mobile applications[4]. Important point is that, to fill the gap with non-volatility. We can take any option to achieve it. No one knows what will be an essential non-volatile technology.

Table 1-1. Typical FoM values and market readiness for established and emerging memory technologies[12].

FoM	SRAM	DRAM	Flash NAND (planer)	ReRAM	FeRAM	PCM	STT-MRAM
Density [bytes per chip]	≈10 MB	1–10 GB	≈10 GB	≈1 GB	≈1 MB	1–10 GB	10–100 MB
Cell size [F^2]	>100	6–10	4–5	6–20	15–40	6–20	35–40
Write time	<10 ns	≈10 ns	≈100 ms	10–100 ns	≈100 ns	10–100 ns	≈10 ns
Program energy per bit	1–10 pJ	1–10 pJ	≈10 nJ	≈10 pJ	1 pJ	0.1–1 nJ	<1 pJ
Retention	Volatile	Volatile 10–100 ms	Nonvolatile >10 years	Nonvolatile >10 years	Nonvolatile >10 years	Nonvolatile >10 years	Nonvolatile >10 years
Endurance	$>10^{15}$	$>10^{15}$	10^2–10^5	10^6–10^9	$>10^{15}$	10^6–10^9	$>10^{15}$
Maturity	Product	Product	Product	Early product	Product	Early product	Early product
Market price [$ per GB]	10–100 k	≈10	≈1	≈1 k	10–100 k	10–100	1–10 k

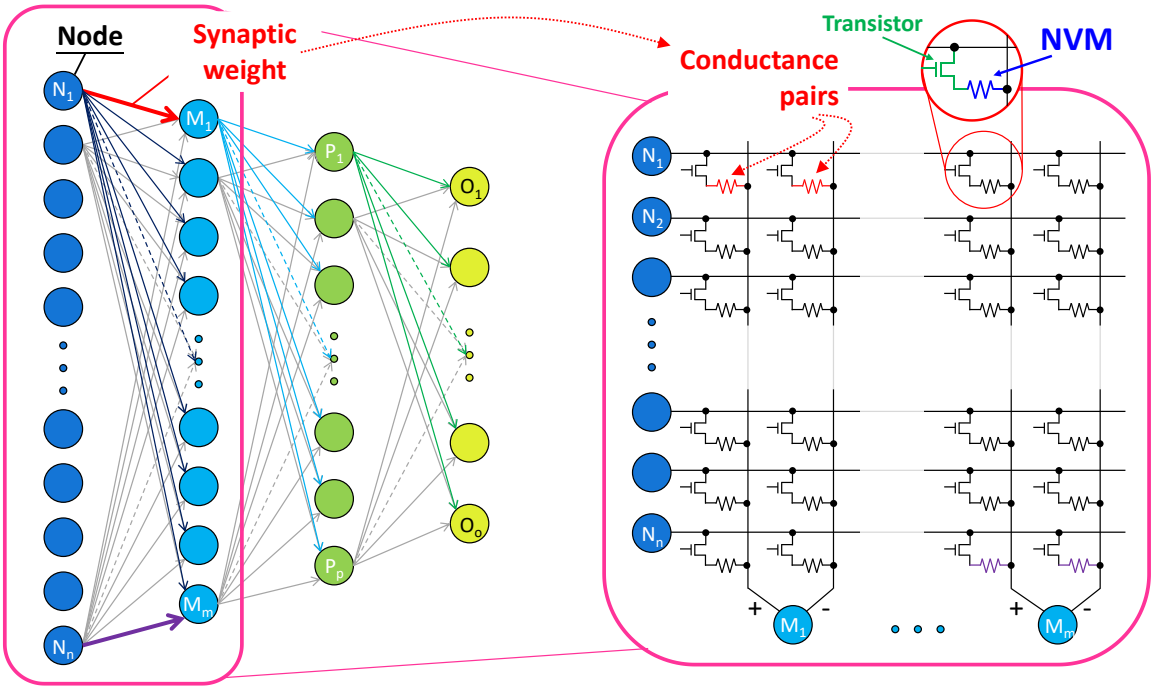


Fig. 1-5 Schematic of typical neural network and its hardware implementation by using non-volatile memory devices[14].

For the future, expansion of the territory of non-volatile memory is easy to be predicted, since many exciting applications have been emerged beyond the classical PCs. One interesting application is the neuromorphic computing. In 2012, great success of deep learning was announced in terms of object recognition[13], which had been strongly supported by the progresses of hardware including memory devices, and the popularization of internet. Deep learning utilizes the neural network which mimic the human brain. Typical neural network and its conceptual hardware implantation is illustrated in **Fig. 1-5**[14] where a neuron corresponds to a node, and synapse which is a connection point between neurons, corresponds to the weighted edge. As shown in the figure, NVM devices are required to store the weight. The purpose of their non-volatility is to save energy in neuromorphic computing. Here, each node performs the product-sum calculation as follows:

$$y_i = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n, \tag{1.1}$$

where, y_i is a calculated result of i-th node in current layer, x_n represents the input from n-

th node in previous layer, and w_n represents the synaptic weight. Therefore, neural network can be implemented by the NVM array since it can perform following calculation:

$$I_i = G_1V_1 + G_2V_2 + \cdots + G_nV_n, \quad (1.2)$$

where, I_i is current as a calculated result of i-th node, G_n represents the conductance of each NVM device as a synaptic weight, and V_n represents the voltage as a input from previous layer. Since machine learning technique including deep learning has versatility, it has been utilized everywhere, and its application will be further expanded, requiring the further improvement of non-volatile memory devices.

1.2 Si-based Flash Memory

In this section, Si-based flash memory technology is overviewed by focusing on a memory cell. Flash memory, which is most popular semiconductor non-volatile memory, has changed the world as mentioned above. Tremendous efforts had been devoted to miniaturize the planer cell for realizing a low bit cost memory chip. Although planer cell finally reached the scaling limit, three dimensional (3D) NAND array which pave the alternative way for lowering the bit cost was proposed[15]. Although 3D NAND is in the mainstream of the recent market, its several challenges have been pointed out[16].

1.2.1 Planer Cell

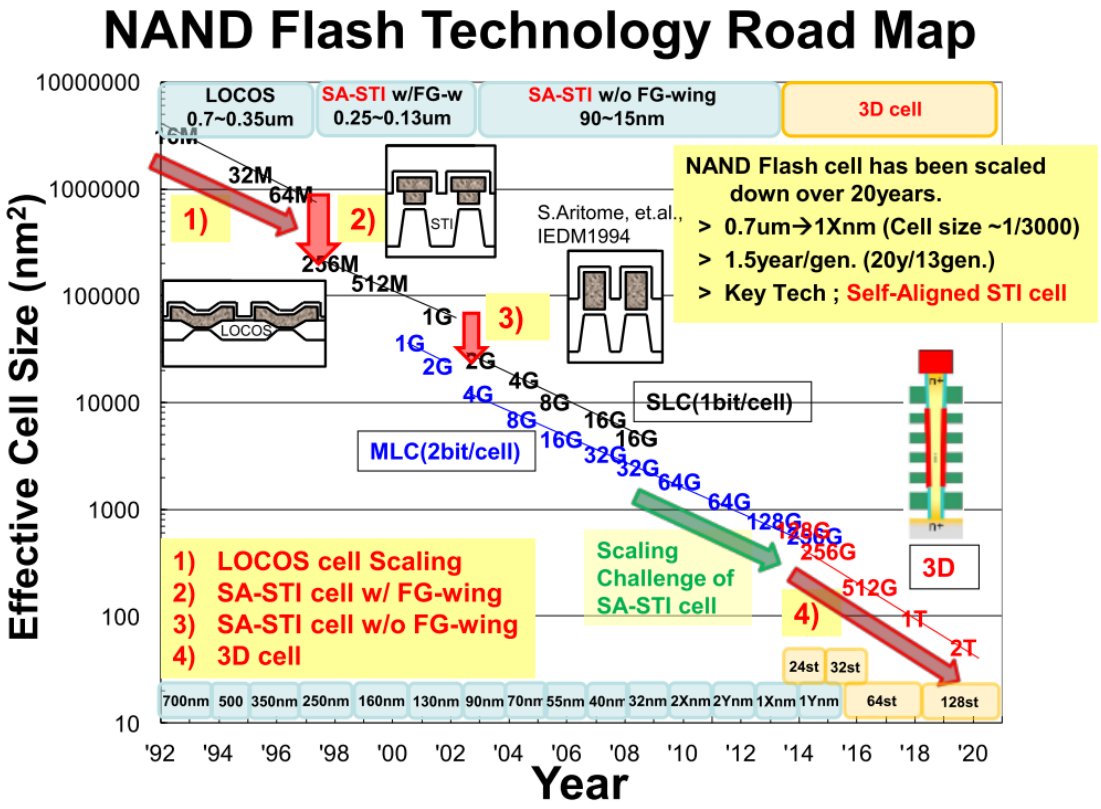


Fig. 1-6 Comprehensive roadmap of NAND flash technology[5].

Since NAND flash memory was on the market in 1992, planer cell had been utilized as a memory cell for a long time. Comprehensive scaling roadmap of NAND flash memory technology by S. Aritome is shown in **Fig. 1-6**[5]. According to the figure, planer cell

technology had been used until the middle of 2010s. In this excellent miniaturization for low bit cost, there are two important requirements. One is the isolation between neighboring cells, and the other is to ensure the capacitive coupling ratio between FG and CG.

Since flash memory cell is operated in a memory array, isolation between neighboring cells is very important. Whereas local oxidation of silicon (LOCOS) was used for isolation at the early stage of production (LOCOS cell in **Fig. 1-6**), shallow trench isolation (STI) replaced the role of LOCOS at the end of 1990s (SA-STI cell in **Fig. 1-6** where SA stands for *self-aligned*).

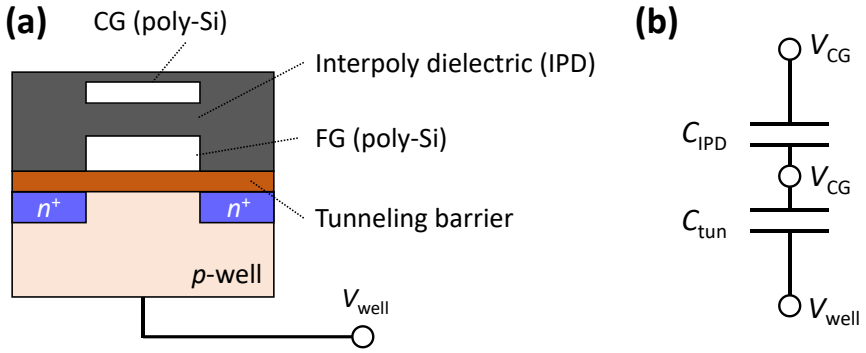


Fig. 1-7 (a) Typical planer cell structure and **(b)** its equivalent circuit.

Capacitive coupling ratio is related to the program and erase (P/E) operation. As schematically shown in **Fig. 1-7(a)**, typical planer cell has FG FET structure, where poly-Si is usually used as CG and FG. Here, insulator between FG and channel is called tunneling barrier or tunneling oxide, while the insulator between CG and FG is usually called interpoly dielectric (IPD). Therefore, as shown in **Fig. 1-7(b)**, gate stack of planer cell equivalently represented by the serial connection of these capacitances, where C_{IPD} and C_{tun} are the capacitance of IPD and tunneling barrier, respectively. When V_{CG} is applied to the CG and *p*-well is grounded (V_{well} = 0 V), potential of FG (V_{FG}) can be determined by following equation:

$$V_{\mathrm{FG}} = \frac{C_{\mathrm{IPD}}}{C_{\mathrm{IPD}} + C_{\mathrm{tun}}} V_{\mathrm{CG}}. \tag{1.3}$$

Since large V_{FG} is required for tunneling between the FG and channel, large capacitive

coupling ratio is helpful for low voltage operation. Its typical value is 0.6[17]. For this purpose, as shown in **Fig. 1-8(a)**, FG is capped by IPD and CG for SA-STI cell with FG-wing. Due to the requirement for further miniaturization, it was updated to SA-STI cell without FG-wing as shown in **Fig. 1-8(b)**. The active area of C_{poly}, highlighted by blue is larger than that of C_{tun}, highlighted by red, to ensure the coupling ratio.

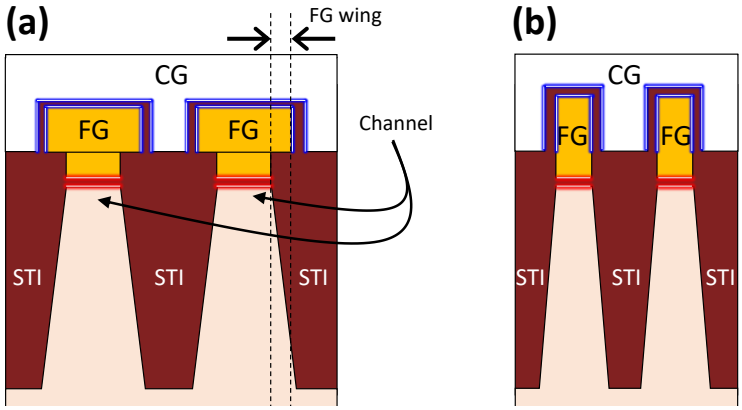


Fig. 1-8 Cross-sectional view of SA-STI cell **(a)** with and **(b)** without FG wing[5].

Although low bit cost is a first priority of memory, single cell-level performance should also be focused on. One important metric is a memory window defined by the difference of threshold voltages. For the further reduction of bit cost, multi-level cell (MLC) technology, i.e., more than one bit are represented by one cell, has been developed. Actually, the word *multi-level cell* often represents the 2 bits/cell, while triple-level cell (TLC), quad-level cell (QLC), and penta-level cell (PLC) represent 3 bits/cell, 4 bits/cell and 5 bits/cell, respectively. **Fig. 1-9** shows one example of V_{th} distributions of multi-level cell[18]. It is clear that the large memory window is required for multi-level cell technology to ensure the appropriate margin between the states.

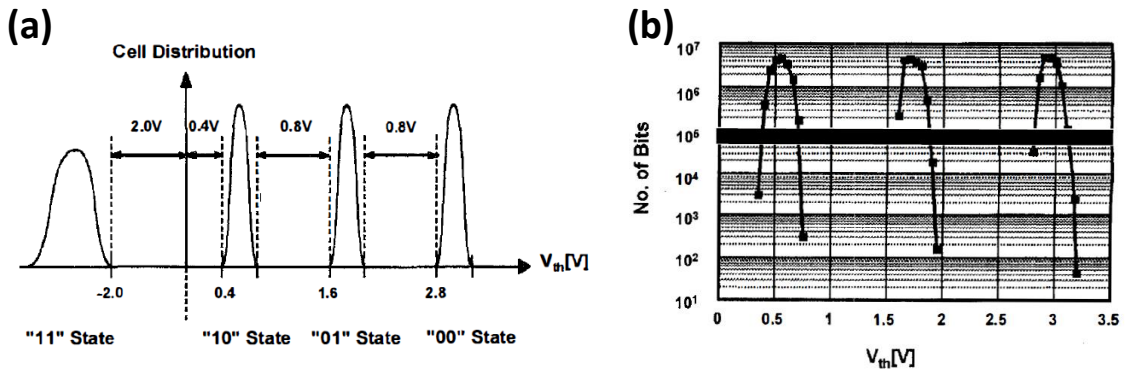


Fig. 1-9 One example of V_{th} distribution of multilevel cell.
(a) Designed and **(b)** measured[18].

P/E speed is also important because it determines the range of applications. Since NAND flash memory has a large capacity while its speed is low (100 μs ~ 1 ms), it is used as a storage rather than working memory. Retention and endurance are related to the reliability of flash memory cell. Retention characteristic represents how robust the V_{th}s of the device against time. Generally, non-volatile memory is required to keep the data for 10 years without power supply. Since 10 years are too long to test, thermal accelerated test is often used for the evaluation[19,20]. Endurance characteristic represents how robust the V_{th}s of the device against P/E cycles. According to the **Table 1-1**, typical value is 10^2~10^5 cycles[12] while it strongly depends on the product.

1.2.2 Scaling Limit of Planer Cell

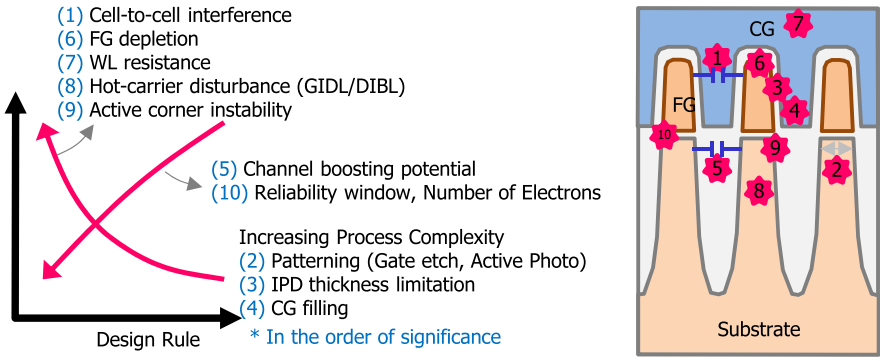


Fig. 1-10 Scaling limitations in planer cell. The number indicates the seriousness. Smaller value means more serious[16].

As well as MOSFET, there are various scaling limitations for planer cell as shown in **Fig. 1-10**[16]. According to the literature, most important issue is that cell-to-cell interference (CTCI). Since the distance between cells is shortened with scaling, neighboring FGs are coupled via parasitic capacitance. Thus, when the FG potential of selected cell is changed by electron injection or ejection, that of neighboring cell is also changed undesirably. As long as the capacitive coupling ratio can be ensured, reducing the FG thickness is one simple solution for CTCI[21].

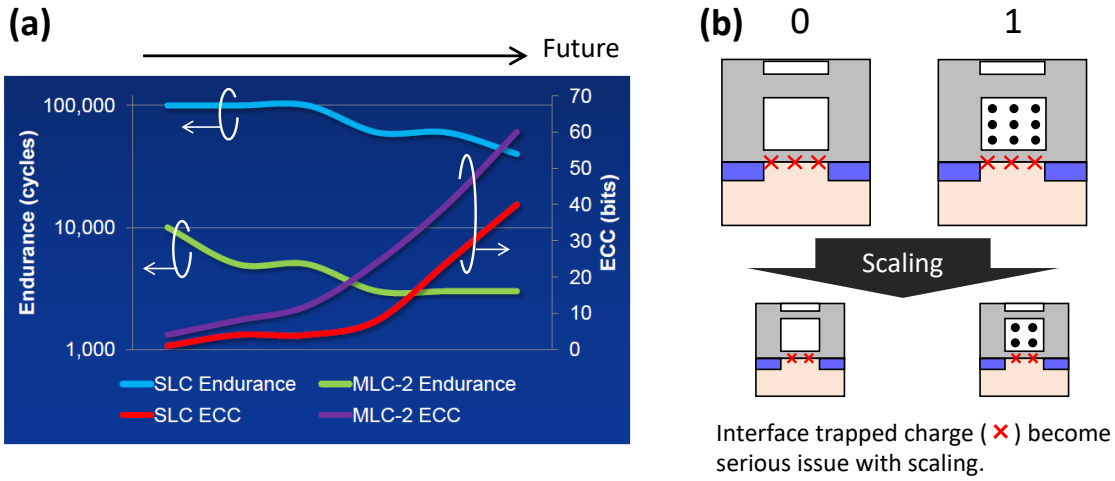


Fig. 1-11 (a) Endurance and ECC bits trends[22]. Corresponding to the degradation of endurance, more ECC bits are required. **(b)** Conceptual illustration of severe effect by the interface trapped charges in a scaled cell.

On the contrary, cell-level reliability issue should also be focused on. As shown in **Fig. 1-11(a)**[22], endurance is predicted to be degraded with scaling. The degradation is more severe for multi-level cell. In addition, as shown in **Fig. 1-11(b)**, the effect of single trapped charge on the V_{th} distribution become severe with the scaling, since the number of charges stored in the FG is decreased. Roughly speaking, the FG is scaled three-dimensionally while the interface is scaled two-dimensionally, resulting in the severe condition for the scaled device. Due to this, "0" and "1" may no longer be distinguished. These *cell-level* reliability issues have been masked with advanced *system-level* techniques such as an error correction by error correction code (ECC) bits and a wear leveling, which is a technique to average the P/E cycles of each cell. Probably, in order to increase the presence of their product on the market, some cell-level reliability issues have been sacrificed instead of lowering bit cost. Although the severe reliability issues shown in **Fig. 1-11(a)** is well known, multi-level scheme is often accepted. Needless to say, however, if the cell-level reliability issues are overcome, the system-level technique can be much simpler. Alternatively, if the system-level techniques are remained, more advanced cell such as TLC, QLC, and PLC cell can be realized.

In case of planer cell, the conflict between the scaling and performance degradation finally reached the insurmountable phase.

1.2.3 Three-dimensional Cell

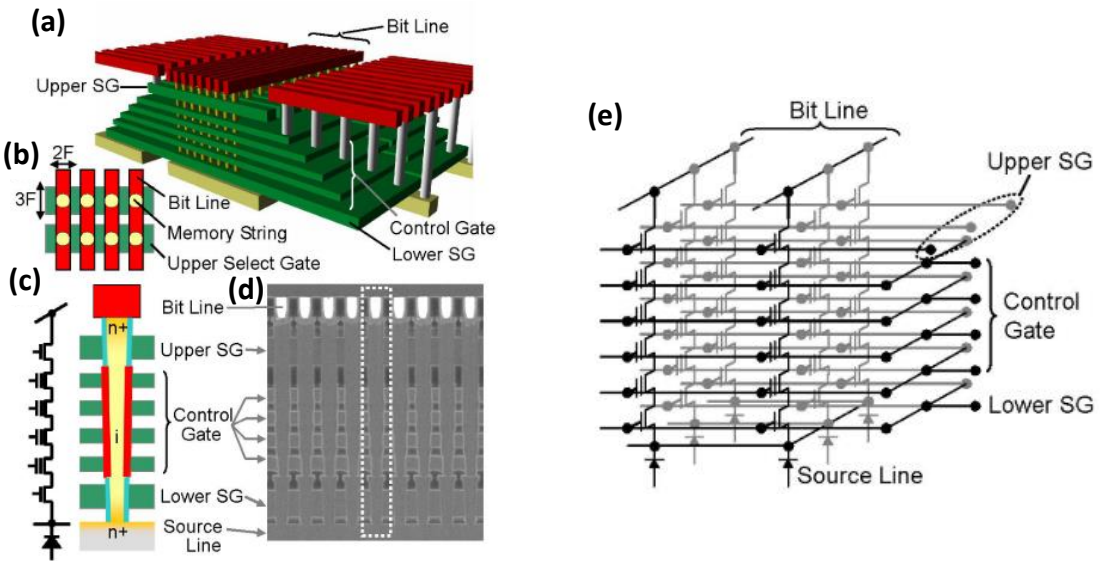


Fig. 1-12 (a) Birds-eye view and **(b)** top-down view of the 3D flash memory. **(c)** Enlarged view of the memory string, and **(d)** cross-sectional SEM image of the 3D flash memory array. **(e)** Equivalent circuit of the 3D flash memory[23].

Paradigm shift was occurred in 2007 by 3D NAND cell array[15,23] shown in **Fig. 1-12**. In 3D NAND flash memory, its capacity can be increased by stacking cell layers, that is, the density of cells can be increased free from lateral scaling. Therefore, design rule can be relaxed, and the space between neighboring cells is easy to be assured. Consequently, various scaling penalty of planer cell could be overcome. One example shown in **Fig. 1-13** which is reported at 2017 IEEE International Memory Workshop (IMW)[24]. In addition, charge trap (CT) type cell[25] is coming into use instead of FG type cell[26], since vertical CT type cells are easy to be fabricated as compared with vertical FG type cell. Typical cross-sectional views of them are shown in **Fig. 1-14**[27]. Nowadays, 3D NAND is in the mainstream, and whether to use CT type or FG type strongly depends on the strategy of each company.

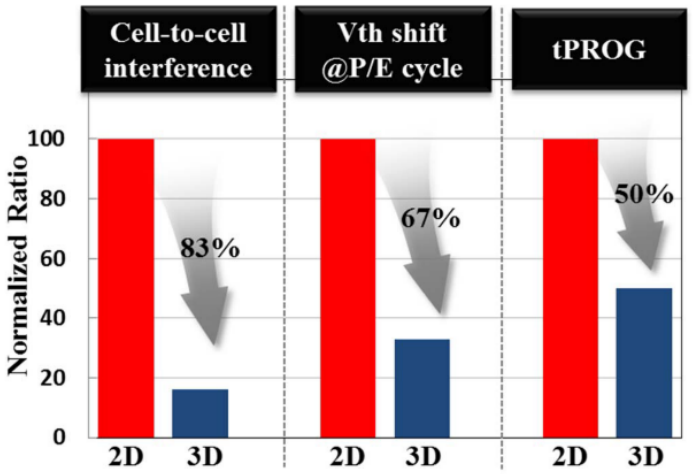


Fig. 1-13 Performance comparison between planer NAND cell and 3D NAND cell[24].

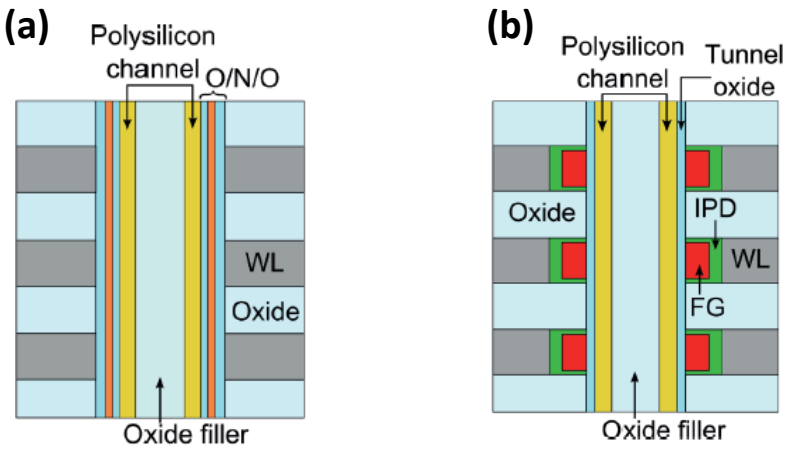


Fig. 1-14 Cross-sectional images of **(a)** CT type and **(b)** FG type 3D NAND flash memory[27]. CT type is sometimes employed for 3D NAND due to its structural simplicity.

1.2.4 Challenges in Three-dimensional Cell

Although some issues which planer cell encountered was temporally overcome, the glory of 3D cell will not last forever. The number of stacking layers has been increased[28], and at present, it has reached 176 layers[29]. However, it must have a limitation. In addition, while lateral scaling (× $1/\alpha$) could increase the density within a fixed area by the square (× α^2), vertical stacking (× β) could increase the density only by the same amount of the multiplication (× β), where α and β are the constants larger than one.

One possible solution is that the scaling of layer thickness. However, decreasing thickness of each layer makes neighboring cells closer, that is, same scaling limitations in planer cell shown in **Figs. 1-10** and **1-11** reappear. If the charge trap layer is cylindrically covered the channel pillar, charge spreading become a serious issue as shown in **Fig. 1-15**[30]. Moreover, poly-Si channel of 3D NAND is often problematic due to its grain boundary. For example, instability of read current has been reported due to the grain boundary[31,32].

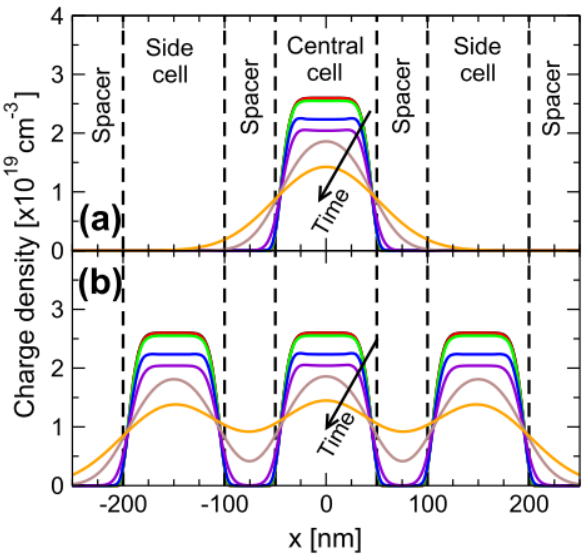


Fig. 1-15 Simulation study of charge spreading in CT type 3D NAND flash memory for **(a)** neutral side cells and **(b)** programmed side cells[30].

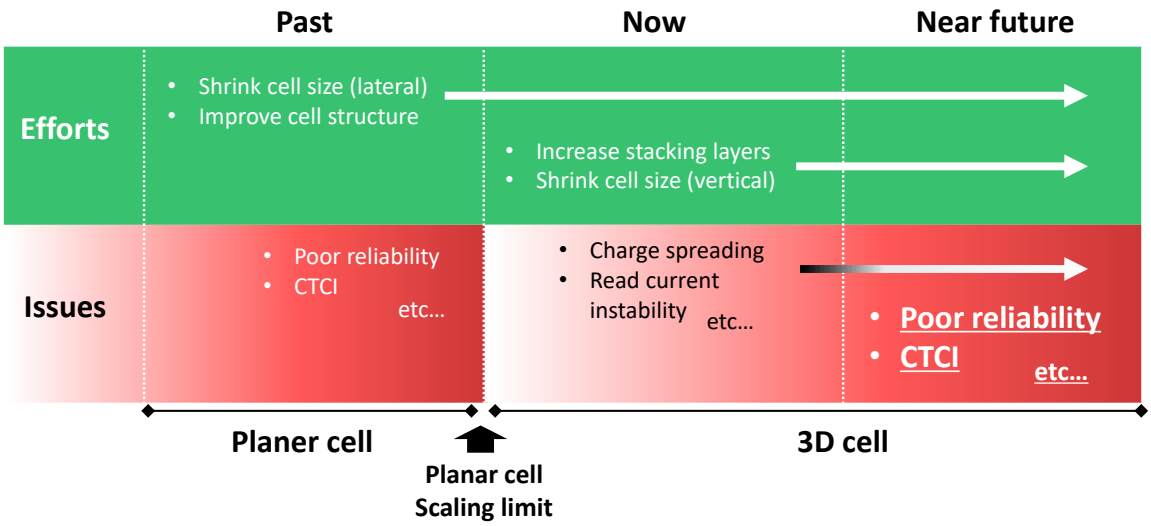


Fig. 1-16 Brief summary of efforts and issues for Si-based flash memories. Although breakthrough was achieved by 3D NAND technology, the glory will not last forever.

Above discussions are summarized in **Fig. 1-16**. Previously, planer cell was in a mainstream of NAND flash memory product. For achieving low bit cost, tremendous efforts, including the change of cell structure, had been devoted to shrink the cell size in lateral dimension. However, many issues including poor reliability (cell-level) and CTCI (array-level) became severe issues with the scaling, and finally reached the scaling limit. Currently, even though the appearance of 3D NAND technology could solve the issues related to the lateral scaling, temporally, other issues resulted from its 3D structure and poly-Si channel have appeared. In addition, other efforts related to vertical direction have been required. In near future, consequently, same problems will be emerged again. Therefore, beyond the conventional Si technology, alternative strategy such as other material systems should be considered.

1.3 2D Materials

2D materials are composed of atomic layers which are bonded by weak van der Waals (vdW) force. Within each layer, strong covalent bonds are formed between neighboring atoms. In 2004, K. S. Novoselov and A. K. Geim et al. reported the experimental characterization of atomically thin graphene flake[33]. The innovative point is that they showed that atomically thin graphene can easily be isolated from bulk crystal by using scotch tape. For this achievement, they won the Nobel Prize in Physics in 2010, being the trigger for the abundant studies of 2D materials. Around 2010 is also a period when the limitation of planer technology was seriously claimed, and new strategy was aggressively explored. Therefore, it is quite natural that the graphene, which has ultra-high carrier mobility (e.g., ~15,000 cm^2 $V^{-1}s^{-1}$ at room temperature for multilayer graphene[33]), was started to be studied as a promising candidate of next generation electronic devices. Although graphene as a channel for MOSFETs has lost researcher's attention due to its zero-bandgap nature, 2D materials such as insulating hexagonal boron nitride (*h*-BN) and semiconducting transition metal dichalcogenides (TMDCs) have attracted much attention as candidates for new electronic/optical devices, since they have unique properties which are not limited to high mobility. Of course, graphene does not lose its importance as a 2D material. There are lots of researches for its novel properties and applications.

In this section, before discussing the 2D materials-based flash memory devices, basic properties and applications of 2D materials are overviewed. In addition, its heterostructure is also discussed.

1.3.1 Graphene

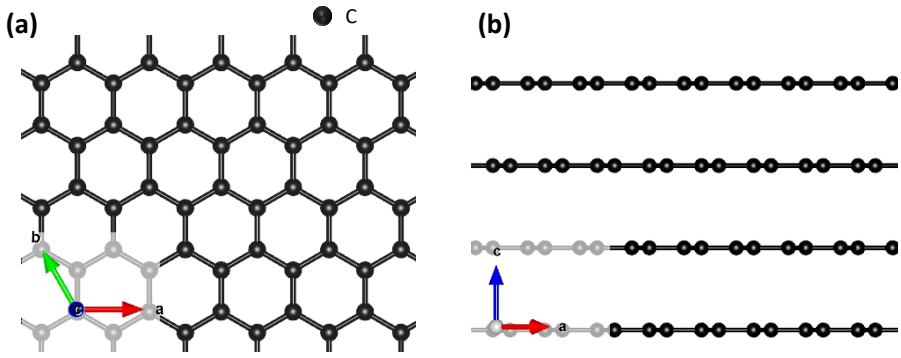


Fig. 1-17 Atomic structure of graphene. **(a)** Top view and **(b)** side view.

Graphene is the most popular 2D materials in the world. As shown in **Fig. 1-17**, monolayer graphene shows a honeycomb structure. $2s$, $2p_x$ and $2p_y$ orbitals of each carbon atom are hybridized to form three sp^2 orbitals, resulting in strong interatomic bonds called σ bond, while remained $2p_z$ orbitals result in weak interatomic bonds called π bond. Since electrons in the $2p_z$ orbitals (π electrons) does not contribute to the covalent σ bond, it can contribute to the electrical conductivity of graphene. Based on the tight-binding approach, the electronic band structure for π electrons in graphene is derived as follows[34]:

$$E(k_x, k_y) = \pm\gamma_0 \sqrt{1 + 4\cos\left(\frac{ak_x}{2}\right)\cos\left(\frac{\sqrt{3}ak_y}{2}\right) + 4\cos^2\left(\frac{ak_x}{2}\right)}, \qquad (1.4)$$

where, E, γ_0, and a are the energy of π electrons, transfer integral between nearest-neighbor carbon atoms, and lattice constant of graphene (0.246 nm), respectively. In addition, $\boldsymbol{k} = (k_x, k_y)$ is the wave vector. The band structure is illustrated in **Fig. 1-18**[34,35]. Here, the point where two cones face each other is called Dirac point. Since Fermi level (E_F) of graphene is located at Dirac point, linear E-k dispersion and corresponding density of states (DOS) shown in **Fig. 1-19**[35] play an important role for possessing many interesting properties.

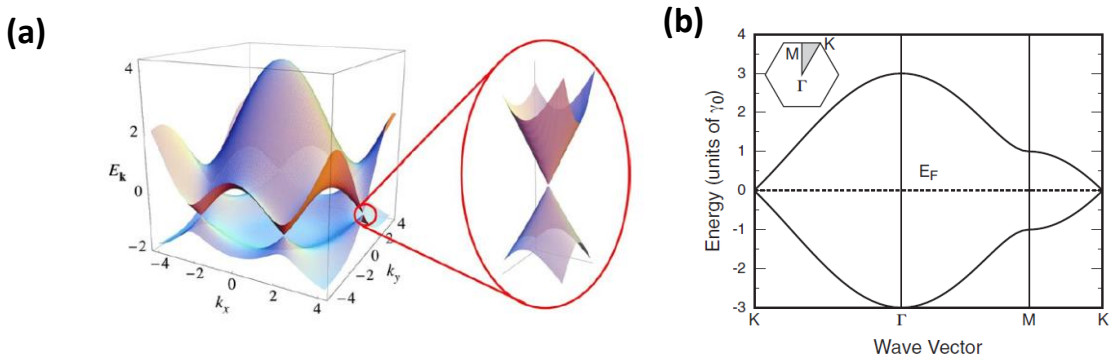


Fig. 1-18 Band structure for π electrons in graphene. **(a)** Three-dimensional view in *k*-space[35] and **(b)** 2D view along the lines with high-symmetry *k*-points[34].

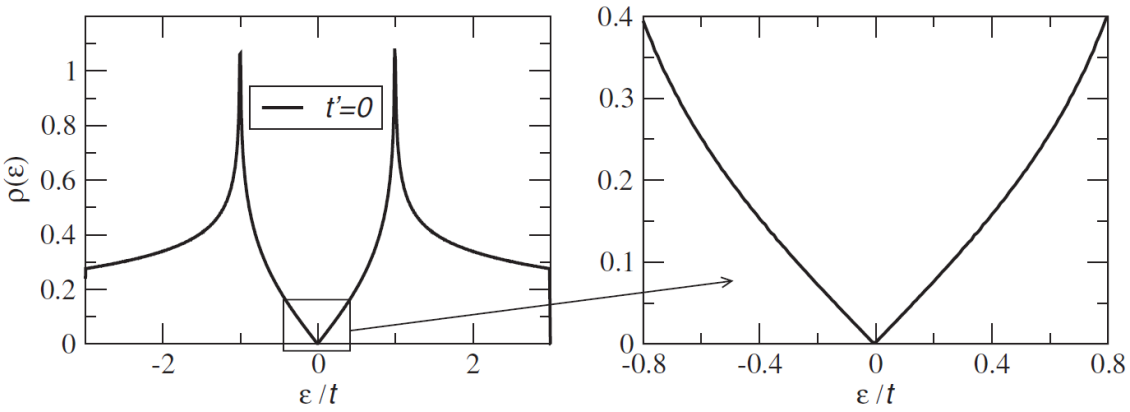


Fig. 1-19 Density of states of graphene as a function of energy of electron when next nearest-neighbor hopping is not considered ($t' = 0$). Energy term is normalized by nearest-neighbor hopping energy t. The details are in the litelature[35].

In terms of FET applications, quantum capacitance (C_Q) is one example of unique property of graphene. Strictly speaking, since C_Q is resulted from low DOS of 2D materials, it is not limited for the graphene but for any 2D materials. When graphene is on the gate stack such as an SiO_2/n^+-Si substrate, the conductivity of graphene can be modulated via the n^+-Si back gate (BG). However, in contrast to 3D materials, graphene cannot screen the electric field from the BG completely due to its low DOS. At this time, E_F of graphene is further modulated to serve sufficient charges in the graphene for screening the electric field, that is, the voltage is dropped across SiO_2 and graphene itself. This can be modeled as the serial connection of two capacitances, C_{ox} for the SiO_2 and C_Q for the graphene. C_Q of graphene has been experimentally investigated[36,37].

At the early stage of 2D research, graphene was treated as a channel material for next generation transistor due to its ultra-high carrier mobility[33,38]. However, since graphene has no bandgap, ON/OFF ratio is not large enough for the channel material. Although the role of graphene was superseded by semiconducting TMDCs as explained below, graphene and graphite are still attracted much attention from many viewpoints such as Cu diffusion barrier[39], vdW contact to TMDC FETs[40], charge trapping layer of memory devices[41–44], and so on.

1.3.2 *h*-BN

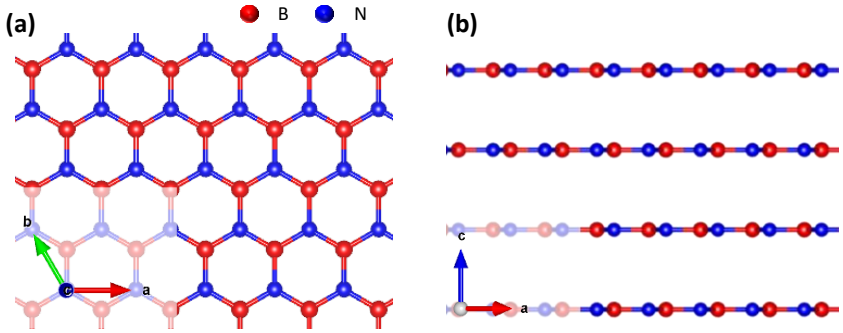


Fig. 1-20 Atomic structure of *h*-BN. **(a)** Top view and **(b)** side view.

***h*-BN** is the most popular 2D insulator with 5.97 eV of band gap[45]. High-quality *h*-BN has been grown by a temperature-gradient method under a high-pressure and high-temperature atmosphere[45]. Crystal structure is very similar to that of the graphene, that is honeycomb structure as shown in **Fig. 1-20**. Because the strong B-N bonding, *h*-BN has a great thermal tolerance up to 850°C[46]. This stable property of *h*-BN allows to be used as an atomically flat substrate[47] and/or passivation film[48]. For example, **Fig. 1-21** clearly shows that the ultimate flatness of *h*-BN results in the uniform charge distribution of graphene[47]. More importantly, since *h*-BN has a SiO_2 comparable dielectric breakdown strength (~12 MV/cm, parallel to c-axis[49]), it is used as a gate dielectric with 2 ~ 4 of relative permittivity[50] for 2D heterostructured electronic devices.

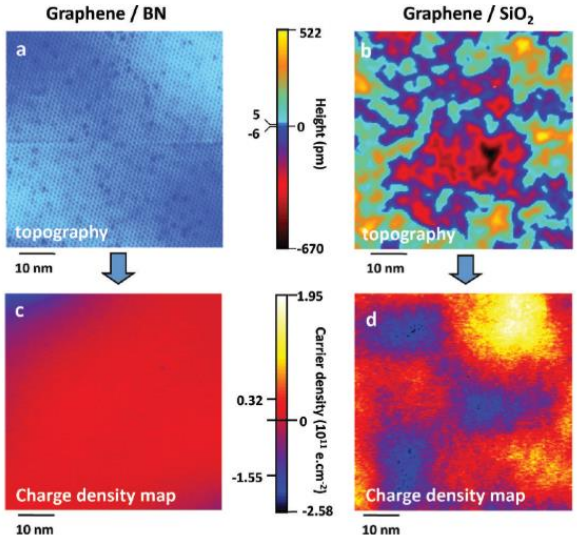


Fig. 1-21 Height and charge density profile of graphene on *h*-BN and on SiO_2 [47].

1.3.3 Transition Metal Dichalcogenides

TMDCs are the famous semiconducting 2D materials. Its generic formula is MX_2, where M represents a transition metal such as molybdenum (Mo) and tungsten (W), and X represents a chalcogen such as sulfur (S), selenium (Se), and tellurium (Te).

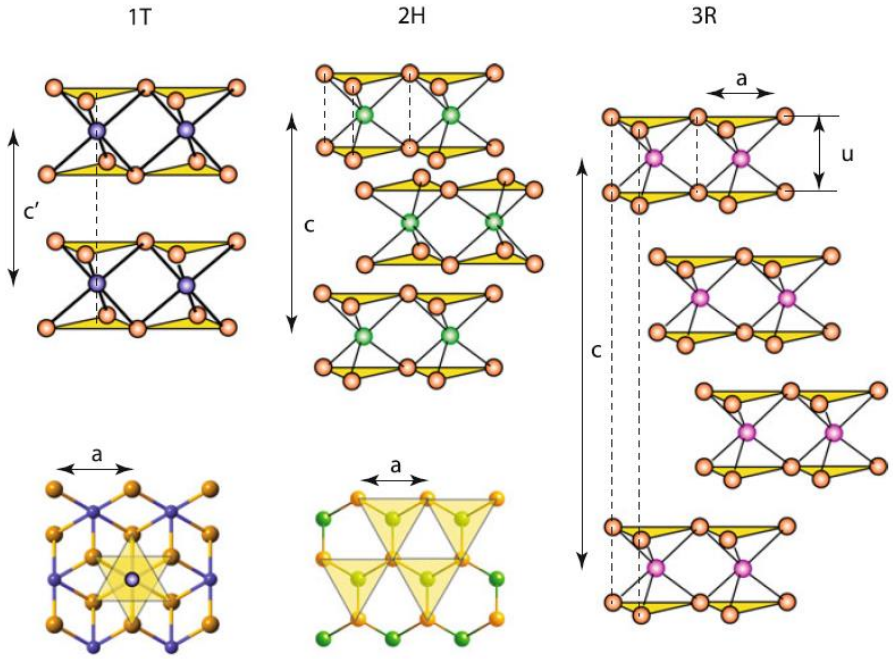


Fig. 1-22 Representative three phases of TMDC crystal[51].

In TMDC crystal, each transition metal atom is covalently bonded with six chalcogen atoms. Representative three phases of TMDC crystal are shown in **Fig. 1-22**[51], where 2H phase is generally used as a semiconductor with 1~2 eV of bandgap[52]. Here, *s*, *p*, *d*-orbitals of transition metal are hybridized to from six-fold coordinate, and each sp^3 hybridized orbital of chalcogen is bonded to them. Interestingly, lone-pair electrons of chalcogen atom plays an important role in the bonding. Although standard covalent bond is formed when two atoms provide one electron per bond, for the covalent bond in TMDC crystals, vacant hybrid orbital of a transition metal is filled by lone-pair electrons in sp^3 hybridized orbital of chalcogen. Hereinafter, compositional formulas such as MoS_2 or WSe_2 represents the 2H phase of them for simplicity.

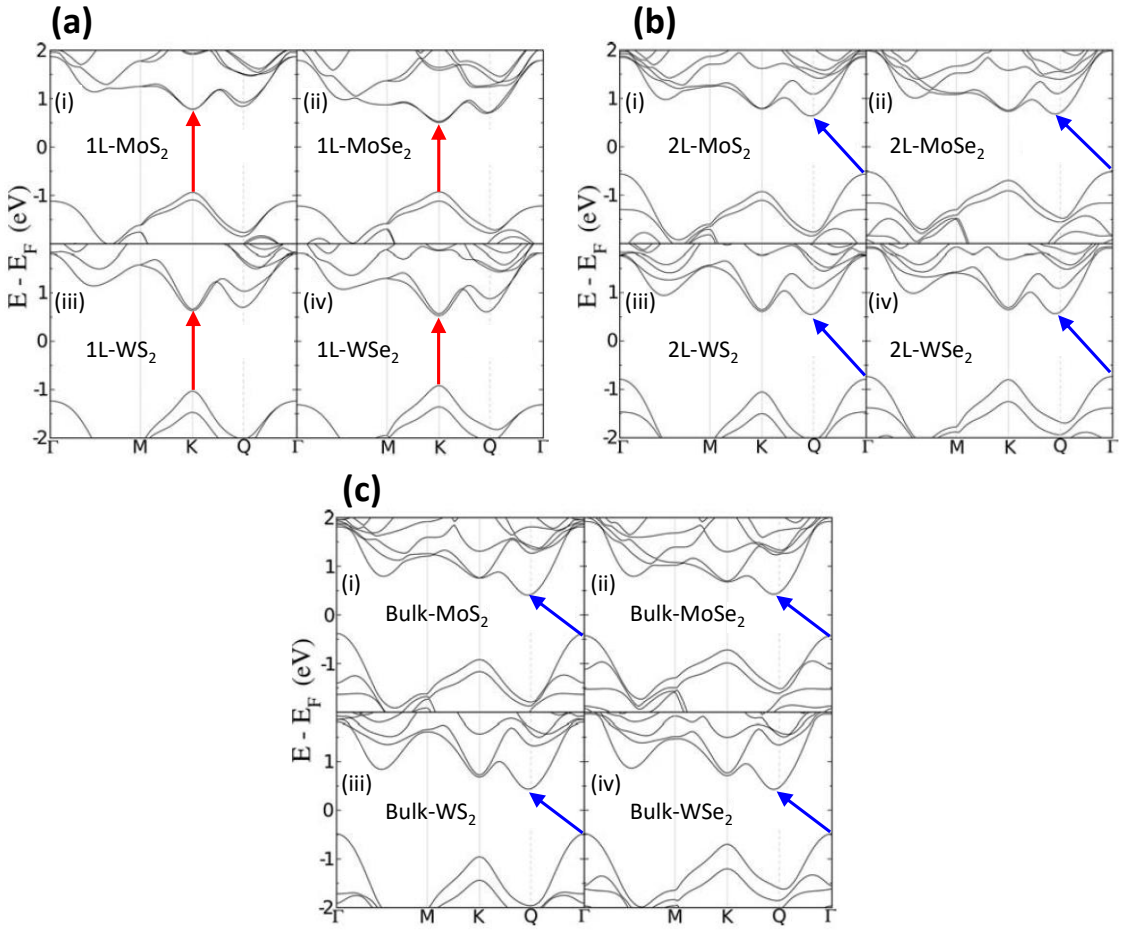


Fig. 1-23 Calculated electronic band structure of **(a)** monolayer, **(b)** bilayer, and **(c)** bulk MoS_2, $MoSe_2$, WS_2 and WSe_2 [53].

One interesting property of TMDCs is that, layer number dependent electronic band structure. **Fig. 1-23** shows the calculated band structures in first Brillouin zone of MoS_2, $MoSe_2$, WS_2, and WSe_2[53]. As indicated by red arrows in **Fig. 1-23(a)**, monolayer of them show the direct bandgap, while bilayer and bulk of them show the indirect bandgap as indicated by blue arrows in **Figs. 1-23(b)** and **(c)**. Interlayer interaction is decreased with decreasing the layer number, resulting in the transition. This is sometimes referred as indirect-to-direct gap transition[51].

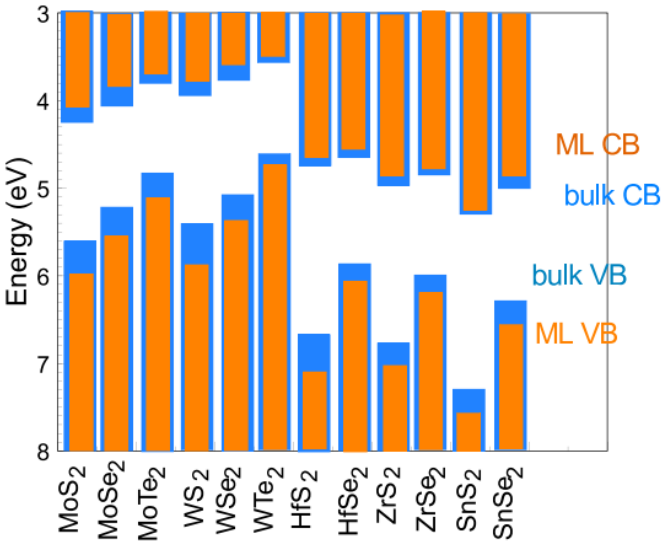


Fig. 1-24 Calculated bandgaps of bulk and monolayer (ML) TMDCs [54].

Fig. 1-24 summarizes the calculated bandgaps of TMDCs[54]. A wide variety of TMDCs provides us the tunability of bandgap by changing the material or layer number. This is one of the advantages of 2D material system which Si technology does not have. In addition, some high-quality TMDC crystals can be grown by a chemical vapor transport (CVT) technique or a physical vapor transport (PVT) technique[55]. These growth techniques enable the experimental study of various TMDCs.

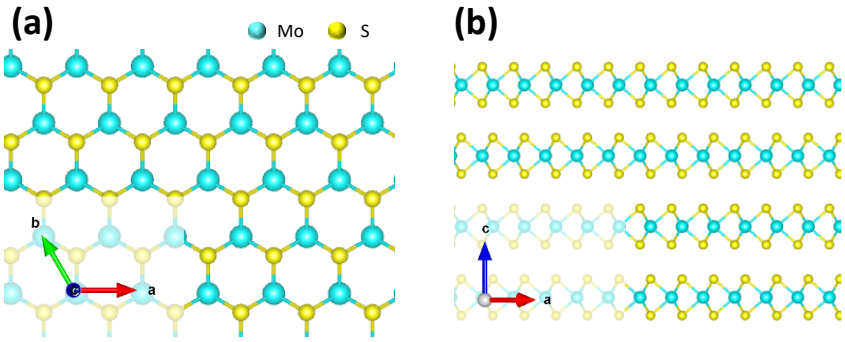


Fig. 1-25 Atomic structure of MoS_2. **(a)** Top view and **(b)** side view.

Among TMDCs, MoS_2 illustrated in **Fig. 1-25** is a popular semiconducting material due to its air stability. Its Brillouin zone is shown in **Fig. 1-26(a)**, and calculated bandgaps and position of the band edge as a function of the number of layers are shown in **Figs. 1-26(b)** and **(c)**, respectively[56]. Here, Λ_{min} point in **Fig. 1-26** is same as Q point in

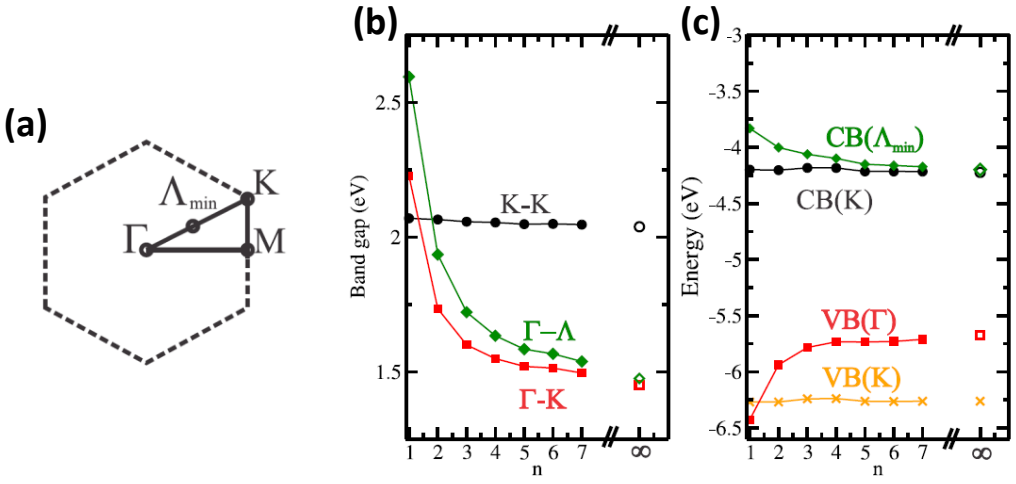


Fig. 1-26 (a) Brillouin zone of MoS_2 with high-symmetry *k*-points. **(b)** Bandgaps and **(c)** position of band edge as a function of the number of layers. Energy value is indicated with respect to the vacuum level[56].

Fig. 1-23. As well as other TMDCs, the bandgap of MoS_2 is increased with decreasing the number of layers. As shown in **Fig. 1-26(c)**, conduction band (CB) edge at K point (CB(K)) and valence band (VB) edge at K point (VB(K)) are insensitive against the number of layers, while CB edge at Λ_{min} point (CB(Λ_{min})) and VB edge at Γ point (VB(Γ)) are sensitive, resulting in different evolutions of bandgap shown in **Fig. 1-26(b)**. This can be explained by different contribution from orbitals of the atoms. Roughly speaking, layer number insensitive edges (CB(K) and VB(K)) are mainly composed of in-plane orbitals of S and Mo atoms, whereas layer number sensitive edges (CB(Λ_{min}) and VB(Γ)) are mainly composed of out-of-plane orbitals of S and Mo atoms. Especially, according to their calculations, VB(Γ) is mainly composed of p_z of S and d_{z^2} of Mo, which are sensitive to the interlayer interaction, while VB(K) is mainly composed of p_x and p_y of S, and d_{xy} and $d_{x^2-y^2}$ of Mo, which are insensitive to the interlayer interaction. As a result, valence band maximum point is changed from VB(K) to VB(Γ). More detailed calculation can be found in the reference[56]. It is noted that, as shown in the reference[56], the difference of the layer number sensitivity cannot be fully understood by above explanations.

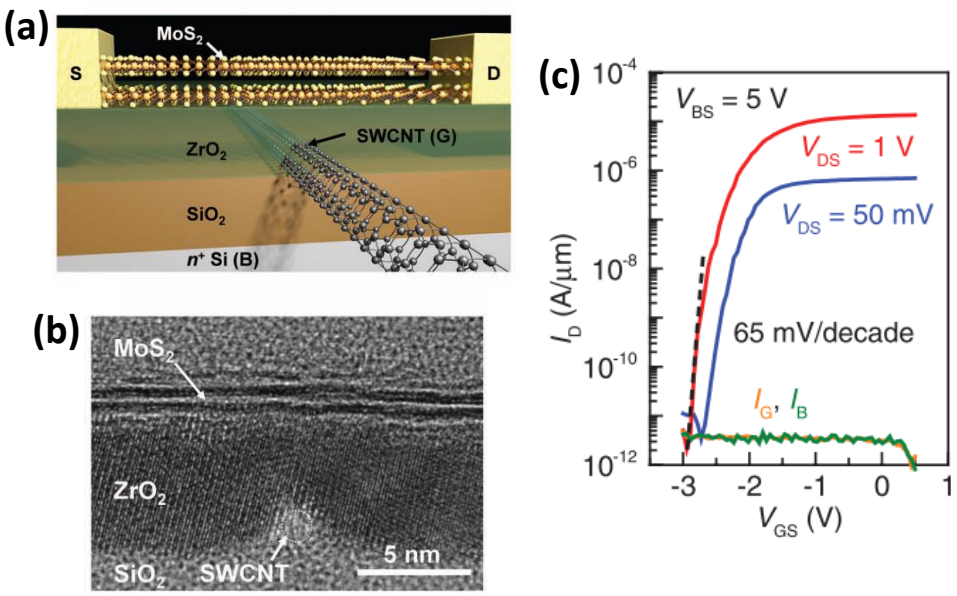


Fig. 1-27 (a) Schematic, **(b)** cross-sectional TEM image, and **(c)** transfer characteristics of 1D2D-FET[58].

In terms of FET application, MoS_2 has been considered as one of promising candidate for channel material[57] due to its strong immunity for the short channel effects. One surprising demonstration in 2016 from Ali Javey group (UC Berkley in US) is shown in **Fig. 1-27**[58]. They have demonstrated the 1D2D-FET schematically illustrated in **Fig. 1-27(a)**, which composed of MoS_2 channel and single-walled carbon nanotube (SWCNT) gate with ZrO_2 gate insulator. Although the gate length is only one nanometer (**Fig. 1-27(b)**), as shown in **Fig. 1-27(c)**, 65 mV/decade of subthreshold swing and ~10^6 of on/off ratio can be achieved with ~10^{-12} A/μm of off current. It should be noted that the effective channel length is about 3.9 nm, based on their simulations. Generally, operation mechanism of 2D FET is understood based of the Schottky barrier FET (SB-FET) model which well describes the ambipolar nature of 2D FET. Representative works for the MoS_2 FET by J. Appenzeller group (Purdue Univ. in US) is shown in **Fig. 1-28**[59]. If MoS_2 and metals are isolated, the band alignment between them is determined as illustrated in **Fig. 1-28(a)**, suggesting that electron injection seems to be dominant for Sc or Ti contact while hole injection seems to be dominant for Ni or Pt contact. However, as shown in **Fig. 1-**

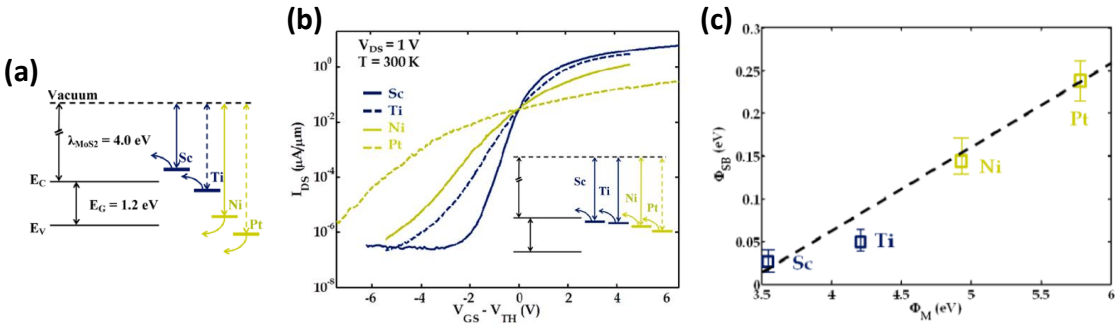


Fig. 1-28 Demonstration of SB-FET. **(a)** Band alignment between MoS_2 and metals when they are isolated. **(b)** Experimental transfer curves. Inset shows the sketch of actual band alignment. **(c)** Extracted Schottky barrier height (Φ_{SB}) as a function of metal work function (Φ_M). Poor slope suggests the strong Fermi level pinning at the MoS_2/metal interface[59].

28(b), *n*-type behavior of MoS_2 FET was obtained for all metals. According to their discussions, it is due to strong Fermi level pinning (FLP) at the metal/MoS_2 interface. **Fig. 1-28(c)** shows the relationship between Schottky barrier height at the interface (Φ_{SB}) as a function of work function of the metal (Φ_M). The slope $d\Phi_{SB}/d\Phi_M$ is around 0.1, indicating that the situation is far from Schottky-Mott limit (slope = 1). Similarly, other TMDC FETs such as WSe_2 FET[60] and $MoTe_2$ FET[61] have been investigated based on the SB-FET model. On the contrary, N. Fang et al. have demonstrated the accumulation mode 2D FET (ACCU-FET), where partial top gate (TG) can effectively modulate the conductance of 2D channel (**Fig. 1-29**)[62]. ACCU-FET model was originally proposed for silicon-on-insulator MOSFET[63,64], where accumulated carriers (majority carriers) contribute the current flow. This is different from MOSFET in which inversion carriers (minority carriers) do so. Based on the fact, the operation mechanism of 2D FET should be understood by the combination of SB-FET and ACCU-FET model.

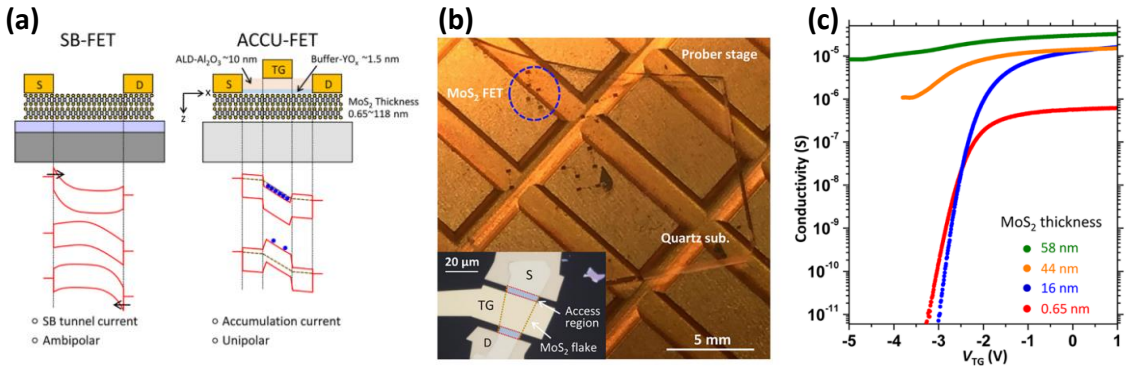


Fig. 1-29 Demonstration of ACCU-FET. **(a)** Conceptual illustration of ACCU-FET model as compared with SB-FET. **(b)** Photo of fabricated device. The margin between TG and S/D electrodes is called access region. **(c)** Transfer curves of MoS_2 FETs. Partial TG effectively modulates the conductivity especially below 16 nm of thickness[62].

1.3.4 2D Heterostructures

Although interesting properties of 2D materials have been predicted and demonstrated as mentioned above, for unveiling the intrinsic property of them, it was necessary to wait for the emergence and development of 2D hetero-stacking technique. In other words, *gate stack* for 2D FETs is the key to access their intrinsic properties.

In conventional 3D material system, Si/SiO_2 is a popular gate stack. Generally, the Si/SiO_2 system is fabricated by thermal oxidation of Si[65], and the improvement of Si/SiO_2 interface quality is one driving force of development of integrated circuits (ICs), since channel of MOSFET is formed underneath the SiO_2. However, volume expansion occurs during the oxidation, resulting in dangling bonds and strained region at the interface[66]. This is the origin of interface states in the bandgap of Si, which degrades the FET performances such as subthreshold swing (SS).

In this context, 2D materials have been attracted much attention since they are ideally free from dangling bonds on their surface. However, since the 2D FET had been fabricated on conventional 3D gate stack such as Si/SiO_2, they were suffered from surface morphology and charged impurity in the 3D insulator, still preventing us to access their intrinsic properties. For example, according to the theoretical study of graphene, DOS and carrier density must be zero at the Dirac point, while current flow has been observed in graphene FET fabricated on a SiO_2/Si substrate. The carrier density which contribute the current flow at the Dirac point is referred as residual carrier density (n^*). At the early stage of graphene research, residual carrier density had been studied, and the origin has been considered to be the charged impurities on/in SiO_2 [36]. For TMDCs, situation is same. Y. H. Lee group (SKKU in Korea) reported the bandgap fluctuation of monolayer MoS_2 as shown in **Fig. 1-30**[67]. Since 2D materials can partially follow the surface morphology due to its strong strain tolerance, local strain causes the local change of band structure, resulting in the fluctuation.

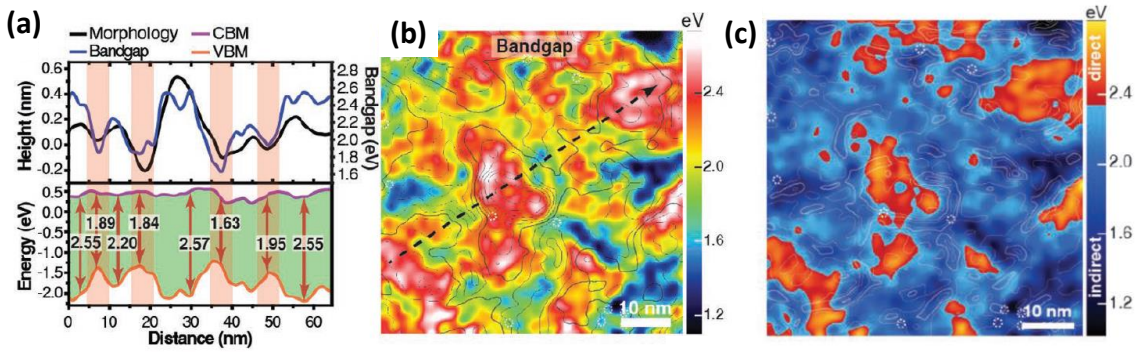


Fig. 1-30 Bandgap fluctuation in monolayer MoS_2. **(a)** Line profiles along the dotted arrow in **(b)** where the value of bandgap is mapped. **(c)** Map of direct and indirect bandgaps in same region of **(b)**[67].

Therefore, it is inevitable that 2D hetero-stacking technique would emerge to obtain dangling bond free and atomically sharp interface in the gate stack. One of the advantages of 2D materials is that, the heterostructure is free from lattice mismatch issue due to their layered nature. Moreover, fortunately, there is a stable insulating 2D material, *h*-BN, which can act as an atomically flat substrate by masking the surface morphology and the effect from charged impurity on/in 3D insulators.

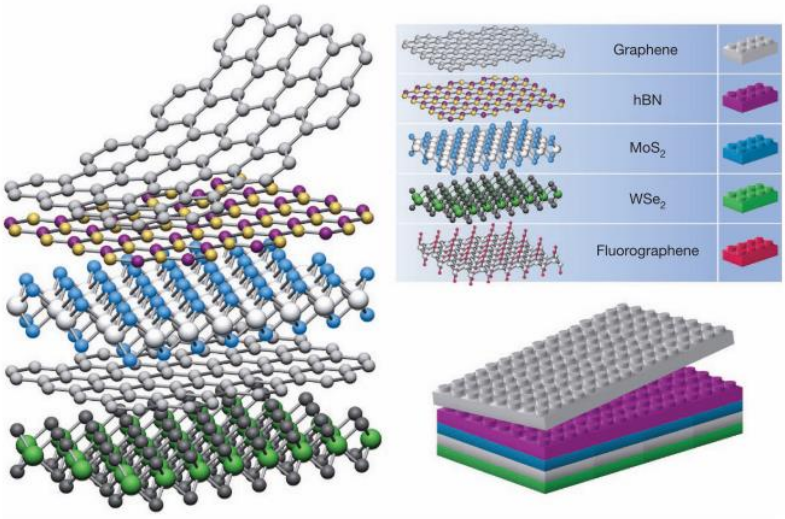


Fig. 1-31 Conceptual illustration of 2D heterostructure[68].

The conceptual illustration of 2D heterostructure is shown in **Fig. 1-31**[68]. To realize the heterostructure, polymer assisted dry transfer technique has been developed[69–72]. As expected, the gate stack with 2D/2D interface(s) enables us to access their intrinsic

properties. The good example is shown in **Fig. 1-32**[73], where the evolution of graphene FET is summarized. Sharper resistivity peak, which represents the intrinsic property of graphene, is achieved with improving the interface quality by 2D hetero-stacking technique. In addition, all 2D heterostructured bilayer graphene FET (graphite TG/*h*-BN/bilayer graphene channel/*h*-BN/graphite BG) has achieved quite sharp ambipolar behavior with less than 0.1 pA of off current[74]. Moreover, 2D heterostructured tunnel FET (TFET) shows less than 60 mV/decade of subthreshold swing due to atomically short *pn* junction at 2D hetero-interface[75]. More recently, moiré superlattice of 2D materials has been attracted much attention because non-trivial properties can be expected due to new and long periodicity of the superlattice[76]. For example, twisted bilayer graphene with about 1.1° of twist angle shows the superconductivity near 0 K[77].

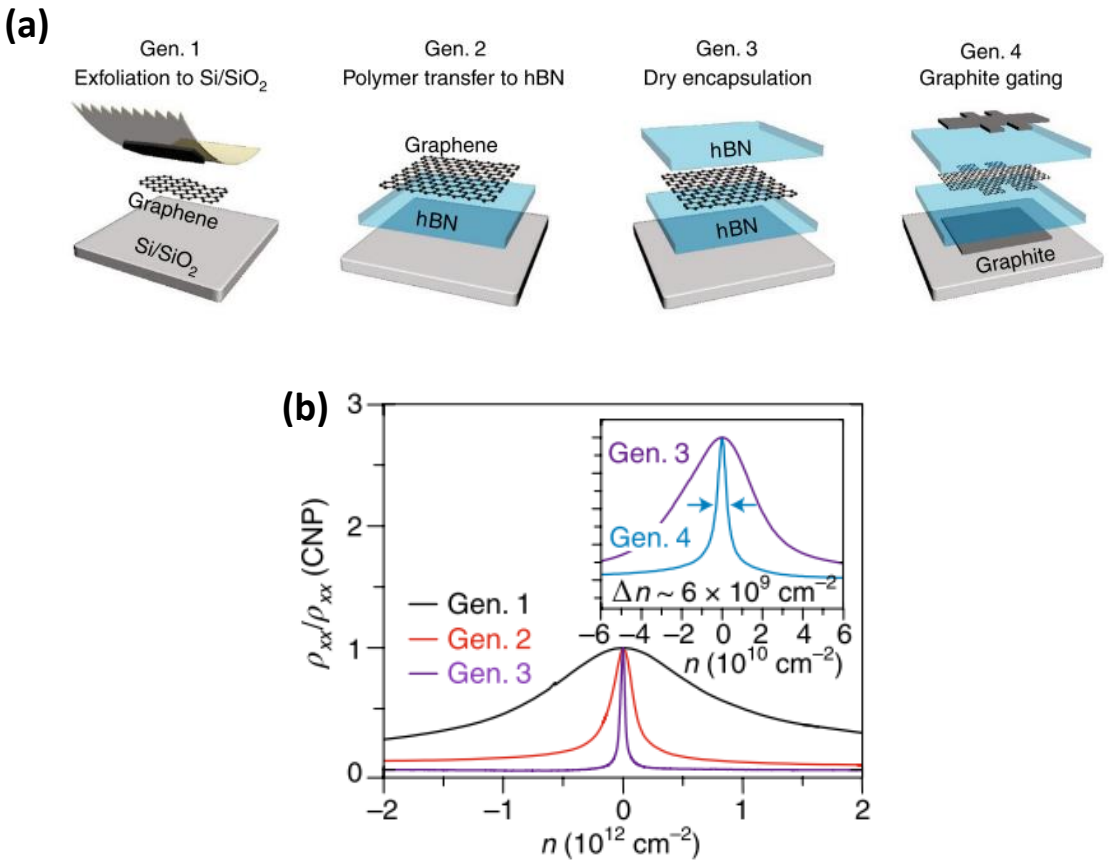


Fig. 1-32 Evolution of **(a)** fabrication technique and **(b)** corresponding electrical characteristics of graphene FET from generation 1 to 4. Sharper peak of normalized resistivity can be obtained with the evolution of generation[73].

1.3.5 Defects in 2D Materials

As mentioned above, whereas 2D materials and their interfaces are often regarded as ideal one, that is, dangling bond free, the theory of thermodynamics requires the existence of defects in their crystals. This perspective is quite important for practical applications since it may be the key of the performance degradation. To understand the defects in 2D crystals leads to the further control and improvement of the performance. In this section, therefore, the defects in 2D materials are discussed from the thermodynamic point of view. Main focus is the defects in MoS_2 since it has been well studied.

According to the thermodynamics, when the reaction occurs spontaneously under constant pressure condition, following equation should be satisfied:

$$\Delta G = G_{\mathrm{f}} - G_{\mathrm{i}} < 0, \tag{1.5}$$

where G_{f} and G_{i} are the Gibbs free energy of final state and initial state, respectively. Here, the theory is utilized to estimate the number of point defects under thermal equilibrium. Consider a perfect crystal without any defects consists of N atoms per unit volume. When n point defects per unit volume are induced in the crystal, possible configuration of the system, W can be described as follows:

$$W = {}_N C_n = \frac{N!}{n!\,(N-n)!}. \tag{1.6}$$

Since the Gibbs free energy can be divided into enthalpy term and entropy term, ΔG can be described regarding to defect formation as follows:

$$\Delta G = nG^f - Tk_{\mathrm{B}} \ln \frac{N!}{n!\,(N-n)!}, \tag{1.7}$$

where, G^f, T and k_{B} are the Gibbs' free energy for one defect formation, temperature, and Boltzmann's constant, respectively. The G^f is often referred as defect formation energy. The ΔG and its first and second terms are illustrated in **Fig. 1-33** as a function of n. While first term is increased with increasing the number of point defects since this is a *penalty* to form the defect, second term is decreased since increasing entropy is favorable for systems. According to the figure, it is clear that point defects should be induced in the crystal system

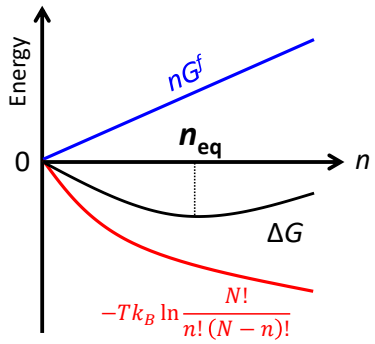


Fig. 1-33 Sketch of the ΔG and its first and second term as a function of n.

for thermodynamic stability. Under thermal equilibrium condition ($\Delta G = 0$), the number of point defects n_{eq} can be described as follows:

$$n_{\mathrm{eq}} = N\exp\left(-\frac{G^f}{k_B T}\right). \tag{1.8}$$

Here, $N >> n$ is assumed. Strictly speaking, pre-factor N should be multiplied N_c which is the number of inequivalent configurations. According to the equation, it is clear that the smaller the value of G^f, the more defects exist under thermal equilibrium. Here, G^f can be described as follows[78]:

$$G^f = F(def) - F(host) + pV^f - \sum_i n_i\mu_i + qE_{\mathrm{F}} + E_{\mathrm{corr}}, \tag{1.9}$$

where, F(def) and F(host) are the Helmholtz free energies of the system with and without the defect, respectively, and p and V^f are the pressure and formation volume of the defect, respectively. The number of required atoms to form the defect is represented by n_i, and its chemical potential per atom is represented by μ_i. Positive n_i means adding the atom and vice versa. The subscript i of n_i and μ_i represents a kind of adding/removing atom such as sulfur. The required charge to form the defect is represented by q. If one electron is required to form the defect, $q = -1$. Additionally, E_{F} is the Fermi level position with respect to the vacuum level, and E_{corr} is the correction term for reasonable calculation.

From the equation, G^f can be considered as a function of chemical potential μ_i and Fermi level E_{F}. Two calculation results for point defects in bulk MoS_2 are shown in **Fig. 1-34**[78], where V^f in **Eq. 1.9** is assumed to be zero. **Fig. 1-34(a)** shows a (defect) formation energy (G^f) as a function of chemical potential of sulfur (μ_S). Since neutral charge state (q

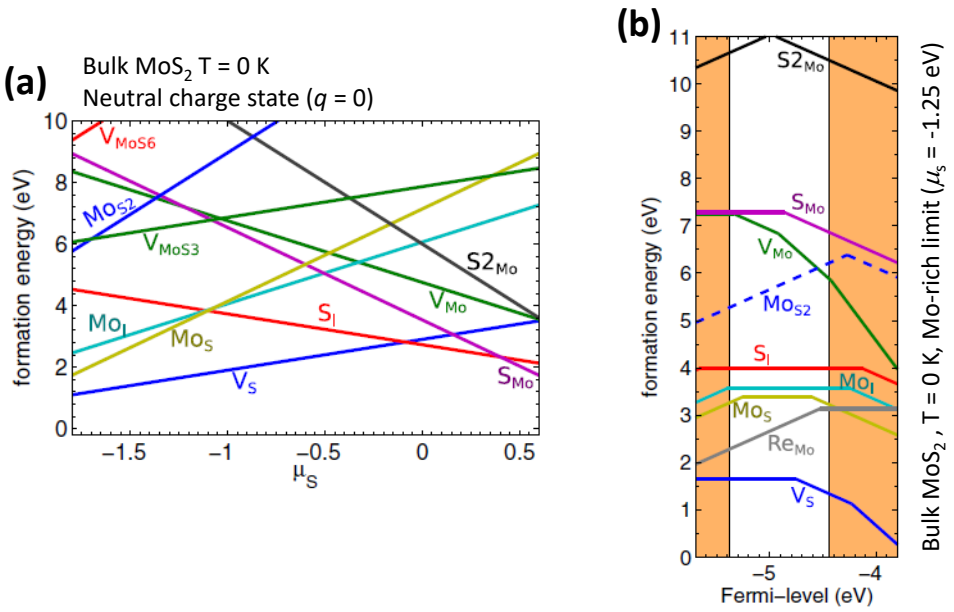


Fig. 1-34 Calculated defect formation energies in bulk MoS_2 as a function of **(a)** chemical potential of sulfur and **(b)** Fermi level[78].

= 0) is considered, formation energies for various defects are simply the functions of μ_S. **Fig. 1-34(b)** shows a formation energy as a function of E_F, where Mo-rich limit (μ_S = –1.25 eV) is considered. While various charge states e.g., q = 0, +1, –1, can be calculated, only most stable state is shown in the figure. Important point of this figure is that, the monosulfur vacancy (V_S), which often referred to simply as sulfur vacancy, has the lowest formation energy for wide range of μ_S and whole range of E_F, suggesting that the sulfur vacancy is most popular point defect in MoS_2. Although the situation is a little different for monolayer case or S-rich limit, this is consistent with experimental studies.

Fig. 1-35(a)[79] shows the histogram of point defects in various MoS_2 monolayers, where ME, PVD and CVD represent mechanical exfoliation, physical vapor deposition, and chemical vapor deposition, respectively. As predicted by previous calculations, V_S is most popular point defect for ME and CVD MoS_2 monolayers while Mo_{S2} (antisite defect where one Mo atom substitutes two sulfur atoms aligned in column) is most for PVD MoS_2 monolayer. According to the figure, ME MoS_2 monolayer has the fewest defects, and this is the reason why many researches use the ME 2D flakes instead of grown ones. The

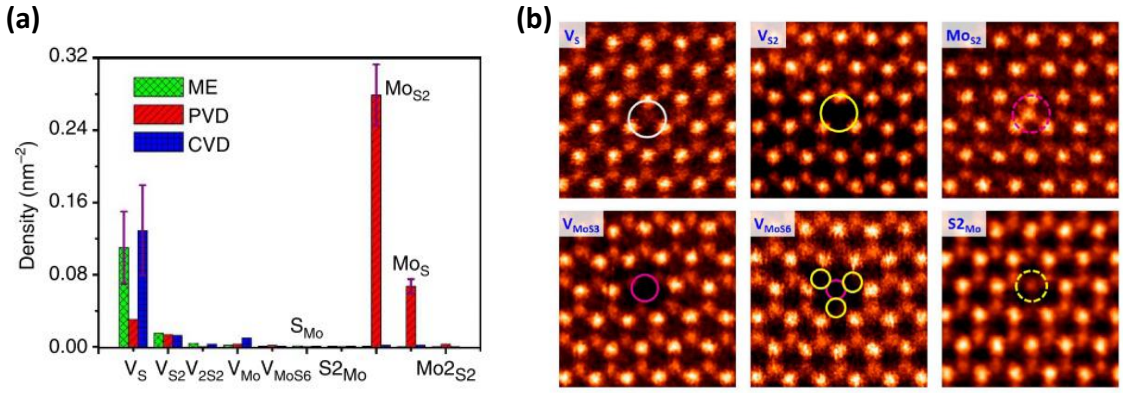


Fig. 1-35 (a) Histogram of various point defects in MoS_2 monolayers[79]. **(b)** STEM-ADF image of various point defects in CVD grown monolayer MoS_2 [81].

density of V_S in ME is around 10^{13} cm^{-2}, and its typical range is 5×10^{12}–5×10^{13} cm^{-2} which has been reported by the litetature[80].**Fig. 1-35(b)**[81] shows the various point defects of CVD grown monolayer MoS_2 observed by annular dark field (ADF) imaging on scanning transmission electron microscope (STEM). Various defects including V_S are clearly observed experimentally. On the other hand, Kish graphite and *h*-BN are well known to be high-quality, as compared with MoS_2. Defect density of Kish graphite is negligibly small[82], and that of *h*-BN grown by Taniguchi group (NIMS in Japan) is about 10^9–10^{10} cm^{-2} according to the litelature[83].

1.4 2D Materials-based Flash Memory

As described above, the period from the middle of 2000s to the early 2010s was a turning point for the electronics in terms of NVM, materials and applications, as follows.

- Emerging NVM technologies has come to be regarded as promising candidates for filling the performance gap in the memory hierarchy.
- In 2007, 3D NAND technology was presented[15,23].
- In 2010, K. S. Novoselov and A. K. Geim were awarded Novel Prize in Physics for their groundbreaking work regarding graphene[33].
- In 2012, deep learning achieved the great success in terms of object recognition[13].

As of 2021, each technology has been energetically promoted, and researches are becoming more interdisciplinary. In this context, 2D materials are increasing their prominence as for next-generation electronic devices[84] due to their high crystallinity and interface quality. According to the International Roadmap for Devices and Systems™ (IRDS™) 2020, 2D materials are considered as one of key for beyond-CMOS (Complementary MOS) technology for the end of 2020s and beyond[85]. Moreover, in recent IEEE International Electron Devices Meeting (IEDM), ultra-scaled 2D FETs have been demonstrated as shown in **Fig. 1-36**[86,87], and its promising future has been depicted[88,89].

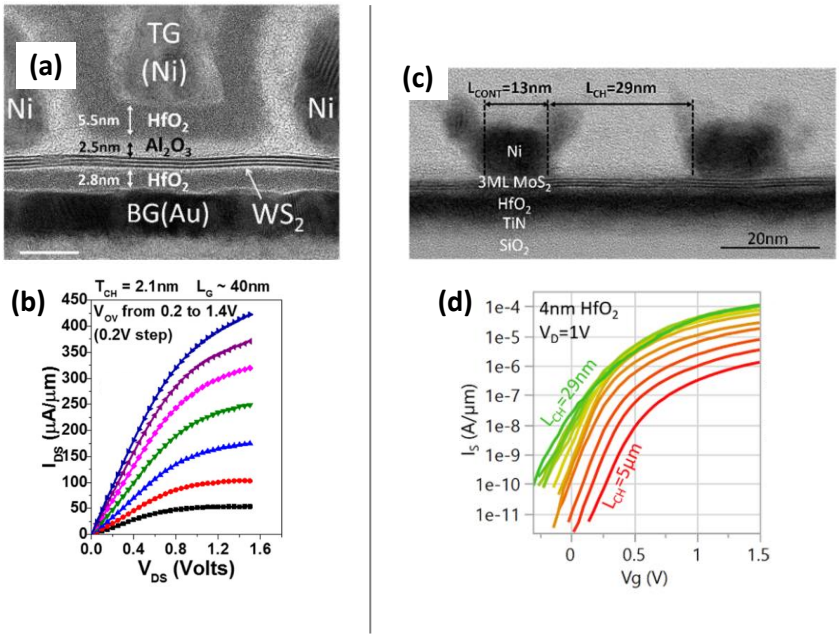


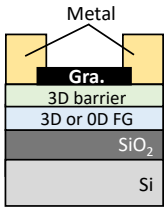
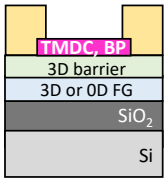
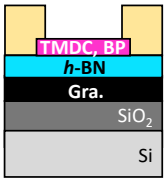
Fig. 1-36 Recent demonstrations of ultra-scaled 2D materials-based MOSFETs. **(a)** Cross-sectional TEM image of double gated WS_2 FET with about 40 nm of channel length. Scale bar represents 10 nm. **(b)** Output characteristics of the WS_2 FET. Overdrive voltage V_{ov} is defined as $V_{gs}–V_{th}$ [86]. **(c)** Cross-sectional TEM image of three monolayer MoS_2 channel FET and **(d)** its transfer curves under various channel length conditions[87].

Therefore, as well as FET applications, it is inevitable that flash memory application of 2D materials has been aggressively studied. Considering the above-mentioned circumstances, future material compatibility can be expected for 2D materials-based flash memories. In addition, atomically sharp and dangling-bond free interface of 2D heterostructure can be expected to improve the memory performance such as fast P/E speed, long retention, and tough endurance. Although single cell-level investigation is a mainstream due to the lack of high-quality and large area production method for 2D materials, there has already been a lot of studies.

In analogy with the evolution of graphene FETs shown in **Fig. 1-32**, the evolution of 2D materials-based flash memory can be summarized as shown in **Table 1-2**, where typical back-gated device structures are also illustrated.

At the early stage of research, generation 1, graphene had played a principal role mainly as a channel, while other materials which compose the gate stack still be 3D or zero

Table 1-2. Evolution of 2D materials-based flash memory devices.

			
	1st generation	**2nd generation**	**3rd generation**
Channel	Graphene	TMDC, BP	TMDC, BP
Tunneling barrier	3D Insulator	3D Insulator	*h*-BN
FG	3D, 0D material	3D, 0D material	Graphene, Graphite, TMDC
S/D electrodes	Metal	Metal	Metal

dimensional (0D) materials. In generation 2, TMDC or BP superseded the role of graphene since they have semiconducting nature. Nevertheless, 3D and 0D materials were used for other parts. As the emergence of 2D hetero-stacking technique, 2D heterostructured NVM devices have been energetically studied. Currently, the research phase is in 3rd generation. Atomically sharp and flat, ideally dangling bond free, and lattice mismatch free 2D/2D interface as mentioned in **section 1.3.4** have been expected in terms of not only performance improvements but also the expansion of the application. In this section, previous studies of 1st and 2nd generations are overviewed in **sections 1.4.1** and **1.4.2**, respectively. Previous studies of 3rd generation are separately overviewed in **sections 1.4.3** and **1.4.4** for standard memory device and advanced functional devices, respectively. Noted that, all of them are not CT type but FG type. Finally, the issues in recent 2D heterostructured NVM device is discussed.

1.4.1 Graphene Channel Devices with Non-2D Materials

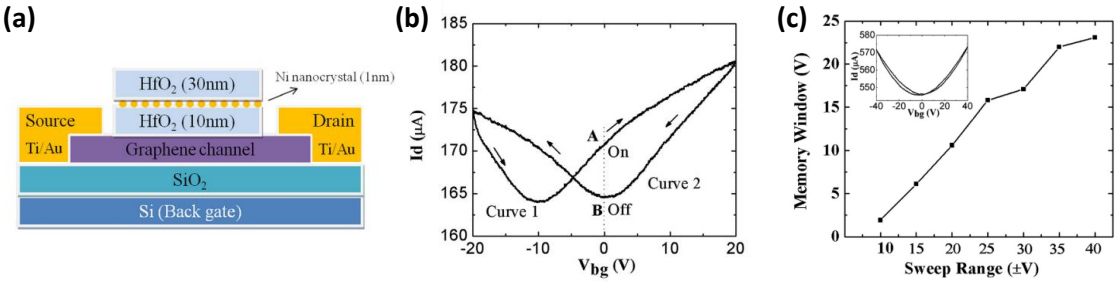


Fig. 1-37 Schematic of typical graphene channel device and its results. **(a)** Device structure of graphene channel device with Ni nanocrystal charge trapping layer. **(b)** *I–V* Round sweep curves with large hysteresis and **(c)** memory window trend against V_{bg} sweep range[90].

As shown in **Fig. 1-37**, N. Zhan et al. have reported the graphene channel FET with HfO_2/Ni nanocrystal/HfO_2 stack, where Ni nanocrystal acts as a charge trapping layer[90]. In that study, over 20 V of memory window when sweep range is ±40 V, and 10 ms operation has been demonstrated. S. A. Imam et al. have also demonstrated the large hysteresis of graphene due to the Si_3N_4 charge trapping layer[91]. In addition, the paper entitled *Graphene Flash Memory* was published on ACS Nano by A. J. Hong et al. where large hysteresis of multi-layer graphene inserted MOS capacitor has been demonstrated in the capacitance-voltage curves[92].

Above three reports were published in 2011, just one year after the groundbreaking research of graphene was awarded the Novel Prize. Although they used graphene, conventional non-2D materials still be used.

1.4.2 Semiconducting 2D Channel Devices with Non-2D Materials

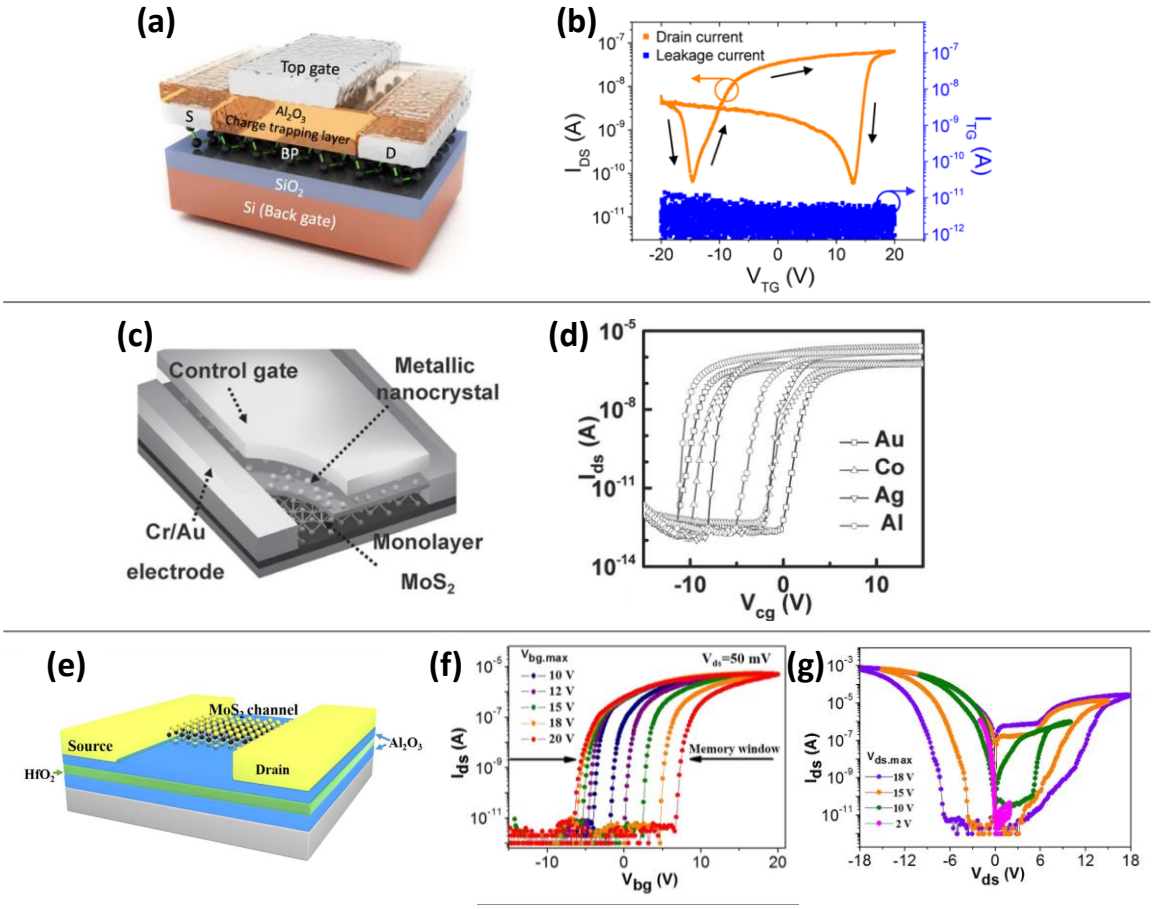


Fig. 1-38 Schematics of typical semiconducting 2D channel devices and their results. **(a)** Black phosphorus (BP) channel device with Al_2O_3 charge trapping layer and **(b)** its *I*–*V* round sweep curves[93]. **(c)** Monolayer MoS_2 channel device with metallic nanocrystal charge trapping layer and **(d)** its *I*–*V* round sweep curves[94]. **(e)** Operation mode switchable device with MoS_2 channel, and its **(f)** three-terminal and **(g)** two-terminal operations[95].

Zero-bandgap nature of graphene is critical to the flash memory application, since the two stable states are difficult to be distinguished due to low on/off ratio. To overcome the issue, TMDCs and other semiconducting 2D materials have inherited the role of channel from graphene. For example, as shown in **Figs. 1-38(a)** and **(b)**, H. Tian et al. have demonstrated the memory operation of black phosphorus (BP) channel FET with Al_2O_3 charge trapping layer[93]. Not only 3D materials, but also 0D materials were used as a charge trapping layer. As shown in **Figs. 1-38(c)** and **(d)**, J. Wang et al. have demonstrated the memory device

with monolayer MoS_2 channel and metal nanocrystal charge trapping layer[94]. They have claimed that the memory window can be increased with increasing the work function of the metal nanocrystal. They have also demonstrated over 10^6 of on/off ratio, 10 ms P/E operation, and 2,000 s of retention characteristic. X. Hou et al. have proposed the operation mode switchable memory device[95]. As shown in **Fig. 1-38(e)**, they used MoS_2 and Al_2O_3 encapsulated HfO_2 as a channel and charge trapping layer, respectively. They have demonstrated not only three-terminal operation by BG pulse, but also two-terminal operation by drain pulse (**Figs. 1-38(g)** and **(h)**).

By using TMDCs and other semiconducting 2D materials as a channel, on/off ratio can be drastically increased. However, non-2D materials still be used in their device[96,97].

1.4.3 2D Heterostructured Devices

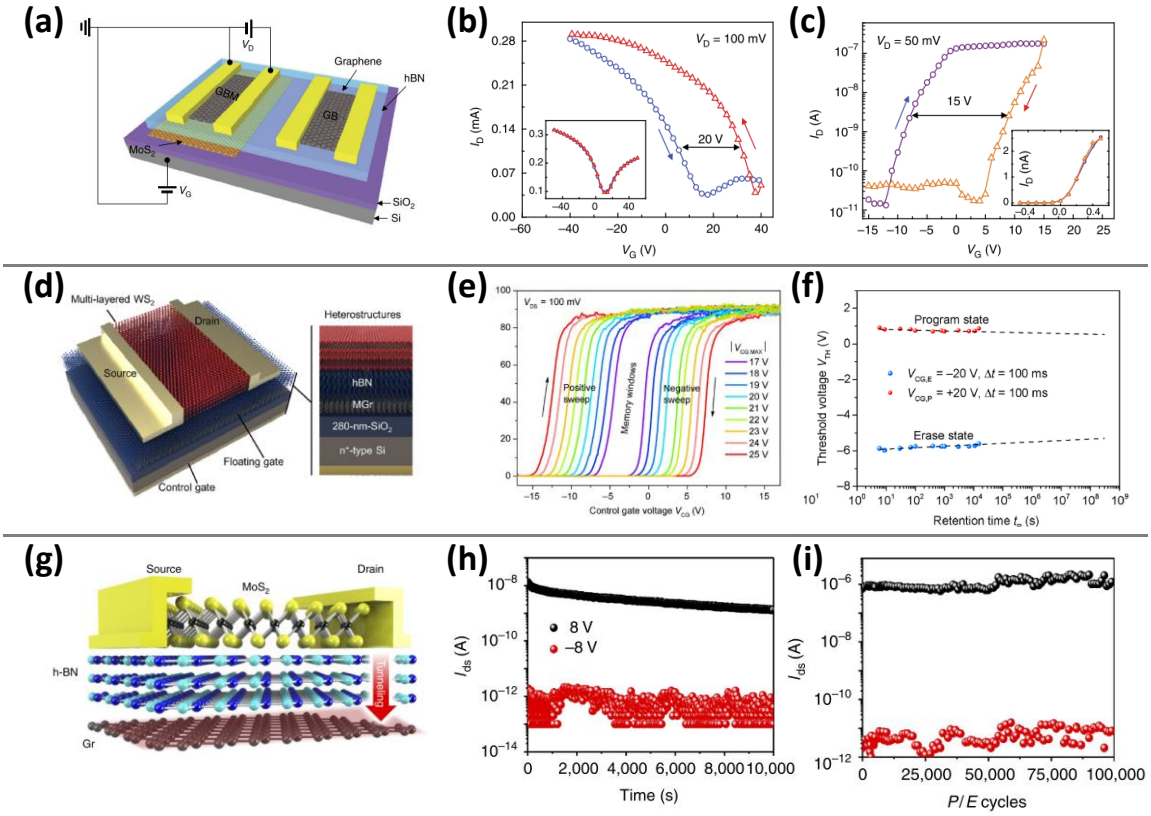


Fig. 1-39 Schematics of typical 2D heterostructured devices and their results. **(a)** Schematic of graphene channel/*h*-BN tunneling barrier/MoS_2 FG heterostructured device and **(b)** its *I*–*V* round sweep curves. **(c)** *I*–*V* round sweep curves of reversely stacked device[41]. **(d)** WS_2 channel/*h*-BN tunneling barrier/multi-layer graphene FG heterostructured device and **(e)** its *I*–*V* round sweep curves and **(f)** retention characteristics[42]. **(g)** Schematic of two-terminal tunneling RAM, and its **(h)** retention and **(i)** endurance characteristics[43].

With the help of the development of dry transfer technique, 2D heterostructured devices have been aggressively studied since 2D/2D interface can be exploited.

One of the pioneer works was given by M. S. Choi et al. in 2013, which. Schematic of their device and representative results are shown in **Figs. 1-39(a)–(c)**[41]. They have demonstrated two kinds of devices. One is the graphene/*h*-BN/MoS_2 heterostructured device where graphene, *h*-BN, and MoS_2 were used as a channel, tunneling barrier, and FG, respectively. The other is the MoS_2/*h*-BN/graphene heterostructured device where MoS_2 and graphene were used as a channel and FG respectively, while the role of *h*-BN

was maintained. In addition to the large on/off ratio, good retention characteristic can be obtained by MoS_2/*h*-BN/graphene device as compared with the reverse stacked device. As shown in **Figs. 1-39(d)–(f)**, D. Qiu et al. have designed the 2D heterostructure for negligible charge loss[42]. As a result of considering the band alignment of 2D heterostructure during retention period, they have claimed that the WS_2/*h*-BN/multi-layer graphene is good stack in terms of non-volatility. Indeed, they have demonstrated only ~13% charge loss after 10 years by extrapolation of the data at room temperature. It is worth noting that, they have also measured the tunneling current between the FG and channel, which is the key of device operation. As shown in **Fig. 1-40**, they fabricated another 2D heterostructure for the measurements, and tunneling current from WS_2 channel to multi-layer graphene via *h*-BN has been measured and analyzed by Fowler-Nordheim (FN) plot. Other TMDCs such as $MoTe_2$[98] and $MoSe_2$[99] have been also used as a channel of 2D heterostructured devices.

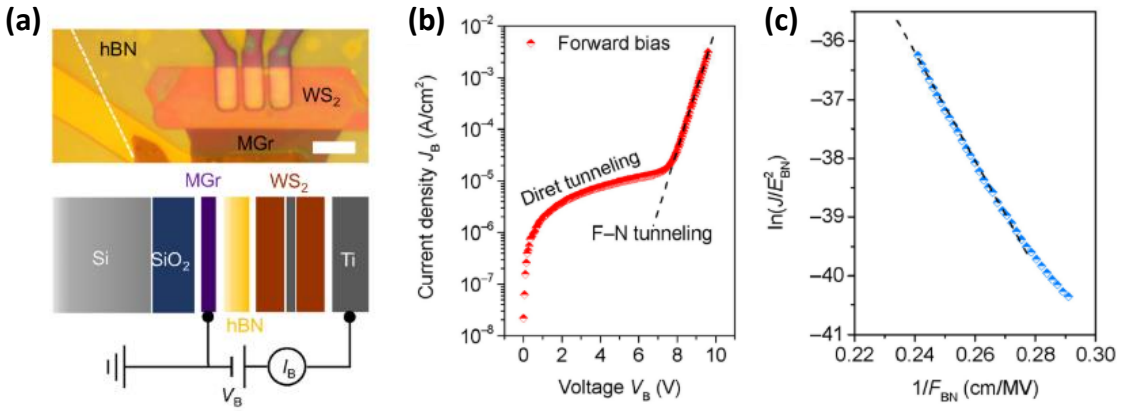


Fig. 1-40 Tunneling current analysis in WS_2/*h*-BN/multi-layer graphene heterostructure. **(a)** Device photo and measurement setup. The scale bar is 10 μm. **(b)** Measured tunneling current density and **(c)** its FN plot[42].

In addition, Q. A. Vu et al. have demonstrated excellent performances of MoS_2/*h*-BN/ graphene heterostructured device, while they claimed that this is not flash memory but *tunneling RAM (TRAM)* since it can be operated by two terminals (only source and drain)[43]. Schematic of the device and its typical results are shown in **Figs. 1-39(g)–(i)**. Long time retention of 10,000 s, 10^5 of endurance, thermal stability at 510 K, and nearly 20 % of strain tolerance have been demonstrated.

1.4.4 Advanced Functional Devices

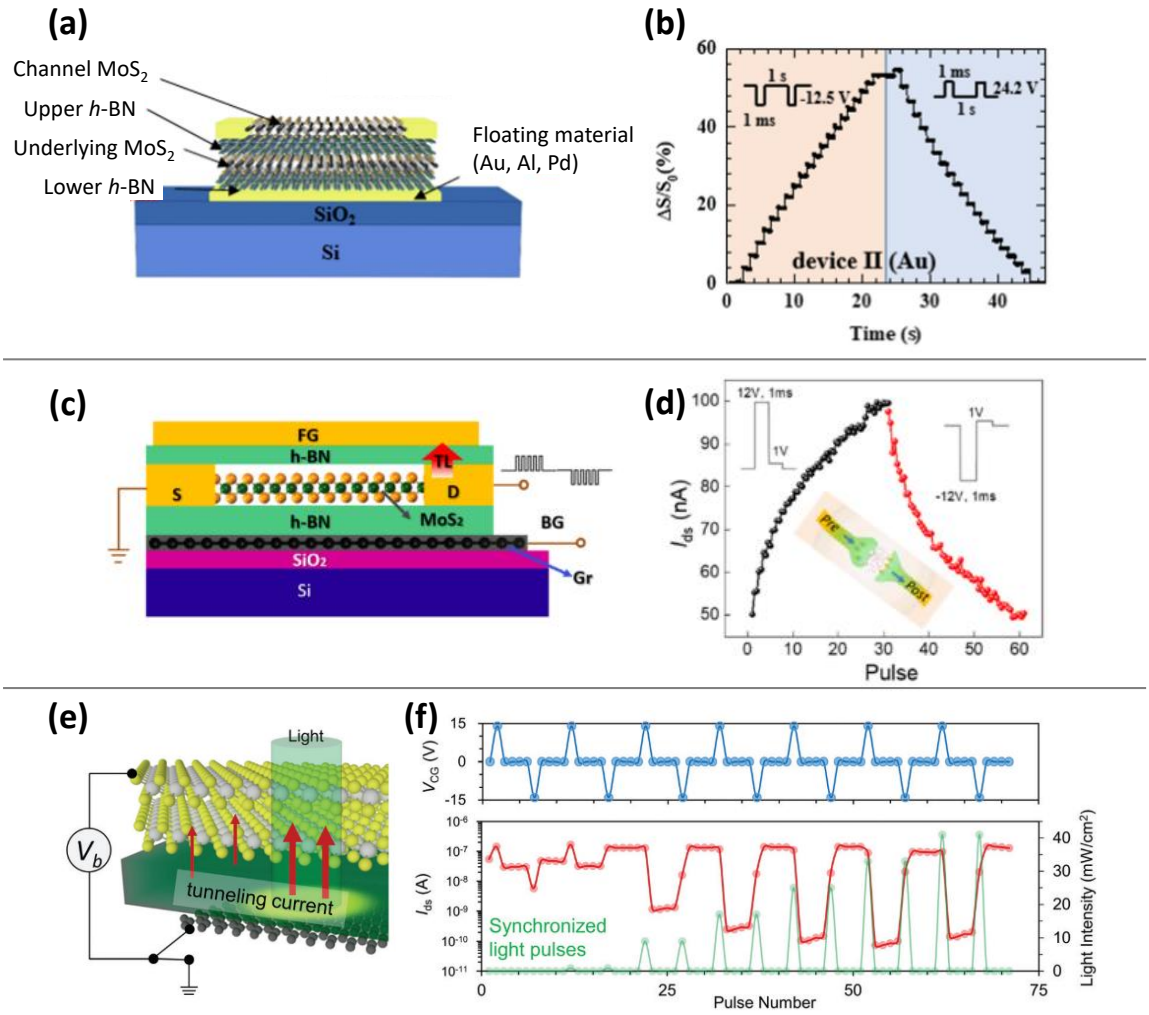


Fig. 1-41 Schematics of typical advanced functional devices and their results.
(a) Double FG device and **(b)** its artificial synaptic behavior. By continuous application of voltage pulse, device conductance almost linearly modulated[100]. **(c)** Artificial synapse based on 2D heterostructure and **(d)** its artificial synaptic behavior[101]. **(e)** Optoelectronic memory device based on PtS_2/*h*-BN/graphene heterostructure and **(f)** its optical controllability of the device conductance assisted by gate voltage pulse[102].

Beyond that, the functionality of 2D heterostructured devices has been expanded. One important functionality is an artificial synaptic behavior. As mentioned above, neural network is composed of the weighted edge and node. Since the weight should be continuously controlled, precise control of the device conductivity is desired. As shown in **Figs. 1-41(a)** and **(b)**, excellent controllability of device conductivity by the number of

pulse has been demonstrated by MoS_2/*h*-BN/MoS_2/*h*-BN/Au heterostructured device[100]. **Fig. 1-41(c)** also shows the artificial synaptic device based on 2D heterostructure. Although *h*-BN encapsulated MoS_2 and graphene heterostructure was used, tunneling was occurred between external FG on the 2D heterostructure and drain electrode. In addition to a good conductance controllability shown in **Fig. 1-41(d)**, conductance change by 50 ns ultra-fast pulse has been demonstrated[101].

On the other hand, optoelectronic memory devices have also been demonstrated. For example, Y. Chen et al. fabricated the PtS_2/*h*-BN/graphene heterostructure as an optoelectronic memory device as shown in **Fig. 1-41(e)**[102]. They have claimed that the asymmetric barrier height at PtS_2/*h*-BN interface enables to separate the generated electron hole pairs under light illumination, resulting in an optical controllability of memory states. Similarly, both electrical and optical controllability for the device conductance has been demonstrated by MoS_2/*h*-BN/graphene heterostructure[103].

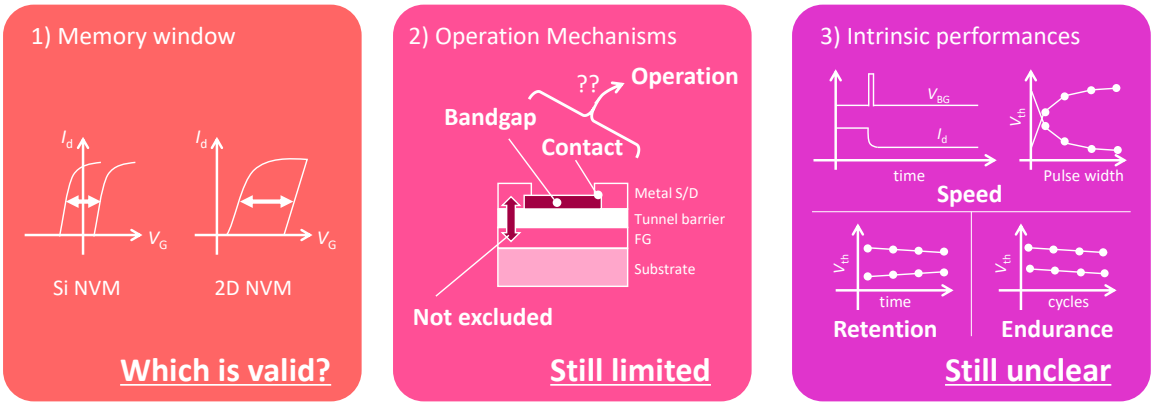


Fig. 1-42 Three issues to be solved for recent 2D heterostructured NVM devices.

However, above mentioned evolutions have not been sufficiently supported by the understandings of operation mechanisms. At present, proof-of-concept works have taken precedence over the understandings, weakening the foundations of 2D materials-based flash memory technology. Specifically, three issues shown in **Fig. 1-42** should be solved. 1) memory window, which is most basic and important metric for memory performance evaluation, is extracted by unusual way, and 2) the effect of the bandgap of 2D materials and metal/2D contact on the memory operation is still unclear. In addition, as exemplified in **Fig. 1-43**[41,102,104], since the source and drain (S/D) metal electrodes have been overlapped on the FG, poor metal/2D interface cannot be excluded from the possible tunneling paths. One simple solution to restrict the tunneling path in 2D heterostructure is to make the *access region* as shown in **Fig. 1-44**. However, as summarized in **Table 1-3**,

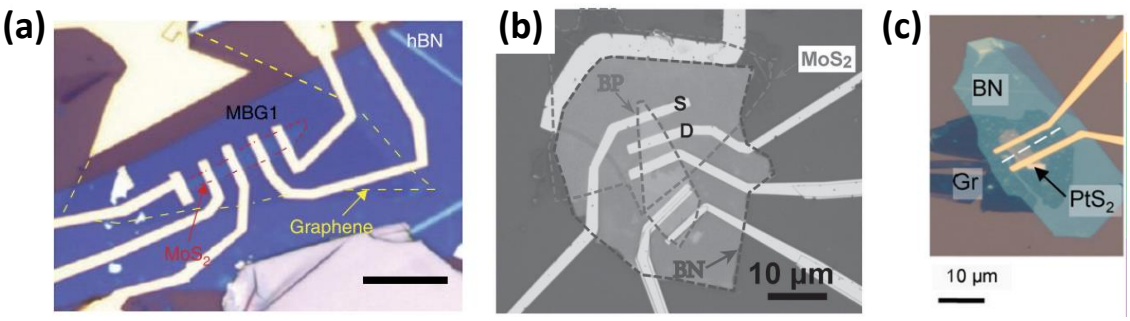


Fig. 1-43 Examples of S/D electrodes overlapping on the FG. **(a)** MoS_2/*h*-BN/graphene heterostructure[41], **(b)** BP/*h*-BN/MoS_2 heterostructure[104], and **(c)** PtS_2/*h*-BN/graphene heterostructure[102].

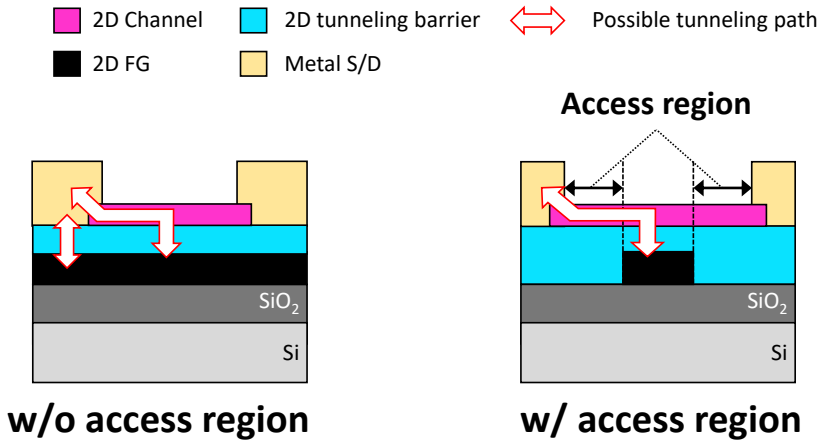


Fig. 1-44 Possible tunneling paths in the device with and without access region.

most of previous studies in 2nd and 3rd generations did not make the access region. Therefore, their claim that *"the 2D heterostructure is the tunneling path"* must be reconsidered. This is quite risky situation because the validity of above results has not been supported. This is also wasteful situation because the fertile variety of 2D materials cannot be utilized effectively. Due to this, 3) the *intrinsic* performances are still unknown. Therefore, the operation mechanisms should be comprehensively supported as soon as possible, and after that, the intrinsic performances should be investigated.

Table 1-3. Summary of the previous studies. Materials, structure and characteristics.

No.	Channel	Tunneling barrier	FG	Access region	I_d plateau† (scale)	Ohmic contact†	Ref.
1	MoS_2	*h*-BN	multi-layer graphene	No	Yes (log)		105
2	ReS_2	*h*-BN	Graphene	No	Yes (linear)	Confirmed	106
3	MoS_2	*h*-BN	graphene	No	Yes (linear)		107
4	PtS_2	*h*-BN	graphene	No	Yes (log)	Confirmed	102
5	$MoTe_2$	*h*-BN	graphene	No	Yes (linear)		98
6	MoS_2	*h*-BN	graphene	No	Yes (linear)		108
7	MoS_2	*h*-BN	metal	No	No	Confirmed	109
8	MoS_2	*h*-BN	MoS_2 & metal	No	Yes (linear)		100
9	$MoSe_2$	*h*-BN	graphene	No	Yes (log)	Confirmed	99
10	BP/graphene hetero-stack	*h*-BN	graphene	No	Yes (log)	Confirmed	110
11	WS_2	*h*-BN	multi-layer graphene	No	Yes (linear)	Confirmed	42
12	BP	*h*-BN	MoS_2	No	Yes (linear)	Confirmed	104
13	MoS_2	*h*-BN	graphene	No	Yes (linear)		41
14	BP	*h*-BN	graphene	No	Yes (log)	Confirmed	111
15	WSe_2	Al_2O_3	CdSe QDs	No	No	Confirmed	96
16	MoS_2	HfO_2	metallic nanocrystal	No	Yes (log)	Confirmed	94
17	MoS_2	HfO_2	few-layer graphene	No	No	Confirmed	97
18	MoS_2	Al_2O_3	HfO_2	No	No	Confirmed	112
19	MoS_2	*h*-BN	graphene	Yes*	No*		113
20	BP	*h*-BN	graphene	Yes	No	Confirmed	111

QDs in this table represent quantum dots and black phosphorus, respectively.
* The device has both top gate and back gate, and FG exists only in top gate side.
† Columns "I_d (drain current) plateau" and "ohmic contact" will be discussed in **chapter 4**.

1.5 Objectives and Organizations of the Thesis

The objective of this study is to reveal the comprehensive operation mechanisms and intrinsic performances of 2D heterostructured non-volatile memory devices. For the purpose, this study was conducted by following four steps.

Step 1: New measurement called *floating gate voltage (V_{FG}) measurement* is proposed.

Step 2: The validity of the memory window extraction method is discussed.

Step 3: Comprehensive understandings of memory operation are provided.

Step 4: Performances are reasonably investigated based on above understandings.

In **chapter 2**, V_{FG} measurement is proposed to unveil the operation mechanisms behind the standard *I–V* round sweep transfer curves. Starting from the device fabrication procedure, characterization of the fabricated devices by atomically force microscopy, Raman spectroscopy and transmission electron microscopy are explained. Following the conceptual explanation of the principle of V_{FG} measurement and its result called *V_{FG} trajectory*, that is experimentally confirmed. V_{FG} trajectory can be described by two regions, i.e., capacitive coupling region with tilted V_{FG} and feedback by tunneling region with pinned V_{FG}. Additionally, the effect of quantum capacitance of graphene FG on the V_{FG} trajectory is discussed.

In **chapter 3**, the validity of memory window extraction method is discussed based on the V_{FG} trajectory. The importance of memory window as a fundamental metrics is emphasized first, two different methods to extract the memory window, which depend on the research fields, are explained. Next, the memory window of fabricated $MoTe_2$ channel memory device is experimentally extracted by the two methods. Difference of obtained two memory windows are discussed based on the V_{FG} trajectory, and it finally be concluded that the memory window extracted by *I–V* round sweep transfer curves are generally overestimated. To expand the generality of the discussion, criterion for memory window overestimation is derived.

In **chapter 4**, operation mechanisms are comprehensively revealed. Proposed V_{FG} measurement is applied for fabricated three devices, i.e., MoS_2, WSe_2 and $MoTe_2$ channel devices. Inherent V_{FG} trajectories are resulted from each device whereas the standard *I–V* round sweep transfer curves are very similar. By analyzing the V_{FG} trajectories in detail, the operation mechanisms are understood in terms of three current limiting paths. Moreover, based on the understandings, V_{FG} trajectory can be controlled as expected, which further supports the validity of the understandings. Tunneling between S/D electrode and FG is experimentally proved for the first time, suggesting that the tunneling path claimed in most of previous study is invalid. Interestingly, the origin of drain current (I_d) plateau in round sweep transfer curves, which is often observed in previous study, can also be explained by the V_{FG} trajectory.

In **chapter 5**, performances of 2D heterostructured non-volatile memory is reasonably investigated as mainly focused on the P/E speed. Surprisingly, 50 ns P/E operation can be achieved by appropriately designed device, which is comparable to the operation speed of storage class memory. Although superior 2D/2D interface seems to be the key, control experiment reveals that it is not the key. Breakdown test by using the pulse generator suggest the strong breakdown strength of *h*-BN under fast voltage pulse stress is the key. In addition, retention and endurance of the MoS_2 channel device are also investigated. There are 10^5 s and 5×10^4 P/E cycles, respectively. They are comparable to previous studies.

In **chapter 6**, above findings are summarized. For future development of 2D heterostructured non-volatile memory technology, some outlooks are discussed.

It is noted that, while 2D heterostructured non-volatile memory devices (2D materials-based FG type flash memory cells, 3rd generation and beyond) are focused on throughout this thesis, most of the findings can be extended to previous generations as well.

Chapter 2
V_{FG} Measurement for 2D Materials-based Flash Memory

Different from conventional Si technology, a wide variety of 2D materials is a great advantage for designing memory devices. Especially, there are infinite combinations for constructing 2D heterostructured non-volatile memory (NVM) devices. In general, semi-metallic graphene or graphite is used as a floating gate (FG), *h*-BN is used as a tunneling barrier, and semiconducting 2D material such as transition metal dichalcogenides (TMDCs) and black phosphorus (BP) is used as a channel. In order to demonstrate its memory operation, drain current (I_d)–gate voltage (V_g) round sweep measurement is always employed, which may be resulted from the history of hysteresis analysis in 2D research field. However, the information obtained from I_d–V_g round sweep measurement is quite limited, e.g., channel polarity, subthreshold swing, on/off ratio and threshold voltages (V_{th}s). Moreover, most of 2D materials-based flash memory device, including 2D heterostructured NVM device, have shown similar I_d–V_g round sweep curves even though stacking materials are different. This obscures us to obtain a clear insight of their operation. In other words, the wide variety of 2D materials has not been properly exploited yet. Therefore, new measurement technique which gives us the clear insight has been strongly demanded.

In this chapter, floating gate voltage (V_{FG}) measurement for 2D materials based-flash memory device is proposed. Since V_{FG} governs the tunneling between the FG and channel, the result called V_{FG} trajectory is expected to unveil the operation mechanisms behind I_d–V_g round sweep transfer curves. Starting from the device fabrication procedure, the principle of the V_{FG} measurement and expected V_{FG} trajectory are explained. Following the experimental proof of the principle, effect of quantum capacitance of graphene FG is discussed, since monolayer graphene is often used as an FG.

2.1 Device Fabrication and Characterization

2.1.1 Device Structure

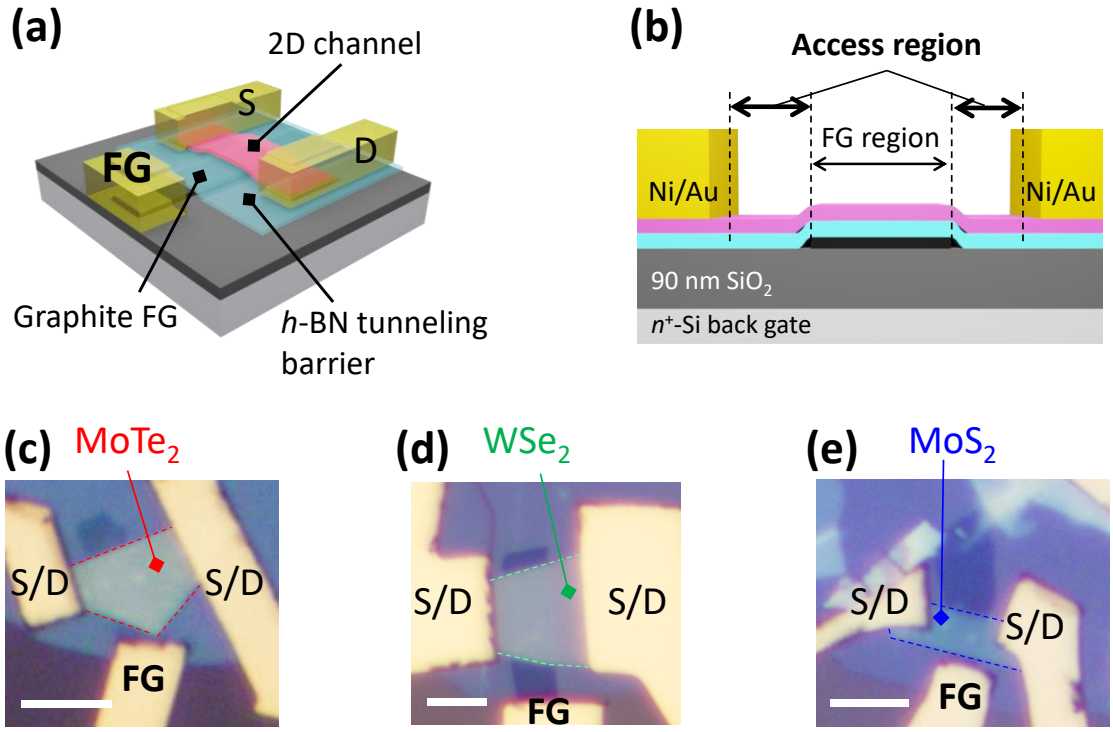


Fig. 2-1 Schematics and photos of 2D heterostructured NVM device. **(a)** Birds eye view and **(b)** cross-sectional view of the device. **(c)–(e)** Photos of fabricated $MoTe_2$, WSe_2, MoS_2 channel devices, respectively. The scale bars represent 5.0 μm.

Figs. 2-1(a) and **(b)** illustrates the schematics of 2D heterostructured NVM device. The heterostructure was fabricated on the 90 nm SiO_2/n^+-Si substrate which acts as a gate insulator and global back gate, respectively. Graphite and *h*-BN were used as an FG and a tunneling barrier, respectively, and TMDC was used as a channel material. The electrodes are composed of thermally evaporated Ni/Au stack. Since this is a standard heterostructure for 2D heterostructured memory, it can be expected that the advantage of proposed V_{FG} measurement over the conventional I_d–V_g measurement will be highlighted. Not only source and drain (S/D) electrodes, but also FG electrode was attached, which enables us to measure the V_{FG} and tunneling current between the FG and channel. As shown in **Fig. 2-1(b)**, the margin between graphite FG and S/D electrodes called access regions were

designed in order to restrict the tunnel path in the 2D heterostructure. The channel region over the FG is denoted as FG region in this study, to distinguish it from access region.

The photos of fabricated devices are shown in **Figs. 2-1(c)–(e)**. In this study, three different TMDCs, MoS_2, WSe_2 and $MoTe_2$ were used as a channel. The thickness of each 2D materials measured by atomic force microscope (AFM) are summarized in **Table 2-1**. Since 2D interface is a focus point of this study, thickness of the flakes is not crucial point, that is, monolayer TMDCs or graphene is not necessarily required. The values of thickness are interpreted that they have bulk properties.

Table 2-1. Thickness of each 2D material measured by AFM.

Device	Channel	Tunneling barrier	FG
$MoTe_2$ channel device	8.3 nm ($MoTe_2$)	15.0 nm (*h*-BN)	6.9 nm (Graphite)
WSe_2 channel device	4.5 nm (WSe_2)	9.6 nm (*h*-BN)	9.2 nm (Graphite)
MoS_2 channel device	3.2 nm (MoS_2)	10.6 nm (*h*-BN)	2.5 nm (Graphite)

Here, important properties of these TMDCs are summarized in **Table 2-2**[54]. According to the table, bulk bandgaps (E_gs) and polarity for Ni contact are different for each TMDC channels, that is, the effect of E_g and polarity of the channel on the device operation will be elucidated based on the appropriate comparison and analysis of the results. This is the reason why these TMDCs are selected as channel materials in this study.

Table 2-2. Important properties of TMDCs used for the channel[54].

	Bulk E_g	Polarity for Ni contact
$MoTe_2$	1.01 eV	Ambipolar
WSe_2	1.33 eV	Ambipolar
MoS_2	1.35 eV	*n*-type

2.1.2 Fabrication Procedure

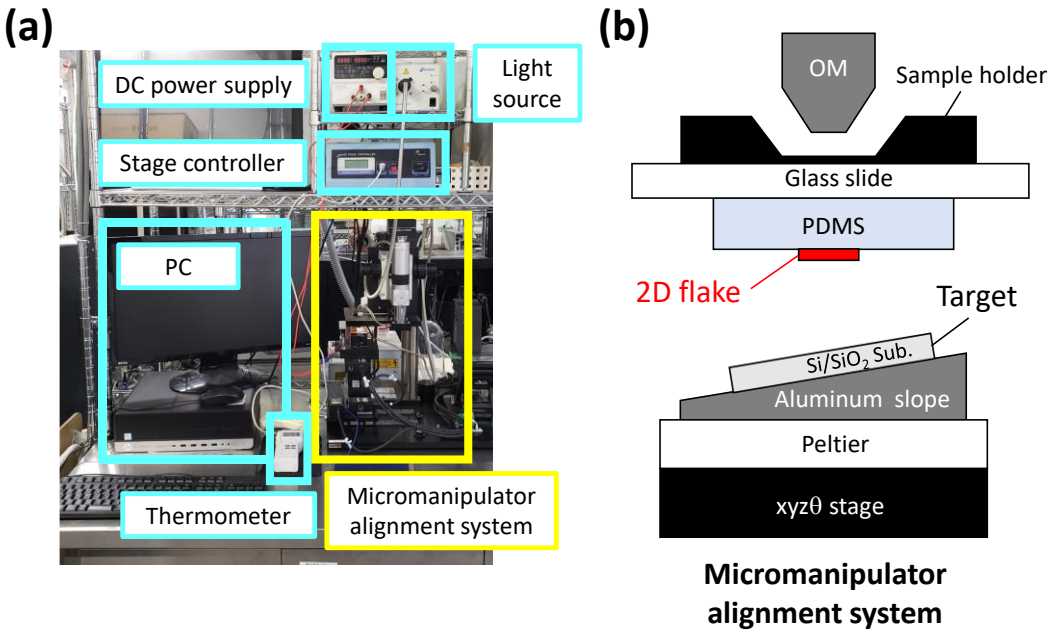


Fig. 2-2 (a) Photo of the dry transfer system and **(b)** schematic of the micromanipulator alignment system highlighted by yellow in **(a)**[70].

Next, fabrication procedure is explained in detail. The dry transfer system shown in **Fig. 2-2(a)** was employed to fabricate the 2D heterostructures. **Fig. 2-2(b)** shows the schematic of micromanipulator alignment system highlighted by yellow in **(a)**. Before the stacking, each 2D material was mechanically exfoliated by a scotch tape and transferred onto the polydimethylsiloxane (PDMS) film. The 90 nm SiO_2/n^+-Si substrate was pre-heated on a hot plate at 200°C for 2 minutes to eliminate the water molecules adsorbed on the surface.

After that, graphite FG, *h*-BN tunneling barrier and TMDC channel were subsequently transferred from PDMS to the substrate. As shown in **Fig. 2-2(b)**, an aluminum slope with ~3° of tilt was used to suppress the bubbles at hetero-interfaces[71] . In addition, substrate was heated up by Peltier element for 110–120°C for graphite and *h*-BN transfer, and for 50–70°C for TMDCs transfer, respectively. The DC supply shown in **Fig. 2-2(a)** is connected to the Peltier element and used for increasing its temperature.

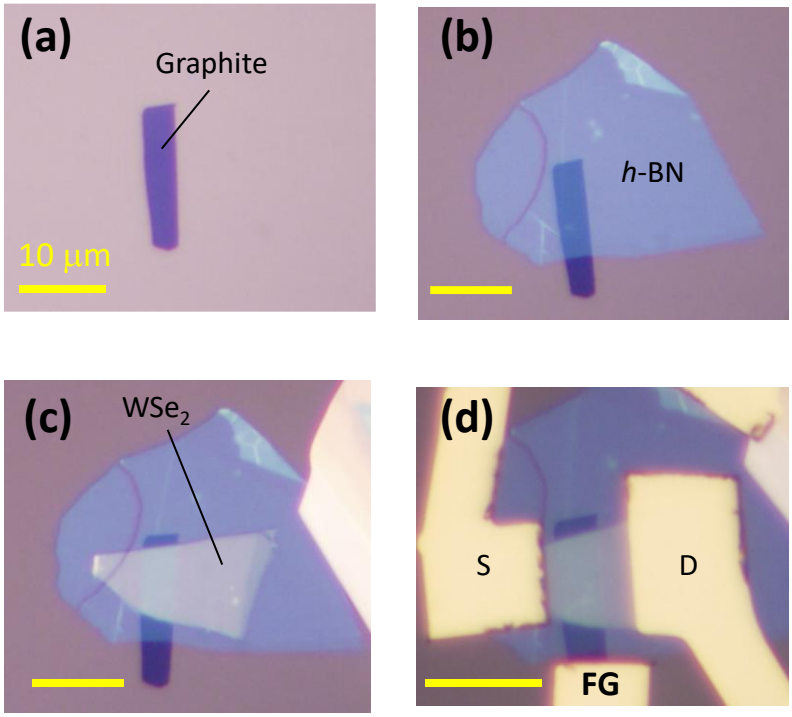


Fig. 2-3 Photos after each process. After **(a)** graphite transfer, **(b)** *h*-BN transfer, **(c)** WSe_2 transfer, and **(d)** electrodes fabrication.

Photos after each process of WSe_2 channel device are shown in **Figs. 2-3(a)–(d)**. Following the 2D dry transfer, standard electron-beam lithography was carried out to define the position of electrodes. For measuring the V_{FG}, not only S/D electrodes, but also FG electrode was designed. Then, Ni and Au were subsequently deposited by vacuum thermal evaporation system with 5×10^{-5}–1×10^{-4} Pa.

2.1.3 Characterization by Raman Spectroscopy

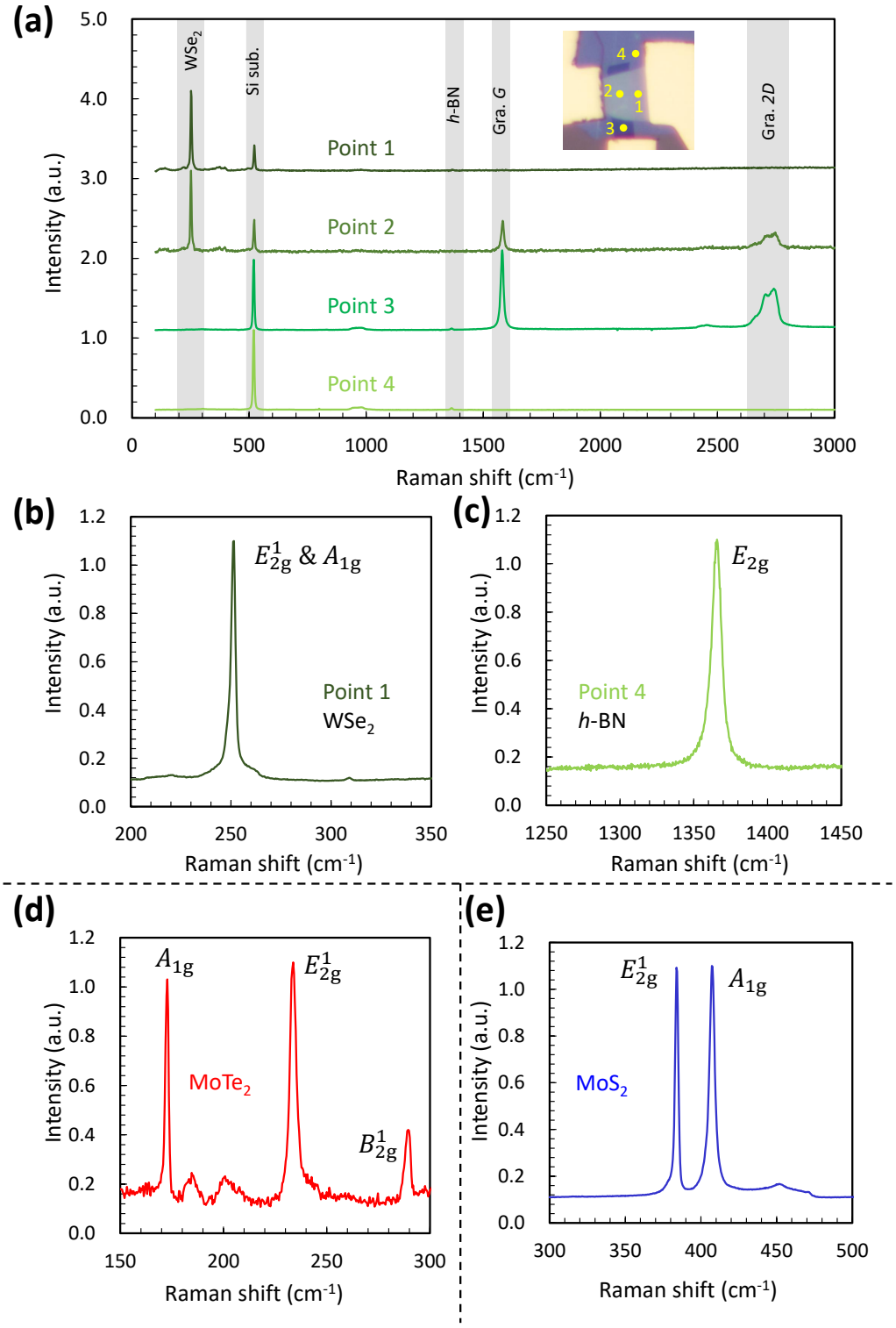


Fig. 2-4 (a) Wide range of Raman spectra of the WSe_2 device. Point 1–4 correspond to each point shown in the inset. **(b)** and **(c)** Enlarged spectra of the WSe_2 at point 1 and *h*-BN at point 4. **(d)** and **(e)** Raman spectra of $MoTe_2$ and MoS_2 used in the devices.

To characterize the crystallinity of each 2D material used in the 2D heterostructured NVM devices, Raman measurement was carried out. **Fig. 2-4(a)** shows the wide range of Raman spectra of the WSe_2 channel device. A position of each point is shown in the inset. The spectrum at each point well represents the stacking materials. **Figs. 2-4(b)** and **(c)** show the enlarged spectra of WSe_2 at point 1 and *h*-BN at point 4. Although E_{2g} peak of *h*-BN is not observed at point 1 and point 2 since there is a WSe_2 flake over the *h*-BN, it is clearly seen in the **Fig. 2-4(c)**. **Figs. 2-4(d)** and **(e)** show the enlarged spectra of $MoTe_2$ and MoS_2. These peak positions are well consistent with previous studies[114–118], and their sharpness represents the good crystallinity of each 2D material used in this study. Measurement condition is as follows. Laser powers are 0.38 mW for points 3 and 4 in **Figs. 2-4(a)** and **(c)**, and 0.06 mW for others, respectively. Laser spot size and wave length are ~1 μm and 488 nm, respectively.

2.1.4 Characterization by AFM and TEM

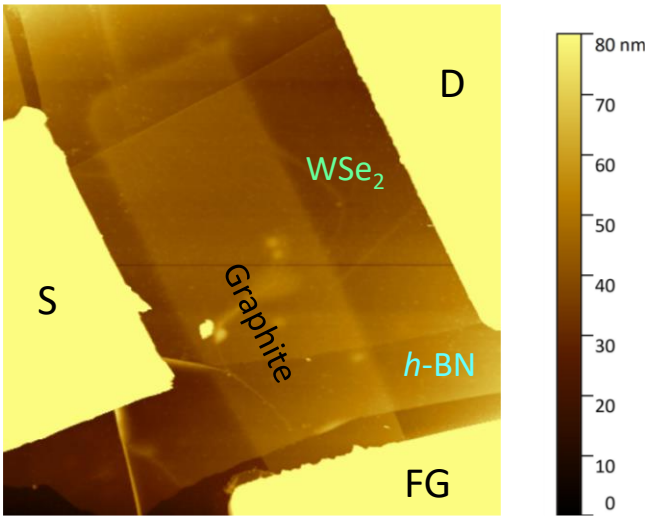


Fig. 2-5 AFM image of the WSe_2 device. Clean 2D/2D interfaces can be achieved by the dry transfer technique explained in **section 2.1.2**.

Next, quality of the 2D heterostructure was confirmed by AFM and transmission electron microscopy (TEM). **s. 2-5** shows the AFM result of the WSe_2 device. Thanks to the dry transfer technique as explained in the **section 2.1.2**, clean 2D/2D interfaces can be achieved.

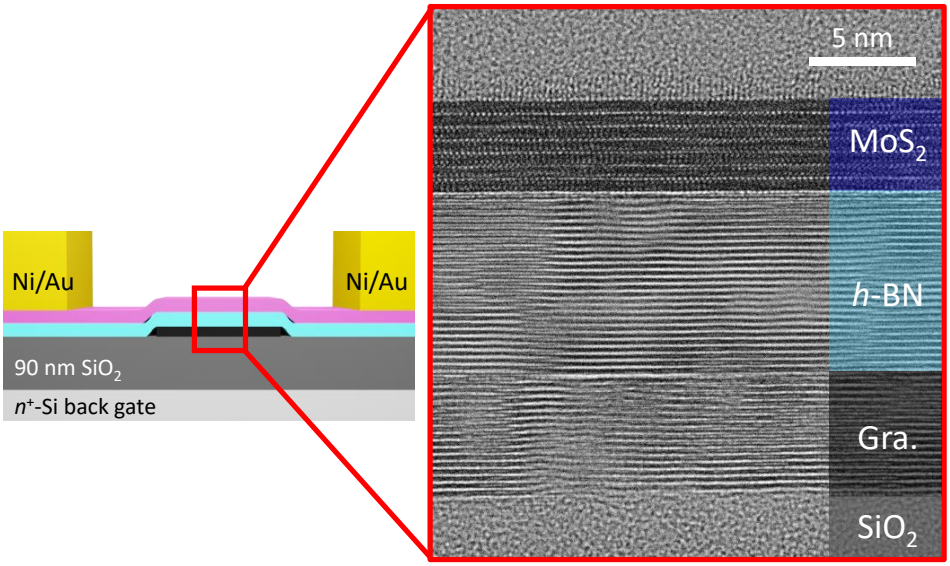


Fig. 2-6 Cross-sectional TEM image of typical MoS_2/*h*-BN/graphite heterostructure on SiO_2/n^+-Si substrate fabricated by the dry transfer technique.

Fig. 2-6 shows the cross-sectional TEM image of typical MoS_2/*h*-BN/graphite heterostructure fabricated by using the transfer system. Atomically sharp interfaces are clearly observed. Therefore, it can be said that the dry transfer technique can fabricate a high-quality 2D heterostructure.

2.2 Principle of V_{FG} Measurement

2.2.1 Measurement Setup

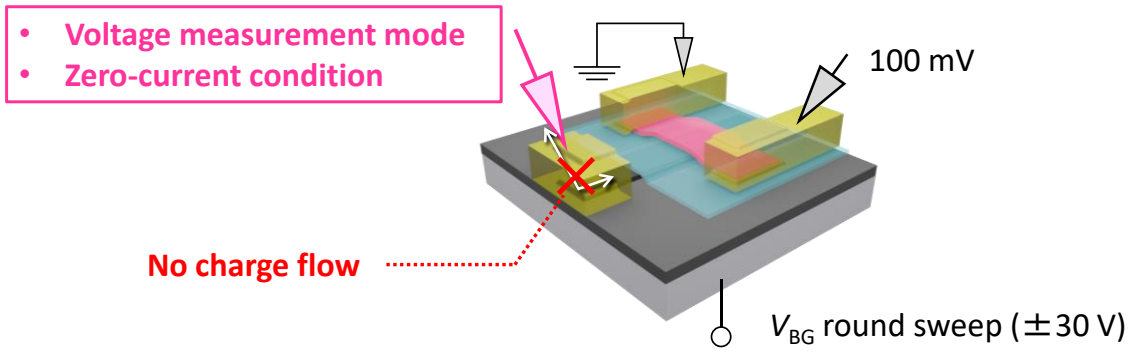


Fig. 2-7 Schematic of the setup for V_{FG} measurement.

Next, the principle of V_{FG} measurement is explained. **Fig. 2-7** shows the setup for V_{FG} measurement. The probe is contacted to the FG electrode and it is set to be voltage measurement mode by using the semiconductor parameter analyzer (Keysight B1500). Since the probe is set to be zero-current condition, no charge flow between the probe and FG during the measurement can be ensured. Thus, V_{FG} and I_d can be measured simultaneously during back gate voltage (V_{BG}) round sweep. Standard round sweep in this study is that, V_{BG} is swept from –30V to +30 V first (positive sweep), and swept from +30 V to –30 V next (negative sweep).

2.2.2 Expected V_{FG} Trajectory

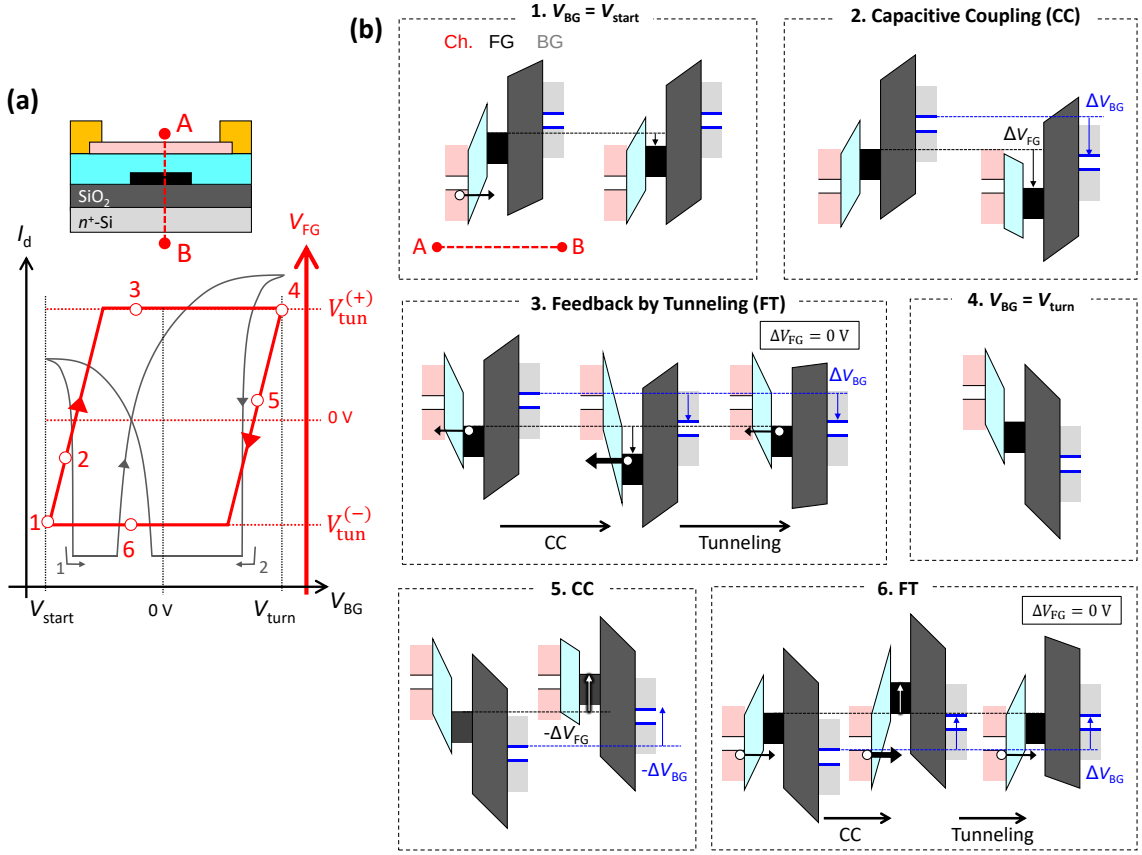


Fig. 2-8 The principle of V_{FG} trajectory. **(a)** Schematic of cross section of the device and expected V_{FG} trajectory superimposed on a sketch of typical I_d–V_{BG} transfer curves. **(b)** Band diagrams along the 2D heterostructure (A–B) shown in **(a)**. Six points correspond to each point of V_{FG} trajectory labeled in **(a)**.

In order to understand the novel result called V_{FG} trajectory, expected one is considered first with band diagrams. **Fig. 2-8(a)** schematically illustrates the cross section of the device and expected V_{FG} trajectory superimposed on typical I_d–V_{BG} transfer curves. **Fig. 2-8(b)** shows the band diagrams along the line A–B shown in **(a)**. The band diagrams are divided into six parts and they correspond to the important points of the trajectory labeled in **(a)**. In this explanation, tunneling carrier is assumed to be hole since hole tunneling is experimentally confirmed as discussed in **chapter 4**. However, electron tunneling is also applicable since electron tunneling from channel to FG is equivalent to hole tunneling from FG to channel, in terms of FG potential modulation.

At first, V_{BG} is swept from V_{start} to V_{turn} as shown in **Fig. 2-8(a)**. Point 1 is the

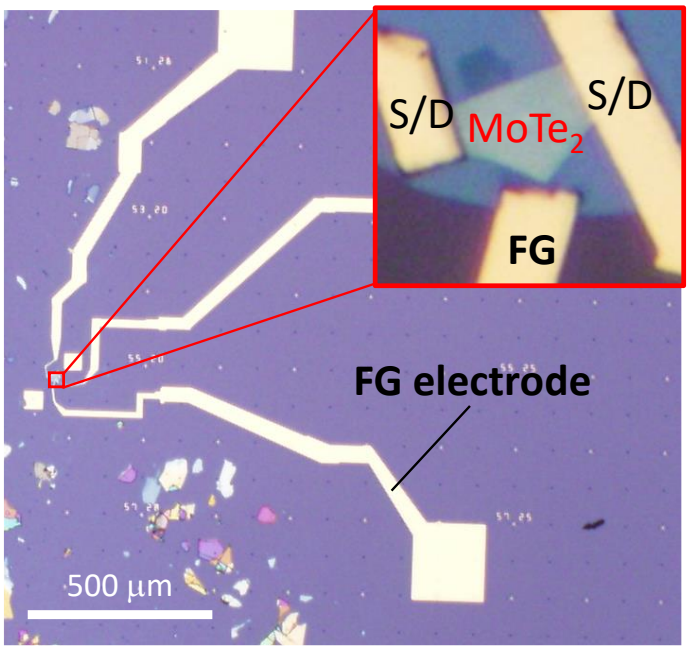


Fig. 2-9 Photo of typical FG electrode. The $MoTe_2$ device is used as an example. Huge FG electrode results in a condition of $C_{ox} >> C_{BN}$.

starting point of V_{BG} positive sweep. Just after the application of V_{start} (<0 V) to the BG, V_{FG} is strongly decreased to negative value due to capacitive coupling between FG and BG, resulting in a hole tunneling from the channel to FG. Since accumulated holes in the FG increase the V_{FG}, hole tunneling is stopped when V_{FG} reaches the negative side of tunnel starting voltage ($V_{tun}^{(-)}$). Therefore, V_{FG} trajectory is started from $V_{tun}^{(-)}$. At the initial stage of V_{BG} positive sweep, V_{FG} is linearly increased with increasing V_{BG} due to the capacitive coupling. This is labeled as point 2, and called capacitive coupling (CC) region in this study. In this region, the slope of V_{FG} can be written by following equation[17]:

$$\frac{dV_{FG}}{dV_{BG}} = \frac{C_{ox}}{C_{ox} + C_{BN}}, \tag{2.1}$$

where, C_{ox} and C_{BN} are the capacitance of SiO_2 between the FG and BG, and the capacitance of *h*-BN between the channel and FG, respectively. Although **Eq. 2.1** is valid when relatively thick graphite is used as an FG since the graphite can completely screen the electric field from BG, monolayer graphene has been often used as an FG (see **table 4-1**). In that case, the role of quantum capacitance become important as discussed in the

section 2.4. Additionally, as shown in the **Fig. 2-9**, the area of FG electrode (>100 μm × 100 μm) is much larger than that of the channel and FG overlap area (~2.0 μm × 2.0 μm), resulting in a condition of $C_{ox} >> C_{BN}$. Therefore, the slope of V_{FG} against V_{BG} is almost one. Beyond the point 2, hole tunneling from graphite FG to the channel is occurred when the V_{FG} reaches the positive side of tunnel starting voltage ($V_{tun}^{(+)}$). As illustrated in point 3 of **Fig. 2-8(b)**, just after the increment of V_{BG} (ΔV_{BG}), V_{FG} is also increased due to capacitive coupling. However, the hole tunneling returns V_{FG} to $V_{tun}^{(+)}$, resulting in a flat trend of V_{FG} pinned at $V_{tun}^{(+)}$. This plateau region is called feedback by tunneling (FT) region in this study. Even at the end of positive sweep, point 4, V_{FG} is remained at $V_{tun}^{(+)}$. Therefore, when sweep direction is switched back at $V_{BG} = V_{turn}$, V_{FG} is immediately decreased with V_{BG} as illustrated in point 5 of **Fig. 2-8(b)**. Since this is also the capacitive coupling region, the slope, dV_{FG}/dV_{BG}, is same as that at the point 2. Finally, V_{FG} reaches $V_{tun}^{(-)}$ and the hole tunneling from the channel to FG leads to the flat trend of V_{FG} as illustrated in point 6 of **Fig. 2-8(b)**. In short, the increase/decrease trend of the V_{FG} is due to capacitive coupling between the FG and BG, whereas flat trend of V_{FG} is due to the feedback mechanism, in which the hole tunneling cancel the V_{FG} modulation. Although the capacitive coupling ratio less than one is a common situation since some devices do not have an FG electrode, above mentioned principle of V_{FG} trajectory is maintained.

2.3 Experimental Result of V_{FG} Measurement

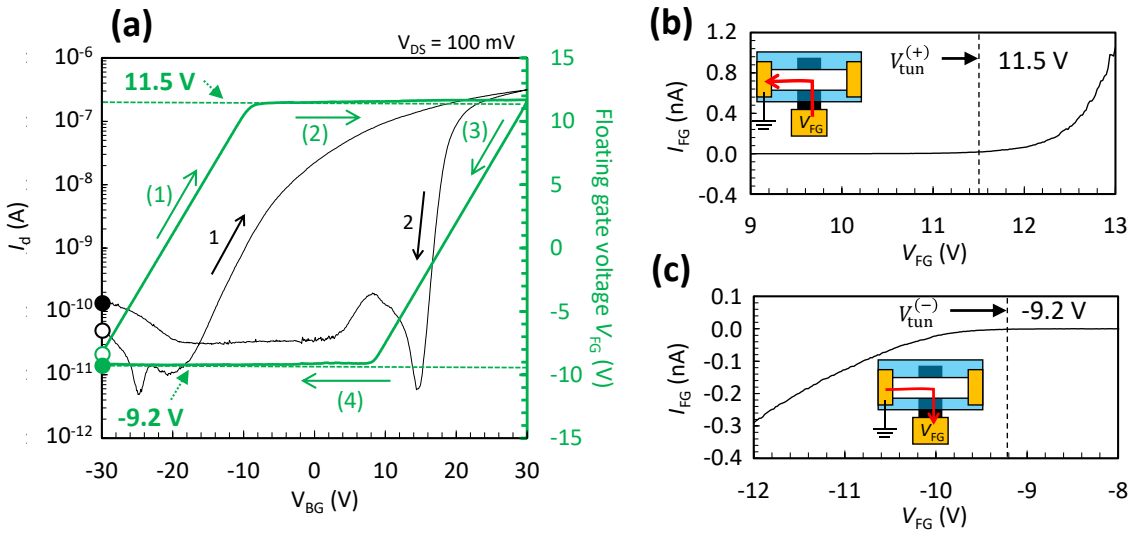


Fig. 2-10 (a) Measured V_{FG} trajectory of $MoTe_2$ device superimposed on corresponding I_d–V_{BG} transfer curves. **(b)**, **(c)** Measured tunneling currents between the FG and channel with grounded source electrode. To measure the current, V_{FG} was manually swept.

In this section, experimental investigation of the principle is explained. Measured V_{FG} trajectory of the $MoTe_2$ device superimposed on its I_d–V_{BG} transfer curves is shown in **Fig. 2-10(a)**. The parallelogram shape trajectory can be obtained as expected in **section 2.2**. To confirm the consistency of tunnel starting voltages $V_{tun}^{(+)}$ and $V_{tun}^{(-)}$, the tunneling current between the FG and channel was measured by manually sweeping the V_{FG}, with grounded source electrode. **Figs. 2-10(b)** and **(c)** show the measured tunneling currents from the FG to channel, and from the channel to FG, respectively. It is clear that +11.5 V of $V_{tun}^{(+)}$ and –9.2 V of $V_{tun}^{(-)}$ shown in **Fig. 2-10(a)** are consistent with the tunnel starting voltages shown in **Figs. 2-10(b)** and **(c)**. These experimental data support the principle mentioned in previous section. In addition, negligible effect of V_{FG} measurement for I_d–V_{BG} round sweep transfer curves was confirmed as shown in **Fig. 2-11**.

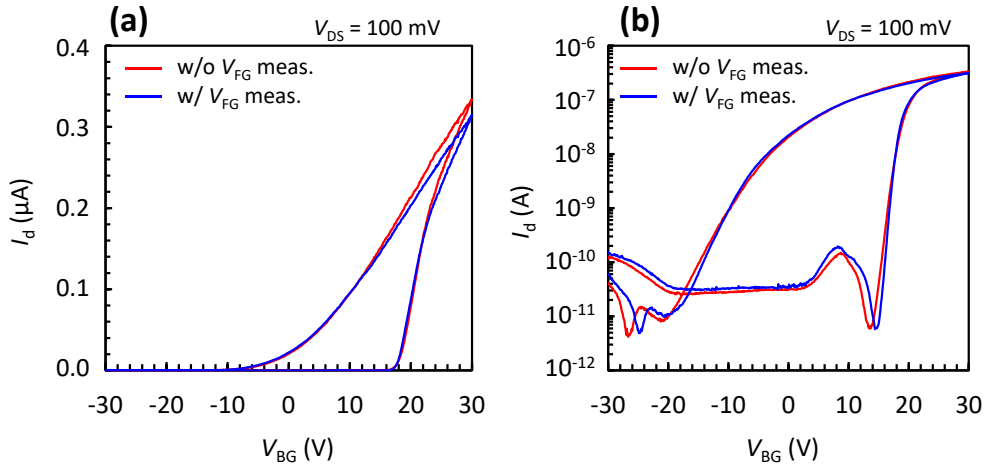


Fig. 2-11 I_d–V_{BG} transfer curves with and without V_{FG} measurement in **(a)** linear scale and **(b)** semi-log scale. Negligible effect of V_{FG} measurement on I_d–V_{BG} measurement was confirmed.

2.4 Effect of Quantum Capacitance of Graphene FG

Finally, the effect of quantum capacitance of monolayer graphene FG is discussed since it has been commonly used in previous studies (see **table 4-1**).

2.4.1 Capacitive Coupling Region

Fig. 2-12 shows the schematic and equivalent circuit of the 2D heterostructured NVM device with monolayer graphene FG. When FG electrode is considered, the capacitance between FG and BG (C_{ox}) can be divided into two parts. One is C_{ox}^{metal} which denotes the capacitance between FG electrode and BG, and the other is C_{ox}^{gra} which denotes the capacitance between monolayer graphene FG and BG. Therefore, C_{ox} can be written as follows:

$$C_{ox} = C_{ox}^{metal} + C_{ox}^{gra}. \tag{2.2}$$

In addition, C_Q^{gra} shown in **Fig. 2-12** is the quantum capacitance of monolayer graphene FG. Each capacitance and its active area are indicated by color in the figure. Here, C' is defined as follows:

$$C' = \frac{C_{BN} C_Q^{gra}}{C_{BN} + C_Q^{gra}}. \tag{2.3}$$

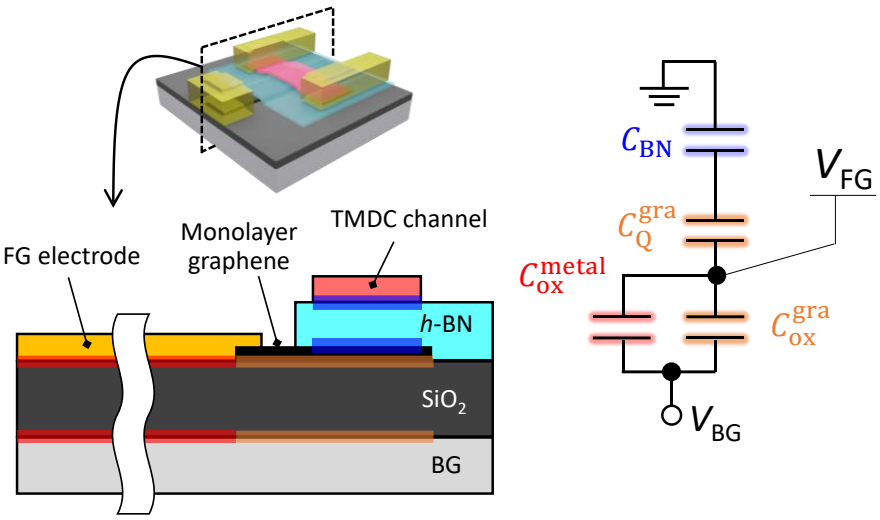


Fig. 2-12 Cross-sectional view of the device with monolayer graphene FG and its equivalent circuit. Each capacitance and its active area are indicated by color.

Therefore, **Eq. 2.1** can be rewritten as follows:

$$\frac{dV_{\mathrm{FG}}}{dV_{\mathrm{BG}}} = \frac{C_{\mathrm{ox}}}{C_{\mathrm{ox}} + C'}. \tag{2.4}$$

Depending on the Fermi level position of the monolayer graphene, two limiting cases can be considered. One is $C_{\mathrm{BN}} \gg C_{\mathrm{Q}}^{\mathrm{gra}}$ when the Fermi level is located near the Dirac point, and the other is $C_{\mathrm{BN}} \ll C_{\mathrm{Q}}^{\mathrm{gra}}$ otherwise. The former leads to $C' \approx C_{\mathrm{Q}}^{\mathrm{gra}}$, and the latter leads to $C' \approx C_{BN}$. However, in both cases, C_{ox} is much larger than C' because of the existence of large FG electrode. Therefore, when large FG electrode is formed as shown in **Fig. 2-9**, there is no influence from the quantum capacitance of monolayer graphene on the capacitive coupling region.

Of course, FG electrode is not always required, even though it can be used for measuring not only V_{FG} trajectory but also tunneling current between the FG and channel. When FG electrode is not fabricated, C_{ox} equals to $C_{\mathrm{ox}}^{\mathrm{gra}}$. As an example, typical situation is considered as follows. A gate dielectric and a tunneling barrier are assumed to be 90 nm SiO_2 with 3.9 of ε_r and 10 nm *h*-BN with 3.0 of ε_r, respectively. Here, ε_r is relative permittivity. Since ε_r of *h*-BN has been reported as 2–4[50], ε_r of *h*-BN is assumed to be 3.0 for following calculations. In addition, active area for C_{ox} and $C_{\mathrm{Q}}^{\mathrm{gra}}$ is assumed to be

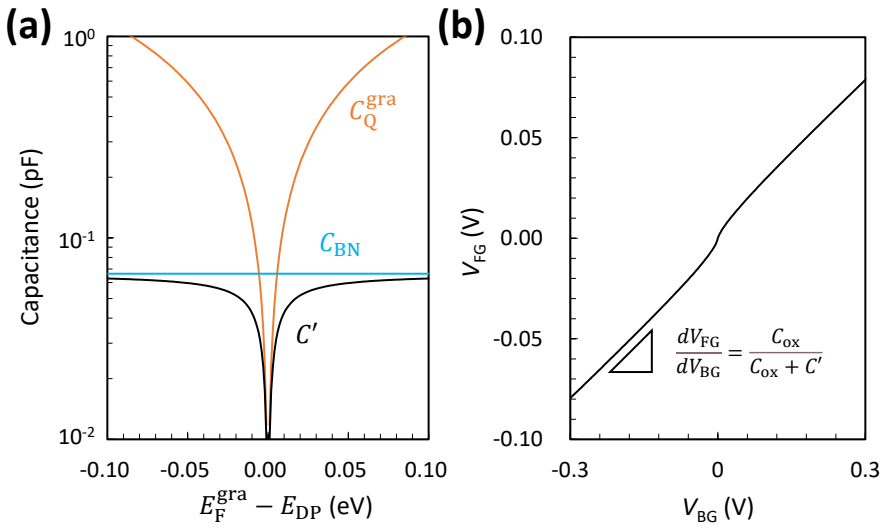


Fig. 2-13 (a) Calculated capacitances and **(b)** V_{FG} as a function of V_{BG} in CC region.

5.0 μm × 10.0 μm, and it for C_{BN} is assumed to be 5.0 μm × 5.0 μm. The former and the latter are indicated by orange and blue in **Fig. 2-12**, respectively. Here, C_Q^{gra} can be calculated as follows:

$$C_Q^{gra} = q^2 DoS_{gra}(E_F^{gra}), \tag{2.5}$$

where, q is a charge of electron and $DoS_{gra}(E_F^{gra})$ is a density of states of graphene as a function of Fermi level of graphene (E_F^{gra}). Density of states of graphene can be calculated as follows[119]:

$$DoS_{gra}(E_F^{gra}) = \frac{2|E_F^{gra} - E_{DP}|}{\pi(\hbar v_F)^2}, \tag{2.6}$$

where, E_{DP}, $\hbar$ and v_F are the energy at Dirac point of graphene, reduced Plank's constant and Fermi velocity of electrons in graphene. **Fig. 2-13(a)** shows the calculated C_Q^{gra} and C' against $E_F^{gra} - E_{DP}$. It is obvious that, C' is deviated from C_{BN} only when $E_F^{gra} - E_{DP}$ is in the energy range of ±0.05 eV, that is, the effect of C_Q^{gra} is negligible except when E_F^{gra} is near E_{DP}. V_{FG} can be calculated as a function of V_{BG} by using above equations and the relationship $E_F^{gra} - E_{DP} = qV_{FG}$. The calculation result is shown in **Fig.**

2-13(b), where $E_F^{gra} = E_{DP}$ when V_{BG} = 0 V is assumed for simplicity. Since typical situation is assumed for the calculation, small influence of C_Q^{gra} on the V_{FG} trajectory in the capacitive coupling region is elucidated even when the device has no FG electrode.

2.4.2 Feedback by Tunneling Region

FT region is considered next. It is quite helpful to start from the discussion of tunneling in the metal/*h*-BN/graphene heterostructure on SiO_2/Si substrate shown in **Fig. 2-14(a)**, which has been investigated by the previous study[120]. The literature has experimentally proved that the barrier height for holes can be modulated by the electric field from BG, when holes are tunneled from the graphene to metal. The band diagram of this system and measured tunneling currents are shown in **Figs. 2-14(b)** and **(c)**, respectively. The application of positive voltage to the BG leads to an increase of Fermi level of the graphene, resulting in higher barrier height for holes, and vice versa. This is clearly observed in left side of **Fig. 2-14(c)**. On the other hand, when holes are tunneled from the metal to graphene, Fermi level of the metal is not modulated by the electric field from BG due to its large density of states. Thus, measured tunneling currents shown in right side of **Fig. 2-14(c)** does not depend on the V_{BG}.

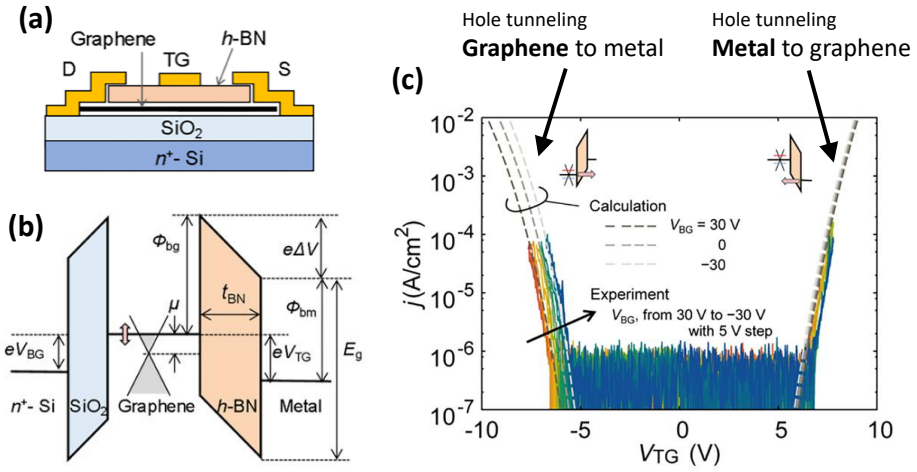


Fig. 2-14 (a) Schematic and **(b)** band diagram of metal/*h*-BN/graphene heterostructure on SiO_2/Si substrate. **(c)** Measured tunneling current in this system[120].

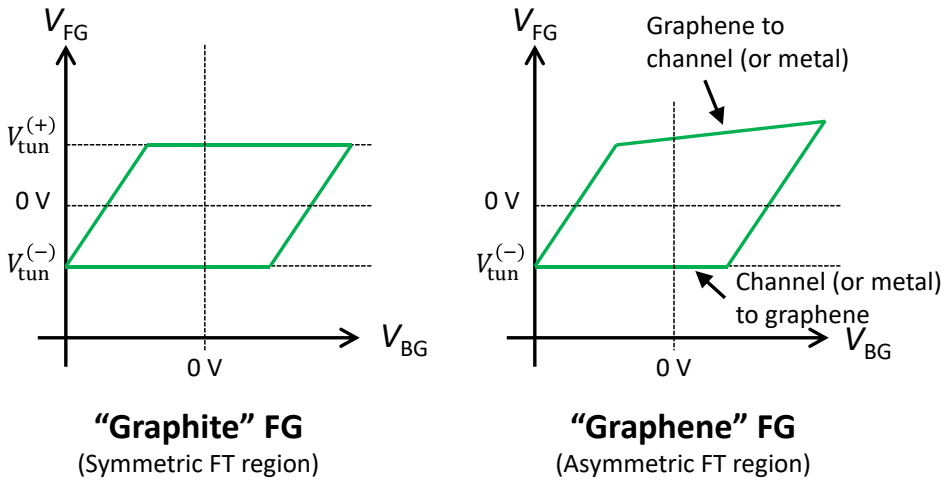


Fig. 2-15 Comparison of the expected V_{FG} trajectory between bulk graphite FG case and monolayer graphene FG case. Asymmetric FT region can be expected due to Fermi level modulation of monolayer graphene when it is used as an FG.

Considering the results mentioned above, the asymmetric FT region can be expected as illustrated in **Fig. 2-14**. In this figure, the simplest case is assumed. The deviation from the principle will be discussed in **chapter 4**. During V_{BG} positive sweep, since holes in graphene FG are tunneled to the channel, V_{FG} in FT region is not absolutely flattened but slightly increased with increasing V_{BG} due to Fermi level modulation of the graphene FG. On the contrary, during V_{BG} negative sweep, the V_{FG} is absolutely flattened because holes are tunneled from the channel or metal S/D electrode, whose quantum capacitance can be neglected due to their large density of states, to the FG.

In other words, higher V_{FG} is required when a program operation is conducted with +30 V of V_{BG}, as compared with the case of graphite FG. Since it may negatively affect the performances such as low program speed or low reliability due to higher required electric field, graphite is used as an FG in this study, instead of monolayer graphene.

2.5 Summary

In this chapter, V_{FG} measurement technique was proposed, and its principle was experimentally demonstrated by measuring the $MoTe_2$ device with an FG electrode. This simple technique plays an important role throughout this study since V_{FG} trajectory unveils the device characteristics behind the I_d–V_g curves.

Chapter 3
Memory Window Overestimation

There are many performance metrics for evaluating memory devices, e.g., program and erase (P/E) speed, retention and endurance. Since there is a long history for Si based-flash memory research, the evaluating methods have been developed for them. This means that, when we start to evaluate new devices, 2D materials based-flash memory devices, the validity of the methods should be checked. To understand the current situation, each research history is looked back. Around 2010, intrinsic property of graphene was eagerly explored. For quantitative evaluation of some imperfectness, the hysteresis in *I–V* round sweep was often evaluated. This made the hysteresis measurement popular in 2D research. As a result, the hysteresis measurement has still been employed for the study of 2D materials based-flash memory, and more importantly, measured hysteresis has been interpreted as a memory window. On the contrary, for Si-based flash memory research, nearly 40 years have passed since 1984 when F. Masuoka proposed the flash memory[3]. In this research field, memory window is extracted from single sweep *I–V* curves. Although hysteresis measurement is sometimes employed, it is for the study of charge trapping[31,121], not for the memory window. Consequently, there is a huge gap in terms of the memory window extraction, resulting in the lack of fair comparison of them.

In this chapter, the importance of memory window is highlighted first, and then, the gap is explained with experimental results. After that, floating gate voltage (V_{FG}) measurement proposed in this study is exploited to discuss the origin of the gap. Memory window overestimation by *I–V* round sweep is finally revealed. Moreover, the criterion of the memory window overestimation is discussed for expanding the generality.

3.1 Memory Window as a Fundamental Metric

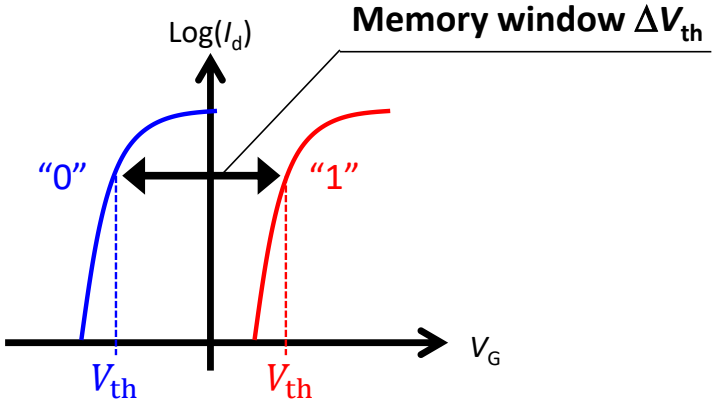


Fig. 3-1 The definition of memory window.

As shown in **Fig. 3-1**, the difference between threshold voltages (V_{th}s) is defined as a memory window (ΔV_{th}). This implies that how clearly "0" and "1" can be distinguished.

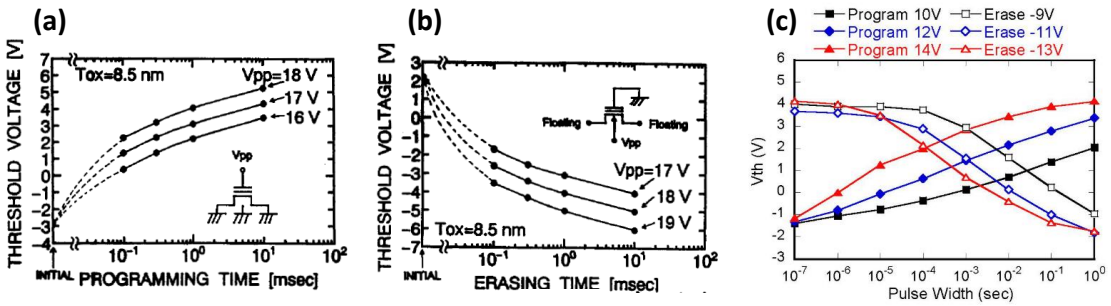


Fig. 3-2 Typical **(a)** program and **(b)** erase characteristics of planer NAND flash memory[122]. **(c)** Typical P/E characteristics of 3D NAND flash memory[23].

Some basic performance metrics such as P/E speed, retention and endurance are described in term of memory window. In case of P/E speed evaluation, as shown in **Fig. 3-2**, the V_{th}s after P/E operation are plotted against P/E pulse width[23,122]. This plot is quite useful for designing the memory, because if the required memory window is determined by the requirement of peripheral circuit, and/or allowable variability of cells, acceptable P/E speed is determined, and vice versa. In case of endurance and retention, the V_{th}s are

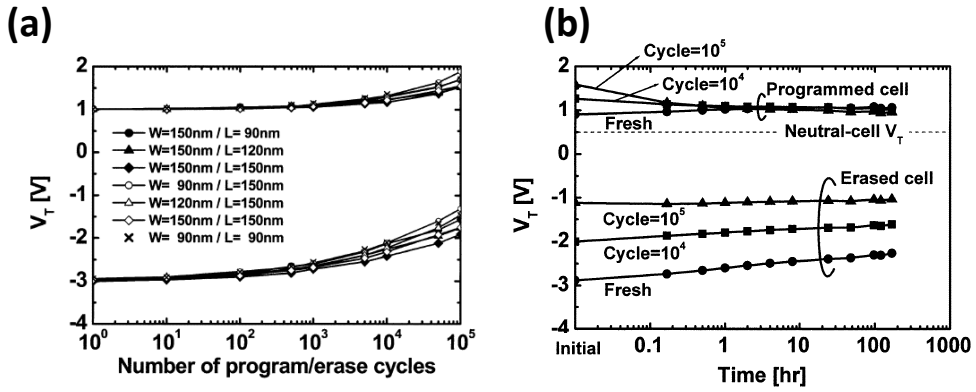


Fig. 3-3 Typical **(a)** endurance and **(b)** retention characteristics of planer NAND flash memory[20].

plotted against P/E cycles and data retention time, respectively. Typical characteristics of Si-based flash memory cell are shown in **Fig. 3-3**[20]. The direction of V_{th} shift (increase/decrease) suggests the polarity of discharged or trapped charge, and the amount of V_{th} shift indicates how many charges are trapped or detrapped, that is, these plots enable us to discuss the origin of the degradation. Since the origin is very important for further improvement, memory window should be discussed in terms of ΔV_{th}. Noted that, the difference of drain current (I_d) at gate voltage (V_g) = 0 V is sometimes considered as a window, since the current is used to distinguish the high resistance state and low resistance state in practical read operation. Of course, the origin cannot be discussed by the current difference. Moreover, the feasibility of the multi-level cell is also evaluated by the memory window. Recently, its application has been expanded towards the artificial synaptic device in which the precise control of channel conductivity is required[100,101,108,113]. Therefore, it is obvious that the memory window is the most fundamental metric.

However, there are two different memory window extraction methods depending on the research field. There is no discussion which method should be used for 2D materials based-flash memory devices. This is quite serious because the validity of other reported values including retention and endurance have not been ensured.

3.2 Memory Window Extraction Methods

3.2.1 *I–V* Round Sweep in 2D Research Field

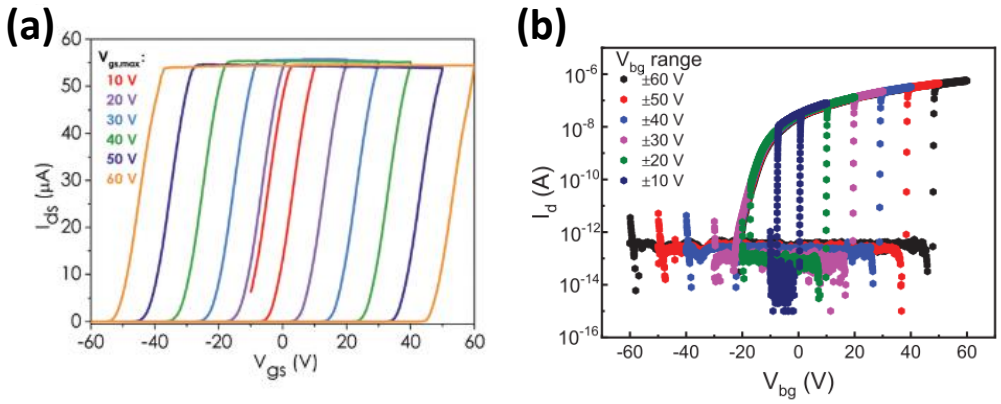


Fig. 3-4 Typical I_d–V_g round sweep curves of 2D heterostructured NVM devices. **(a)** ReS_2/*h*-BN/graphene heterostructured device[106], and **(b)** *h*-BN/MoS_2/*h*-BN heterostructured device with metal floating gate[109].

Fig. 3-4 shows the typical I_d–V_g round sweep curves of 2D heterostructured non-volatile memory (NVM) device[106,109]. The memory window has been defined as the margin of V_{th}s extracted from positive and negative sweep curves. Although this is the same as hysteresis measurement, it is commonly used in 2D research field. Interestingly, the I_d–V_g round sweep curves have often resulted in relatively large window. For example, D. Qiu et al. have reported the over 20 V of memory window with only ±25 V of V_g sweep range[42]. They measured the WS_2/*h*-BN/multi-layer graphene heterostructured NVM device on 280 nm SiO_2/Si substrate. R. Cheng et al. have reported that over 140 V of quite large memory window can be achieved when V_g sweep range is ±90 V with $MoSe_2$/*h*-BN/graphene heterostructured NVM device on 300 nm SiO_2/Si substrate[99]. In addition, linear relationship between memory window and V_g sweep range has often been reported in previous reports[98,99,106,109]. However, the origin of these huge memory windows has not been revealed yet.

3.2.2 *I–V* Single Sweep in Si Research Field

Different from the 2D research field, *I–V* single sweep is a standard method to extract the memory window in Si research field. At first, program operation is carried out by applying relatively high gate voltage to the control gate (typically, 16–20 V[17]). After the program operation, I_d–V_g single sweep is carried out with small V_g range to prevent the additional program. Consequently, V_{th} after program operation can be extracted. Similarly, V_{th} after erase operation can be extracted. After measuring both V_{th}s, memory window can be calculated by the difference of them.

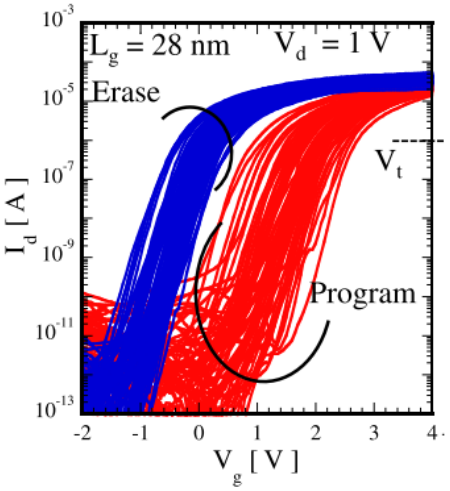


Fig. 3-5 Typical I_d–V_g singe sweep curves of Si based-flash memory cell after P/E operation[123].

Fig. 3-5 shows the typical I_d–V_g single sweep curves after P/E operation[123]. Although +16 V and –16 V were used for P/E operation respectively, memory window is about 2 V, which is quite smaller than that of 2D materials based-flash memory device.

3.3 Analysis by V_{FG} Trajectory

3.3.1 Extracted Memory Windows of the $MoTe_2$ Channel Device

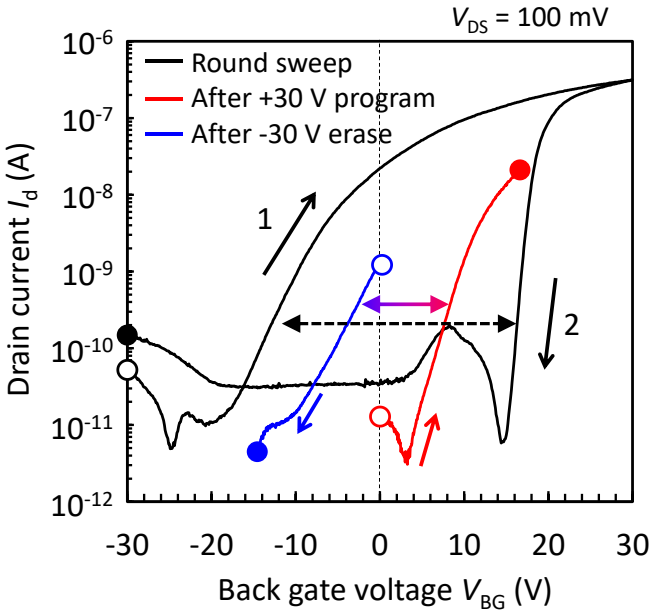


Fig. 3-6 Round sweep and single sweep I_d–V_{BG} curves of the $MoTe_2$ device. Open and fill symbols indicate the start and end points of the sweep.

In this section, difference between these two memory windows and its origin are experimentally investigated. As the test device, fabricated $MoTe_2$ channel device was employed. Since the hole tunneling in metal/*h*-BN/graphene heterostructure has been reported[120], that can also be expected for the fabricated devices with graphite floating gate (FG). Although the MoS_2 is a popular 2D semiconductor in 2D research field, it has been reported that strong Fermi level pinning at the metal/MoS_2 interface prevents the hole injection from metal to MoS_2[59]. This may obscure the focus point. On the contrary, ambipolar nature of $MoTe_2$ is quite helpful to focus on the memory window, since the hole injection from metal to $MoTe_2$ is available. At first, memory window of the $MoTe_2$ device was extracted by above mentioned two methods. **Fig. 3-6** shows the three kinds of I_d–back gate voltage (V_{BG}) curves, in which the round sweep curves are represented by black, and single sweep curves after P/E operation are represented by red and blue, respectively. The P/E operation was conducted by applying ±30 V to the BG for 10 s to completely charge

or discharge the FG.

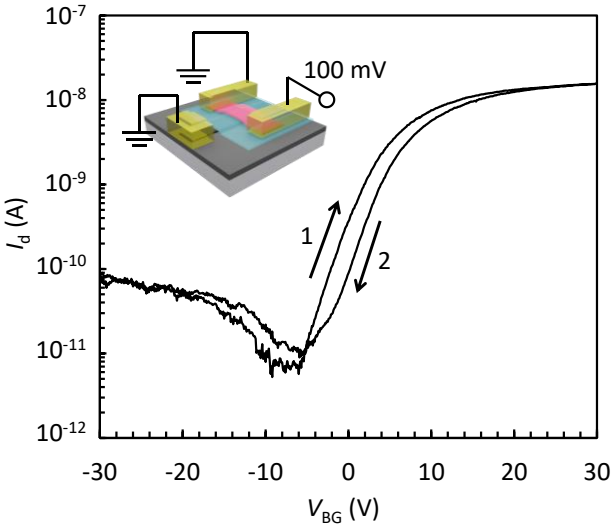


Fig. 3-7 I_d–V_{BG} round sweep curves with grounded FG.

In the round sweep curves, ambipolar nature of $MoTe_2$ and large memory window (~30 V) are clearly observed, whereas the set of single sweep curves results in +11.6 V of small memory window. Moreover, in the case of round sweep, the slope of negative sweep is steeper than that of positive sweep. This is strange, because the both slopes should be the same if the charges stored in the FG shift the I_d-V_{BG} curve horizontally. It can be considered that there is another mechanism to change the slope. Of course, a negligible hysteresis due to interface trapped charges or adsorbed water molecules[124], has been confirmed as shown in **Fig. 3-7**, where I_d–V_{BG} round sweep curves with grounded FG are shown. Negligible hysteresis means that the large memory window shown in **Fig. 3-6** is attributed to the charges stored in the FG.

3.3.2 *I–V* Round Sweep Case

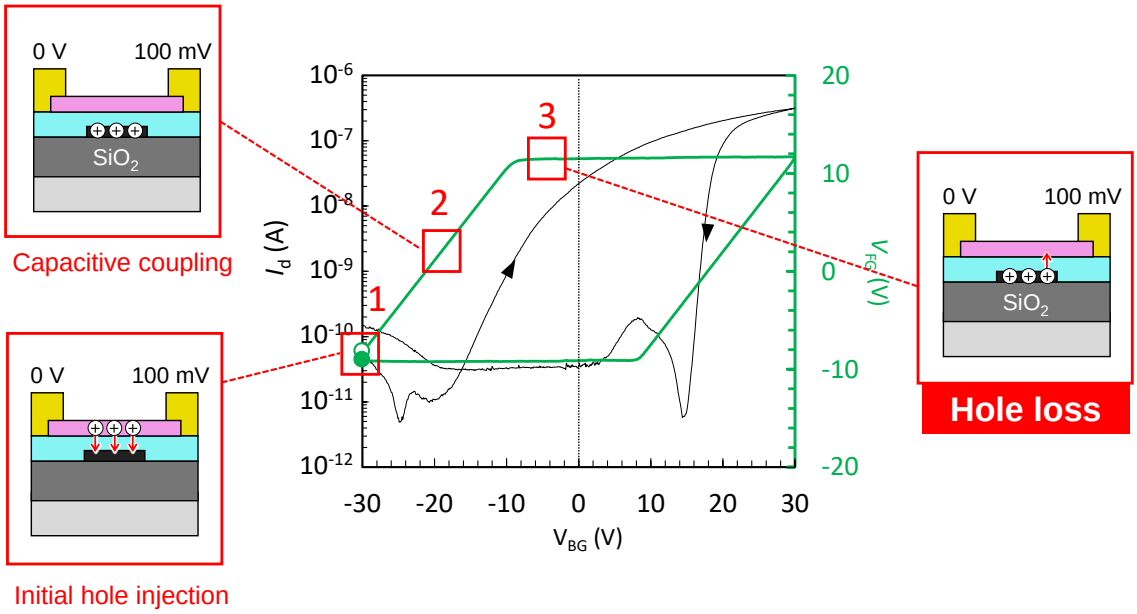


Fig. 3-8 Hole loss during I_d–V_{BG} round sweep. Initially stored holes are kept in FG in CC region, while they are tunneled from the FG to channel in FT region.

Next, the V_{FG} trajectory is employed to elucidate the origin of round sweep based-window. Measured V_{FG} trajectory of the $MoTe_2$ device shown in **Fig. 2-10(a)** is again shown as **Fig. 3-8**. This figure also illustrates the operation of the device at each important point.

1) At the starting point of the round sweep (indicated by the open symbol), holes are tunneled from the $MoTe_2$ channel to the FG, resulting in a negative side of tunnel starting voltage ($V_{tun}^{(-)}$) of V_{FG}. Holes are initially stored in the FG. 2) At the initial stage of positive sweep, holes are kept in the FG since the tunneling does not occur between the FG and channel in capacitive coupling (CC) region of V_{FG} trajectory. 3) However, before V_{BG} reaches the 0 V, V_{FG} reaches the positive side of tunnel starting voltage ($V_{tun}^{(+)}$), resulting in a hole loss from the FG to channel. In other words, feedback by tunneling (FT) region is started.

Interestingly, the V_{FG} trajectory has also revealed the origin of the difference in slopes between positive and negative sweeps. As illustrated in **Fig. 3-9**, the channel conductance can be divided into two parts, i.e., the conductance of access region (G_{access})

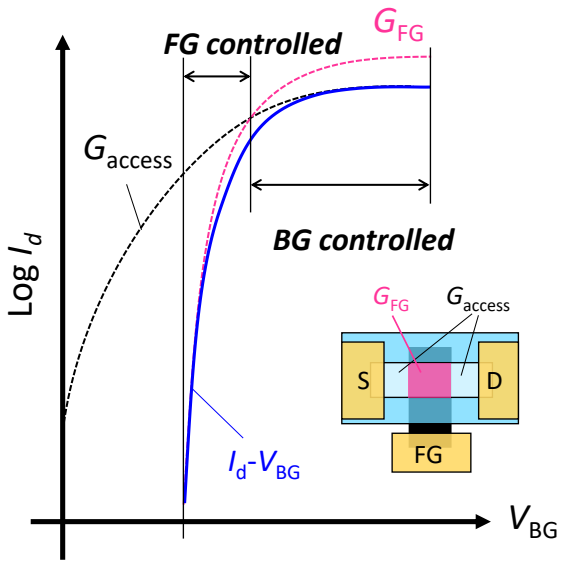


Fig. 3-9 Schematic of FG controlled mode and BG controlled mode. Dotted lines schematically represent the conductance of FG region and access region. Total channel conductance is determined by the smaller of G_{access} and G_{FG}.

and that of the FG region (G_{FG}). While G_{access} is modulated by the BG, G_{FG} is modulated by the FG. In addition, the access region is far from the BG since 15.0 nm *h*-BN and 90 nm SiO_2 are inserted between the channel and BG, whereas the FG region is close to the BG since only 15.0 nm *h*-BN is inserted between the channel and FG. Large FG electrode attached to the device results in the condition of $dV_{FG}/dV_{BG} \approx 1$ (see **section 2.2.2**). Therefore, the FG has more controllability than the BG during V_{BG} round sweep. As illustrated in **Fig. 3-9**, G_{FG} is rapidly increased with increasing V_{BG} (magenta) as compared with G_{access} (black). Since total channel conductance is determined by the smaller of G_{access} and G_{FG}, the I_d–V_{BG} curve can be drawn as the blue curve. In other words, I_d is gradually increased/decreased with V_{BG} if the total channel conductance is mainly controlled by the BG, while I_d is rapidly increased/decreased with V_{BG} if the conductance is mainly controlled by the FG. As labeled in **Fig. 3-9**, the former is called *BG controlled mode* and the latter is called *FG controlled mode* in this study.

In the case of positive sweep, I_d is gradually increased with increasing V_{BG}, and corresponding V_{FG} is high enough to make FG region strong *n*-type. Therefore, the I_d–V_{BG} curve is in BG controlled mode. On the other hand, in the case of negative sweep, I_d is

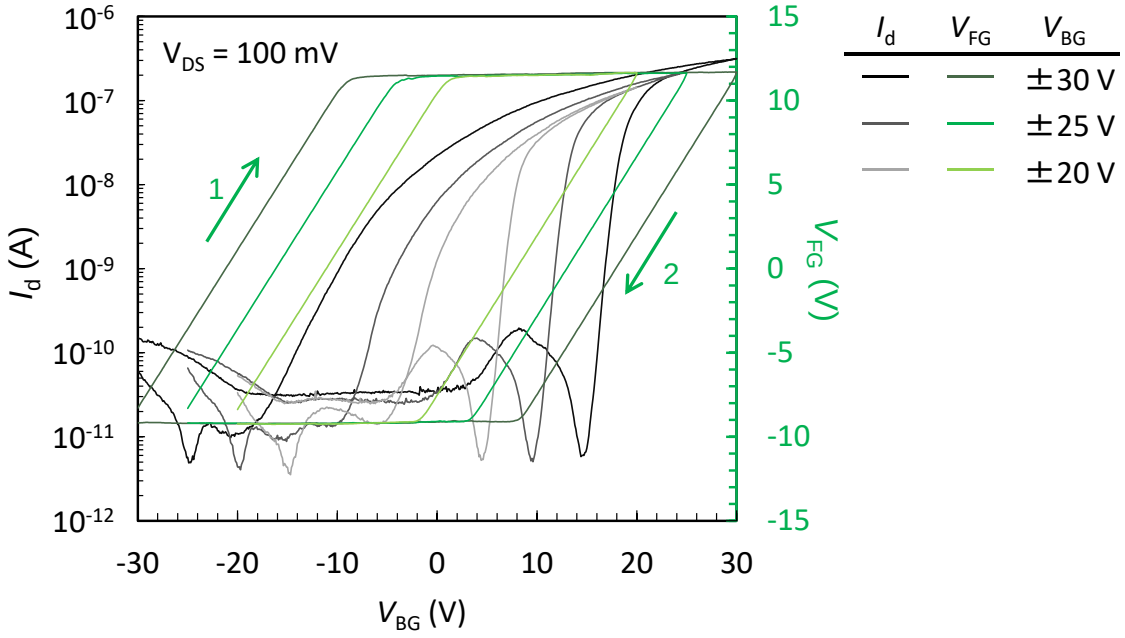


Fig. 3-10 V_{FG} trajectories with various V_{BG} sweeping range superimposed on corresponding I_d–V_{BG} round sweep curves.

rapidly decreased with decreasing V_{BG}, and corresponding V_{FG} is also decreased by the same amount as V_{BG}. Therefore, the I_d–V_{BG} curve is in FG controlled mode. In other words, for the negative sweep, the channel is immediately turned off by the FG. This is quite artificial, and I_d–V_{BG} curve in negative sweep can be shifted as desired regardless of the number of charges stored in the FG. V_{FG} trajectories with various V_{BG} sweeping range shown in **Fig. 3-10** obviously represents the artificiality in negative sweep. The amount of V_{BG} sweeping range shift coincide with the shift of I_d–V_{BG} curve in negative sweep. This consistency may be resulted from the high-quality 2D heterostructure ($MoTe_2$/*h*-BN/graphite), and supports the above-mentioned discussion. Although the existence of the 2D/3D interface (*h*-BN/SiO_2) may leads a little inconsistency in positive sweep (they should be overlapped each other because of BG controlled mode), gradual slope of the BG controlled mode is observed in all three I_d–V_{BG} curves.

3.3.3 *I*–*V* Single Sweep Case

Unfortunately, a V_{FG} trajectory cannot be measured with I_d–V_{BG} single sweep curve after P/E operation. Since the measurement mode is manually switched from P/E operation mode to DC measurement mode, charges stored in the FG are moved to the probe connected to the FG electrode during the interval, resulting in a 0 V of V_{FG} before I_d–V_{BG} single sweep measurement. However, fortunately, the V_{FG} trajectory during I_d–V_{BG} single sweep can be easily expected based on above discussions.

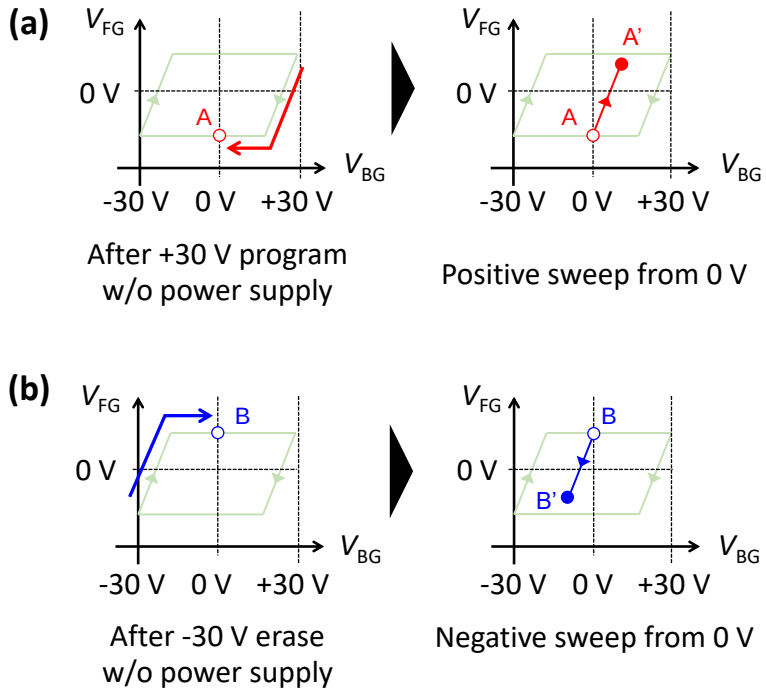


Fig. 3-11 Expected V_{FG} trajectories which correspond to I_d–V_{BG} single sweep curves.

The expected V_{FG} trajectories are shown in **Fig. 3-11**, where V_{FG} trajectory in round sweep is also illustrated for comparison. As illustrated in **Fig. 3-11(a)**, after +30 V program, V_{FG} reaches the point A when without power supply. (No voltage is applied to the device.) When the V_{BG} is swept from 0 V to positive direction, V_{FG} is increased from point A to A'. This means that, the I_d–V_{BG} curve measured in this way is determined by the number of charges stored in the FG at V_{BG} = 0 V. Similarly, after –30 V erase, V_{FG} reaches the point B when without power supply as illustrated in **Fig. 3-11(b)**. When the V_{BG} is swept from 0 V to negative direction, V_{FG} is decreased from point B to B'. The I_d–V_{BG}

curve measured in this way is also determined by the number of charges stored in the FG at V_{BG} = 0 V. Therefore, the memory window extracted from the set of I_d–V_{BG} single sweep curves represents the difference of the number of charges stored in the FG at V_{BG} = 0 V.

3.3.4 Memory Window Overestimation by *I*–*V* Round Sweep

Finally, appropriate method for extracting the memory window of 2D materials based-flash memory devices is discussed. Since flash memory is a *non-volatile* memory, the number of charges stored in FG at V_{BG} = 0 V should be focused on, instead of initially stored charges in the FG at the starting point of V_{BG} round sweep. Therefore, according to above mentioned discussions, it can be concluded that the memory window extracted from I_d–V_{BG} round sweep is overestimated. The I_d-V_{BG} curve in positive sweep (from –30 V to +30 V) corresponds to the device with initially stored holes. These holes are lost before V_{BG} reaches 0 V, that is, the number of holes exceeds the capacity of the device. Here, *capacity* refers to the number of charges which the device can store in the FG without power supply. On the contrary, the single sweep-based window is exactly resulted from the difference of the number of charges stored in the FG at V_{BG} = 0 V.

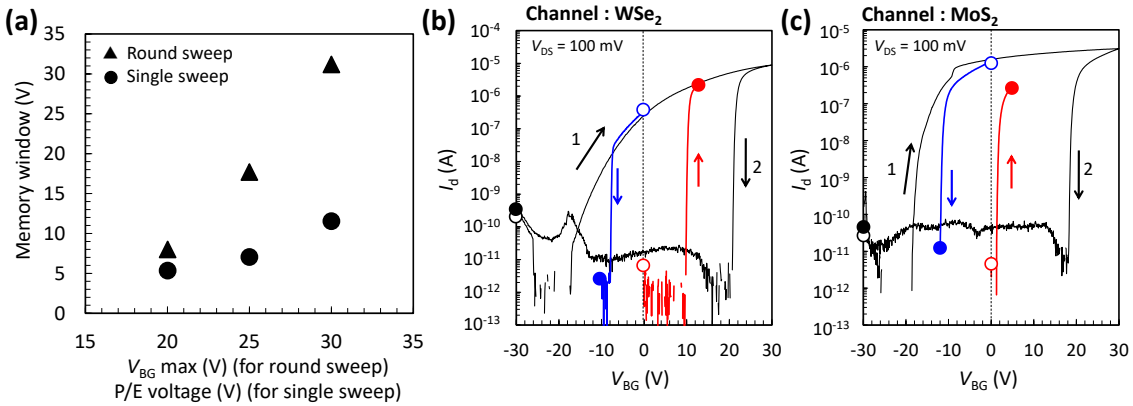


Fig. 3-12 The generality of memory window overestimation. **(a)** Comparison of both memory windows with various voltage conditions. **(b)** and **(c)** I_d–V_{BG} round sweep curves (black) and the set of I_d–V_{BG} single sweep curves (red and blue) for the WSe_2 and MoS_2 devices, respectively. Open and filled symbols represent the start and end points of the sweep, respectively.

Fig. 3-12(a) shows the comparison of both memory windows under various voltage conditions. Horizontal axis indicates the maximum V_{BG} for round sweep or P/E voltage for single sweep. Memory window overestimation is clearly observed in the range of 20 V to 30 V. Interestingly, as some previous studies have claimed[98,99,106,109], linear trend of memory window against the maximum V_{BG} is observed when round sweep is used. This can be roughly understood by the FG controlled I_d–V_{BG} curves, which is shifted by the same amount as the increase of V_{BG} sweep range (see **Fig. 3-10**). Probably, the situation has been the same in the previous studies. Indeed, memory window overestimation is also observed for the fabricated WSe_2 and MoS_2 device as shown in **Figs. 3-12(b)** and **(c)**, suggesting that the memory window overestimation is not unique for the $MoTe_2$ device but common for any 2D materials based-flash memory devices.

Noted that, according to above discussions, the value of I_d when V_{BG} is 0 V should be the same for round sweep and single sweep. However, the I_d-V_{BG} curves of the $MoTe_2$ device shown in **Fig. 3-6** and that of the MoS_2 device shown in **Fig. 3-12(c)** are not consistent, while that of the WSe_2 device shown in **Fig. 3-12(b)** seems to be reasonable. This suggests the short-term retention characteristic is poor, that is, some stored charges are lost before the single sweep measurements. It is about 10 s from the end of P/E operation to the start of the single sweep.

3.4 Criterion for Memory Window Overestimation

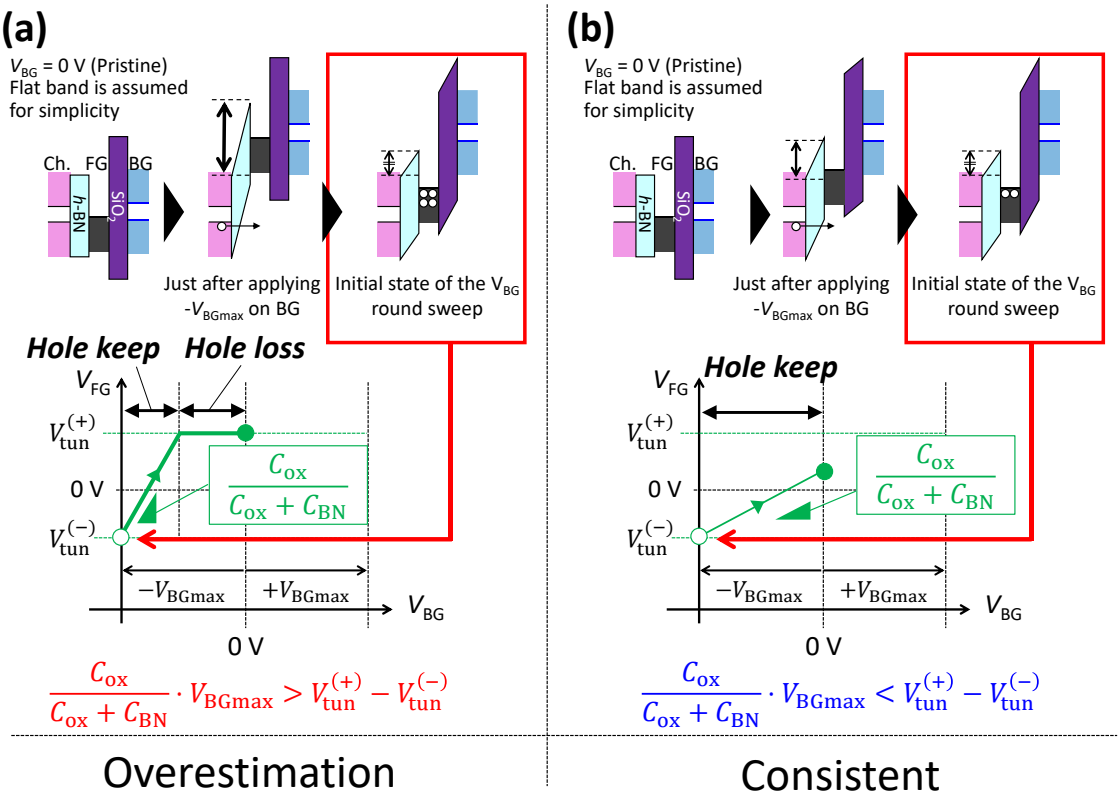


Fig. 3-13 Schematics of generalized **(a)** overestimation case and **(b)** consistent case. Band diagrams show the initial hole injection from the channel to FG and initially stored holes in the FG. The critical situation for the overestimation is that, the loss of the initially stored charges before V_{BG} reaches 0 V.

Since some parameters are different among studies, more generalized explanation for overestimation is required. In this context, the criterion for the overestimation is derived in this section. The conclusion is that, the memory window overestimation by I_d–V_g round sweep is not always but commonly encountered.

As discussed in previous section, the critical situation for the overestimation is that, the loss of initially stored charges before V_{BG} reaches 0 V. In other words, memory window is not overestimated by I_d–V_{BG} round sweep if initially stored charges are kept until V_{BG} reaches 0 V. Both situations are illustrated in **Fig. 3-13**, where symmetrical V_{BG} sweeping range is assumed. In this figure, band diagrams are also shown to illustrate the

initial hole injection and initially stored holes in the FG. According to the figure, the memory window is overestimated when following equation is satisfied:

$$\frac{C_{\mathrm{ox}}}{C_{\mathrm{ox}}+C_{\mathrm{BN}}}\cdot V_{\mathrm{BGmax}} > V_{\mathrm{tun}}^{(+)} - V_{\mathrm{tun}}^{(-)}, \tag{3.1}$$

where, C_{ox}, C_{BN} and V_{BGmax} are the capacitance of SiO_2, that of *h*-BN, and the maximum V_{BG} in round sweep, respectively. The criterion is composed of three parameters, i.e., tunnel starting voltages ($V_{\mathrm{tun}}^{(+)}$, $V_{\mathrm{tun}}^{(-)}$), V_{BG} sweeping range ($\pm V_{\mathrm{BGmax}}$), and capacitive coupling ratio ($C_{\mathrm{ox}}/(C_{\mathrm{ox}} + C_{\mathrm{BN}})$). In **Fig. 3-13**, high and low capacitive coupling ratio is assumed for the overestimation and consistent case, respectively. Although the V_{FG} at starting point of round sweep is the same, the number of initially stored holes is different due to the difference of capacitive coupling ratio. Indeed, low capacitive coupling ratio seems to be one solution for the overestimation issue.

However, as overviewed in **section 1.2**, increasing the capacitive coupling ratio is standard strategy for Si-based flash memory cell to achieve the effective P/E operation. In Si-based flash memory cell, typical value is 0.6[17]. Assuming the value of V_{BGmax}, $V_{\mathrm{tun}}^{(+)}$ and $V_{\mathrm{tun}}^{(-)}$ are +30 V, +7 V and −7 V, respectively, 0.6 of capacitive coupling ratio can satisfy the **Eq. 3.1**. Since assumed values are typical, and 0.6 of capacitive coupling ration is also easy to be achieved by 2D materials-based flash memory devices, **Eq. 3.1** can easily be satisfied. Therefore, I_{d}–V_{g} single sweep must be used for extracting the memory window of 2D materials based-flash memory devices.

3.5 Summary

In this chapter, the memory window overestimation by I_d–V_g round sweep measurement was clarified by the analysis of V_{FG} trajectory. The key point is that, initially stored holes are lost before V_g reaches the 0 V during round sweep. The number of stored charges at V_g = 0 V (i.e., without power supply) is important to guarantee the *non-volatility* of NVM devices. According to the derived criterion for the overestimation, to decrease the capacitive coupling ratio is one solution to avoid the overestimation issue by round sweep. However, it is not realistic, since to increase the ratio is standard strategy for flash memory. Therefore, I_d–V_g single sweep measurement must be used to extract the memory window of 2D materials based-flash memory devices, as well as Si based-flash memory cells. This understanding enables the fair comparison of their memory windows.

In this chapter, the power of V_{FG} measurement has been clarified in terms of the validity check of the memory window extraction method. In next chapter, its appearance will be expanded to the operation mechanisms.

Chapter 4
Operation Mechanisms

As mentioned in previous chapters, a rich variety is an advantage of 2D materials. Especially for 2D heterostructured non-volatile memory (NVM) devices, infinite number of combinations can be accessed by the stacking technique. Therefore, many 2D heterostructured NVM devices have been demonstrated as summarized in **Table 4-1**. This is same as **Table 1-3**, shown again. More importantly, the scope of 2D materials based-flash memory device is now expanding toward more advanced functional devices. However, these progresses have not been sufficiently supported by the understandings of their operation mechanisms. The key issues are listed below.

1. *The effect of metal/2D contact on their operation is still unclear.*
   Metal/ 2D contact is one of hottest topics in 2D field-effect-transistor (FET) research, since it governs the performance of 2D FET. However, there is a lack of discussion for NVM applications.

2. *The effect of the bandgap (E_g) of 2D channel on their operation is still unclear.*
   Each semiconducting 2D material has inherent E_g, and which generally depends on the layer number in atomically thin regime (see **section 1.3.3**). It is better for us to utilize the variety for NVM device design.

3. *Metal to 2D tunneling path has NOT been excluded.*
   2D heterostructure, which includes the superior 2D/2D interfaces, has often been claimed as a tunneling path. However, due to the lack of access region (see **Table 4-1**), poor metal/2D interface has not been excluded from the tunneling path.

Thus, there is a lack of device design policy. For the future development, the comprehensive understanding of the operation mechanisms is strongly required.

In this chapter, floating gate voltage (V_{FG}) trajectory is employed again to elucidate the operation mechanisms. First, the limited information from *I–V* transfer curves is briefly explained, and the inherent V_{FG} trajectories are highlighted. Interestingly, some trajectories are deviated from the principle explained in **chapter 2**, enabling us to access the operation mechanisms behind *I-V* transfer curves. The operation mechanisms can be clarified in terms of the three current limiting path. Next, the trajectory control based on the understandings is demonstrated. It further supports the validity of the understandings. Temperature dependences of all three V_{FG} trajectories are also discussed. Finally, the origin of drain current (I_d) plateau, which is often observed in previous studies as shown in **Table 4-1**, is revealed.

Table 4-1. Summary of the previous studies. Materials, structure and characteristics.

No.	Channel	Tunneling barrier	FG	Access region	I_d plateau (scale)	Ohmic contact	Ref.
1	MoS_2	*h*-BN	multi-layer graphene	No	Yes (log)		[105]
2	ReS_2	*h*-BN	Graphene	No	Yes (linear)	Confirmed	[106]
3	MoS_2	*h*-BN	graphene	No	Yes (linear)		[107]
4	PtS_2	*h*-BN	graphene	No	Yes (log)	Confirmed	[102]
5	$MoTe_2$	*h*-BN	graphene	No	Yes (linear)		[98]
6	MoS_2	*h*-BN	graphene	No	Yes (linear)		[108]
7	MoS_2	*h*-BN	metal	No	No	Confirmed	[109]
8	MoS_2	*h*-BN	MoS_2 & metal	No	Yes (linear)		[100]
9	$MoSe_2$	*h*-BN	graphene	No	Yes (log)	Confirmed	[99]
10	BP/graphene hetero-stack	*h*-BN	graphene	No	Yes (log)	Confirmed	[110]
11	WS_2	*h*-BN	multi-layer graphene	No	Yes (linear)	Confirmed	[42]
12	BP	*h*-BN	MoS_2	No	Yes (linear)	Confirmed	[104]
13	MoS_2	*h*-BN	graphene	No	Yes (linear)		[41]
14	BP	*h*-BN	graphene	No	Yes (log)	Confirmed	[111]
15	WSe_2	Al_2O_3	CdSe QDs	No	No	Confirmed	[96]
16	MoS_2	HfO_2	metallic nanocrystal	No	Yes (log)	Confirmed	[94]
17	MoS_2	HfO_2	few-layer graphene	No	No	Confirmed	[97]
18	MoS_2	Al_2O_3	HfO_2	No	No	Confirmed	[112]
19	MoS_2	*h*-BN	graphene	Yes*	No*		[113]
20	BP	*h*-BN	graphene	Yes	No	Confirmed	[111]

QDs and BP in this table represent quantum dots and black phosphorus, respectively.

* The device has both top gate and back gate, and FG exists only in top gate side.

4.1 Limited Information from *I*–*V* Transfer Curves

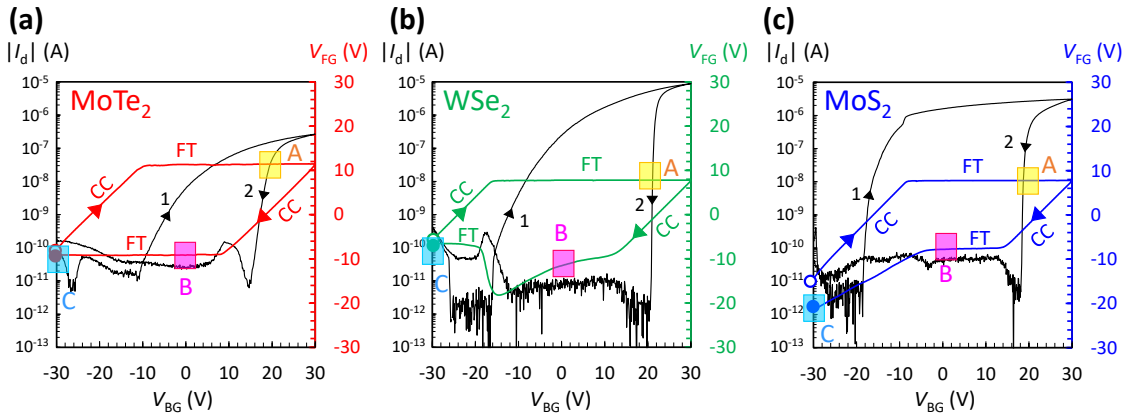


Fig. 4-1 Measured V_{FG} trajectories superimposed on I_d–V_{BG} round sweep transfer curves for the **(a)** $MoTe_2$ device, **(b)** WSe_2 device, and **(c)** MoS_2 device. Highlighted regions, i.e., region A, B, and C are the focused regions discussed in **section 4.3**. The open and filled circles represent the start and end points of the V_{FG} trajectory, respectively. The drain-source voltage (V_{ds}) is 100 mV for all measurements.

As mentioned in **chapter 2**, three devices were fabricated in this study, i.e., $MoTe_2$, WSe_2 and MoS_2 channel devices. Their I_d–back gate voltage (V_{BG}) round sweep curves are shown in **Fig. 4-1** with each V_{FG} trajectory. Although channel materials are different for each device, very similar curves are observed with large hysteresis. Since $MoTe_2$ and WSe_2 have an ambipolar nature (see. **Table 2-2**), *p*-branch is observed, while only *n*-branch is observed for the MoS_2 device. This is consistent with previous studies[59–61]. Although some metrics related to FET performance, e.g., subthreshold swing and on/off ratio can be extracted from the I_d–V_{BG} curves, how the channel material affects the memory operation is impossible to extract.

4.2 Inherent V_{FG} Trajectories

However, each device has its own V_{FG} trajectory as shown in **Fig. 4-1**. Interestingly, except the $MoTe_2$ device, measured V_{FG} trajectories are deviated from the principle explained in **chapter 2**. The deviation is observed only in negative sweep. In the case of WSe_2 device, capacitive coupling (CC) region is the same as the principle while the V_{FG} further decreases

even after it reaches the negative side of tunnel starting voltage ($V_{\mathrm{tun}}^{(-)}$). At the starting point of *p*-branch, V_{FG} goes back to the $V_{\mathrm{tun}}^{(-)}$ and keeps its value, which is the feedback by tunneling (FT) region. In the case of MoS_2 device, the trend in CC region is also the same as the principle and it is fixed at $V_{\mathrm{tun}}^{(-)}$. However, V_{FG} is decreased again when V_{BG} is further decreased to –30 V. For better understandings, "CC" and "FT" are placed on the corresponding regions as shown in **Fig. 4-1**. Noted that, the slopes of V_{FG} in the deviated regions are not as steep as that in CC region, suggesting that the insufficient feedback by relatively low tunneling current. When V_{BG} sweeping rate is decreased to ensure the feedback, as show in **Fig. 4-2** where V_{FG} trajectories of the MoS_2 device are shown, the moderate slope become flattened. This means that, there is a gradation between pure FT mode and pure CC mode as shown in the inset of **Fig. 4-2**. In this study, all measurements were carried out with standard sweep rate of 0.86 V/s except the measurement shown in **Fig. 4-2**. Since V_{FG} determines the tunneling between the FG and channel, the inherent V_{FG} trajectories may represent the character of each channel material. The correspondence of tunnel starting voltages was experimentally confirmed as shown in **Fig. 4-3**. Noted that, for the $MoTe_2$ device, I_{d}–V_{BG}, V_{FG} trajectory and tunneling current is not exactly same as previous chapters. This is due to the difference of measurement date. However, of course, the difference is negligible for the discussions in the study. The discussions in this chapter is based on the results shown in this chapter. To obtain more insight about the operation mechanisms, three regions in which the difference is noticeable were focused as indicated in **Fig. 4-1**. In region A where V_{BG} = +20 V and V_{FG} equals to positive side of tunnel starting voltage ($V_{\mathrm{tun}}^{(+)}$), all three V_{FG} trajectories are in the FT region, that is, tunneling between the channel and FG is occurred for all devices in this region. In region B with V_{BG} = 0 V, only WSe_2 device shows the moderate slope of V_{FG} trajectory while other two are in the FT region. In region C with V_{BG} = –30 V, only MoS_2 device shows the deviated trajectory.

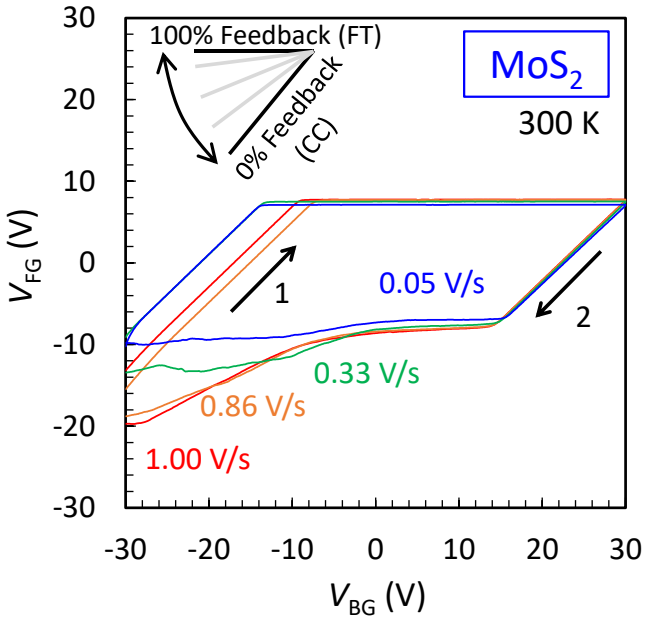


Fig. 4-2 V_{FG} trajectories of the MoS_2 device with various V_{BG} sweeping rates. Since the feedback mechanism is ensured with decreasing the sweeping rate, V_{FG} trajectory become flattened. All DC measurements in this study, except this measurement, are carried out with standard sweeping rate, 0.86 V/s.

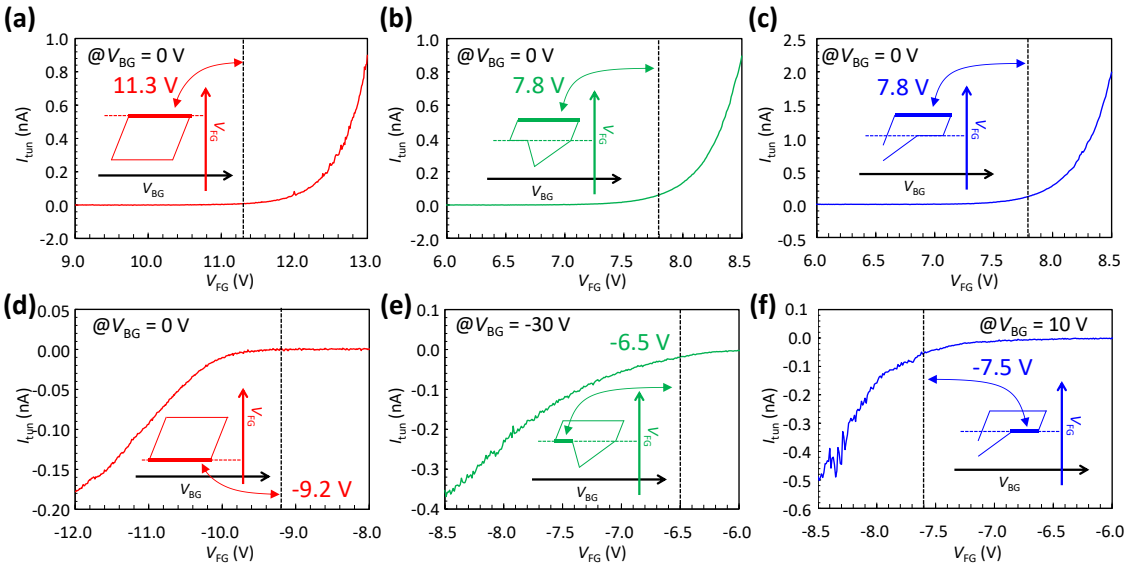


Fig. 4-3 The correspondence of the V_{FG}s in FT region and measured tunnel starting voltages. The insets show the rough sketch of each V_{FG} trajectory. **(a)–(c)** The measured tunneling currents from graphite FG to $MoTe_2$, WSe_2, and MoS_2 channel, respectively. **(d)–(f)** The measured tunneling currents from $MoTe_2$, WSe_2, and MoS_2 channel to graphite FG, respectively.

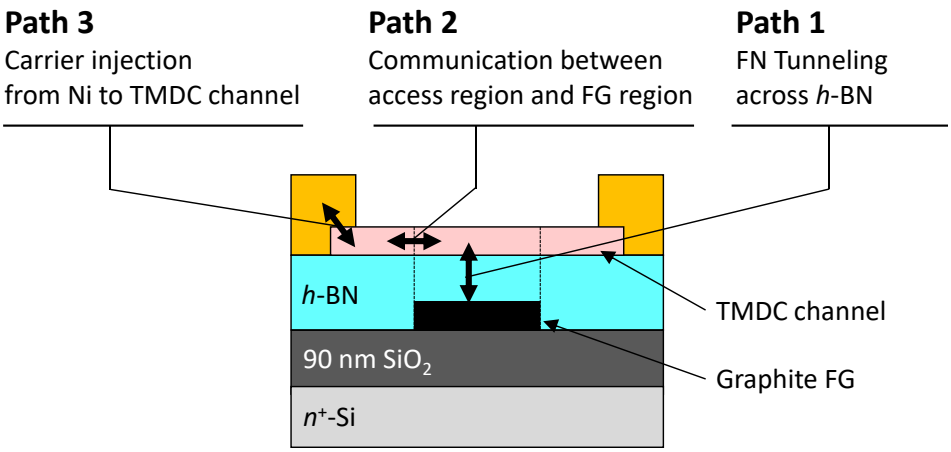


Fig. 4-4 Three tunneling paths in 2D heterostructured NVM device with access region. Not only path 1, but also all paths should be *open* for flowing tunneling current between the FG and channel.

In addition, tunneling path in 2D heterostructured NVM device should be carefully considered for the understanding of the operation, since the device has no substrate electrode which Si based-flash memory device has[17,125]. **Fig. 4-4** shows the three kinds of tunneling paths in 2D heterostructured NVM device. Path 1 is the Fowler-Nordheim (FN) tunneling across *h*-BN tunneling barrier, where previous studies have generally considered. Since the access region is fabricated to utilize the 2D heterostructure as a tunneling path, path 2, that is the communication between access region and FG region, should be considered. This path is important when *pn* junction is formed in the channel under certain combinations of V_{BG} and V_{FG}. Path 3 is the carrier injection from metal to the 2D channel. Since the important role of metal/2D contact on the 2D-FET is well known[59–61], path 3 should also be focused on in terms of memory operation. The important point is that, not only path 1, but all paths should be *open* for flowing the tunneling current.

4.3 Three Current Limiting Paths

4.3.1 Path 1: FN Tunneling across *h*-BN

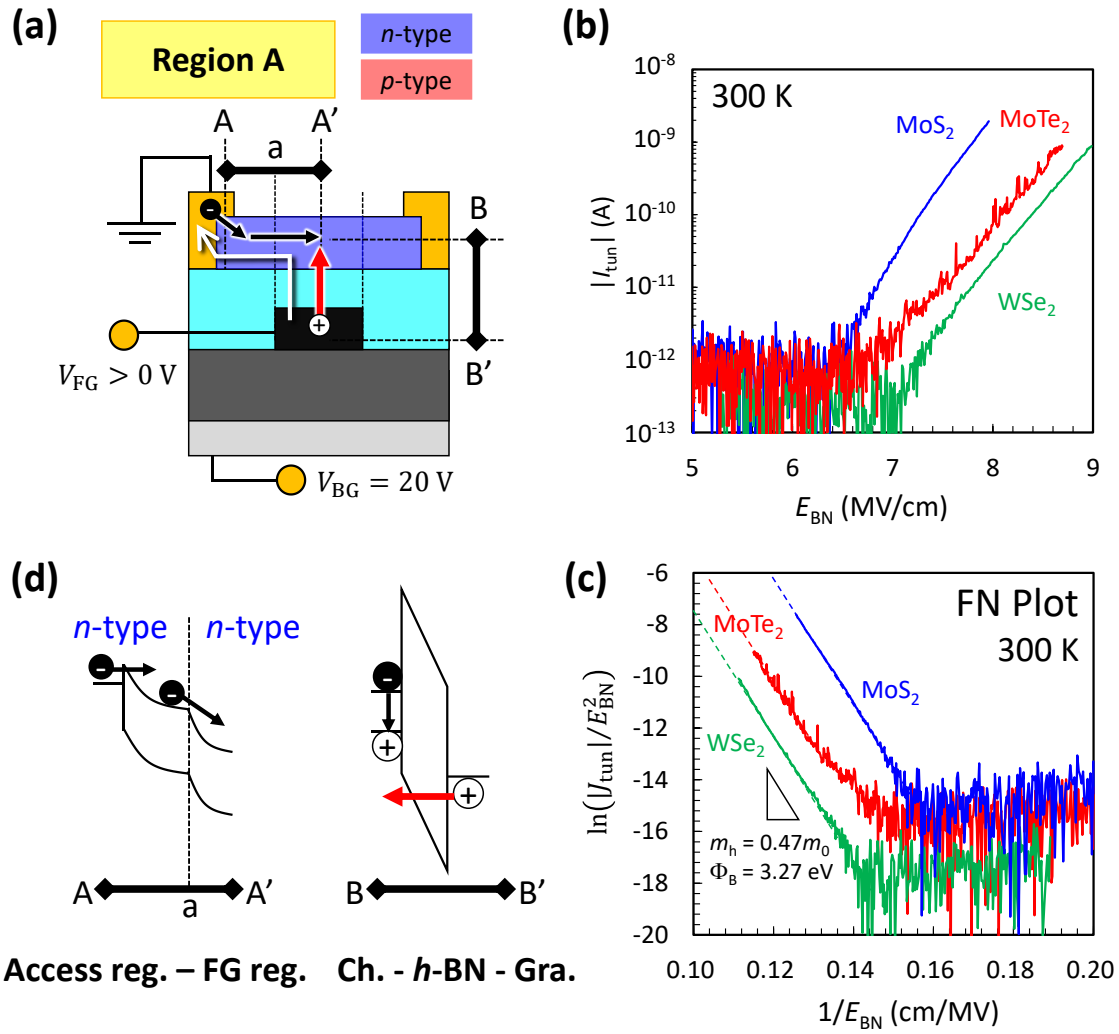


Fig. 4-5 Schematics of operation mechanism and important results in region A. **(a)** Schematic of channel polarity (*n*-type), tunneling current direction (white arrow), and the carrier flow in region A. Red arrow indicates the carrier flow which limits the total tunneling current (FN tunneling across *h*-BN). **(b)** Measured tunneling current and **(c)** its FN plot for each device. **(d)** Band diagrams along the A–A' (lateral direction in the channel) and B–B' (vertical direction in the 2D heterostructure) in **(a)**.

Here, the operation mechanisms in each region are discussed in terms of current limiting path, which limits the total tunneling current. Schematics of operation mechanism and

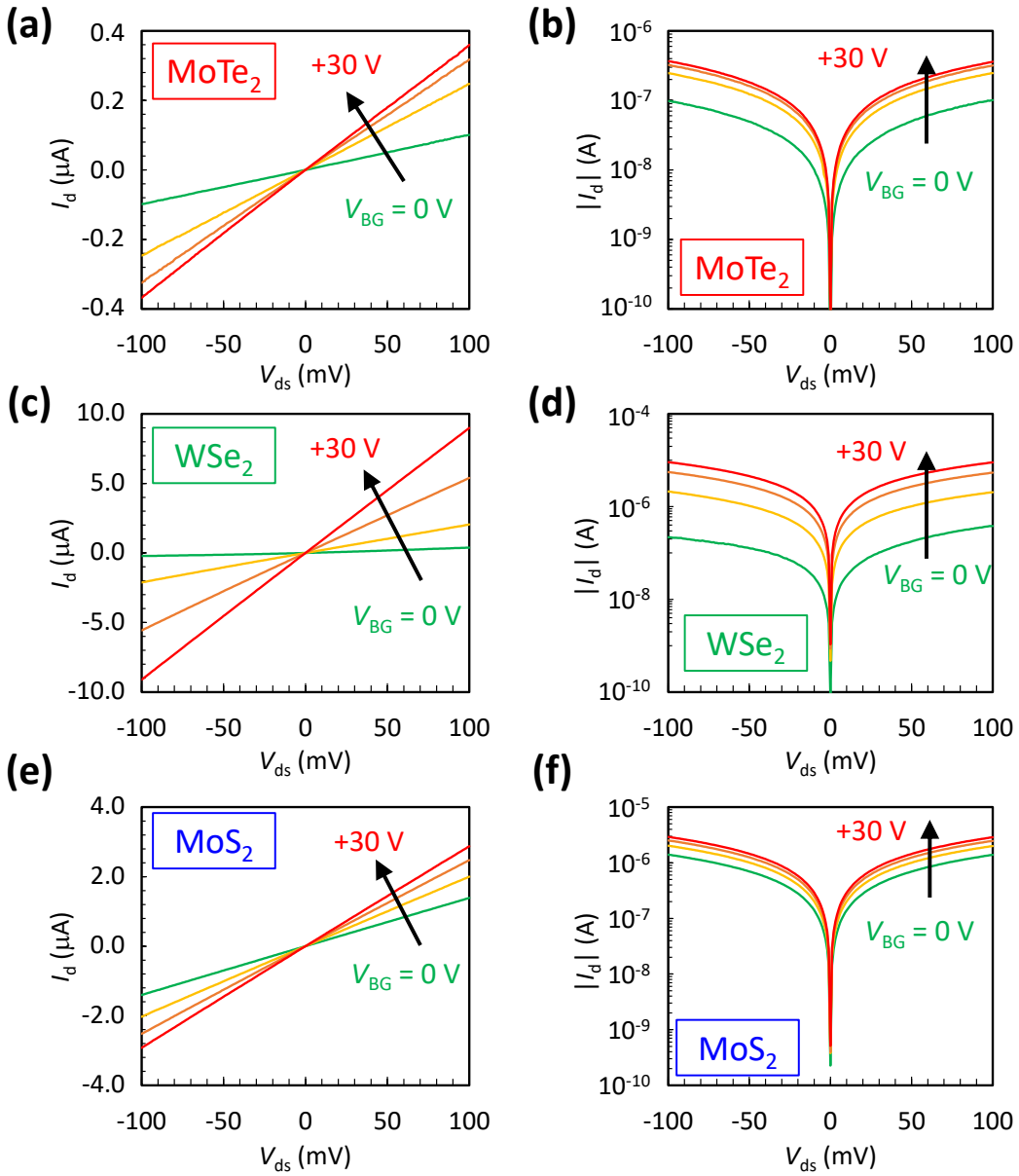


Fig. 4-6 The ohmic conduction of each device. **(a)**, **(b)** Linear and semi-log plot of I_d–V_{ds} for the $MoTe_2$ device. **(c)**, **(d)** Linear and semi-log plot of I_d–V_{ds} for the WSe_2 device. **(e)**, **(f)** Linear and semi-log plot of I_d–V_{ds} for the MoS_2 device. The step of V_{BG} is 10 V.

important results in region A are summarized in **Fig. 4-5**. As indicated by a white arrow, the tunneling current flows from the FG to channel. In this region, voltage conditions (V_{BG} = +20 V and $V_{FG} = V_{tun}^{(+)}$) result in whole *n*-type channel for all three devices as illustrated in **Fig. 4-5(a)**. Therefore, path 2 is no need to consider as a current limiting path. In addition, the effect of path 3 is negligible since electron injection from Ni to these three transition

metal dichalcogenides (TMDCs) is quite easy[59–61]. Ohmic contact of the devices were experimentally confirmed as shown in **Fig. 4-6**. Here, tunneling current from the FG to channel was measured by sweeping V_{FG} with grounded source. The absolute value of measured tunneling current ($|I_{tun}|$) is plotted as a function of the electric field across *h*-BN (E_{BN}) in **Fig. 4-5(b)**. Here, E_{BN} is calculated by dividing the V_{FG} by the thickness of *h*-BN (t_{BN}). Since ohmic contact has been confirmed and no *pn* junction is formed in the channel, almost all of V_{FG} should be applied across the *h*-BN. The observation of tunneling current for all devices is consistent with the flat trend of V_{FG} trajectories in region A. An amount of tunneling current is large enough to ensure the feedback mechanism. **Fig. 4-5(b)** can be replotted as an FN plot, as shown in **Fig. 4-5(c)**. Since FN tunneling current density (J_{tun}) is described by following equation, linear relationship in FN plot means that the current is FN tunneling current.

$$J_{\mathrm{tun}} \propto E_{\mathrm{BN}}^2 \exp\left(-\frac{4\sqrt{2m^*}\Phi_{\mathrm{B}}^{3/2}}{3q\hbar}\cdot\frac{1}{E_{\mathrm{BN}}}\right), \tag{4.1}$$

where, m^*, Φ_B, q and $\hbar$ are the effective mass of the tunneling carrier in *h*-BN, barrier height for the tunneling carrier, the charge of electron and reduced Plank's constant, respectively. In **Fig. 4-5(c)**, the slope of FN plot calculated by using the barrier height for holes (Φ_B = 3.27 eV[45,126,127]) and hole effective mass in the *h*-BN ($m_h = 0.47m_0$[128], where m_0 is free electron mass) is also shown. The agreement of the slope and measurement results represents that hole is a tunneling carrier for all three devices.

According to **Eq. 4.1**, FN tunneling current should not have temperature dependence ideally. Therefore, the tunneling current was measured under various temperature as shown in **Fig. 4-7**. For the WSe_2 and MoS_2 devices, almost no temperature dependence is clearly observed in the range of 10 K to 300 K, while the dependence is observed for the $MoTe_2$ device only below 125 K. This may be attributed to the effect of Schottky barrier at the metal/2D contact, and will be discussed in **section 4.5**. The temperature dependence of tunneling currents shown in **Fig. 4-7** further supports the tunneling current in region A is FN tunneling current.

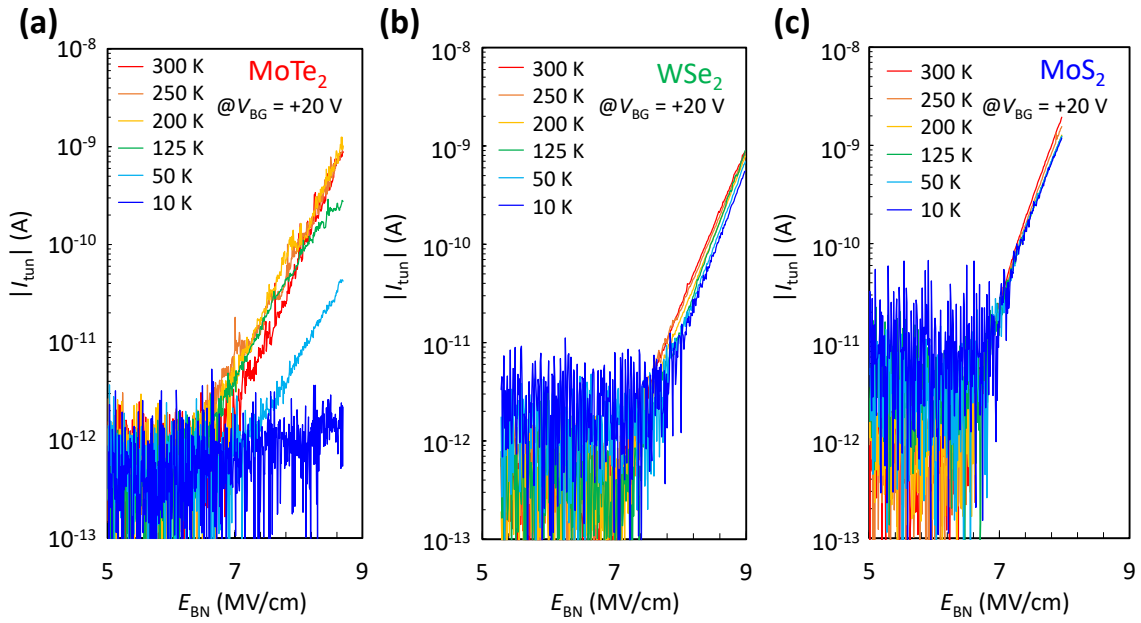


Fig. 4-7 Temperature dependence of tunneling currents in region A for **(a)** the $MoTe_2$ device, **(b)** the WSe_2 device, and **(c)** the MoS_2 device.

Additionally, the validity of the calculations for FN plot is discussed in detail. For the FN plot, J_{FN} was calculated by dividing I_{tun} by the whole area of FG region (S_{FG}) as shown in **Fig. 4-8(a)**. This assumes the maximum area for the tunneling current. On the other hand, as shown in **Fig. 4-8(b)**, there are two possible conduction modes. One is the partial conduction mode where the tunneling current flow in the part of FG region, and the other is the edge conduction mode where the tunneling current flow at the edge of FG region. However, as discussed below, the variation of the conduction area does not matter as long as the slope of FN plot is considered. **Eq. 4.1** can be rewritten as follows,

$$\ln\left(\frac{J_{\text{tun}}}{E_{\text{BN}}^2}\right) = -\frac{4\sqrt{2m^*}\Phi_{\text{B}}^{3/2}}{3q\hbar}\cdot\frac{1}{E_{\text{BN}}} + C, \tag{4.2}$$

where C is a constant. Since $J_{tun} = I_{tun}/S_{FG}$, following equation can be derived.

$$\ln\left(\frac{I_{\text{tun}}}{E_{\text{BN}}^2}\right) - \ln(S_{\text{FG}}) = -\frac{4\sqrt{2m^*}\Phi_{\text{B}}^{3/2}}{3q\hbar}\cdot\frac{1}{E_{\text{BN}}} + C. \tag{4.3}$$

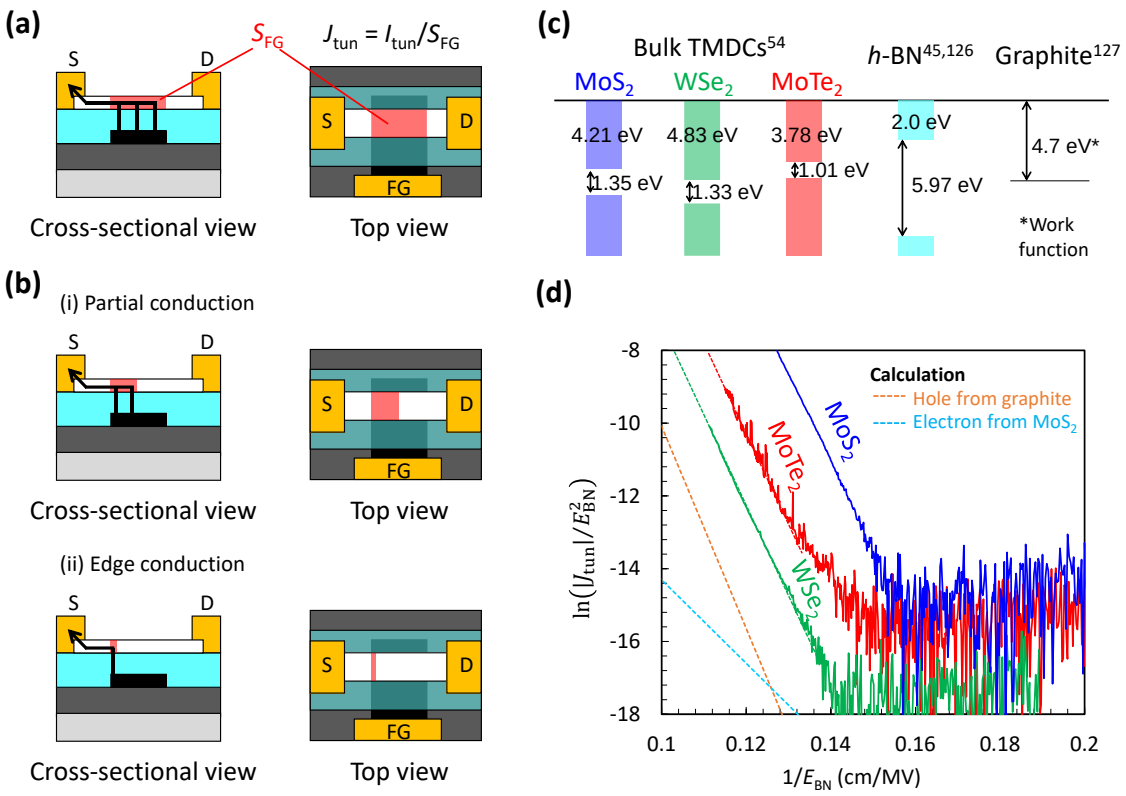


Fig. 4-8 The discussions for the validity of the calculations for FN plot. **(a)** Assumed tunneling area for tunneling current density (J_{tun}) calculation. **(b)** Two possible conduction modes of tunneling: partial conduction and edge conduction. **(c)** Band alignment of the 2D materials[45,54,126,127]. **(d)** Calculated slopes by assuming hole or electron tunneling. Experimental results shown in **Fig. 4-5(c)** are also shown for comparison.

Therefore, regardless of the value of S_{FG}, the slope of the plot does not change. Moreover, we could reproduce the experimentally obtained slope of FN plot by using **Eq. 4.1**. There are two possible situations for the tunneling. One is the hole tunneling from graphite FG to TMDC channel, and the other is the electron tunneling from TMDC channel to graphite FG. According to the band alignment shown in **Fig. 4-8(c)**[45,54,126,127], barrier heights for each case can be calculated. Here, hole effective mass ($m_h = 0.47m_0$) and electron effective mass ($m_e = 0.26m_0$) in h-BN[128] were used for hole and electron tunneling current calculations, respectively. Two calculation results are shown in **Fig. 4-8(d)**, where MoS_2 was assumed as a channel for the electron tunneling. It is noted that calculated results are vertically shifted for the ease of viewing the slope. For the comparison, experimental results are shown again. The slope of the hole tunneling case well agrees the experimental

results, rather than the electron tunneling case. Moreover, the same slopes are observed for experimental results even though different TMDCs were used as channels, further supporting the hole tunneling from graphite FG. Although the electron tunneling from TMDC channel is reasonable according to the band alignment shown in **Fig. 4-8(c)**, the hole tunneling from graphite FG via *h*-BN is consistent with the previous report[120].

From above discussions, it can be concluded that the current limiting path in region A is the path 1 (indicated by a red arrow in **Fig. 4-5(a)**). The operation mechanism in region A is explained as follows. As illustrated in **Figs. 4-5(a)** and **(d)**, holes are tunneled from graphite FG to TMDC channel, and it is recombined with electrons in the channel which easily injected from the Ni electrode.

4.3.2 Path 2: Communication between Access Region and FG Region

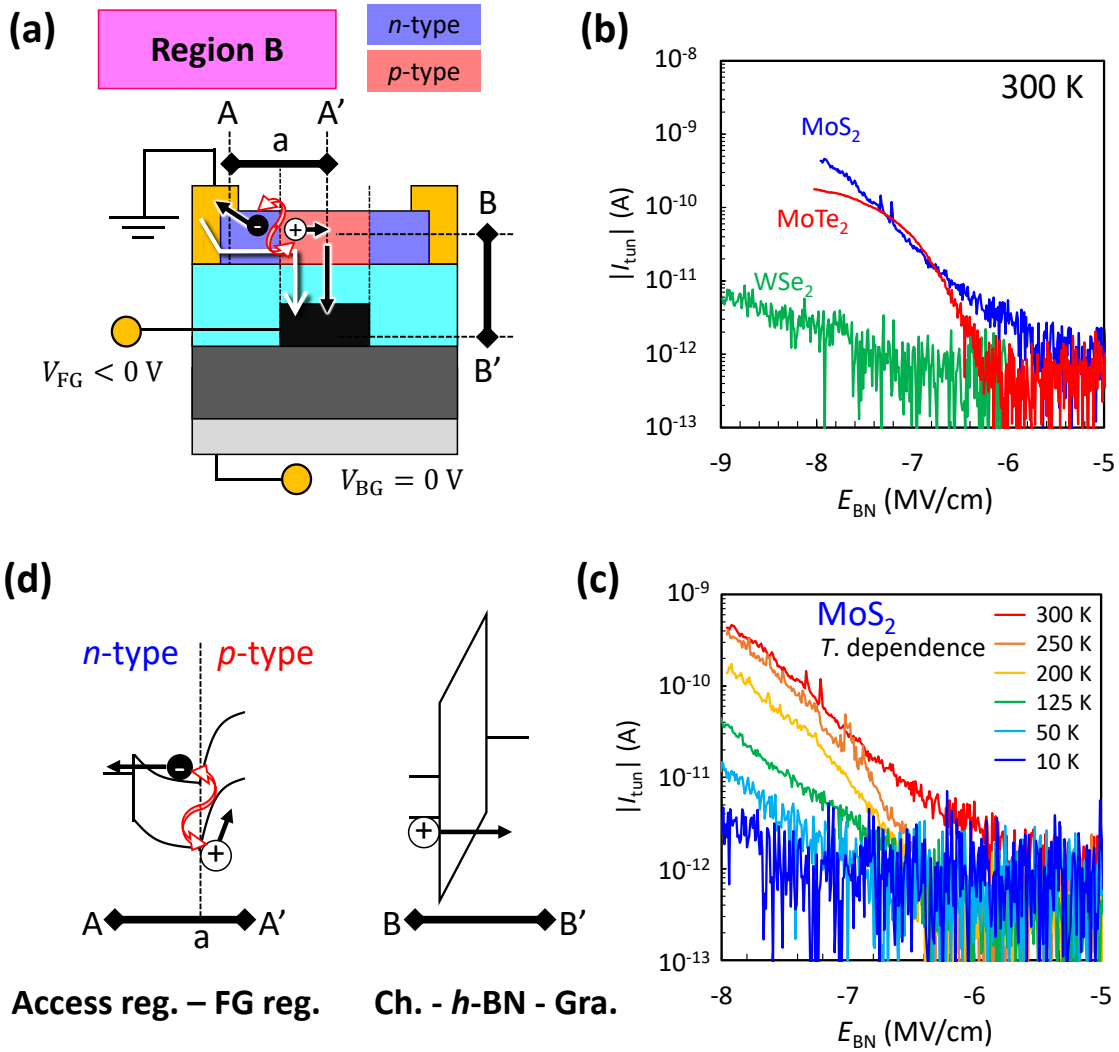


Fig. 4-9 Schematics of operation mechanism and important results in region B. **(a)** Schematic of channel polarity (*n*-type for access region and *p*-type for FG region), tunneling current direction (white arrow), and the carrier flow in region B. Red arrows indicate the carrier flow which limits the total tunneling current. **(b)** Measured tunneling current of each device. **(c)** Temperature dependence of tunneling current for the MoS_2 device. **(d)** Band diagrams along A–A' and B–B' in **(a)**.

Next, the operation mechanism in region B is discussed. Schematics of the operation mechanism and important results are summarized in **Fig. 4-9**. According to the positive sweep I_d–V_{BG} curves shown in **Fig. 4-1**, the TMDCs used in this study have normally-on nature. Therefore, the access region is *n*-type, while the FG region is *p*-type in region B, since V_{FG} is negatively large enough (which equals to $V_{tun}^{(-)}$ for $MoTe_2$ and MoS_2 devices,

and below $V_{\mathrm{tun}}^{(-)}$ for WSe_2 device). Although *p*-type conduction of MoS_2 is difficult to observe due to the huge barrier height for hole injection at metal/MoS_2 contact, the previous study has proved that the inversion layer of the MoS_2 is easy to be formed by exploiting the capacitance–voltage measurement technique[62]. As a result, in region B, *pn* junction formed in the channel should be considered. Of course, as well as region A, electron injection from Ni to the TMDC channel does not limit the tunneling current. The tunneling current direction is illustrated as a while arrow in **Fig. 4-9(a)**, which is same as the reverse current direction of *pn* junction. **Fig. 4-9(b)** shows the measured tunneling current of the three devices. Relatively large tunneling current was observed for the MoS_2 and $MoTe_2$ device, whereas only small tunneling current was observed for the WSe_2 device. The result is consistent with the trend of V_{FG} trajectory in region B.

Here, the bulk E_{g} and the polarity for Ni contact of each TMDC channel are considered. As summarized in **Table 2-2**, the bulk E_{g} of $MoTe_2$, WSe_2 and MoS_2 are 1.01, 1.33 and 1.35 eV, respectively[54]. Moreover, since it is well known that the MoS_2 shows strong *n*-type behavior due to the strong Fermi level pinning, abrupt *pn* junction in the channel can be expected even though the E_{g} is relatively large. Based on the above discussions, sufficiently large tunneling current shown in **Fig. 4-9(b)** can be attributed to the electron-hole pair generation at the boundary of access region and FG region. Relatively large bulk E_{g} and gradual *pn* junction of WSe_2 may lead to the low tunneling current. To confirm this, temperature dependence of the tunneling current for the MoS_2 device was measured as shown in **Fig. 4-9(c)**. The tunneling current was strongly decreased with decreasing temperature from 300 K to 10 K. The temperature dependence for other two devices are shown in **Fig. 4-10** in which the same trends were observed. Since the FN tunneling current should not have temperature dependence ideally, these trends suggests that the current limiting path is NOT path 1. This supports the generation based tunneling in region B since the number of thermally generated carriers is decreased with decreasing temperature. FN plots of the tunneling currents shown in **Figs. 4-9(b)** and **(c)**

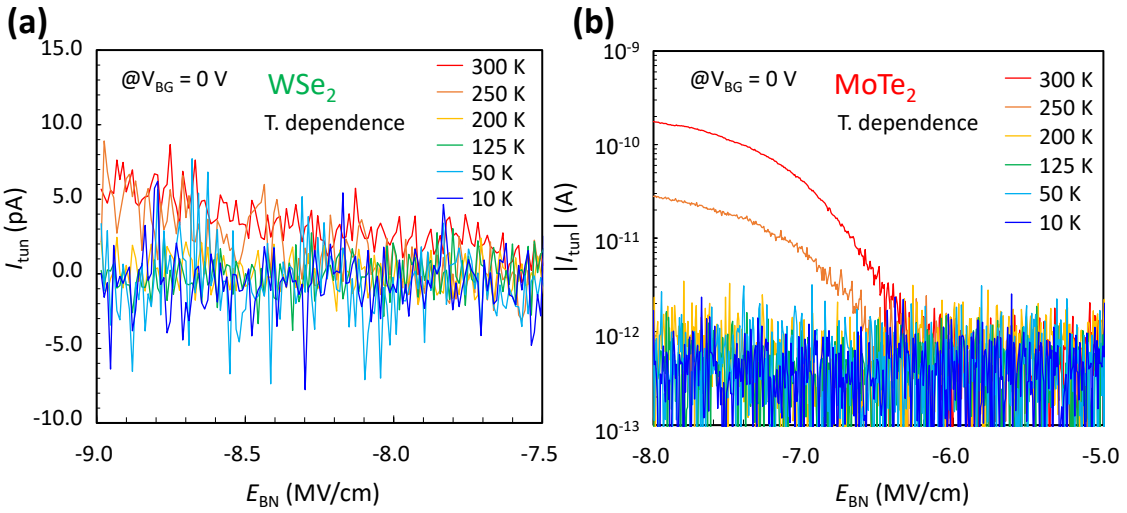


Fig. 4-10 Temperature dependence of tunneling currents in region B for **(a)** the WSe_2 device and **(b)** the $MoTe_2$ device.

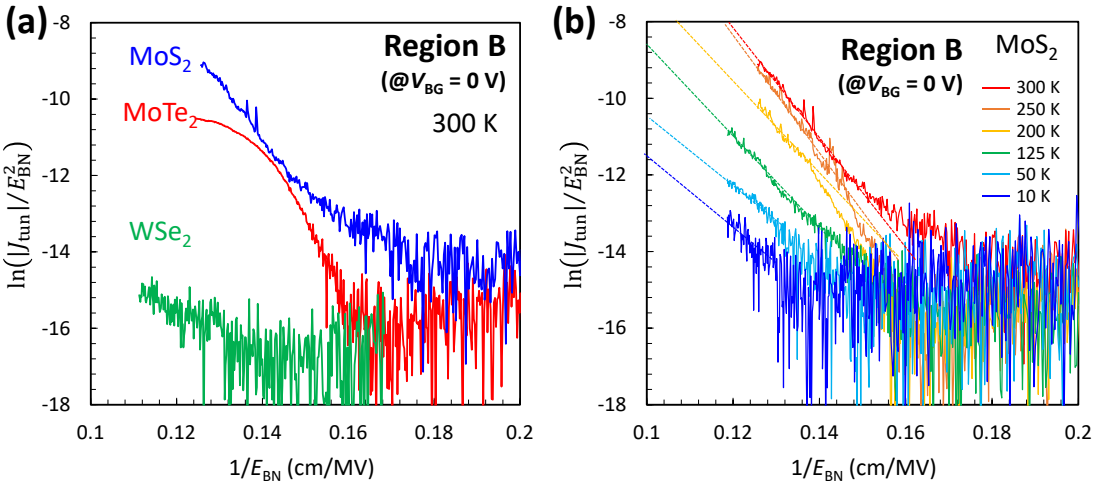
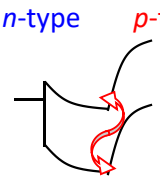
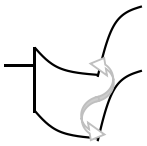
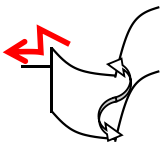
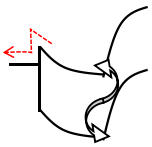
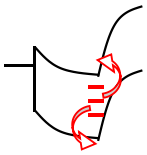
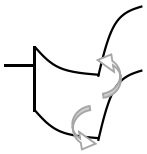


Fig. 4-11 (a) FN plot of tunneling current for each device at 300 K shown in **Fig. 4-9(b)**. **(b)** FN plot of the temperature dependence of tunneling current for MoS_2 device shown in **Fig. 4-9(c)**.

are shown in **Figs. 4-11(a)** and **(b)**, respectively. By contrast with region A, non-linearity is clearly observed for the $MoTe_2$ device, and for the MoS_2 device, linearity is poor for all temperatures. Dotted lines are also added in **Fig. 4-11(b)** for eye-guide, in which the inconsistency of the slopes is clear. These are clear evidences that the current limiting path in region B is NOT path 1.

Table 4-2. Three mechanisms related to electron-hole pair generation in region B.

Mechanism \ Mode	FT mode (Flat)	CC mode (Tilted)	Transition
1. Thermal activation	*n*-type *p*-type  High thermal energy	 Low thermal energy	FT→CC Decreasing T.
2. Thermionic emission	 Large TE current	 Small TE current	FT→CC Decreasing T.
3. Trap assisted generation	 w/ defect state	 w/o defect state	CC→FT Degradation

Of course, the thermal generation of electron-hole pair is not the only temperature dependent mechanism. For further discussions, three mechanisms related to the generation in region B are summarized in **Table 4-2**. Decreasing the temperature leads to not only the reduction of thermal energy to generate electron-hole pair (mechanism 1), but also the reduction of thermionic emission (TE) current (mechanism 2), that is, the number of electrons which can overcome the potential barrier from the TMDC channel to Ni is decreased. This further reduce the tunneling current.

Mechanism 3 is a trap assisted generation. The transition from CC mode to FT mode in region B is shown in **Fig. 4-12**. In this figure, **Fig. 4-1(c)** is shown again as **Fig. 4-12(b)** for comparison. Within 13 days, V_{FG} trajectory in region B was changed from CC mode to FT mode, suggesting the degradation of the MoS_2 device. Although the device was stored in vacuum, point defect such as sulfur vacancy might be formed since it has

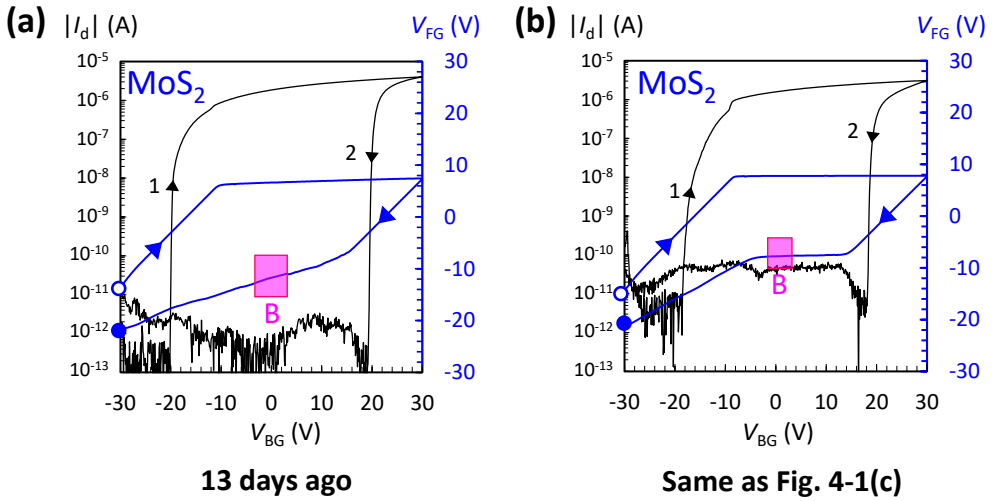


Fig. 4-12 The comparison of the V_{FG} trajectories of MoS_2 device. Within 13 days, the shape of V_{FG} trajectory is changed from CC mode to FT mode.

relatively low formation energy[78,79,81], resulting in the formation of defect states in the E_g. This enhances the generation rate. This is a possible mechanism of the transition of V_{FG} trajectory with time, and it further supports that the electron-hole pair generation limits the tunneling current in region B.

In short, at 300 K, tunnel current limiting path in region B is not path 1 nor path 3, and all of above mentioned discussions support the generation based tunneling in region B. The current limiting path, generation at the boundary of access region and FG region, is indicated by red arrows in **Figs. 4-9(a)** and **(d)**. Generated holes are moved to FG region and tunneled across *h*-BN, while generated electrons are moved to the electrode. When the temperature is decreased, not only the electron-pair generation but also TE current is suppressed, resulting in the temperature dependent tunneling current.

4.3.3 Path 3: Carrier Injection from Metal Electrode to 2D Channel

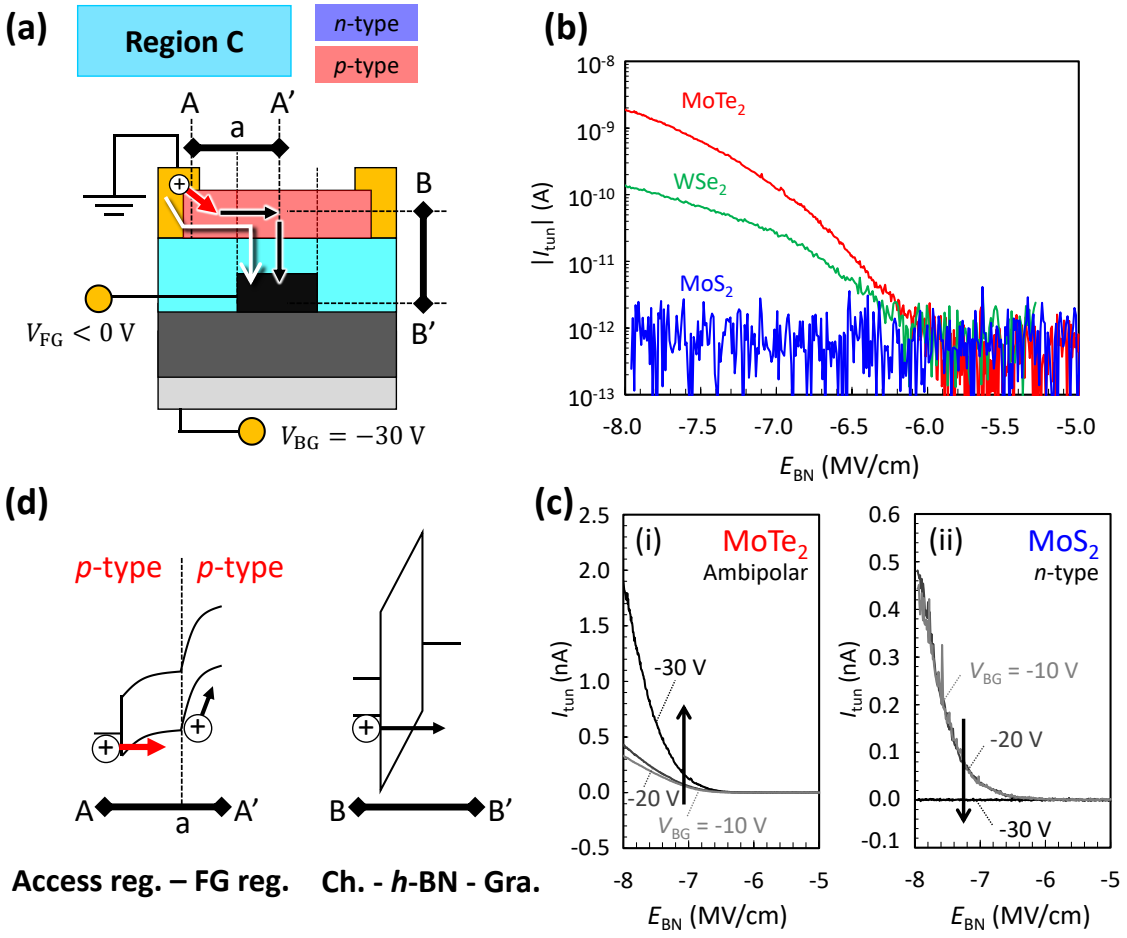


Fig. 4-13 Schematics of operation mechanism and important results in region C. **(a)** Schematic of channel polarity (*p*-type), tunneling current direction (white arrow), and the carrier flow in region C. Red arrow indicates the carrier flow which limits the total tunneling current (hole injection from Ni). **(b)** Measured tunneling current for each device. **(c)** Measured tunneling currents with various V_{BG}. Opposite trends are shown for ambipolar $MoTe_2$ device and *n*-type MoS_2 device. **(d)** Band diagrams along A–A' and B–B' in **(a)**.

Finally, the operation mechanism in region C is discussed. As well as regions A and B, schematics of operation mechanism and important results are summarized in **Fig. 4-13**. In region C, as shown in **Fig. 4-13(a)**, whole channel region is *p*-type since V_{BG} and V_{FG} are negative enough, suggesting that the effect of path 2 is negligible. In addition, the tunneling current flow from TMDC channel to graphite FG as indicated by a white arrow. As shown in **Fig. 4-13(b)**, tunneling current was observed only for the ambipolar channel devices,

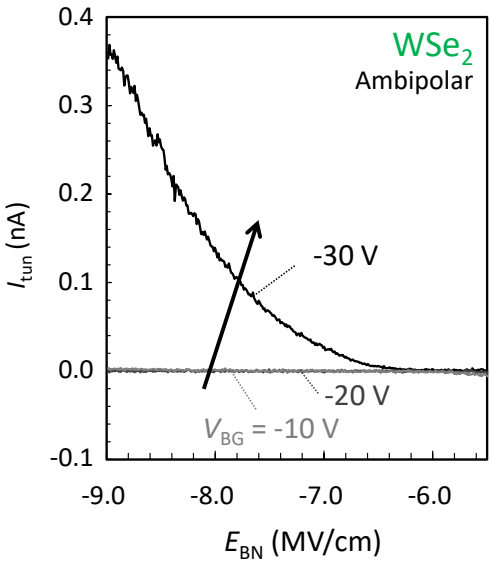


Fig. 4-14 Tunneling current of the WSe_2 device with various V_{BG}.

which is consistent with the flat trend of V_{FG} trajectory (see. **Figs. 4-1(b)** and **(c)**). Here, tunneling currents under various V_{BG} condition shown in **Fig. 4-13(c)** are considered. In the case of the $MoTe_2$ device, tunneling current was increased with decreasing V_{BG} from –10 V to –30 V. Since decreasing V_{BG} in this range implies the enhancement of *p*-type of the access region, the increase of tunneling current can be attributed to the enhancement of hole injection from Ni electrode. This is confirmed by the observation of opposite trend for the MoS_2 device. Indeed, increase of tunneling current with decreasing V_{BG} was also observed for the WSe_2 device as shown in **Fig. 4-14**.

Path 1 can be excluded from the current limiting path. One evidence is the non-linearity in FN plot as shown in **Fig. 4-15**. Another evidence is that the existence of temperature dependence in the tunneling currents as shown in **Fig. 4-16**. These results indicate the tunneling current in region C is not FN tunneling current. The possible mechanism of the temperature dependence is the suppression of TE from Ni to TMDC channel. As well as electron injection by TE, hole injection by TE is also suppressed by decreasing the temperature, resulting in a low doped *p*-type access region. Since voltage drop across the access region is increased, more V_{FG} is required for the tunneling with decreasing temperature.

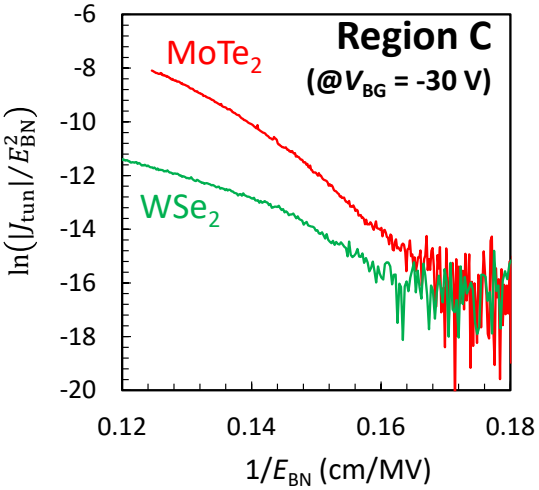


Fig. 4-15 FN plot of the tunneling currents in region C.

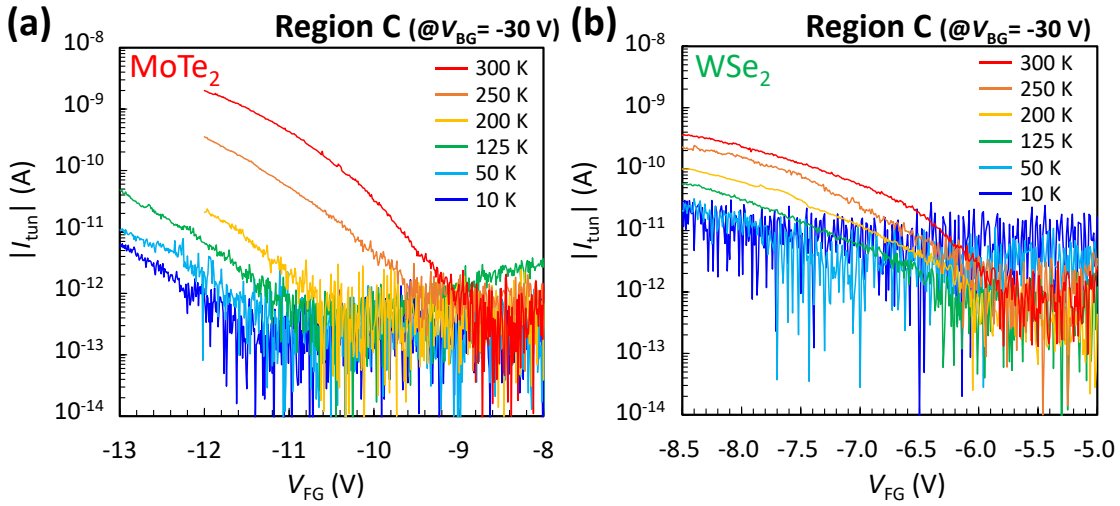


Fig. 4-16 Temperature dependence of tunneling currents in region C for **(a)** the $MoTe_2$ device and **(b)** the WSe_2 device.

In short, the tunneling current in region C is not limited by path 1 nor path 2. According to the above discussions, the only possible current limiting path in region C is the path 3 as indicated by a red arrow in **Fig. 4-13(a)**. As illustrated in **Figs. 4-13(a)** and **(d)**, holes are injected from Ni electrode and pass through to FG region since there is no *pn* junction in the channel. They are tunneled from TMDC channel to graphite FG.

Noted that, V_{FG} which is the FG potential relative to the source electrode is composed of three voltage terms, i.e., voltage drop across path 1 (V_{path1}), path 2 (V_{path2}), and path 3 (V_{path3}). In region A, V_{FG} is reduced to V_{path1} since the contributions from path 2

and path 3 are negligible, while V_{FG} cannot be reduced to V_{path1} in region B and C. Thus, strictly speaking, E_{BN} calculated by dividing V_{FG} by the thickness of *h*-BN shown in **Fig. 4-5** and **Fig. 4-13** are no longer correct, and therefore, roughly 10 nm *h*-BN was not electrically broken down even when V_{FG} reached approximately –20 V for the WSe_2 and MoS_2 devices.

The quick summary of the discussions in **Section 4.3** is as follows. The current limiting paths in regions A, B and C are the path 1 (FN tunneling across *h*-BN), path 2 (communication between access region and FG region), and path 3 (hole injection from Ni to TMDC channel), respectively. This gives a clear and comprehensive point of view for the operation mechanisms of 2D materials-based flash memory devices.

4.4 V_{FG} Trajectory Control

4.4.1 Region B Control by Temperature

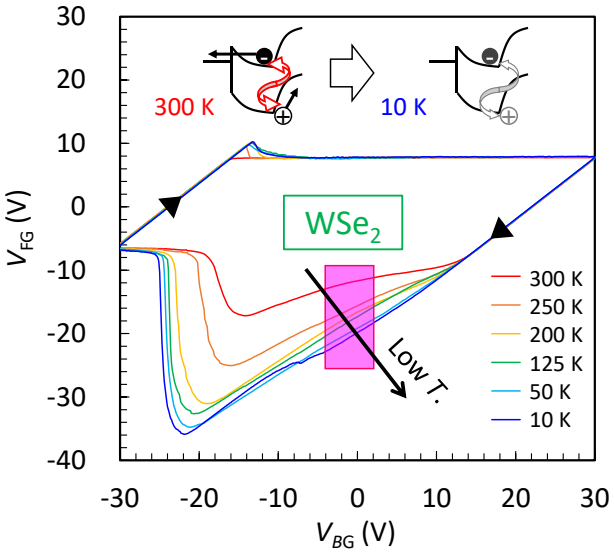


Fig. 4-17 Temperature dependence of V_{FG} trajectory of the WSe_2 device. The trend of V_{FG} in region B can be controlled by decreasing temperature to suppress the electron-hole pair generation.

Next, based on the understandings as mentioned above, V_{FG} trajectory is experimentally controlled as desired. **Fig. 4-17** shows the temperature dependence of V_{FG} trajectory for the WSe_2 device. In region B, the transition toward pure CC mode was observed with

decreasing temperature. As schematically shown in the inset, generation of electron hole pair is suppressed with decreasing temperature, resulting in the transition since path 2 is a current limiting path in region B. This can be interpreted as region B control by temperature, and supports the validity of above understandings.

4.4.2 Region C Control by Device Structure

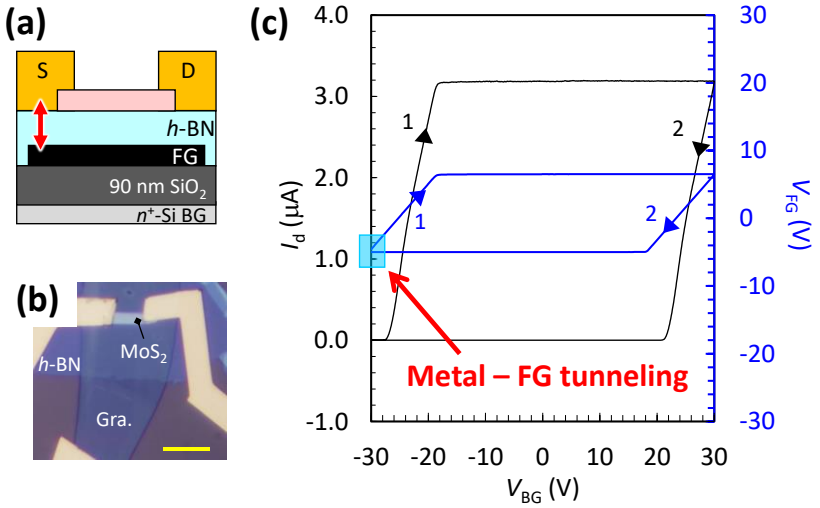


Fig. 4-18 (a) Schematic of metal-FG tunneling current in the device without access region. **(b)** Photo of fabricated device without access region. **(c)** V_{FG} trajectory of the device shown in **(b)**. Flat trend of V_{FG} in region C clearly represents the metal-FG tunneling.

As summarized in **Table. 4-1**, almost all of previous studies have not designed the access region, that is, there is a possibility that tunneling current flow between metal electrode and the FG as shown in **Fig. 4-18(a)**, whereas previous studies have focused only on the 2D heterostructure as a tunneling path. To reveal the true tunneling path, the device without access region was additionally fabricated as shown in **Fig. 4-18(b)**. Fabrication procedure is same as other devices. For the device, MoS_2, *h*-BN and graphite were used as a channel, tunneling barrier and FG, respectively. As well as other devices, Ni/Au were used for electrodes. The measured V_{FG} trajectory of the device without access region is shown in **Fig. 4-18(c)**. While moderate slope was observed in region C for the MoS_2 device with

access region, it was vanished for the MoS_2 device without access region. Since current limiting path in region C is hole injection from the metal electrode, the flattened V_{FG} trajectory suggests the formation of additional tunneling path, that is, metal-FG tunneling path. Although some previous reports has claimed the metal-2D tunneling path by a device simulation[43,101], this is the first experimental evidence of it.

Additionally, the change of V_{FG} trajectory shown in **Fig. 4-2** can also be interpreted as region C control by the V_{BG} sweeping rate. **Fig. 4-2** shows the change of V_{FG} trajectory for the MoS_2 device with decreasing V_{BG} sweeping rate. While moderate slope of V_{FG} was observed in region C when the sweep rate was 1.00 V/s, that was almost vanished when the sweep rate was 0.05 V/s.

4.5 Temperature Dependence of V_{FG} Trajectories

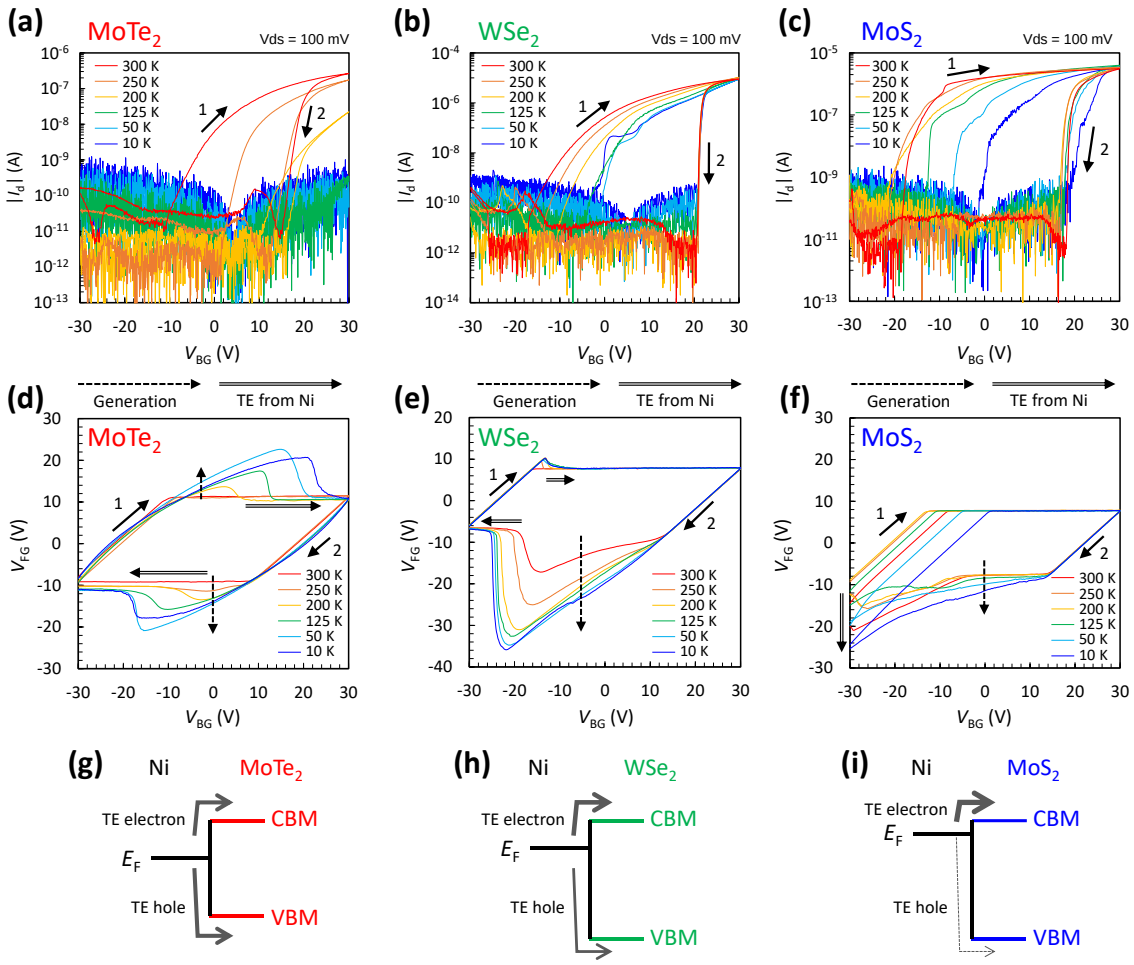


Fig. 4-19 Temperature dependence of I_d–V_{BG} round sweep transfer curves and V_{FG} trajectories for **(a)**, **(d)** the $MoTe_2$ device, **(b)**, **(e)** the WSe_2 device, and **(c)**, **(f)** the MoS_2 device. **(g)**–**(i)** Schematics of band alignment at the interface of Ni electrode and TMDC channels of each device expected from the temperature dependence shown in **(a)**–**(c)**.

The temperature dependence of V_{FG} trajectories and corresponding I_d–V_{BG} curves are discussed in this section. Measurement results are summarized in **Fig. 4-19**. As shown in **Figs. 4-19(a)**–**(c)**, for all devices, *n*-branch was more clearly observed than *p*-branch, suggesting that the Fermi level of Ni is pinned near the conduction band minimum (CBM) rather than the valence band maximum (VBM). In addition, *n*-branch of the $MoTe_2$ device is more sensitive to the temperature than that of the WSe_2 device. As a result, the schematics of band alignment between Ni and TMDC channel can be illustrated as **Figs. 4-19(g)**–**(i)**.

In **Figs. 4-19(d)–(f)**, the changes of V_{FG} trajectory with decreasing temperature are indicated by two kinds of arrows. One is a dotted arrow which indicates the change due to the suppression of generation current, the other is a double-line arrow which indicates the change due to the reduction of TE current. In this section, the details of the temperature dependence are explained in two subsections: positive sweep and negative sweep. In the following explanation, the polarity combination of access region/FG region/access region is simply denoted as *n*/*n*/*n*, *n*/*p*/*n*, and so on.

4.5.1 Positive Sweep

V_{FG} trajectory is flattened when *n*/*n*/*n* channel is formed since sufficiently large tunneling current can flow between the FG and channel to activate the feedback mechanism. Since the threshold voltages for all three devices are located below 0 V of V_{BG} at 300 K, *n*/*n*/*n* is easy to be achieved. For this reason, the FT region is wide in positive sweep as compared with in negative sweep. However, flat V_{FG} is possible even before the *n*/*n*/*n* channel formation because of the generation current in *p*/*n*/*p* channel at high temperature such as 300 K. To trace the transition from the *p*/*p*/*p* channel to the *n*/*n*/*n* channel via *p*/*n*/*p* channel is the key to understand the temperature dependence. As mentioned in previous section, TE of electron from Ni to TMDC channel and electron-hole pair generation are suppressed with decreasing the temperature. Since the band alignment at the metal/TMDC interface and bulk E_g depend on the channel material, different temperature dependence was observed as follows.

$MoTe_2$ device: At the starting point of the round sweep (V_{BG} = –30 V) where channel is *p*/*p*/*p*, V_{FG} was roughly independent of the temperature as shown in **Fig. 4-19(d)**. This attributes to the availability of the path 3 (hole injection from Ni to $MoTe_2$) in the range of 10 K to 300 K. According to the **Fig. 4-19(a)**, however, the starting point of *n*-branch was drastically shifted to the positive side of V_{BG} with decreasing the temperature, indicating that a delay of transition point from *p*/*n*/*p* channel to *n*/*n*/*n* channel. In addition, thermal

energy to generate electron-hole pair in *p*/*n*/*p* channel is also decreased. Consequently, as shown in **Fig. 4-19(d)**, the hump (CC mode) appeared in V_{FG} trajectory, and it was grown up with decreasing temperature due to insufficient tunneling current for the feedback. In short, the reduction of generation current works to maintain the CC mode during the *p*/*n*/*p* channel period even after the V_{FG} exceeds the tunnel starting voltage (dotted arrow), while the reduction of the TE current delays the point where V_{FG} become flattened (double-line arrow).

WSe_2 device: Similar to the $MoTe_2$ device, as shown in **Fig. 4-19(e)**, V_{FG} at the starting point of round sweep was roughly independent because of the availability of the path 3. Even at 300 K, generation of electron-hole pair at *pn* junction in the channel is quite limited. In order to flatten the V_{FG} trajectory, *n*/*n*/*n* channel is necessary. Therefore, small hump appeared in the V_{FG} trajectory with decreasing the temperature, which corresponds to the positive shift of the starting point of *n*-branch shown in **Fig. 4-19(b)**.

MoS_2 device: Different from other two ambipolar channel devices, as shown in **Fig. 4-19(f)**, V_{FG} at the starting point of round sweep was shifted down with decreasing the temperature. It attributes to the lack of the feedback mechanism by drastic reduction of TE of hole. In addition, according to **Figs. 4-19(c)** and **(f)**, *n*/*n*/*n* channel is already achieved when V_{FG} reaches the tunnel starting voltage, resulting in no hump in V_{FG} trajectory.

4.5.2 Negative Sweep

As well as the positive sweep, the key point to understand the temperature dependence is tracing the transition from *n*/*n*/*n* channel to *p*/*p*/*p* channel via *n*/*p*/*n* channel. However, *p*/*p*/*p* channel is more difficult to achieve than *n*/*n*/*n* channel, resulting in asymmetric V_{FG} trajectories especially for the WSe_2 and MoS_2 devices shown in **Figs. 4-19(e)** and **(f)**. Roughly symmetric V_{FG} trajectories for the $MoTe_2$ device shown in **Fig. 4-19(d)** might be attributed to the location of Fermi level of Ni which is close to the midgap of $MoTe_2$, as

illustrated in **Fig. 4-19(g)**. Generally, narrow FT region was obtained in negative sweep. Temperature dependence is explained as follows.

$MoTe_2$ device: No hump was observed at 300 K since the tunneling current is large enough to active the feedback mechanism regardless of the channel polarity. However, the hump appeared and it was grown up with decreasing the temperature, as well as positive sweep.

WSe_2 device: As mentioned in **section 4.4.1**, V_{FG} trajectory in region B became pure CC mode with decreasing temperature due to the suppression of electron-hole pair generation. In addition, starting point of the *p*-branch was slightly left shifted with decreasing temperature, which is obvious in the range of 300 K to 200 K as shown in **Fig. 4-19(b)**. This is attributed to the reduction of TE of holes from Ni to WSe_2, and causes the left shift of the point where V_{FG} becomes flat.

MoS_2 device: Although the contribution of generation is larger than the WSe_2 device due to the abrupt *pn* junction in the MoS_2 channel, as shown in **Fig. 4-19(f)**, the mode of V_{FG} trajectory in region was slightly changed from FT to CC, especially below 50 K. Therefore, when V_{FG} is further swept to the –30 V of V_{BG}, the final point of V_{FG} is shifted down. As sweeping the V_{BG} to –30 V, the main contributor to the shifting down is changed from the suppression of generation in region B to the reduction of TE of holes in region C.

Although large noise current at low temperature obscured the starting point of *n*- and *p*-branch in I_d–V_{BG} curves especially for the $MoTe_2$ device as shown in **Fig. 4-19(a)**, the temperature dependence of V_{FG} trajectories can be understood by tracing the channel polarity transition during positive and negative sweep. At this stage, above understandings are quite useful and reasonable.

4.6 The Origin of I_d Plateau

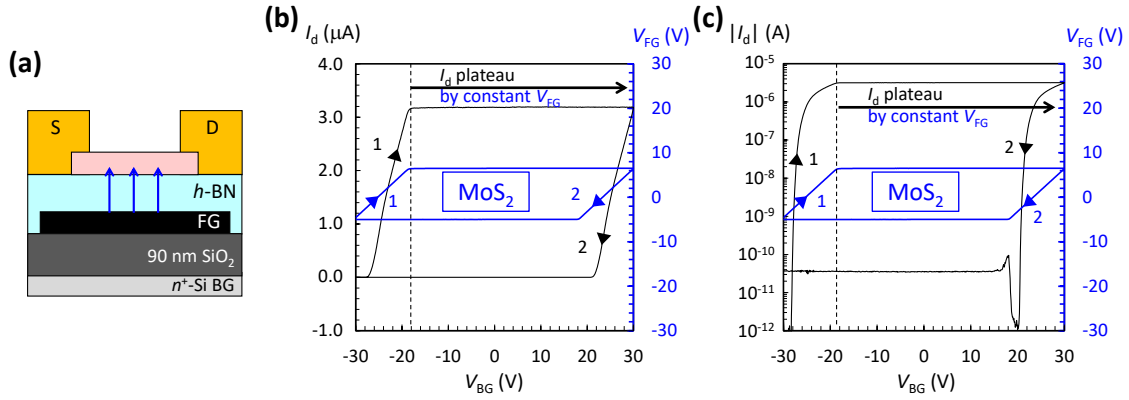


Fig. 4-20 The origin of I_d plateau. **(a)** Schematic of the device without access region. Channel conductivity is fully controlled by V_{FG}. **(b)** Linear scale and **(c)** semi-log scale of round sweep transfer curves with V_{FG} trajectory. The starting point of I_d plateau is well consistent with that of FT region.

As shown in **Figs. 4-20(b)** and **(c)**, plateau region was observed in the I_d–V_{BG} curves resulted from the device without access region, regardless of the I_d scale. Interestingly, as summarized in **Table 4-1**, the I_d plateau has been often observed in previous studies. The V_{FG} trajectory clearly explains the origin of the I_d plateau.

If contact resistance is quite high, it limits the drain current when channel conductance is relatively small. However, as mentioned with **Fig. 4-6**, ohmic contact at Ni/MoS_2 interface has been experimentally confirmed in this study as well as some previous studies summarized in **Table 4-1**. Therefore, the contact resistance is not the reason why the observation of I_d plateau. On the other hand, since there is no access region, the conductivity of the channel is completely controlled by the FG as illustrated in **Fig. 4-20(a)**. Here, V_{FG} is determined by the capacitive coupling between FG and BG, and pinned at one side of tunnel starting voltage, once the feedback mechanism is activated. According to **Figs. 4-20(b)** and **(c)**, the starting point of I_d plateau is well consistent with that of FT region. Consequently, constant V_{FG} in FT region results in the I_d plateau. The observation of I_d plateau in previous studies might be attributed to the constant V_{FG} in FT region of the trajectory, since most of their devices have no access region.

It should be mentioned that, the previous study[111] has also prepared access region for their black phosphorus/*h*-BN/graphene heterostructured NVM device, whereas V_{FG} measurement was not employed. Their round sweep transfer curves can be reasonably explained by the above understanding, which further supports its validity.

4.7 Summary

In this chapter, comprehensive understandings of operation mechanisms of 2D materials-based flash memory were unveiled by the analysis of the V_{FG} trajectories of the $MoTe_2$, WSe_2 and MoS_2 devices. The operation mechanisms were clarified by focusing on three kinds of current limiting paths, i.e., FN tunneling across *h*-BN (path 1) for region A, communication between access region and FG region (path 2) for region B, and hole injection from electrode to channel (path 3) for region C. In addition, for the first time, metal-2D tunneling path in the device without access region was experimentally proved. Although further studies will be required to fully quantize the theory, at this point, the understandings are quite useful to design the materials and structure of 2D heterostructured NVM devices. In addition, the performances of the device can also be predicted, which will be discussed in next chapter.

Chapter 5
Performances

As mentioned in previous chapters, the validity of the benchmarks presented in previous literatures has not been reasonably supported, since memory window extracted by *I–V* round sweep measurement is generally overestimated. In addition, the performances of the device with superior 2D/2D heterostructured tunneling path are still unknown since most of previous studies have investigated the device with 2D/metal heterostructured tunneling path.

In this chapter, therefore, the performances of 2D heterostructured non-volatile memory (NVM) devices are reasonably investigated as mainly focused on program and erase (P/E) speed. This is because, P/E speed determines the range of applications.

5.1 Program and Erase Speed

5.1.1 Preparation for High-speed Measurement

For the high-speed measurement, the pulse generator should be carefully selected first. Regarding the Keysight B1500A semiconductor parameter analyzer system, there are two modules for the high speed (<1 μs) measurement as summarized in **Table 5-1**[129]. One is the B1530A waveform generator/fast measurement unit (WGFMU) module with B1531A remote-sense and switch unit (RSU) module for output, and the other is the B1525A high voltage semiconductor pulse generator unit (HV-SPGU) module. Although WGFMU is specified for ultra-fast *I–V* measurement, maximum output voltage is not large enough for the P/E operation. Since ±30 V of output is required in this study, the HV-SPGU is selected. Measurement system, including a probe station and cables, is the same as DC measurement.

Table 5-1. Comparison of capabilities between WGFMU and HV-SPGU[129].

	WGFMU	**HV-SPGU**
±30 V output	Impossible	Possible
Voltage measurement	Possible*1	Impossible*2
Current measurement	Possible	Impossible

*1 Only output voltage from the WGFMU module can be monitored. Voltage measurement for arbitrary points like an oscilloscope cannot be performed.
*2 Although HV-SPGU has the voltage measurement capability, it is not accessible via standard measurement software *EasyEXPERT*.

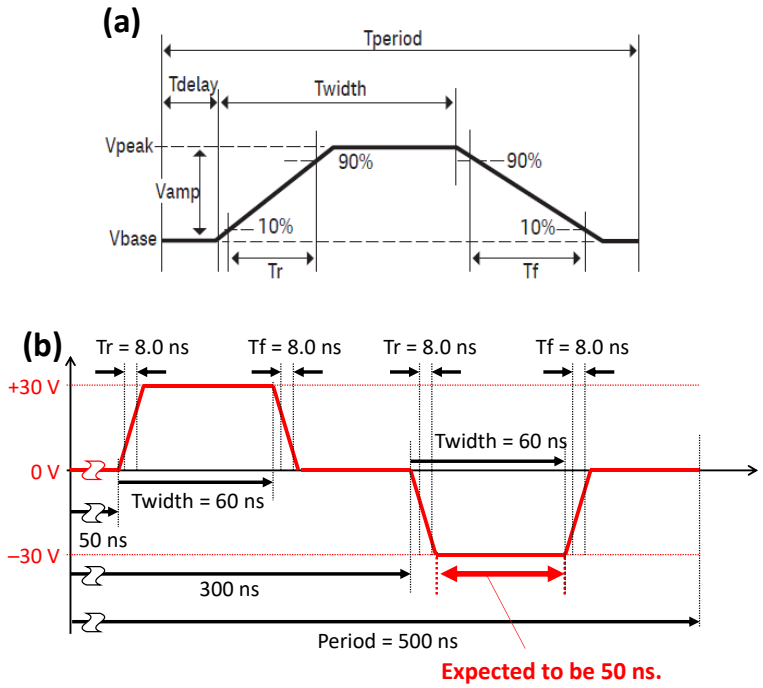


Fig. 5-1 (a) Definition of each parameter which constructs the output pulse[129].
(b) Illustration of the designed waveform for testing the pulse application.
The values of *Tr*, *Tf* and *Twidth* are the minimum values in this study.

Next, settings for HV-SPGU are described. **Fig. 5-1(a)** shows the definition of each parameter which constructs the output pulse[129]. It should be noted that, the *Twidth* includes *Tr*, that is, the duration of voltage plateau is less than *Twidth*. According to the data sheet shown in **Table 5-2**[129], minimum value of *Twidth* is 50 ns when *Vamp* is larger than 10 V, and minimum programmable value of *Tr*/*Tf* is 8 ns. In order to ensure the rectangle shape of the pulse, minimum value of *Twidth* in this study was set to 60 ns. Consequently, the fastest voltage pulse in this study is designed as shown in **Fig. 5-1(b)**.

Of course, this periodic waveform shown in **Fig. 5-1(b)** is only for testing the pulse application. For actual program or erase operation, single pulse of +30 V or –30 V is applied to the back gate (BG) stage.

Table 5-2. Pulse range and pulse parameter[*1] of HV-SPGU[129].

Frequency range		0.1 Hz to 33 MHz
Pulse period	Programmable range	30 ns to 10 s
	Resolution	10 ns
	Minimum	100 ns[*3]
	Accuracy	±1 % (±0.01 %[*2])
Width	Programmable range	10 ns to (period–10 ns)
	Resolution	2.5 ns (Tr and Tf ≤ 8 μs) 10 ns (Tr or Tf > 8 μs)
	Minimum	**50 ns** (25 ns, typical)[*3]
	Accuracy	±(3% + 2 ns)
Transition time (Tr/Tf)[*5]	Programmable range	**8 ns to 400 ms (10–90%)**
	Resolution	2 ns (Tr and Tf ≤ 8 μs) 8 ns (Tr or Tf > 8 μs)
	Minimum (typical)	< 15 ns[*3]
	Minimum	20 ns (Vamp ≤ 10 V) 30 ns (Vamp ≤ 20 V) **60 ns (Vamp > 20 V)**
	Accuracy (10 –90%)	–5% to 5% + 10 ns (Vamp ≤ 10 V) –5% to 5% + 20 ns (Vamp ≤ 20 V)
Output relay switching time[*4]	Open/close	<100 μs

*1. Unless otherwise stated, all specifications assume a 50 Ω termination.
*2. Supplemental characteristics.
*3. This is specified at Vamp ≤ 10 V.
*4. Minimum value in which timing accuracy can be applied.
*5. The time from 10% to 90% of Vamp which is the amplitude of output pulse.

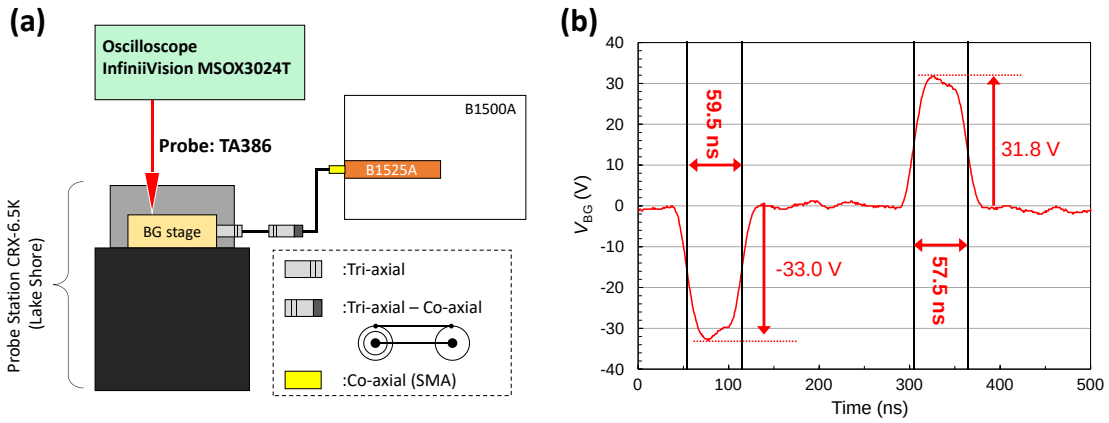


Fig. 5-2 (a) Measurement setup for high-speed voltage pulse. **(b)** Waveform of V_{BG} measured by the oscilloscope when HV-SPGU is set to apply the voltage pulse shown in **Fig. 5-1(b)**.

Since the parasitic impedance of the measurement system generally distorts the pulse shape, actual output was checked by the oscilloscope (InfiniiVision MSOX3024T, 200 MHz, 5 GSa/s) with the 200 MHz passive probe (TA386). Measurement setup is illustrated in **Fig. 5-2(a)**. The oscilloscope proved the BG stage. **Fig. 5-2(b)** shows the measured voltage (V_{BG}) waveform when the settings of HV-SPGU is as shown in **Fig. 5-1(b)**. Although the pulse was a little distorted, sufficiently narrow pulse application to the BG could be confirmed. For the –30 V pulse, actual amplitude was –33.0 V and full width at half maximum (FWHM) was 59.5 ns, and for the +30 V pulse, actual amplitude was +31.8 V and FWHM was 57.5 ns. In this study, these pulses are treated as –30 V, 50 ns pulse and +30 V, 50 ns pulse, respectively since the duration of voltage plateau can be expected to be 50 ns as shown in **Fig. 5-1(b)**. Similarly, the word *XX seconds pulse* in this study is defined as the voltage pulse where expected duration of plateau is XX seconds.

5.1.2 P/E Speed of the Devices with Access Region

Next, P/E speed of the fabricated three devices, i.e., the $MoTe_2$, WSe_2, and MoS_2 channel devices without access region was evaluated. Before the evaluation, the P/E speed was roughly predicted based on the understandings mentioned in **chapter. 4**. Since program operation is conducted by +30 V of V_{BG} pulse, corresponding current limiting path is path 1: Fowler-Nordheim (FN) tunneling across *h*-BN. Since relatively large tunneling current (~2 nA) can be observed for all three devices as shown in **section 4.3.1**, high program speed can be expected. On the other hand, erase operation is conducted by –30 V of V_{BG} pulse, that is, corresponding current limiting path is path 3: hole injection from Ni electrode to the transition metal dichalcogenide (TMDC) channel. As mentioned in **section 4.3.3**, the ability of hole injection is quite poor for *n*-type MoS_2, whereas the hole injection is possible for ambipolar WSe_2 and $MoTe_2$. Therefore, ultra-low erase speed can be expected for the MoS_2 device while the WSe_2 and $MoTe_2$ device will show moderate erase speed.

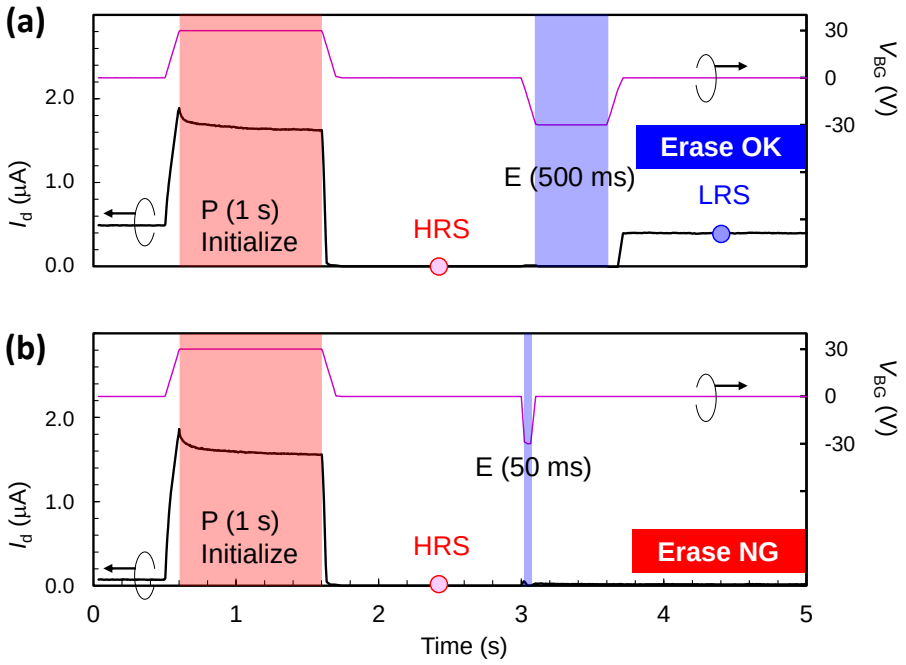


Fig. 5-3 Operation waveforms of the MoS_2 device without access region. After initialize by +30 V and 1 s of V_{BG} pulse, **(a)** –30 V, 500 ms pulse can erase the device while **(b)** –30 V, 50 ms erase cannot.

Fig. 5-3 shows the typical P/E operation waveforms of the MoS_2 device. First, the device was initialized to high resistance state (HRS) by +30 V and 1 s of V_{BG} pulse. Here, drain-source voltage (V_{ds}) was kept at 100 mV throughout the P/E operation. Read operation was conducted with 0 V of V_{BG} after programming. Then, erase operation was conducted by –30 V of V_{BG} pulse. As shown in **Fig. 5-3(a)**, when erase operation was conducted with 500 ms pulse width, the state can be flipped from HRS to low resistance state (LRS), while the state cannot be flipped when the pulse width was 50 ms, as shown in **Fig. 5-3(b)**. In other words, erase operation can be performed in 500 ms, while not in 50 ms. Similarly, P/E speed of the WSe_2 device was investigated. It should be noted that, for the $MoTe_2$ device, P/E speed was investigated in terms of threshold voltage (V_{th}) shift since there is no drain current (I_d) difference at V_{BG} = 0 V as shown later (**Fig. 5-5(a)**). The results are summarized in **Fig. 5-4**. As predicted above, the MoS_2 device showed very slow erase speed of 100 ms, while ambipolar WSe_2 and $MoTe_2$ device show moderate erase speed of 20 ms and 5 ms, respectively. This is qualitatively consistent with the amount of

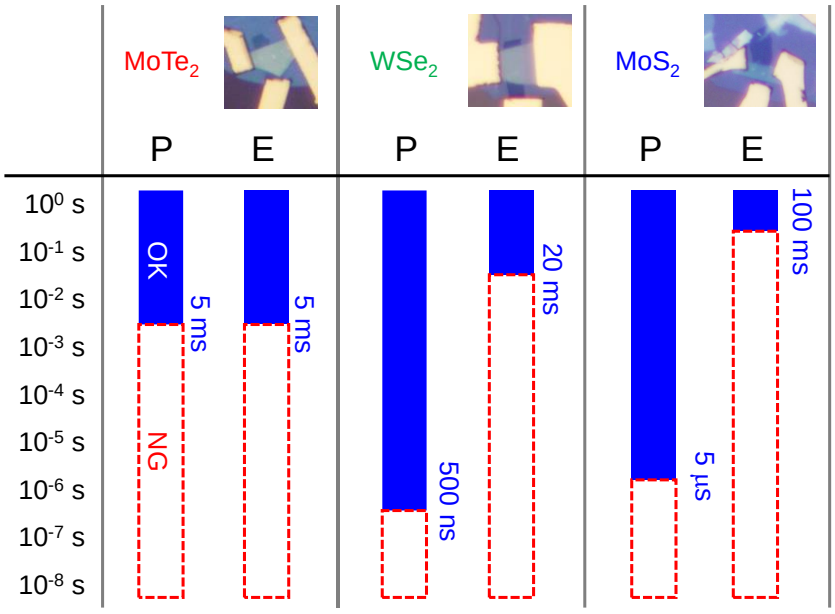


Fig. 5-4 Measured P/E speed of the $MoTe_2$, WSe_2 and MoS_2 devices without access region. The WSe_2 and MoS_2 devices show remarkable programming speed as for flash memory device.

tunneling current in region C as described in **section 4.3.3**, that is, the larger the tunneling current, the faster the erase speed. This clearly shows the usefulness of the provided understandings. However, the results also suggest the limitation of hole injection in terms of the erase speed. On the other hand, regarding to the program speed which was predicted to be fast, there was a variation. While the program speed was 500 ns for the WSe_2 device, that is only 5 ms for the $MoTe_2$ device.

To reveal the origin, operation mechanism in region A is reconsidered. In region A, whole channel region is *n*-type, and holes are tunneled from the FG to channel. The holes are recombined with electrons in the channel. Here, I_d–V_{BG} round sweep transfer curves just before speed test are shown in **Fig. 5-5**. Since many experiments such as low temperature measurement have passed, they are slightly different from **Fig. 4-1**, but sufficiently similar. The values of I_d at V_{BG} = +30 V are also labeled in **Fig. 5-5**. Due to the difference of device geometry and electron density in the channel, the amount of I_d was varied. Interestingly, the larger the I_d, the faster the program speed, suggesting that the program speed can be fastened by increasing the number of recombined electrons. Another possible origin is the thickness of *h*-BN. As summarized in **Table 2-1**, the thickness of *h*-BN is 15.0 nm for the $MoTe_2$ device, 9.6 nm for the WSe_2 device, and 10.6 nm for the MoS_2 device. The thinner the *h*-BN thickness, the faster the program speed. Since the program voltage is the same (+30 V) for all devices, thinner *h*-BN may result in the larger tunneling current during program operation.

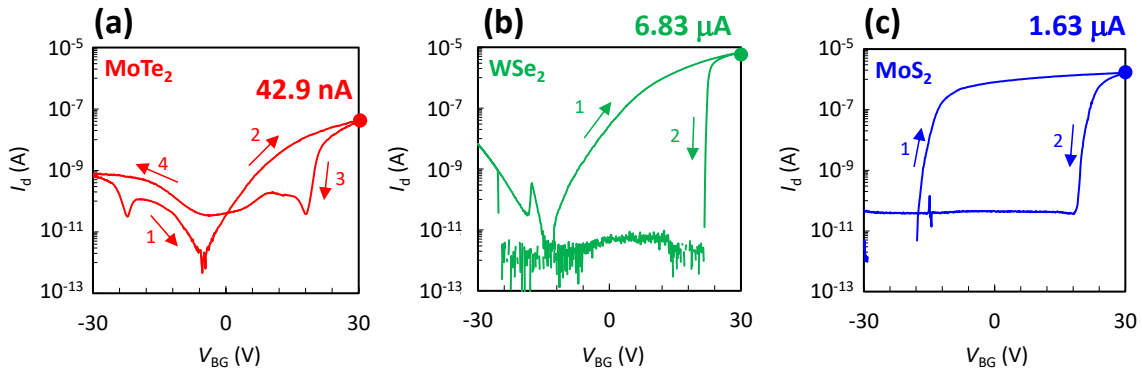


Fig. 5-5 I_d–V_{BG} round sweep curves just before the speed test for **(a)** the $MoTe_2$ device, **(b)** the WSe_2 device, and **(c)** the MoS_2 device.

Fast program operation of the WSe_2 and MoS_2 device is emphasized here. Since typical write speed of flash memory is known as 100 μs[12], 500 ns (the WSe_2 device) and 5 μs (the MoS_2 device) of program speed are remarkably fast, even though the large parasitic capacitance exists due to global BG structure, suggesting that the ultra-fast P/E operation will be potentially achieved by 2D heterostructured NVM device.

5.1.3 Device Design for High-speed and Robust P/E Operation

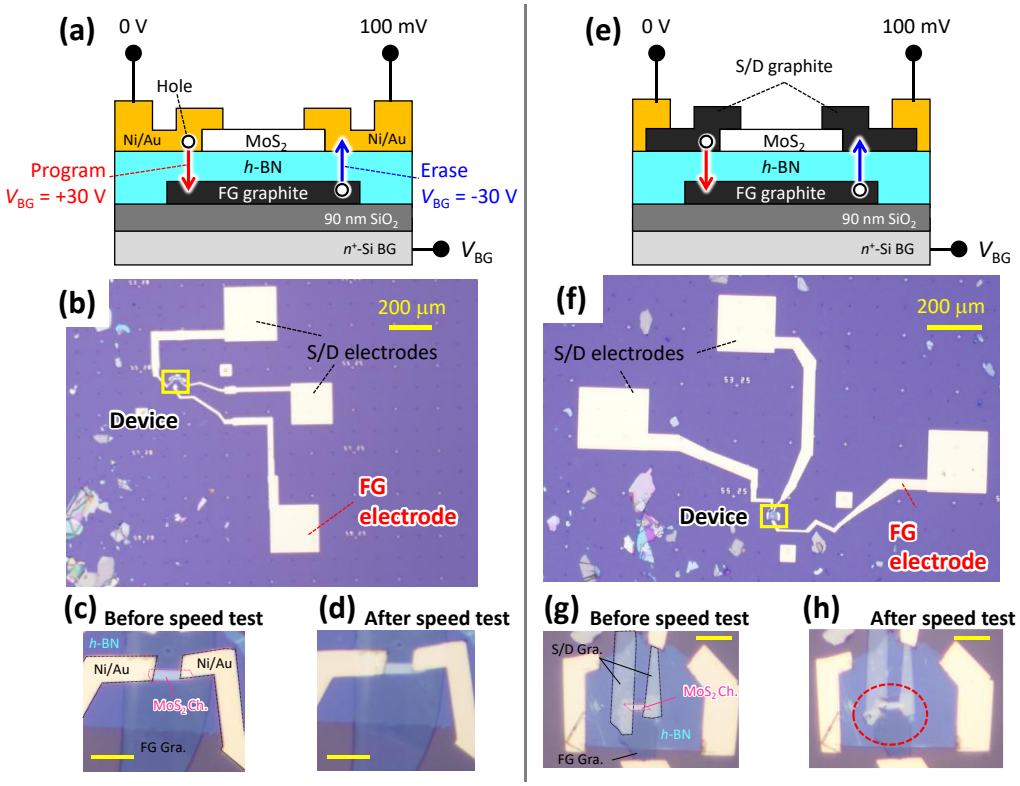


Fig. 5-6 Schematics and photos of the devices with huge FG electrode. Access region is eliminated for employing the tunneling between the FG and S/D. **(a)** Schematic and operation mechanism of the metal contact device. Its **(b)** overall photo and **(c)**, **(d)** enlarged photos before and after speed test, respectively. **(e)** Schematic and operation mechanism of the graphite contact device. Its **(f)** overall photo and **(g)**, **(h)** enlarged photos before and after speed test, respectively.

For further improvement of the P/E speed, the devices were designed. The important point for high-speed operation is that, to inject/eject the holes into/from the FG. As mentioned in previous section, hole injection from Ni to the channel may limit the erase speed to the order of millisecond. Therefore, regarding the P/E speed, the device structure without access region can be considered again since the holes are directly injected to the FG from source/drain (S/D) electrodes (ejected from the FG to S/D electrodes), not via the channel. As mentioned in **section 4.4.2**, the device without access region has been already fabricated. In addition, another device was fabricated with graphite S/D electrodes for realizing the 2D heterostructured tunneling path. Schematics and photos of the devices are shown in **Fig.**

5-6. Except the S/D material, the same materials were used. Here, graphite S/D electrodes were transferred by polydimethylsiloxane (PDMS) assisted dry transfer technique, after the MoS_2/*h*-BN/FG graphite heterostructure fabrication. The expected operation is illustrated in **Figs. 5-6(a)** and **(e)**. Since metal (Ni/Au) or bulk graphite was used as the S/D, large tunneling current by a lot of holes can be expected. For the program operation, holes are tunneled from the FG to the source electrode since V_{FG} is positive, while for the erase operation, holes are tunneled from the drain electrode to the FG since V_{FG} is negative. Generally, the quality of interface between evaporated metal and 2D is poor due to the damage by evaporated metal atoms with high kinetic energy[130]. It was also observed in this study as shown in **Fig. 5-7**. Therefore, the superiority of 2D/2D interfaces in terms of the P/E speed will be extracted by the comparison of their speed.

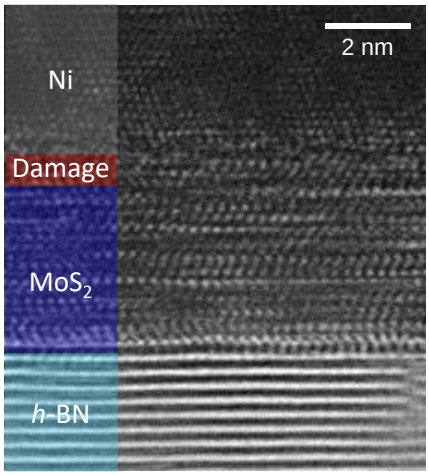


Fig. 5-7 Typical transmission electron microscope (TEM) image of 2D heterostructure underneath the thermally evaporated Ni. Damage layer is clearly seen between the MoS_2 and Ni.

However, unfortunately, both devices were broken just after the speed test was performed. Although 100 ns of program speed can be achieved for both devices, the insulating property of *h*-BN was drastically degraded. For the graphite contact device, as compared with **Fig. 5-6(g)**, **Fig. 5-6(h)** clearly indicates the unrecovered damage at the 2D heterostructured tunneling path, suggesting that the destruction is resulted from the breakdown due to large tunneling current. For the metal contact device, even though the

damage is not obvious in **Fig. 5-6(d)** as compared with **Fig. 5-6(c)**, insulating property of *h*-BN was severely degraded as shown in **Fig. 5-8**, where I_{tun} represents the tunneling current from the FG to the source electrode.

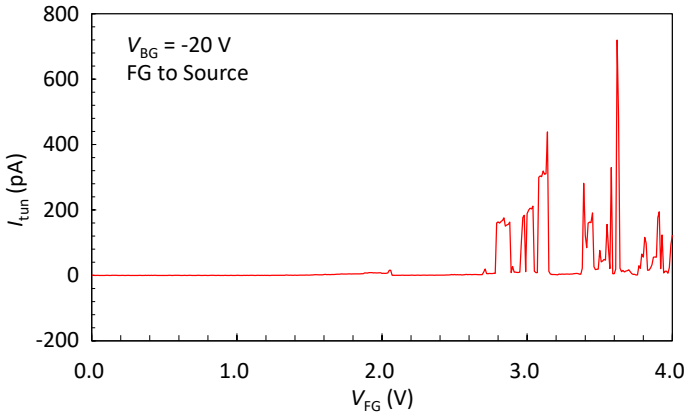


Fig. 5-8 Typical tunneling current characteristic after the degradation of *h*-BN for the metal contact device.

As shown in **Figs. 5-6(b)** and **(f)**, both devices have large FG electrode and pad (~ 200 μm × 200 μm), which may lead to the large tunneling current to charge or discharge the FG. Here, the typical value of tunneling current is estimated. Firstly, change in the amount of charge stored in the FG (ΔQ) is calculated by following equation:

$$\Delta Q = C_{ox}\Delta V_{th}, \tag{5.1}$$

where, C_{ox} and ΔV_{th} are the capacitance between the FG and BG, and the change of V_{th} caused by ΔQ. For C_{ox}, 90 nm SiO_2 (relative permittivity is 3.9) with the area of 200 μm × 200 μm is assumed. If ΔV_{th} is 10 V, $\Delta Q = 1.52 \times 10^{-10}$ C is resulted, which corresponds to around 10^9 holes. Since the holes should be tunneled to the FG within 100 ns, tunneling current is calculated to be 1.52 mA, which is too large as a tunneling current. Therefore, for the ultra-fast and robust P/E operation, the area of FG pad should be small enough to prevent the breakdown, even at the expense of V_{FG} and tunneling current measurement capability.

Before the destruction, tunneling current in graphite/*h*-BN/graphite system had been measured by the graphite contact device shown in **Fig. 5-6(f)**. Therefore, before the

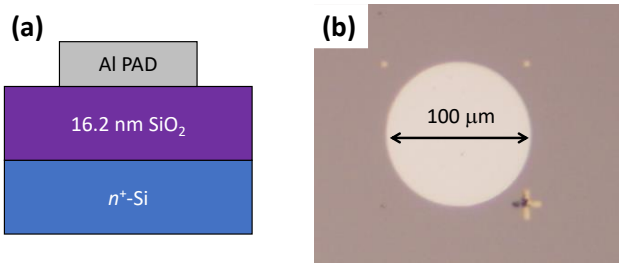


Fig. 5-9 (a) Schematic and **(b)** photo of fabricated MIS capacitor.

next device design and fabrication, the potential of 2D heterostructured tunneling path was confirmed in terms of the amount of tunneling current. For the purpose, Al/SiO_2/n^+-Si metal-insulator-semiconductor (MIS) capacitor was additionally fabricated as shown in **Fig. 5-9**. The thickness of SiO_2 measured by an ellipsometer is 16.2 nm, and diameter of Al pad is 100 μm. The tunneling current was measured in accumulation condition, i.e., positive voltage was applied to the Al pad. Therefore, only the electron tunneling current from n^+-Si to Al pad is considered as a tunneling current in 3D system.

Fig. 5-10 shows the comparison of tunneling current density (J_{tun}) in the 2D system and the 3D system as a function of electric field across the insulator (E_{ox}). For the 2D system, V_{BG} was fixed at –30 V when tunneling current flows from the graphite S/D to graphite FG, while it was fixed at +30 V when the current flows in the opposite direction, in order to mimic the erasing and programming situations, respectively. When the tunneling current flowed from the graphite S/D to graphite FG (blue), J_{tun} was much larger

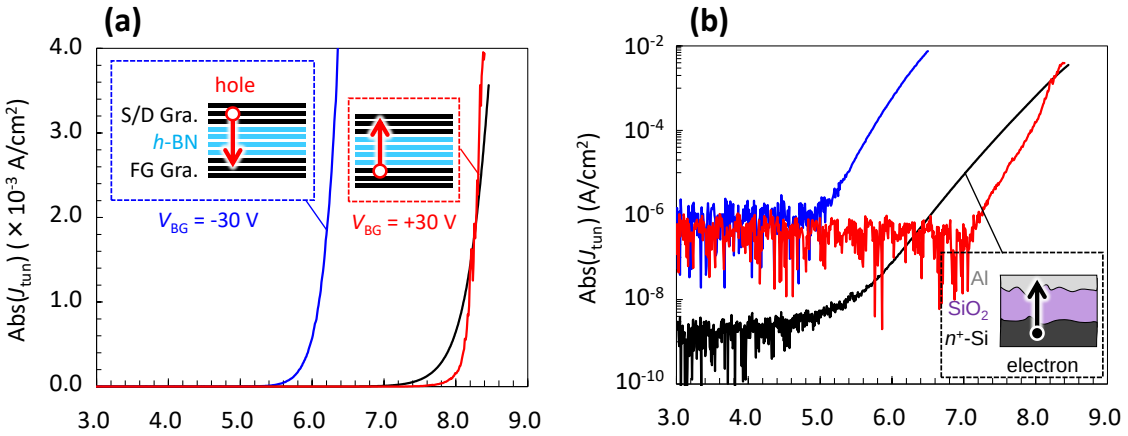


Fig. 5-10 The comparison of tunneling current in 2D system (graphite/*h*-BN/graphite) and 3D system (Al/SiO_2/n^+-Si). **(a)** Linear scale and **(b)** semi-log scale.

than that of 3D system. On the other hand, for the tunneling current flowed from graphite FG to graphite S/D (red), larger field was required to start the tunneling than 3D system. However, the J_{tun} in 2D system was finally larger than that of 3D system due to its higher sensitivity to E_{ox}. As a result, the potential of 2D heterostructured tunneling path for ultra-fast P/E operation could be confirmed.

Since the 2D heterostructure is symmetric (graphite/*h*-BN/graphite), the tunnel starting fields seem to be the same, which is not consistent to the measurement results. This can be explained by focusing on the Fermi level (E_F) shift of *hole source* graphite. For the case of the tunneling from FG to S/D, hole source graphite is the FG as shown in **Fig. 5-11(a)**. Here, E_F shift of the FG and S/D graphite is explained step by step, by using band diagrams shown in **Fig. 5-11(b)**. First, (i) the situation with V_{FG} = 0 V is considered. By applying +30 V to the BG, E_F of FG graphite is shifted upward while its potential (0 V) is maintained. Next, (ii) V_{FG} is swept from 0 V to positive direction. (iii) Due to positive V_{FG}, E_F of S/D graphite is shifted upward while its potential (0 V) is maintained. (iv) The holes in the FG are hard to tunnel since the E_F of FG is initially shifted upward. The situation is completely opposite for the case of the tunneling from S/D to FG. As shown in **Fig. 5-11(c)**, *hole source* graphite is the S/D. Since E_F of S/D graphite is shifted downward due to negative V_{FG} as shown in **Fig. 5-11(d)**, hole is easy to tunnel, resulting in the low tunnel starting field. This is the possible explanation for the difference of tunnel starting voltages.

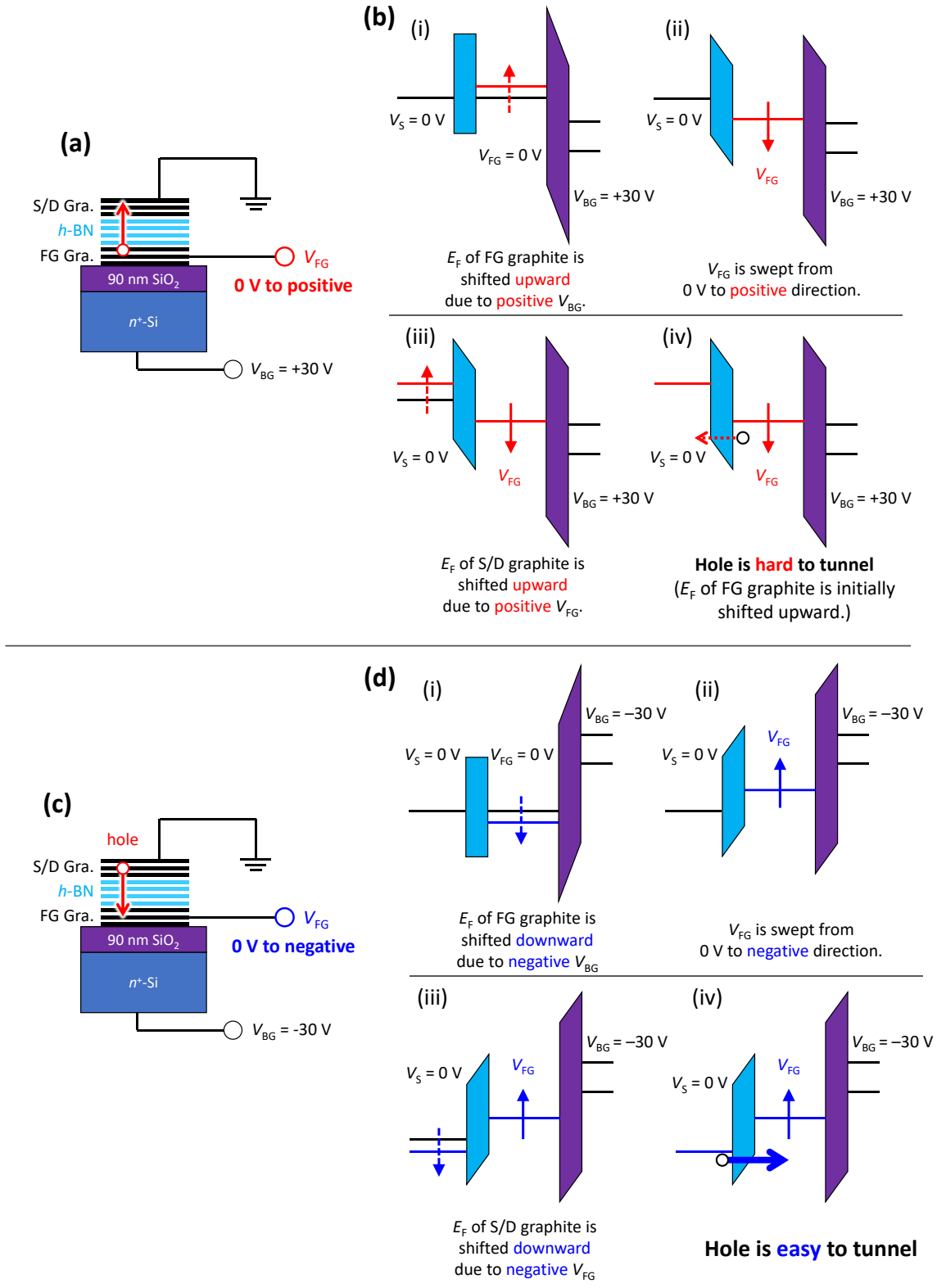


Fig. 5-11 Possible origin of the difference of tunnel starting field for tunneling currents in 2D system. Bias conditions of the tunneling **(a)** from FG to S/D, and **(c)** from S/D to FG. **(b)**, **(d)** Band diagrams which explain the E_F shift of FG and S/D graphite for the case of **(a)** and **(c)**, respectively.

For the next device fabrication, the area of FG was carefully designed. Although small or no FG electrode is required to prevent the large tunneling current during ultra-fast P/E operation, large FG electrode is required to ensure the capacitive coupling ratio. If the program operation is performed by +30 V of V_{BG}, V_{FG} can be calculated as follows:

$$V_{FG} = \frac{C_{ox}}{C_{ox} + C_{BN}} \times 30, \tag{5.2}$$

where, C_{BN} is the capacitance between the channel and FG. On the other hand, according to the results shown in **section 4.3.1**, required electric field for the tunneling across *h*-BN is around 7.0 MV/cm. Therefore, for the successful program operation, following equation should be satisfied:

$$\frac{C_{ox}}{C_{ox} + C_{BN}} \times 30 > 7.0 \ \ (\mathrm{MV/cm}) \times t_{BN}, \tag{5.3}$$

where, t_{BN} is the thickness of *h*-BN. From the **Eq. 5.3**, the importance of large capacitive coupling ratio (C_{ox} / (C_{ox} + C_{BN})) is clear. By assuming the thickness of SiO_2 is 90 nm, left side and right side of **Eq. 5.3** were calculated as a function of t_{BN}, as shown in **Fig. 5-12**. For the calculation of left side, three kinds of area ratio (S_{ox}/S_{BN}) are assumed, where S_{ox} and S_{BN} represent the active area for C_{ox} and C_{BN}, respectively. Inset illustrates the cross section of the device (for more information, please see **Fig. 2-12**). According to the figure, for example, when t_{BN} is 8.0 nm, it can be understood that the S_{ox} should be larger than $2S_{BN}$ for successful program operation.

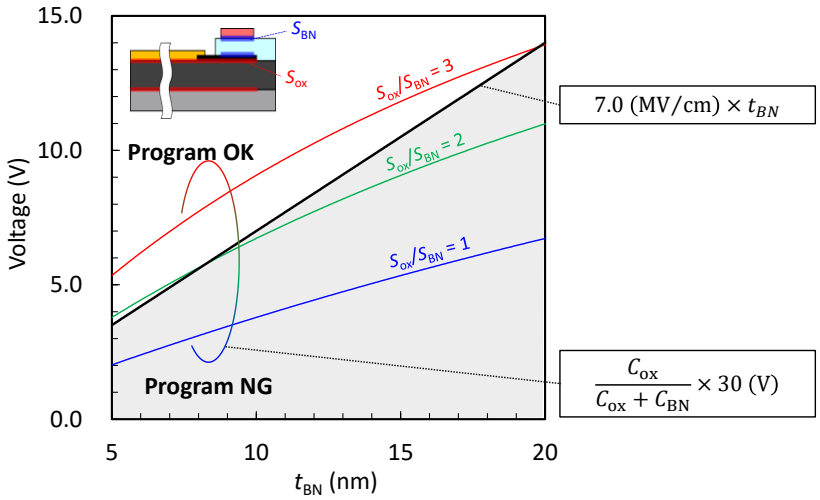


Fig. 5-12 The design of coupling ratio of the device.

5.1.4 Fifty Nanoseconds P/E Operation and Its Possible Origin

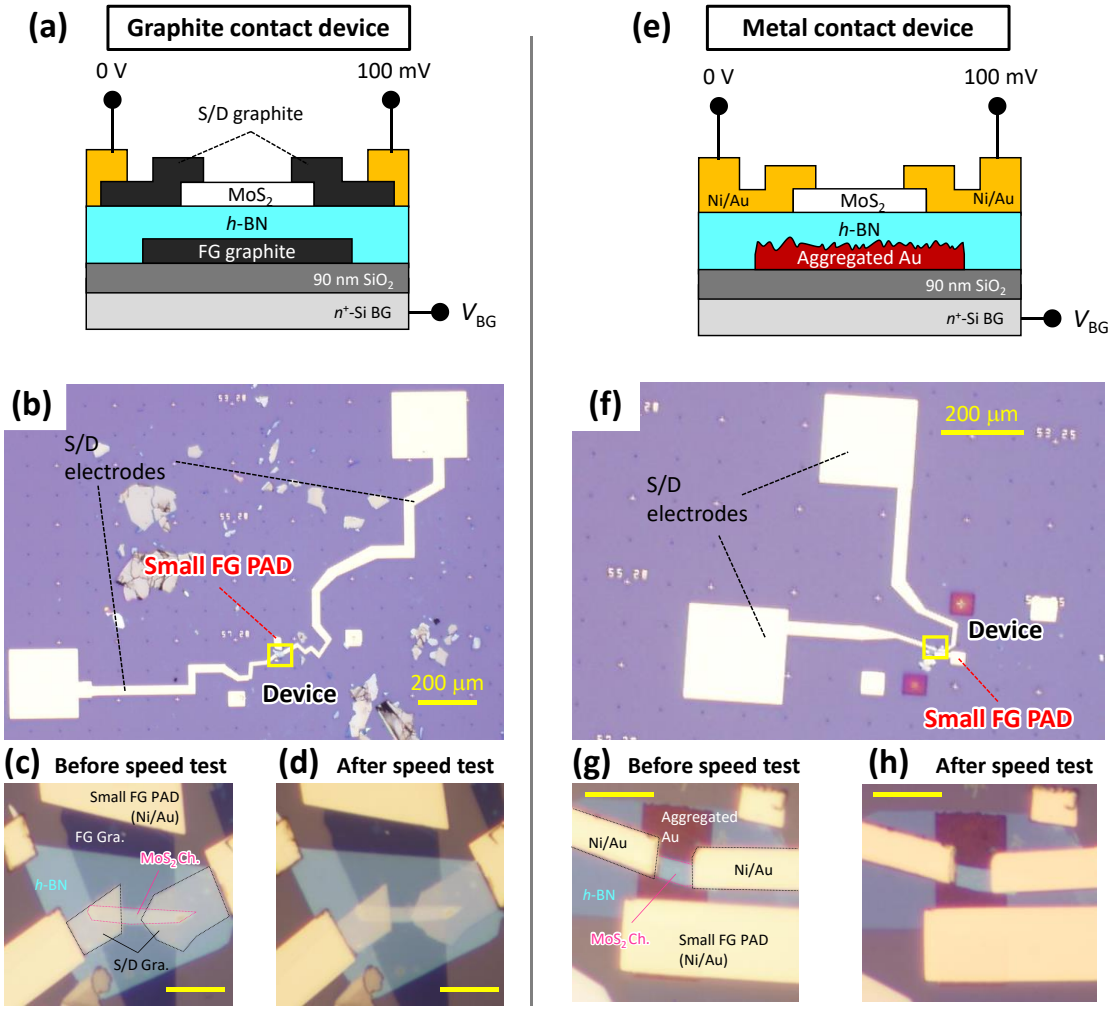


Fig. 5-13 Schematics and photos of the device with small FG pad. **(a)** Schematic of the graphite contact device. Its **(b)** overall photo and **(c)**, **(d)** enlarged photos before and after speed test, respectively. **(e)** Schematic of the metal contact device with aggregated Au FG. Its **(f)** overall photo and **(g)**, **(h)** enlarged photos before and after speed test, respectively.

Based on the above design policy, two devices were fabricated again with small FG pad (less than 50 μm × 50 μm) as shown in **Fig. 5-13**. One is the graphite contact and graphite FG device shown in **Figs. 5-13(a)–(d)**, the other is the metal contact and metal FG device shown in **Figs. 5-13(e)–(h)**. For the later, thermally aggregated Au was used as an FG for increasing roughness at hetero-interface. This was prepared by heating the substrate on a hot plate (200°C, 2 min) following thin Au evaporation. Hereinafter, the graphite contact and graphite FG device, and the metal contact and metal FG device are simply called the *graphite contact device* and the *metal contact device*, respectively.

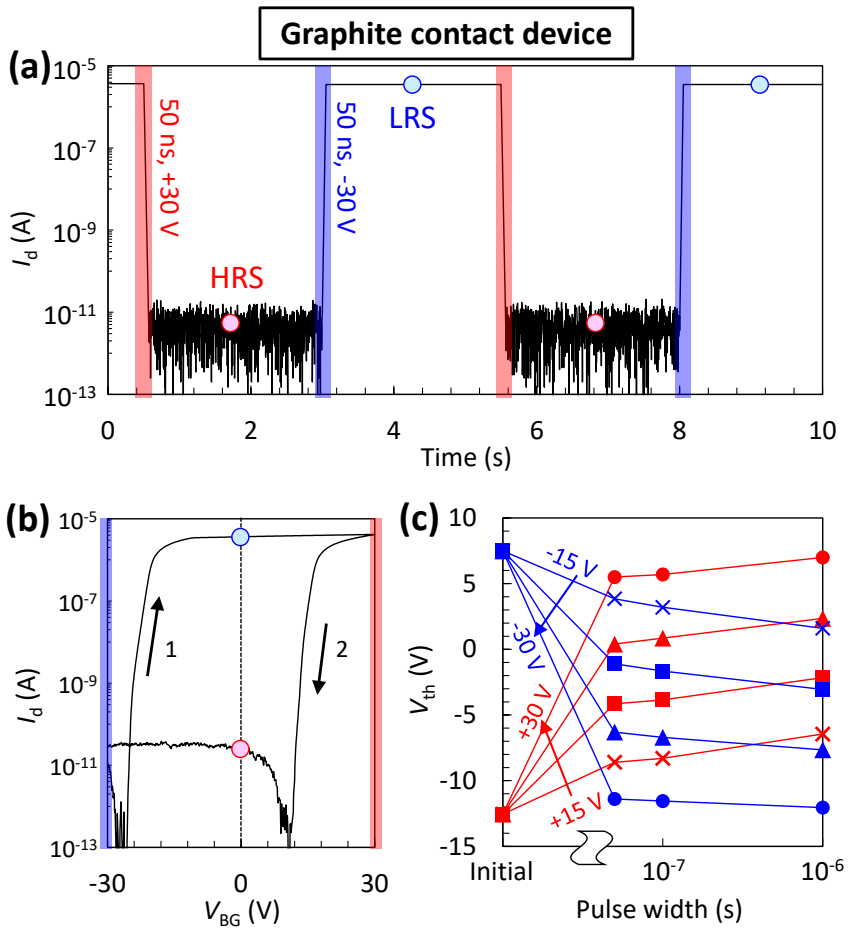


Fig. 5-14 Speed test results of the graphite contact device. **(a)** Drain current waveform with 50 ns P/E pulses. **(b)** I_d–V_{BG} round sweep curves. **(c)** V_{th} shift after P/E operation as a function of pulse width.

At first, the speed test was conducted for the graphite contact device. Remarkably, the device showed 50 ns ultra-fast P/E operation as shown in **Fig. 5-14(a)**. This is consistent with the above mentioned tunneling current comparison. HRS and LRS are also indicated in I_d–V_{BG} round sweep curves shown in **Fig. 5-14(b)**. In addition, V_{th} shift after P/E operation is summarized in **Fig. 5-14(c)** as a function of pulse width. Even though ±25 V and 50 ns voltage pulses were used to P/E operation, 6.7 V of memory window can be achieved. This clearly indicates that a sufficiently large number of holes are transferred between the FG and electrode during very short period, 50 ns.

Next, the effect of roughness at the hetero-interfaces in tunneling path was investigated. Surprisingly, the metal contact device also showed the 50 ns P/E operation as

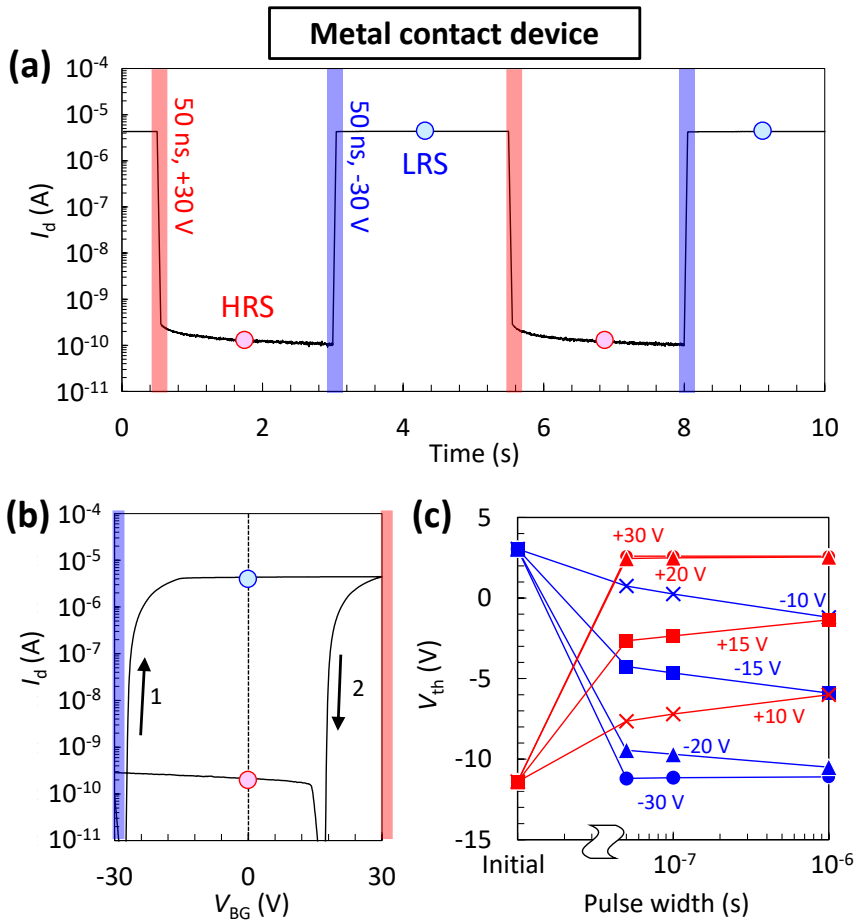


Fig. 5-15 Speed test results of the metal contact device. **(a)** Drain current waveform with 50 ns P/E voltage pulses. **(b)** I_d–V_{BG} round sweep curves. **(c)** V_{th} shift after P/E operation as a function of pulse width.

shown in **Fig. 5-15(a)**. Round sweep I_d–V_{BG} curves of the device is also shown in **Fig. 5-15(b)**. Moreover, similar trend of V_{th} shift was observed as shown in **Fig. 5-15(c)**. Additionally, it is worth noting that, P/E speed test can be repeatedly performed thanks to the small FG pad. **Figs. 5-13(d)** and **(h)** clearly show no degradation of the device even after P/E speed test. According to the results, it is revealed that the 2D/2D interface is not the key of the ultra-fast P/E operation as opposed to the expectation. It should be emphasized again that, the 50 ns pulse is the fastest value in this study to ensure the rectangle shape of the pulse in the DC measurement system. As shown in **Fig. 5-14(c)** and **Fig. 5-15(c)**, 50 ns is long enough for ensuring the memory window when the V_{BG} is ±30 V, suggesting that the 2D heterostructured NVM devices can be programmed/erased even

below 50 ns, as long as the system is specified for high-speed measurement.

More recently, H. -J. Gao group (CAS in China) has demonstrated the 21 ns P/E operation of 2D heterostructured NVM device (InSe channel/*h*-BN tunneling barrier/Multilayer graphene FG)[131], and they have claimed that the 2D/2D interface is the key, which is opposed to above results. In addition, P. Zhou group (Fudan Univ. in China) has also demonstrated the 20 ns P/E operation of 2D heterostructured NVM device (MoS_2 channel/*h*-BN tunneling barrier/Multilayer graphene FG)[132], and they have claimed that the *double-barrier modified FN tunneling* seems to be the key. While they have postulated that the electron tunneling from MoS_2 to the FG when the programming ($V_{BG} > 0$ V), this is not consistent with the discussion in **section 4.3.1**, where hole tunneling from the FG to channel has been proved. Interestingly, 2D heterostructured neuromorphic device has also achieved the large enough conductance modulation with 50 ns voltage pulse, even though the tunneling path includes 2D/3D interfaces[101]. It is obvious that the ultra-fast nature of 2D heterostructured NVM devices will attract huge attention since the speed is much higher than conventional Si-based flash memories (~100 μs[12]). Therefore, it should be supported by reasonable mechanisms.

The point is quite simple, that is, high tunneling current results in high-speed. Although the comparison of tunneling current shown in **Fig. 5-10** confirmed the potential of 2D heterostructured NVM device for ultra-fast P/E speed, the results revealed that 2D/2D interface is not necessary. Therefore, *h*-BN itself can be considered as the key. Although the breakdown strength of *h*-BN (~12 MV/cm[49]) is similar to that of SiO_2 (13~14 MV/cm[133]), these values are determined by DC measurement. In fact, for SiO_2, the relationship between applied electric field and time to breakdown has been investigated in ultra-fast time scale[134,135], which was motivated by the desire to protect the circuit from electrostatic discharge. As shown in **Fig. 5-16**[135], as long as stress time is short enough, ultra-thin SiO_2 (< 5 nm) is not broken down even when the electric field is higher than 14 MV/cm. Based on the results, they have claimed that the trend follows the 1/E model[136], which is one of the most popular models for time-dependent dielectric breakdown (TDDB).

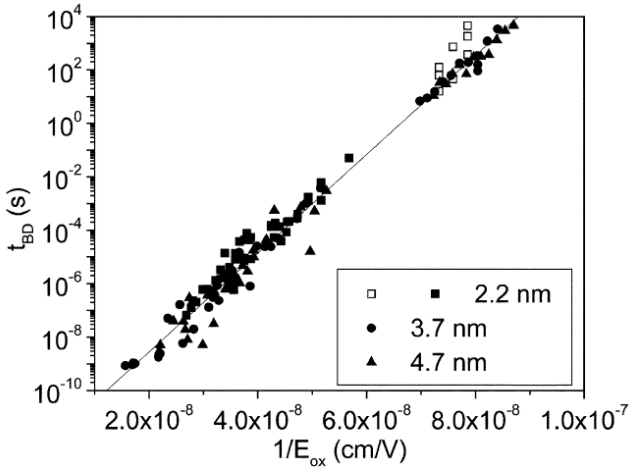


Fig. 5-16 Relationship between time to breakdown and applied electric field of ultra-thin SiO_2[135]. In ultra-fast time scale, the trend follows the 1/E model.

Their experimental results can be fitted by following equation:

$$t_{BD} = C_1 \exp\left(\frac{C_2}{E_{ox}}\right), \tag{5.4}$$

where, C_1 and C_2 are 5.58×10^{-13} s and 431 MV/cm, respectively, and t_{BD} represents the time to breakdown. This suggests that the *h*-BN may possess the breakdown strength more than 12 MV/cm as long as in ultra-fast time scale, which allows the high tunneling current.

Again, the V_{FG} measurement seemed to be the powerful tool to access the mechanisms. If the transient V_{FG} can be measured during program/erase operation in sub-microsecond regime as shown in **Fig. 5-17**, actual electric field across *h*-BN during the operations can be estimated. Since the feedback by tunneling current needs time, the V_{FG} may exceed the tunnel starting voltage just after raising the V_{BG} as indicated by dotted circle in **Fig. 5-17**, resulting in high field and current. However, unfortunately, standard oscilloscope cannot measure the transient response of V_{FG}, since internal resistance of an oscilloscope is low (10 MΩ when 10:1 passive prove is used). This is much lower than that of source/measure unit (SMU) which is used for V_{FG} measurement in DC. The SMU has over 10^{13} Ω of internal resistance[129] which prevents the charge loss during V_{FG} measurement. As a result, the stored charges in the FG move into the oscilloscope during V_{FG} measurement as shown in **Fig. 5-18(a)**. Typical transient response of V_{FG} is shown in **Fig. 5-18(b)**, where the value of V_{FG} is quite low.

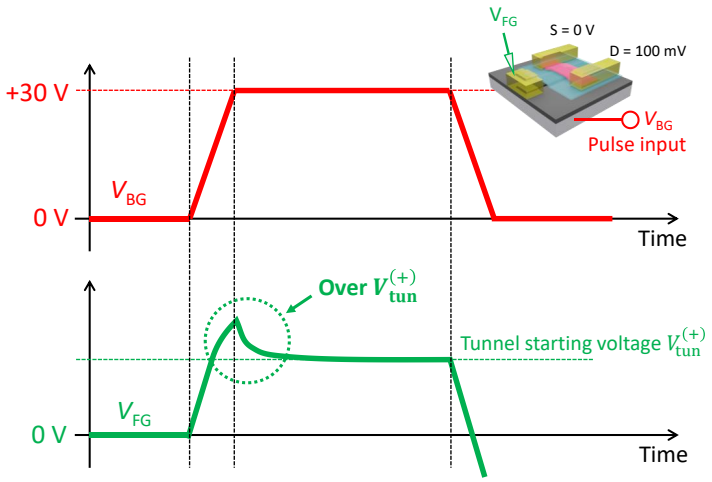


Fig. 5-17 Illustration of expected V_{FG} response in ultra-fast time scale.

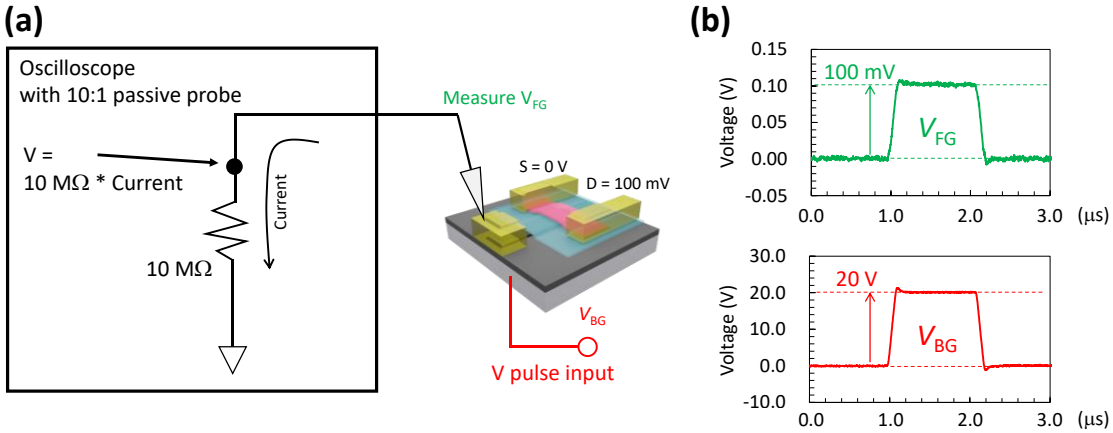


Fig. 5-18 Unmeasurable V_{FG} response due to low internal resistance of an oscilloscope.
(a) Illustration of the charge loss during V_{FG} measurement by an oscilloscope.
(b) Typical waveforms of V_{FG} and V_{BG}.

As an alternative experiment, by using the HV-SPGU, breakdown strength of *h*-BN under fast voltage pulse was evaluated. For the purpose, metal/*h*-BN/metal stacks were fabricated as shown in **Figs. 5-19(a)–(c)**, where top and bottom electrodes are thermally evaporated Ni/Au. First, the voltage pulse was applied to the bottom electrode of sample 1. The pulse width was fixed at 100 ms, and the amplitude was increased subsequently. The sample 1 was finally broken down at 18.4 MV/cm. The sequence is indicated by the cyan arrow in **Fig. 5-19(d)**, and summarized in **Table 5-3**. Next, the same test was conducted for sample 2 while pulse width was fixed at 10 ms. Surprisingly, sample 2 was not broken down even after 26.1 MV/cm, 10 ms voltage stress. Here, required electric field

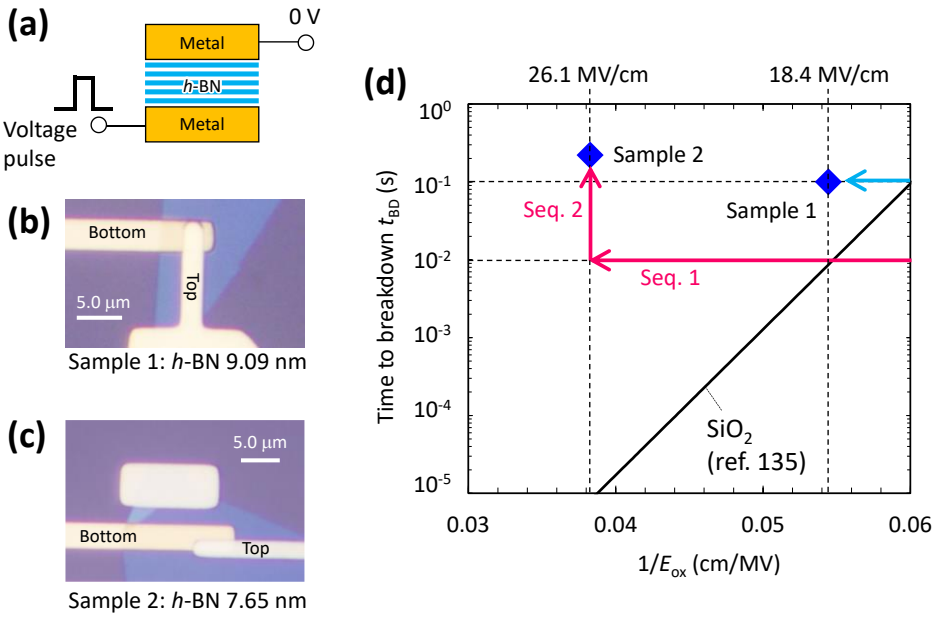


Fig. 5-19 Breakdown strength of *h*-BN under fast voltage pulse stress. **(a)** Schematic of fabricated metal/*h*-BN/metal stack. Top electrode is grounded while the voltage pulse is applied to the bottom electrode. **(b)**, **(c)** Photos of fabricated samples. **(d)** Comparison of time to breakdown between *h*-BN and SiO_2[135]. The details of experimental procedure are described in main text.

for 50 ns P/E operation is estimated by assuming typical situation: the device is fabricated on a 90 nm SiO_2/n^+-Si substrate, memory window is 10 V, the area of FG including FG pad is 50 μm×50 μm, and the tunneling area is 30 μm^2. The required tunneling current density is calculated to be 633 A/cm^2. By substituting the hole effective mass in *h*-BN ($0.47m_0$[128]) and the barrier height at *h*-BN/graphite interface (3.27 eV[45,126,127]) into the FN formula, required electric field can be estimated to be about 21 MV/cm, that is, 26.1 MV/cm is large enough for 50 ns P/E operation. Therefore, the test was continued by applying the 26.1 MV/cm, 10 ms pulse repeatedly, whose purpose is to reveal how many times the *h*-BN can withstand the severe stress. The sample 2 was finally broken down after 22nd application of 26.1 MV/cm, 10 ms pulse. These two sequences are indicated by the magenta arrows in **Fig. 5-19(d)** and summarized in **Table 5-4**. Both samples possessed stronger breakdown strength than SiO_2, suggesting that this is the key of the ultra-fast nature of 2D heterostructured NVM devices.

Table 5-3. Experimental procedure in **Fig. 5-19(d)** for sample 1.

Pulse No.	Field (MV/cm)	Result
1	7.7	OK
2	8.8	OK
3	9.9	OK
4	11.0	OK
5	12.2	OK
6	13.8	OK
7	15.7	OK
8	18.4	**Breakdown**

*Pulse width is fixed at 100 ms.

Table 5-4. Experimental procedure in **Fig. 5-19(d)** for sample 2

	Pulse No.	Field (MV/cm)	Result
Sequence 1	1	7.7	OK
	2	8.8	OK
	3	9.9	OK
	4	11.0	OK
	5	12.2	OK
	6	13.8	OK
	7	15.7	OK
	8	18.4	OK
Sequence 2	9 (1)	26.1	OK
	10 (2)	26.1	OK
	11 (3)	26.1	OK
	…	…	…
	29 (21)	26.1	OK
	30 (22)	26.1	**Breakdown**

*Pulse width is fixed at 10 ms.

It should be emphasized that, the comparison is unfavorable for *h*-BN. In the previous study[134,135], they applied the single voltage pulse to the SiO_2 for evaluating its t_{BD}, while in this study, voltage pulses were applied to the *h*-BN repeatedly. This is resulted from the difference of measurement system (high-speed voltage/current measurement system versus only high-speed pulse generator) and facility of sample preparation

(sophisticated Si process versus exfoliation and dry transfer). For the previous SiO_2 study, ultra-fast current response could be measured to estimate the t_{BD}, and many samples for the breakdown could be easily prepared. The situation is opposite to this *h*-BN study. Moreover, thickness of *h*-BN is thicker than SiO_2. It is well known that the breakdown strength in DC is increased with decreasing thickness, especially below 10 nm[137].

Interestingly, **Fig. 5-19(d)** suggests that the *h*-BN may not follow the 1/E model. While the 1/E model is valid for amorphous SiO_2 whose breakdown mechanism is explained by percolation[138], its validity for layered *h*-BN whose breakdown mechanism is explained by layer-by-layer breakdown[137] is not obvious. In addition, the previous literature[137] has also reported that the parameter of β in Weibull plot is increased with increasing the *h*-BN thickness as opposed to the SiO_2 case, suggesting that the completely different model will be required to describe the breakdown phenomena of *h*-BN. Although the comparison between *h*-BN and SiO_2 in fast time scale (< 1 s) should be sophisticated in the future, the comparison shown in **Fig. 5-19(d)** is quite important as the first step, since it has not been studied yet. The breakdown phenomenon of *h*-BN in fast time scale should be studied for the future to reasonably support the superiority of 2D heterostructured NVM devices.

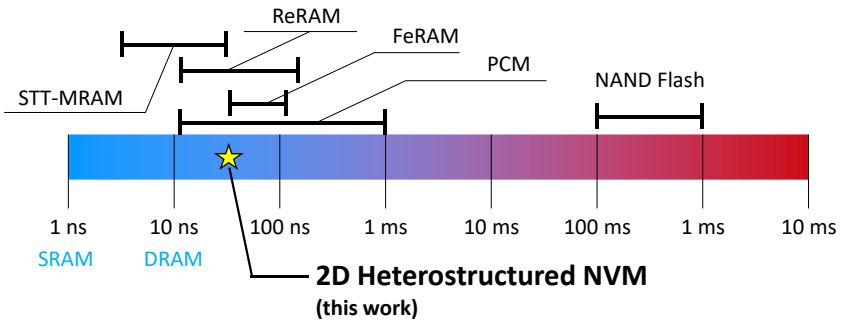


Fig. 5-20 Comparison of the achievement in this work and other memory technologies[12,139].

The achievement is compared with other NVM technologies[12,139] in **Fig. 5-20**. The speed of 2D heterostructured NVM devices is in the range of storage class memory which is now strongly desired.

5.2 Retention and Endurance

Reliability of 2D heterostructured NVM device was also investigated in terms of retention and endurance. Before the endurance test, retention test was carried out since endurance test will break the device. These tests were performed for the MoS_2 device with access region and large FG pad, which is the same as the device shown in **Fig. 2-1(e)**.

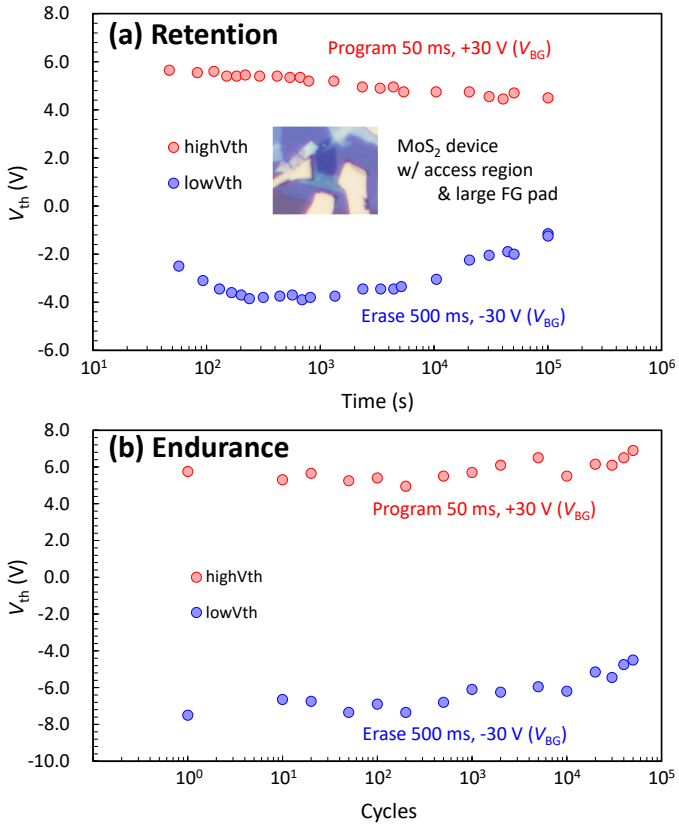


Fig. 5-21 (a) Retention and **(b)** endurance characteristics of the MoS_2 device with access region and huge FG electrode.

Fig. 5-21(a) shows the retention characteristics at room temperature. Program and erase operations were conducted with +30 V, 50 ms voltage pulse, and –30 V, 500 ms voltage pulse, respectively. Over 5.0 V of memory window can be achieved even after 10^5 s, that is the same as the maximum test time in 2D heterostructured NVM device research[109], at this time.

Fig. 5-21(b) shows the endurance characteristics at room temperature. Program and erase pulses are the same as retention study. Both V_{th}s were stable until 5×10^4 P/E

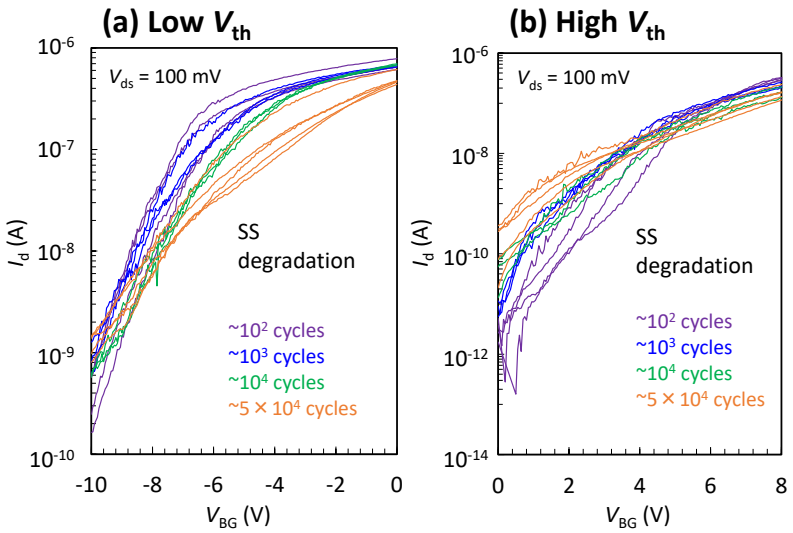


Fig. 5-22 Transition of I_d–V_{BG} transfer curves during endurance test. **(a)** Low V_{th} side and **(b)** high V_{th} side.

cycles which is the largest value among 2D heterostructured NVM studies. However, after 10^5 cycles, the insulating property of *h*-BN was strongly degraded and P/E operation could not be performed. To discuss the origin of degradation, I_d–V_{BG} curves for reading V_{th} were examined. According to **Figs. 5-22(a)** and **(b)**, subthreshold swing (SS) is increased with increasing P/E cycles. When the geometrical capacitances and the temperature are constant, SS degradation can be attributed to the increase of the interface trap capacitance (C_{it}) due to the increase of interface trap density (D_{it}). Although the 2D/2D interface is ideally dangling bond free, it is well known that the sulfur vacancy is easy to be formed in MoS_2 even at room temperature[78,80,81,140]. Since *h*-BN and graphite are stable materials, the SS degradation may be due to the increase of D_{it} in MoS_2 resulted from the defect formation during P/E cycling. However, as well as the speed test, *h*-BN is finally degraded and lost its insulating property, suggesting that the further improvement can be expected with small FG pad.

5.3 Summary

In this chapter, the intrinsic performances of 2D heterostructured NVM device were investigated as mainly focused on the P/E speed. Remarkably, 50 ns ultra-fast P/E operation could be achieved by the device with small FG pad, which prevents the device destruction due to large tunneling current. The P/E speed was in the range of storage class memory. Interestingly, the controlled experiment has proved the superior 2D/2D interface is not the key for the ultra-fast nature. Strong breakdown strength of *h*-BN under fast voltage pulse stress may be the key. Further study for this phenomenon will be required. Reliability of the 2D heterostructured NVM device was also investigated in terms of retention and endurance. Although the examined MoS_2 device has passed the many experiments including low temperature measurement at 10 K, it showed 10^5 s retention and 5×10^4 P/E cycles endurance which are comparable to previous studies. As well as speed test, further improvement can be expected by reducing the FG pad size. In short, according to the results, the *h*-BN itself seems to be the key to great performances of 2D heterostructured NVM, such as high speed, long-time retention and tough endurance.

Chapter 6
Summary and Outlook

6.1 Summary

In this study, operation mechanism and intrinsic performance of 2D heterostructured non-volatile memory (NVM) devices were investigated. Due to its ultimately high-quality interfaces and future material compatibility, 2D heterostructured NVM devices have been considered as promising candidates for next-generation electronic devices.

The key achievements of this study are summarized as follows.

1) **New and useful measurement technique called floating gate voltage (V_{FG}) measurement was proposed for unveiling the operation mechanisms behind *I–V* transfer curves.** Since the measurement can be performed just by fabricating the additional electrode onto the floating gate (FG), it is applicable to any 2D materials based-flash memory devices. Capacitive coupling (CC) region with tilted V_{FG} and feedback by tunneling (FT) region with pinned V_{FG} are the keys for understanding the result, V_{FG} trajectory. The effect of quantum capacitance (C_Q) of graphene FG was also discussed. (**Chapter 2**)

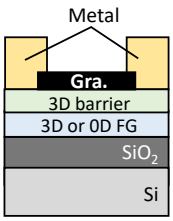
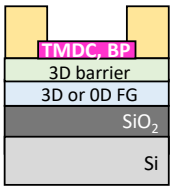
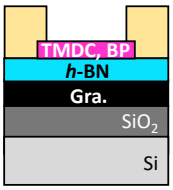
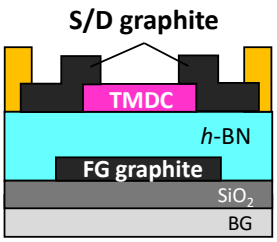
2) **Memory window overestimation by *I–V* round sweep, which has been generally utilized in 2D research field, was revealed based on the V_{FG} trajectory analysis.** When back gate voltage (V_{BG}) is swept from –30 V to positive direction, initially stored holes in the FG are lost before V_{BG} reaches 0 V. In other words, round sweep-based window represents a huge number of carriers that exceeds the capacity of the device. The criterion for the overestimation was also derived. (**Chapter 3**)

3) **Comprehensive operation mechanisms ware revealed based on the V_{FG} trajectory analysis.** Fabricated three devices with different channel materials ($MoTe_2$, WSe_2, MoS_2) showed the inherent V_{FG} trajectories like fingerprint, enabling to access the operation mechanisms behind *I-V* curves. The operation mechanisms have been clarified in terms of three current limiting paths. The metal-2D tunneling path in the device without access region was experimentally proved for the first time. The origin of drain current (I_d) plateau was also clarified by the pinned V_{FG} in FT region. (**Chapter 4**)

4) **Ultra-fast program and erase (P/E) operation of 50 ns was demonstrated, and its possible origin was revealed to be the strong breakdown strength of *h*-BN under fast voltage pulse stress.** In order to ensure the large tunneling current, the device structure without access region was focused on, where the tunneling between source and drain (S/D) electrodes and the FG can be employed. For robust P/E operation, small FG pad is the key. Controlled experiment proved the 2D/2D interface is not the key, and breakdown test for *h*-BN under fast voltage pulse suggested the strong breakdown strength of *h*-BN as compared with SiO_2. It can be the key since large tunneling current can be allowed by strong breakdown strength. Additionally, the 10^5 s retention and 5×10^4 P/E cycles endurance of the MoS_2 device with access region was demonstrated. (**Chapter 5**)

As the conclusion, 2D heterostructured NVM device with graphite contact is proposed as *4th generation* device as shown in **Table 6-1**. Although the 2D/2D interface is not the key of ultra-fast operation, it can still be expected as the key of reliability, since atomically flat interface may be free from the electric field concentration which degrades the breakdown strength.

Table 6-1. The 4th generation device as the conclusion of this study.

				
	1st generation	**2nd generation**	**3rd generation**	**4th generation**
Channel	Graphene	TMDC, BP	TMDC, BP	TMDC
Tunneling barrier	3D Insulator	3D Insulator	*h*-BN	*h*-BN
FG	3D, 0D material	3D, 0D material	Graphene, Graphite, TMDC	Graphite
S/D electrodes	Metal	Metal	Metal	Graphite

6.2 Outlook

Towards the practical applications, future outlook is discussed here. Whereas exfoliated flake based-single cell-level study is a mainstream at present, middle-term and long-term prospects after the establishment of sophisticated 2D process are discussed briefly.

Short-term:

- For reasonably sporting the ultra-fast nature, breakdown of *h*-BN should be investigated comprehensively, especially in the ultra-fast time scale. The breakdown study for *h*-BN in DC time scale[49,137,141] will give the insights. Stress induced leakage current (SILC), which is one of the reliability issue of conventional Si-based flash memory[17], should also be investigated.
- Performances at high temperature (>300 K) should be investigated since the joule heating is unavoidable in practical applications. The V_{FG} measurement will be quite useful at high temperature as well as this study.
- The barrier height engineering is available for 2D heterostructured NVM devices. A plenty of metallic 2D metallic transition metal dichalcogenides (TMDCs) shown in **Fig. 6-1**[142] enables performance optimization via barrier height engineering as schematically shown in **Fig. 6-2**. Although graphite is good candidate from the viewpoint of stability, important tradeoffs such as speed versus retention, and speed versus power consumption should be investigated.

Middle-term:

- The investigation of scalability will be important. For the scaled device, edge of 2D materials become dominant. Although the surface of 2D material has no dangling-bond, the edge has it, suggesting that the undesired performance degradation. This study requires the precise control of the device geometry.

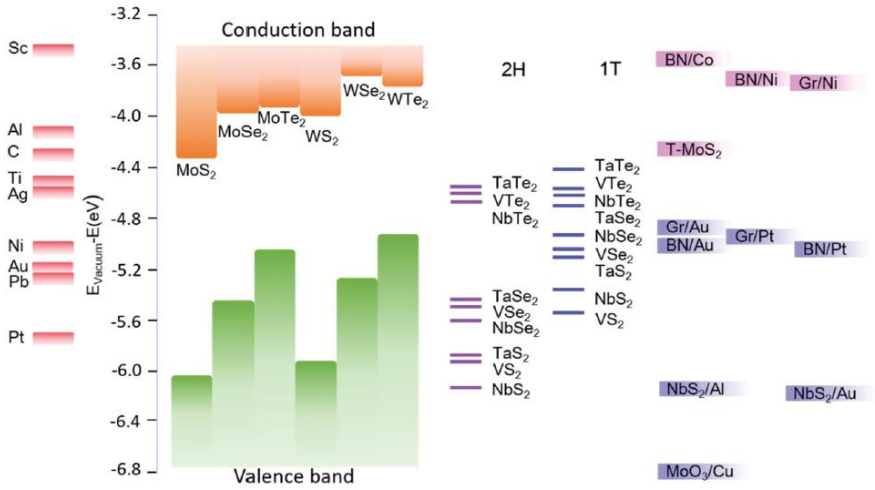


Fig. 6-1 Band alignment of various TMDCs, elemental contact metals and metal/buffer contacts[142].

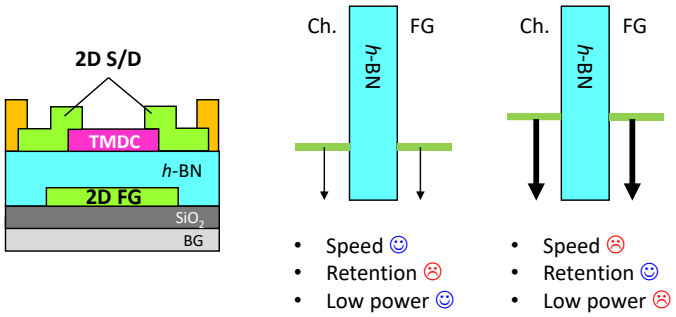


Fig. 6-2 Barrier height engineering for 4th generation device.

Long-term:

- NAND or NOR array configuration of the cells should be demonstrated.
- Statistical information such as V_{th} distribution should be investigated. Probably, 2D heterostructured NVM device will result in the sharper distribution of V_{th} than conventional technologies due to its superior interfaces.
- Since the suppression of cell-to-cell interference issue by 2D FG has been simulated[21], it will be demonstrated experimentally.
- New kinds of integration scheme such as van der Waals integration[143] can be explored.

This study will be the important milestone for above future studies.

Appendix 1
List of Symbols

The list in this appendix summarizes the symbols and their typical units used in this thesis.

NOTE

- The symbols temporally defined for the explanation (e.g., V_{start} and V_{turn} in **chapter 2**) or used only in a quoted equation (e.g., F(def), F(host), and E_{corr} in **Eq. 1.9**) are basically excluded. However, some of them which are commonly used in related studies are included. For example, F for feature size and C_{it} for interface trap capacitance.
- Sometimes, material is specified by superscript, e.g., C_{Q}^{gra} for quantum capacitance of graphene (see **chapter 2**).
- For the FN tunneling current calculation, the unit of energy terms should be the joule (J) instead of electron volt (eV).

Symbol	Description	Unit
C_{BN}	Capacitance of *h*-BN	F
C_{IPD}	Capacitance of interpoly dielectric	F
C_{it}	Interface trap capacitance per unit area	F/cm^2
C_{ox}	Capacitance of SiO_2	F
C_Q	Quantum capacitance	F
C_{tun}	Capacitance of tunneling barrier	F
D_{it}	Interface trap density	$/cm^2 \cdot eV$
E_{BN}	Electric field across *h*-BN	MV/cm
E_{DP}	Energy of Dirac point of graphene	eV
E_F	Fermi level	eV
E_g	Bandgap	eV
E_{ox}	Electric field across SiO_2 or an insulator (type is not specified)	MV/cm
F	Feature size	nm
G^f	Defect formation energy	eV
h	Plank's constant (4.135×10^{-15} eV·s)	eV·s
$\hbar$	Reduced Plank's constant (6.582×10^{-16} eV·s)	eV·s
I_d	Drain current	A
I_{tun}	Tunneling current	A
J_{tun}	Tunneling current density	A/cm^2
k_B	Boltzmann's constant (8.612×10^{-5} eV/K)	eV/K
m_0	Free electron mass (9.1×10^{-31} kg)	kg
m^*	Effective mass	kg
m_e	Electron mass in *h*-BN (**chapter 4**)	kg
m_h	Hole mass in *h*-BN (**chapter 4**)	kg
q	Charge for defect formation (**chapter 1**) Charge of electron (1.6×10^{-19} C) (otherwise)	– C
T	Temperature	K

Symbol	Description	Unit
t_{BD}	Time to breakdown	s
t_{BN}	Thickness of *h*-BN	nm
V_{BG}, V_{bg}	Back gate voltage	V
V_{CG}	Control gate voltage	V
V_{DS}, V_{ds}	Drain-source voltage	V
v_F	Fermi velocity of electron in graphene (10^6 cm/s)	cm/s
V_{FG}	Floating gate voltage / potential of floating gate	V
V_g	Gate voltage (gate type is not specified.)	V
V_{th}	Threshold voltage	V
$V_{tun}^{(+)}$	Positive side of tunnel starting voltage (see **chapter 2**)	V
$V_{tun}^{(-)}$	Negative side of tunnel starting voltage (see **chapter 2**)	V
V_{well}	Well voltage	V
ΔV_{th}	Threshold voltage difference	V
ΔQ	Charge difference which corresponds to ΔV_{th}	C
ε_r	Relative permittivity	—
Φ_B	Barrier height	eV

Appendix 2

Abbreviations and Acronyms

ACCU (-FET)	**Accu**mulation mode (-FET)
ADF	**A**nnular **D**ark **F**ield
AFM	**A**tomic **F**orce **M**icroscope
BG	**B**ack **G**ate
BP	**B**lack **P**hosphorus
CB	**C**onduction **B**and
CBM	**C**onduction **B**and **M**inimum
CBRAM	**C**onductive **B**ridge **R**andom **A**ccess **M**emory
CC (region)	**C**apacitive **C**oupling (region)
CG	**C**ontrol **G**ate
CMOS	**C**omplementally **MOS**
CT	**C**harge **T**rap
CTCI	**C**ell-**t**o-**c**ell **I**nterference
CVD	**C**hemical **V**apor **D**eposition
CVT	**C**hemical **V**apor **T**ransport
DC	**D**irect **C**urrent
DOS	**D**ensity **o**f **S**tates
DRAM	**D**ynamic **R**andom **A**ccess **M**emory
ECC	**E**rror **C**orrection **C**ode

FET	**F**ield **E**ffect **T**ransistor
FG	**F**loating **G**ate
FLP	**F**ermi **L**evel **P**inning
FN	**F**owler-**N**ordheim
FoM	**F**igure **o**f **M**erit
FT (region)	**F**eedback by **T**unneling (region)
FWHM	**F**ull **W**idth at **H**alf **M**aximum
h-BN	**h**exagonal **B**oron **N**itride
HDD	**H**ard **D**isc **D**rive
HRS	**H**igh **R**esistance **S**tate
HV-SPGU	**H**igh **V**oltage **S**emiconductor **P**ulse **G**enerator **U**nit
IC	**I**ntegrated **C**ircuit
ICT	**I**nformation and **C**ommunication **T**echnology
IPD	**I**nter**p**oly **D**ielectric
LOCOS	**Loc**al **O**xidation of **S**ilicon
LRS	**L**ow **R**esistance **S**tate
ME	**M**echanical **E**xfoliation
MIS	**M**etal-**I**nsulator-**S**emiconductor
MLC	**M**ulti-**L**evel **C**ell
MOS	**M**etal-**O**xide-**S**emiconductor
MTJ	**M**agnetic **T**unnel **J**unction
NVM	**N**on-**V**olatile **M**emory

PC	**P**ersonal **C**omputer
PCRAM	**P**hase **C**hange **R**andom **A**ccess **M**emory
PDMS	**P**oly**d**i**m**ethyl**s**iloxane
P/E	**P**rogram and **E**rase
PLC	**P**enta-**L**evel **C**ell
PVD	**P**hysical **V**apor **D**eposition
PVT	**P**hysical **V**apor **T**ransport
QD	**Q**uantum **D**ot
QLC	**Q**uad-**L**evel **C**ell
ReRAM	**Re**sistive **R**andom **A**ccess **M**emory
RSU	**R**emote-sense and **S**witch **U**nit
SA-STI (cell)	**S**elf-**A**ligned **S**hallow **T**rench **I**solation (cell)
SB (-FET)	**S**chottky **B**arrier (-FET)
SCE	**S**hort **C**hannel **E**ffect
S/D	**S**ource and **D**rain
SILC	**S**tress **I**nduced **L**eakage **C**urrent
SMU	**S**ource/**M**easure **U**nit
SS	**S**ubthreshold **S**wing
SSD	**S**olid **S**tate **D**rive
STEM	**S**canning **T**ransmission **E**lectron **M**icroscope
STI	**S**hallow **T**rench **I**solation
STT-MRAM	**S**pin-**T**ransfer-**T**orque **M**agnetoresistive **R**andom **A**ccess **M**emory
SRAM	**S**tatic **R**andom **A**ccess **M**emory

TDDB	**T**ime-**D**ependent **D**ielectric **B**reakdown
TE	**T**hermionic **E**mission
TEM	**T**unneling **E**lectron **M**icroscope
TFET	**T**unnel **F**ield **E**ffect **T**ransistor
TG	**T**op **G**ate
TLC	**T**riple-**L**evel **C**ell
TMDC	**T**ransition **M**etal **D**i**c**halcogenide
VB	**V**alence **B**and
VBM	**V**alence **B**and **M**aximum
vdW	**v**an **d**er **W**aals
WGFMU	**W**aveform **G**enerator/**F**ast **M**easurement **U**nit
0D	**Zero D**imensional
1D	**One D**imensional
2D	**Two D**imensional
3D	**Three D**imensional

Appendix 3

Fowler-Nordheim Formula

The derivation of Fowler-Nordheim (FN) formula is given. One polite derivation can be found in the dissertation by Andreas Gehring (Technische Universität Wien, 2003)[144]. In this section, starting from the simple rectangular barrier case, which is good example to understand what is a transmission coefficient. Although detailed numerical explanation of Wentzel-Kramers-Brillouin (WKB) approximation is omitted, the important conclusion, the transmission coefficient for the arbitral $V(x)$, is given. Finally, FN formula is derived by following the derivation in the reference[144].

Transmission coefficient

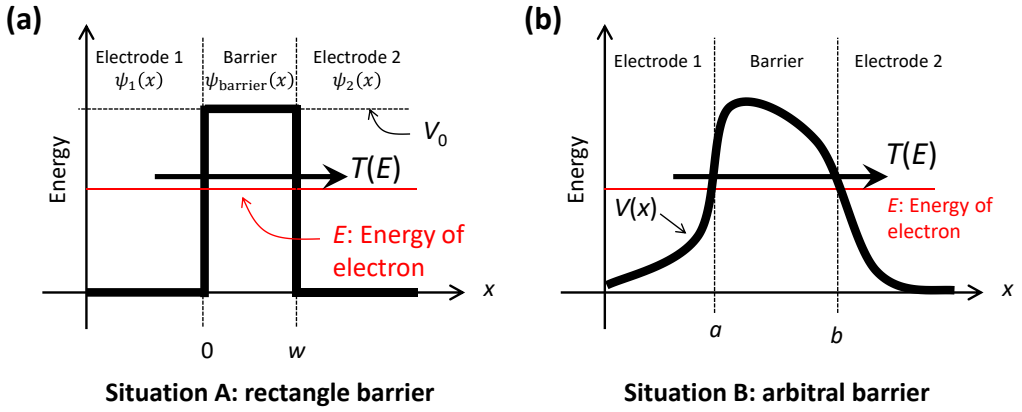


Fig. A3-1 Illustration of two potential barriers.
(a) Simple rectangle barrier and **(b)** arbitral barrier.

Two situations shown in **Fig. A3-1** are considered here. One has a simple rectangle barrier and the other has an arbitrary shape barrier. Electron is tunneled from an electrode one (left) to two (right). Starting with time-independent 1D Schrödinger equation as follows:

$$\left[-\frac{\hbar}{2m}\frac{d^2}{dx^2}+V(x)\right]\psi(x)=E\psi(x), \qquad \text{(A3.1)}$$

where, $\hbar$, m, E are the reduced Plank's constant, mass and energy of electron, respectively, and $V(x)$, $\psi(x)$ are the potential and wavefunction of electron as a function of position, respectively.

For the situation A (**Fig. A3-1(a)**), the rectangle barrier is described as follows:

$$V(\mathrm{x}) = \begin{cases} V_0 & (0 \le x \le w) \\ 0 & (otherwise) \end{cases}, \tag{A3.2}$$

where $V_0 > E$, and w is width of the barrier. The **Eq. A3.1** can be solved in each region as follows:

$$\begin{matrix} \psi_1(x) = A\exp(ikx) + B\exp(-ikx) \\ \psi_{\mathrm{barrier}}(x) = C\exp(-\gamma x) + D\exp(\gamma x) \\ \psi_2(x) = F\exp(ikx) \end{matrix} , \tag{A3.3}$$

where, $k, \gamma \in \mathbb{R}$ are

$$k = \frac{\sqrt{2mE}}{\hbar}, \qquad \gamma = \frac{\sqrt{2m(V_0 - E)}}{\hbar}, \tag{A3.4}$$

respectively. A, B, C, D, and F are constants. For the electrode one, first and second terms represent incident wave and reflected wave, respectively. In the barrier, since γ is a real number, and first term (incident wave) is larger than second term (reflected wave), wave function $\psi_{\mathrm{barrier}}(x)$ represents the exponential decay with position x. In electrode two, there is only transmitted wave. In order to build a bridge from the wave function to the flux, *probability current density J* defined by following equation is used.

$$J = \frac{\hbar}{2mi}\{\psi^* \nabla\psi - (\nabla\psi^*)\psi\}. \tag{A3.5}$$

Probability current density of incident wave J_i and transmitted wave J_t can be calculated as follows,

$$J_\mathrm{i} = \frac{\hbar^2}{2mi}|A|^2, \qquad J_\mathrm{t} = \frac{\hbar^2}{2mi}|F|^2 \tag{A3.6}$$

Therefore, transmission coefficient T can be calculated as $T = J_\mathrm{t}/J_\mathrm{i} = |F|^2/|A|^2$. Here, wavefunctions should satisfy following equations to guarantee the continuity.

$$\psi_1(0) = \psi_{\mathrm{barrier}}(0), \qquad \psi_{\mathrm{barrier}}(w) = \psi_2(w), \tag{A3.7}$$

$$\left.\frac{d\psi_1(x)}{dx}\right|_{x=0} = \left.\frac{d\psi_{\mathrm{barrier}}(x)}{dx}\right|_{x=0}, \qquad \left.\frac{d\psi_{\mathrm{barrier}}(x)}{dx}\right|_{x=w} = \left.\frac{d\psi_2(x)}{dx}\right|_{x=w}. \tag{A3.8}$$

From **Eq. A3.6**–**A3.8**, the transmission coefficient can be derived without any constant, and it is a function of energy as follows:

$$T(E) = \left[1 + \frac{V_0^2}{4E(V_0 - E)} \sinh^2\left(\frac{\sqrt{2m(V_0 - E)}}{\hbar} w\right)\right]^{-1}. \quad \text{(A3.9)}$$

For the situation B (**Fig. A3-1(b)**), potential barrier is described just $V(x)$. Since Schrödinger equation (**Eq. A3.1**) cannot be solved analytically in this situation, WKB approximation is employed. One great explanation for the approximation is given by Prof. Kazuo Muto (Tokyo Institute of Technology, –2016) as his lecture note (2011)[145].

Here, wavefunction $\psi(x)$ is assumed as follows:

$$\psi(x) = G\exp\left(\frac{iS(x)}{\hbar}\right), \quad \text{(A3.10)}$$

where, G is a constant, and $S(x)$ is assumed to be expanded by $\hbar$ as follows:

$$S(x) = S_0(x) + S_1(x)\hbar + S_2(x)\hbar^2 \cdots. \quad \text{(A3.11)}$$

By substituting **Eq. A3.10** and **Eq. A3.11** into **Eq. A3.1**, we can obtain the differential equations for zeroth-order of $\hbar$, first-order of $\hbar$, and so on. In general, the wavefunction is approximated up to the first-order of $\hbar$. This is a brilliant point of WKB approximation, that is, the approximation regards the (reduced) Plank's constant as a key parameter of quantum mechanics. Since $\hbar$ is quite small (1.05×10^{-34} J·s), it is only important in quantum scale. Since the terms above second-order of $\hbar$ are neglected, WKB approximation is also known as *semiclassical* approximation. It should be noted that, WKB approximation is not valid in the vicinity of the position where $V(x) = E$. The position is called *turning point*. Therefore, in the vicinity of turning point, linear approximation of the potential $V(x)$ is used to obtain an exact solution of Schrödinger equation. Finally, similar to the rectangular barrier case, transmission coefficient can be derived as follows:

$$T(E) = \exp\left(-\frac{2}{\hbar}\int_a^b \sqrt{2m(V(x) - E)}dx\right), \quad \text{(A3.12)}$$

where, a and b are the left and right turning points as shown in **Fig. A3-1(b)**. The **Eq. A3.12** is called *Gamov factor*.

FN tunneling formula derivation

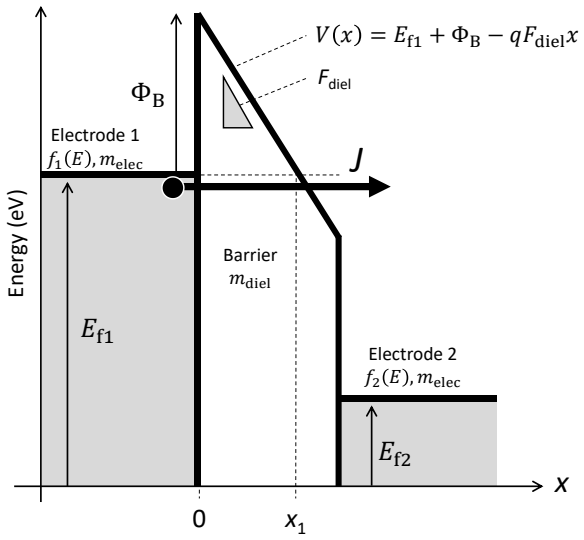


Fig. A3-2 Illustration of triangle barrier and FN tunneling current.

Since FN tunneling is a phenomenon under high electric field, triangle barrier shown in **Fig. A3-2** is considered. The derivation is started from following equations:

$$J_{1\to2} = q\int_0^{+\infty}\int_0^{+\infty}\int_0^{+\infty} g(k_x,k_y,k_z)\cdot f_1(E)T(E)\big(1-f_2(E)\big)\cdot v_x \; dk_x dk_y dk_z, \quad \text{(A3.13)}$$

$$J_{2\to1} = q\int_0^{+\infty}\int_0^{+\infty}\int_0^{+\infty} g(k_x,k_y,k_z)\cdot f_2(E)T(E)\big(1-f_1(E)\big)\cdot v_x \; dk_x dk_y dk_z, \quad \text{(A3.14)}$$

where, $J_{1\to2}$ and $J_{2\to1}$ are the current density from electrode one to two, and from electrode two to one, respectively. The $g(k_x,k_y,k_z)$ is the three-dimensional density of states in the momentum space, k_x, k_y, k_z are the components of wave vector $\boldsymbol{k} = (k_x, k_y, k_z)$, $f_1(E)$ and $f_2(E)$ are the Fermi-Dirac distribution function of electrode one and two, and v_x is a velocity of electron perpendicular to the barrier, respectively. These equations can be easily interpreted by analogy with the Drude model ($J = qnv$, where n is a carrier density). Since these are based on quantum mechanics, the number of tunneled carriers should be calculated by weighting the probability.

As well as the derivation in the reference[144], followings are assumed.

1. Parabolic *E-k* relationship: $E = \frac{\hbar^2 k^2}{2m^*}$ (m^* is effective mass).
2. Conservation of the momentum parallel to the barrier: This means that only *x*-direction (perpendicular to the barrier) is considered.
3. Transmission coefficient only depends on the energy perpendicular to the barrier.
4. Temperature is zero Kelvin. ($T \to 0$ K)
5. Electrode one and two are the same material, that is, there is no work function difference between them.

Then, the net tunneling current density $J = J_{1\to2} - J_{2\to1}$ is,

$$J = \frac{4\pi q m_{\mathrm{elec}}}{h^3}\left[\int_{-\infty}^{E_{\mathrm{f2}}} T(E_x)(E_{\mathrm{f1}} - E_{\mathrm{f2}})dE_x + \int_{E_{\mathrm{f2}}}^{E_{\mathrm{f1}}} T(E_x)(E_{\mathrm{f1}} - E_x)dE_x\right], \qquad \text{(A3.15)}$$

where E_{x}, E_{f1} and E_{f2} are the energy of electron perpendicular to the barrier, Fermi energy of electrode one and two, respectively. Effective mass of the electron in the electrodes are denoted as m_{elec}. Here, first term in **Eq. A3.15** is negligible because the contribution from electrons which has an energy lower than E_{f2} is quite small.

By substituting **Eq. A3.12** to **Eq. A3.15**,

$$J = \frac{4\pi q m_{\mathrm{elec}}}{h^3}\left[\int_{E_{\mathrm{f2}}}^{E_{\mathrm{f1}}} \exp\left(-\frac{2}{\hbar}\int_0^{x_1}\sqrt{2m_{\mathrm{diel}}(V(x) - E_x)}\,dx\right)(E_{\mathrm{f1}} - E_x)dE_x\right], \quad \text{(A3.16)}$$

where, m_{diel} is effective mass of the electron in the dielectric. The definition of the x_1 is illustrated in **Fig. A3–2**. In addition. according to the **Fig. A3–2**,

$$V(x) = E_{\mathrm{f1}} + \Phi_{\mathrm{B}} - qF_{\mathrm{diel}}x, \qquad \text{(A3.17)}$$

where Φ_{B} and F_{diel} are barrier height and electric field across the dielectric, respectively. From **Eq. A3.16** and **Eq. A3.17**, we can derive,

$$J = \frac{4\pi q m_{\mathrm{elec}}}{h^3}\left[\int_{E_{\mathrm{f2}}}^{E_{\mathrm{f1}}} \exp\left(-4\frac{\sqrt{2m_{\mathrm{diel}}}}{3q\hbar F_{\mathrm{diel}}}\{\Phi_B - (E_x - E_{\mathrm{f1}})\}^{\frac{3}{2}}\right)(E_{\mathrm{f1}} - E_x)dE_x\right]. \quad \text{(A3.18)}$$

Here, the term $\{\Phi_B - (E_x - E_{\mathrm{f1}})\}^{\frac{3}{2}}$ is assumed to be expanded as follows,

$$\{\Phi_{\mathrm{B}} - (E_x - E_{\mathrm{f1}})\}^{\frac{3}{2}} \approx \Phi_B^{3/2} - \frac{3}{2}(E_x - E_{\mathrm{f1}})\Phi_{\mathrm{B}}^{1/2}. \quad \text{(A3.19)}$$

This means that $|E_x - E_{\mathrm{f1}}| \ll \Phi_{\mathrm{B}}$, that is, main contributor to the FN tunneling current is the electron with the energy near the E_{f1}.

Noted that, the sign for second term may be mistaken in the reference[144] (Eq. A.11 in p. 118). Therefore, following derivation in the reference has some trivial mistakes while the conclusion is correct. The derivation is corrected as follows.

Substituting **Eq. A3.19** to **Eq. A3.18**,

$$J = \frac{4\pi q m_{\mathrm{elec}}}{h^3}\left[\int_{E_{\mathrm{f2}}}^{E_{\mathrm{f1}}} \exp\left(-4\frac{\sqrt{2m_{\mathrm{diel}}}}{3q\hbar F_{\mathrm{diel}}}\Phi_{\mathrm{B}}^{3/2}\right)\exp\left(\frac{2\sqrt{2m_{\mathrm{diel}}}}{q\hbar F_{\mathrm{diel}}}(E_x - E_{\mathrm{f1}})\Phi_{\mathrm{B}}^{1/2}\right)(E_{\mathrm{f1}} - E_x)dE_x\right]. \quad \text{(A3.20)}$$

Let $\lambda = \frac{2\sqrt{2m_{\mathrm{diel}}}}{q\hbar F_{\mathrm{diel}}}\Phi_{\mathrm{B}}^{1/2}$, $\epsilon = E_x - E_{\mathrm{f1}}$, and $a = -4\frac{\sqrt{2m_{\mathrm{diel}}}}{3q\hbar F_{\mathrm{diel}}}\Phi_{\mathrm{B}}^{3/2}$, then

$$J = \frac{4\pi q m_{\mathrm{elec}}}{h^3}\exp(a)\int_{E_{\mathrm{f2}}-E_{\mathrm{f1}}}^{0} \exp(\lambda\epsilon)\,(-\epsilon)d\epsilon\,. \quad \text{(A3.21)}$$

Since $\int \epsilon\exp(\lambda\epsilon)\,d\epsilon = \frac{1}{\lambda^2}\exp(\lambda\epsilon)\,(\lambda\epsilon - 1)$,

$$\begin{aligned}\int_{E_{\mathrm{f2}}-E_{\mathrm{f1}}}^{0} \exp(\lambda\epsilon)\,(-\epsilon)d\epsilon &= -\left[\frac{1}{\lambda^2}\exp(\lambda\epsilon)\,(\lambda\epsilon - 1)\right]_{E_{\mathrm{f2}}-E_{\mathrm{f1}}}^{0} \\ &= -\frac{1}{\lambda^2}\{\exp(0)\,(0-1) + \exp\big(\lambda(E_{\mathrm{f2}} - E_{\mathrm{f1}})\big)\,(\lambda(E_{\mathrm{f2}} - E_{\mathrm{f1}}) - 1)\} \\ &= \frac{1}{\lambda^2}\{1 - \exp\big(\lambda(E_{\mathrm{f2}} - E_{\mathrm{f1}})\big)\,(\lambda(E_{\mathrm{f2}} - E_{\mathrm{f1}}) - 1)\}. \quad \text{(A3.22)}\end{aligned}$$

Here, $E_{\mathrm{f1}} \gg E_{\mathrm{f2}}$ is assumed, then the second term of **Eq. A3.22** to be zero.

Finally, we can obtain

$$\begin{aligned}J &= \frac{4\pi q m_{\mathrm{elec}}}{h^3}\exp(a)\frac{1}{\lambda^2} \\ &= \frac{q^3 m_{\mathrm{elec}}}{8\pi m_{\mathrm{diel}} h\Phi_{\mathrm{B}}}\exp\left(-\frac{4\sqrt{2m_{\mathrm{diel}}}}{3q\hbar F_{\mathrm{diel}}}\Phi_{\mathrm{B}}^{3/2}\right). \quad \text{(A3.23)}\end{aligned}$$

This is the Fowler-Nordheim formula.

References

(1) David, R.; John, G.; John, R. *Data Age 2025, The Digitization of the World From Edge to Core, An IDC White Papaer*; 2018.

(2) Iwasaki, S.; Ouchi, K. Co-Cr Recording Films with Perpendicular Magnetic Anisotropy. *IEEE Trans. Magn.* **1978**, *14* (5), 849–851.

(3) Masuoka, F.; Asano, M.; Iwahashi, H.; Komuro, T.; Tanaka, S. A New Flash E^2PROM Cell Using Triple Polysilicon Technology. In *1984 International Electron Devices Meeting*; 1984; pp 464–467.

(4) Meena, J. S.; Sze, S. M.; Chand, U.; Tseng, T.-Y. Overview of Emerging Nonvolatile Memory Technologies. *Nanoscale Res. Lett.* **2014**, *9* (1), 526.

(5) Aritome, S. NAND Flash Memory Revolution. In *2016 IEEE 8th International Memory Workshop (IMW)*; 2016; pp 1–4.

(6) Sze, S. M.; Ng, K. K. *Physics of Semiconductor Devices, 3rd Ed.*; John WILEY & Sons: New Jersey, 2007.

(7) Raoux, S.; Burr, G. W.; Breitwisch, M. J.; Rettner, C. T.; Chen, Y.-C.; Shelby, R. M.; Salinga, M.; Krebs, D.; Chen, S.-H.; Lung, H.-L.; Lam, C. H. Phase-change Random Access Memory: A Scalable Technology. *IBM J. Res. Dev.* **2008**, *52* (4.5), 465–479.

(8) Zahoor, F.; Azni Zulkifli, T. Z.; Khanday, F. A. Resistive Random Access Memory (RRAM): An Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications. *Nanoscale Res. Lett.* **2020**, *15* (1), 90.

(9) Hosomi, M.; Yamagishi, H.; Yamamoto, T.; Bessho, K.; Higo, Y.; Yamane, K.; Yamada, H.; Shoji, M.; Hachino, H.; Fukumoto, C.; Nagao, H.; Kano, H. A Novel Nonvolatile Memory with Spin Torque Transfer Magnetization Switching: Spin-RAM. In *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest.*; 2005; pp 459–462.

(10) Huai, Y. Spin-Transfer Torque MRAM (STT-MRAM): Challenges and Prospects. *AAPPS Bull.* **2008**, *18*.

(11) Arimoto, Y.; Ishiwara, H. Current Status of Ferroelectric Random-Access Memory. *MRS Bull.* **2004**, *29* (11), 823–828.

(12) Bertolazzi, S.; Bondavalli, P.; Roche, S.; San, T.; Choi, S.-Y.; Colombo, L.; Bonaccorso, F.; Samorì, P. Nonvolatile Memories Based on Graphene and Related 2D Materials. *Adv. Mater.* **2019**, *31* (10), 1806663.

(13) Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115* (3), 211–252.

(14) Burr, G. W.; Shelby, R. M.; Sebastian, A.; Kim, S.; Kim, S.; Sidler, S.; Virwani, K.; Ishii, M.; Narayanan, P.; Fumarola, A.; Sanches, L. L.; Boybat, I.; Le Gallo, M.; Moon, K.; Woo, J.; Hwang, H.; Leblebici, Y. Neuromorphic Computing Using Non-volatile Memory. *Adv. Phys. X* **2017**, *2* (1), 89–124.

(15) Tanaka, H.; Kido, M.; Yahashi, K.; Oomura, M.; Katsumata, R.; Kito, M.; Fukuzumi, Y.; Sato, M.; Nagata, Y.; Matsuoka, Y.; Iwata, Y.; Aochi, H.; Nitayama, A. Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory. In *2007 IEEE Symposium on VLSI Technology*; 2007; pp 14–15.

(16) Lee, J.; Jang, J.; Lim, J.; Shin, Y. G.; Lee, K.; Jung, E. A New Ruler on the Storage Market: 3D-NAND Flash for High-density Memory and Its Technology Evolutions and Challenges on the Future. In *2016 IEEE International Electron Devices Meeting (IEDM)*; 2016; pp 11.2.1–11.2.4.

(17) Aritome, S. *NAND Flash Memory Technologies*; John WILEY & Sons: New Jersey, 2016.

(18) Jung, T.-S.; Choi, Y.-J.; Suh, K.-D.; Suh, B.-H.; Kim, J.-K.; Lim, Y.-H.; Koh, Y.-N.; Park, J.-W.; Lee, K.-J.; Park, J.-H.; Park, K.-T.; Kim, J.-R.; Yi, J.-H.; Lim, H.-K. A 117-mm^2 3.3-V Only 128-Mb Multilevel NAND Flash Memory for Mass

Storage Applications. *IEEE J. Solid-State Circuits* **1996**, *31* (11), 1575–1583.

(19) Aritome, S.; Riichiro, S.; Koji, S.; Fujio, M. Data Retention Characteristics of Flash Memory Cells after Write and Erase Cycling. *IEICE Trans. Electron.* **1994**, *E77-C* (8), 1287–1295.

(20) Lee, J.-D.; Choi, J.-H.; Park, D.; Kim, K. Data Retention Characteristics of Sub-100 nm NAND Flash Memory Cells. *IEEE Electron Device Lett.* **2003**, *24* (12), 748–750.

(21) Cao, W.; Kang, J.; Bertolazzi, S.; Kis, A.; Banerjee, K. Can 2D-nanocrystals Extend the Lifetime of Floating-gate Transistor Based Nonvolatile Memory? *IEEE Trans. Electron Devices* **2014**, *61* (10), 3456–3464.

(22) Abraham, M. *NAND Flash Architecture and Specification Trends*; Santa Clala, CA, 2012.

(23) Fukuzumi, Y.; Katsumata, R.; Kito, M.; Kido, M.; Sato, M.; Tanaka, H.; Nagata, Y.; Matsuoka, Y.; Iwata, Y.; Aochi, H.; Nitayama, A. Optimal Integration and Characteristics of Vertical Array Devices for Ultra-high Density, Bit-cost Scalable Flash Memory. In *2007 IEEE International Electron Devices Meeting*; 2007; pp 449–452.

(24) Kim, H.; Ahn, S.; Shin, Y. G.; Lee, K.; Jung, E. Evolution of NAND Flash Memory: From 2D to 3D as a Storage Market Leader. In *2017 IEEE International Memory Workshop (IMW)*; 2017; pp 1–4.

(25) Choi, E.; Park, S. Device Considerations for High Density and Highly Reliable 3D NAND Flash Cell in near Future. In *2012 International Electron Devices Meeting*; 2012; pp 9.4.1–9.4.4.

(26) Whang, S.; Lee, K.; Shin, D.; Kim, B.; Kim, M.; Bin, J.; Han, J.; Kim, S.; Lee, B.; Jung, Y.; Cho, S.; Shin, C.; Yoo, H.; Choi, S.; Hong, K.; Aritome, S.; Park, S.; Hong, S. Novel 3-dimensional Dual Control-gate with Surrounding Floating-gate (DC-SF) NAND Flash Cell for 1Tb File Storage Application. In *2010 International Electron Devices Meeting*; 2010; pp 29.7.1–29.7.4.

(27) Compagnoni, C. M.; Goda, A.; Spinelli, A. S.; Feeley, P.; Lacaita, A. L.; Visconti, A. Reviewing the Evolution of the NAND Flash Technology. *Proc. IEEE* **2017**, *105* (9), 1609–1633.

(28) Compagnoni, C. M.; Spinelli, A. S. Reliability of NAND Flash Arrays: A Review of What the 2-D–to–3-D Transition Meant. *IEEE Trans. Electron Devices* **2019**, *66* (11), 4504–4516.

(29) Park, J.-W.; Kim, D.; Ok, S.; Park, J.; Kwon, T.; Lee, H.; Lim, S.; Jung, S.-Y.; Choi, H.; Kang, T.; Park, G.; Yang, C.-W.; Choi, J.-G.; Ko, G.; Shin, J.; Yang, I.; Nam, J.; Sohn, H.; Hong, S.-I.; Jeong, Y.; Choi, S.-W.; Choi, C.; Shin, H.-S.; Lim, J.; Youn, D.; Nam, S.; Lee, J.; Ahn, M.; Lee, H.; Lee, S.; Park, J.; Gwon, K.; Jeong, W.; Choi, J.; Kim, J.; Jin, K.-W. A 176-stacked 512Gb 3b/Cell 3D-NAND Flash with 10.8Gb/mm^2 Density with a Peripheral Circuit Under Cell Array Architecture. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)*; 2021; Vol. 64, pp 422–423.

(30) Maconi, A.; Arreghini, A.; Monzio Compagnoni, C.; Van den bosch, G.; Spinelli, A. S.; Van Houdt, J.; Lacaita, A. L. Comprehensive Investigation of the Impact of Lateral Charge Migration on Retention Performance of Planar and 3D SONOS Devices. *Solid. State. Electron.* **2012**, *74*, 64–70.

(31) Tsai, W.-J.; Lin, W. L.; Cheng, C. C.; Ku, S. H.; Chou, Y. L.; Liu, L.; Hwang, S. W.; Lu, T. C.; Chen, K. C.; Wang, T.; Lu, C.-Y. Polycrystalline-silicon Channel Trap Induced Transient Read Instability in a 3D NAND Flash Cell String. In *2016 IEEE International Electron Devices Meeting (IEDM)*; 2016; pp 11.3.1–11.3.4.

(32) Kang, H.; Jeong, M.; Joe, S.; Seo, J.; Park, S.; Jin, S. H.; Park, B.; Lee, J. Effect of Traps on Transient Bit-line Current Behavior in Word-line Stacked Nand Flash Memory with Poly-Si Body. In *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*; 2014; pp 1–2.

(33) Novoselov, K. S.; Geim, A. K.; Morozov, S. V; Jiang, D.; Zhang, Y.; Dubonos, S. V; Grigorieva, I. V; Firsov, A. A. Electric Field Effect in Atomically Thin Carbon

Films. *Science (80-.).* **2004**, *306* (5696), 666 LP–669.

(34) Ando, T. Theory of Electronic States and Transport in Carbon Nanotubes. *J. Phys. Soc. Japan* **2005**, *74* (3), 777–817.

(35) Castro Neto, A. H.; Guinea, F.; Peres, N. M. R.; Novoselov, K. S.; Geim, A. K. The Electronic Properties of Graphene. *Rev. Mod. Phys.* **2009**, *81* (1), 109–162.

(36) Nagashio, K.; Nishimura, T.; Toriumi, A. Estimation of Residual Carrier Density near the Dirac Point in Graphene through Quantum Capacitance Measurement. *Appl. Phys. Lett.* **2013**, *102* (17), 173507.

(37) Xia, J.; Chen, F.; Li, J.; Tao, N. Measurement of the Quantum Capacitance of Graphene. *Nat. Nanotechnol.* **2009**, *4* (8), 505–509.

(38) Geim, A. K.; Novoselov, K. S. The Rise of Graphene. *Nat. Mater.* **2007**, *6* (3), 183–191.

(39) Hong, J.; Lee, S.; Lee, S.; Han, H.; Mahata, C.; Yeon, H.-W.; Koo, B.; Kim, S.-I.; Nam, T.; Byun, K.; Min, B.-W.; Kim, Y.-W.; Kim, H.; Joo, Y.-C.; Lee, T. Graphene as an Atomically Thin Barrier to Cu Diffusion into Si. *Nanoscale* **2014**, *6* (13), 7503–7511.

(40) Liu, Y.; Wu, H.; Cheng, H.-C.; Yang, S.; Zhu, E.; He, Q.; Ding, M.; Li, D.; Guo, J.; Weiss, N. O.; Huang, Y.; Duan, X. Toward Barrier Free Contact to Molybdenum Disulfide Using Graphene Electrodes. *Nano Lett.* **2015**, *15* (5), 3030–3034.

(41) Sup Choi, M.; Lee, G.-H.; Yu, Y.-J.; Lee, D.-Y.; Hwan Lee, S.; Kim, P.; Hone, J.; Yoo, W. J. Controlled Charge Trapping by Molybdenum Disulphide and Graphene in Ultrathin Heterostructured Memory Devices. *Nat. Commun.* **2013**, *4*, 1624.

(42) Qiu, D.; Lee, D. U.; Lee, K. S.; Pak, S. W.; Kim, E. K. Toward Negligible Charge Loss in Charge Injection Memories Based on Vertically Integrated 2D Heterostructures. *Nano Res.* **2016**, *9* (8), 2319–2326.

(43) Vu, Q. A.; Shin, Y. S.; Kim, Y. R.; Nguyen, V. L.; Kang, W. T.; Kim, H.; Luong, D. H.; Lee, I. M.; Lee, K.; Ko, D.-S.; Heo, J.; Park, S.; Lee, Y. H.; Yu, W. J. Two-terminal Floating-gate Memory with van der Waals Heterostructures for Ultrahigh

on/off Ratio. *Nat. Commun.* **2016**, *7*, 12725.

(44) Sasaki, T.; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Material and Device Structure Designs for 2D Memory Devices Based on the Floating Gate Voltage Trajectory. *ACS Nano* **2021**, *15* (4), 6658–6668.

(45) Watanabe, K.; Taniguchi, T.; Kanda, H. Direct-bandgap Properties and Evidence for Ultraviolet Lasing of Hexagonal Boron Nitride Single Crystal. *Nat. Mater.* **2004**, *3* (6), 404–409.

(46) Li, L. H.; Cervenka, J.; Watanabe, K.; Taniguchi, T.; Chen, Y. Strong Oxidation Resistance of Atomically Thin Boron Nitride Nanosheets. *ACS Nano* **2014**, *8* (2), 1457–1462.

(47) Decker, R.; Wang, Y.; Brar, V. W.; Regan, W.; Tsai, H.-Z.; Wu, Q.; Gannett, W.; Zettl, A.; Crommie, M. F. Local Electronic Properties of Graphene on a BN Substrate via Scanning Tunneling Microscopy. *Nano Lett.* **2011**, *11* (6), 2291–2295.

(48) Vu, Q. A.; Fan, S.; Lee, S. H.; Joo, M.-K.; Yu, W. J.; Lee, Y. H. Near-zero Hysteresis and Near-ideal Subthreshold Swing in h-BN Encapsulated Single-layer MoS_2 Field-Effect Transistors. *2D Mater.* **2018**, *5* (3), 31001.

(49) Hattori, Y.; Taniguchi, T.; Watanabe, K.; Nagashio, K. Anisotropic Dielectric Breakdown Strength of Single Crystal Hexagonal Boron Nitride. *ACS Appl. Mater. Interfaces* **2016**, *8* (41), 27877–27884.

(50) Jung, S.; Myoung, N.; Park, J.; Jeong, T. Y.; Kim, H.; Watanabe, K.; Taniguchi, T.; Ha, D. H.; Hwang, C.; Park, H. C. Direct Probing of the Electronic Structures of Single-layer and Bilayer Graphene with a Hexagonal Boron Nitride Tunneling Barrier. *Nano Lett.* **2017**, *17* (1), 206–213.

(51) Kolobov, A.; Tominaga, J. *Two-dimensional Transition-metal Dichalcogenides, Springer Series in Materials Science, Vol. 239*; Springer International Publishing: Switzerland, 2016.

(52) Chaves, A.; Azadani, J. G.; Alsalman, H.; da Costa, D. R.; Frisenda, R.; Chaves, A. J.; Song, S. H.; Kim, Y. D.; He, D.; Zhou, J.; Castellanos-Gomez, A.; Peeters, F.

M.; Liu, Z.; Hinkle, C. L.; Oh, S. H.; Ye, P. D.; Koester, S. J.; Lee, Y. H.; Avouris, P.; Wang, X.; Low, T. Bandgap Engineering of Two-dimensional Semiconductor Materials. *npj 2D Mater. Appl.* **2020**, *4*, 29.

(53) Roldán, R.; Silva-Guillén, J. A.; López-Sancho, M. P.; Guinea, F.; Cappelluti, E.; Ordejón, P. Electronic Properties of Single-layer and Multilayer Transition Metal Dichalcogenides MX_2 (M = Mo, W and X = S, Se). *Ann. Phys.* **2014**, *526* (9–10), 347–357.

(54) Guo, Y.; Robertson, J. Band Engineering in Transition Metal Dichalcogenides: Stacked versus Lateral Heterostructures. *Appl. Phys. Lett.* **2016**, *108*, 233104.

(55) Ueno, K. Introduction to the Growth of Bulk Single Crystals of Two-dimensional Transition-metal Dichalcogenides. *J. Phys. Soc. Japan* **2015**, *84*, 121015.

(56) Li, T.; Galli, G. Electronic Properties of MoS_2 Nanoparticles. *J. Phys. Chem. C* **2007**, *111* (44), 16192–16196.

(57) Radisavljevic, B.; Radenovic, A.; Brivio, J.; Giacometti, V.; Kis, A. Single-layer MoS_2 Transistors. *Nat. Nanotechnol.* **2011**, *6* (3), 147–150.

(58) Desai, S. B.; Madhvapathy, S. R.; Sachid, A. B.; Llinas, J. P.; Wang, Q.; Ahn, G. H.; Pitner, G.; Kim, M. J.; Bokor, J.; Hu, C.; Wong, H.-S. P.; Javey, A. MoS_2 Transistors with 1-nanometer Gate Lengths. *Science (80-.).* **2016**, *354* (6308), 99 LP–102.

(59) Das, S.; Chen, H. Y.; Penumatcha, A. V.; Appenzeller, J. High Performance Multilayer MoS_2 Transistors with Scandium Contacts. *Nano Lett.* **2013**, *13*, 100–105.

(60) Das, S.; Appenzeller, J. WSe_2 Field Effect Transistors with Enhanced Ambipolar Characteristics. *Appl. Phys. Lett.* **2013**, *103*, 103501.

(61) Nakaharai, S.; Yamamoto, M.; Ueno, K.; Tsukagoshi, K. Carrier Polarity Control in α-$MoTe_2$ Schottky Junctions Based on Weak Fermi-Level Pinning. *ACS Appl. Mater. Interfaces* **2016**, *8* (23), 14732–14739.

(62) Fang, N.; Nagashio, K. Accumulation-mode Two-dimensional Field-Effect

Transistor: Operation Mechanism and Thickness Scaling Rule. *ACS Appl. Mater. Interfaces* **2018**, *10* (38), 32355–32364.

(63) Colinge, J. P.; Flandre, D.; Van de Wiele, F. Subthreshold Slope of Long-channel, Accumulation-mode p-channel SOI MOSFETs. *Solid. State. Electron.* **1994**, *37* (2), 289–294.

(64) Joachim, H.-O.; Yamaguchi, Y.; Inoue, Y.; Nishimura, T.; Tsubouchi, N. Analytical Modeling of Short-channel Behavior of Accumulation-mode Transistors on Silicon-on-Insulator Substrate. *Jpn. J. Appl. Phys.* **1994**, *33* (Part 1, No. 1B), 558–562.

(65) Deal, B. E.; Grove, A. S. General Relationship for the Thermal Oxidation of Silicon. *J. Appl. Phys.* **1965**, *36* (12), 3770–3778.

(66) Kageshima, H.; Shiraishi, K. First-principles Study of Oxide Growth on Si(100) Surfaces and at SiO_2/Si(100) Interfaces. *Phys. Rev. Lett.* **1998**, *81* (26), 5936–5939.

(67) Shin, B. G.; Han, G. H.; Yun, S. J.; Oh, H. M.; Bae, J. J.; Song, Y. J.; Park, C.-Y.; Lee, Y. H. Indirect Bandgap Puddles in Monolayer MoS_2 by Substrate-induced Local Strain. *Adv. Mater.* **2016**, *28* (42), 9378–9384.

(68) Geim, A. K.; Grigorieva, I. V. Van der Waals Heterostructures. *Nature* **2013**, *499* (7459), 419–425.

(69) Ma, X.; Liu, Q.; Xu, D.; Zhu, Y.; Kim, S.; Cui, Y.; Zhong, L.; Liu, M. Capillary-force-assisted Clean-stamp Transfer of Two-dimensional Materials. *Nano Lett.* **2017**, *17* (11), 6961–6967.

(70) Uwanno, T.; Hattori, Y.; Taniguchi, T.; Watanabe, K.; Nagashio, K. Fully Dry PMMA Transfer of Graphene on h-BN Using a Heating/Cooling System. *2D Mater.* **2015**, *2*, 041002.

(71) Toyoda, S.; Uwanno, T.; Taniguchi, T.; Watanabe, K.; Nagashio, K. Pinpoint Pick-up and Bubble-free Assembly of 2D Materials Using PDMS/PMMA Polymers with Lens Shapes. *Appl. Phys. Express* **2019**, *12*, 055008.

(72) Masubuchi, S.; Morimoto, M.; Morikawa, S.; Onodera, M.; Asakawa, Y.; Watanabe, K.; Taniguchi, T.; Machida, T. Autonomous Robotic Searching and Assembly of

Two-dimensional Crystals to Build van der Waals Superlattices. *Nat. Commun.* **2018**, *9* (1), 1413.

(73) Rhodes, D.; Chae, S. H.; Ribeiro-Palau, R.; Hone, J. Disorder in van der Waals Heterostructures of 2D Materials. *Nat. Mater.* **2019**, *18* (6), 541–549.

(74) Uwanno, T.; Taniguchi, T.; Watanabe, K.; Nagashio, K. Electrically Inert h-BN/Bilayer Graphene Interface in All-two-dimensional Heterostructure Field Effect Transistors. *ACS Appl. Mater. Interfaces* **2018**, *10* (34), 28780–28788.

(75) Nakamura, K.; Nagamura, N.; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nagashio, K. All 2D Heterostructure Tunnel Field-Effect Transistors: Impact of Band Alignment and Heterointerface Quality. *ACS Appl. Mater. Interfaces* **2020**, *12* (46), 51598–51606.

(76) He, F.; Zhou, Y.; Ye, Z.; Cho, S.-H.; Jeong, J.; Meng, X.; Wang, Y. Moiré Patterns in 2D Materials: A Review. *ACS Nano* **2021**, *15* (4), 5944–5958.

(77) Cao, Y.; Fatemi, V.; Fang, S.; Watanabe, K.; Taniguchi, T.; Kaxiras, E.; Jarillo-Herrero, P. Unconventional Superconductivity in Magic-angle Graphene Superlattices. *Nature* **2018**, *556* (7699), 43–50.

(78) Komsa, H.-P.; Krasheninnikov, A. V. Native Defects in Bulk and Monolayer MoS_2 from First Principles. *Phys. Rev. B* **2015**, *91* (12), 125304.

(79) Hong, J.; Hu, Z.; Probert, M.; Li, K.; Lv, D.; Yang, X.; Gu, L.; Mao, N.; Feng, Q.; Xie, L.; Zhang, J.; Wu, D.; Zhang, Z.; Jin, C.; Ji, W.; Zhang, X.; Yuan, J.; Zhang, Z. Exploring Atomic Defects in Molybdenum Disulphide Monolayers. *Nat. Commun.* **2015**, *6* (1), 6293.

(80) Vancsó, P.; Magda, G. Z.; Pető, J.; Noh, J.-Y.; Kim, Y.-S.; Hwang, C.; Biró, L. P.; Tapasztó, L. The Intrinsic Defect Structure of Exfoliated MoS_2 Single Layers Revealed by Scanning Tunneling Microscopy. *Sci. Rep.* **2016**, *6* (1), 29726.

(81) Zhou, W.; Zou, X.; Najmaei, S.; Liu, Z.; Shi, Y.; Kong, J.; Lou, J.; Ajayan, P. M.; Yakobson, B. I.; Idrobo, J.-C. Intrinsic Structural Defects in Monolayer Molybdenum Disulfide. *Nano Lett.* **2013**, *13* (6), 2615–2622.

(82) Nagashio, K. Understanding Interface Properties in 2D Heterostructure FETs. *Semicond. Sci. Technol.* **2020**, *35* (10), 103003.

(83) Wong, D.; Velasco, J.; Ju, L.; Lee, J.; Kahn, S.; Tsai, H.-Z.; Germany, C.; Taniguchi, T.; Watanabe, K.; Zettl, A.; Wang, F.; Crommie, M. F. Characterization and Manipulation of Individual Defects in Insulating Hexagonal Boron Nitride Using Scanning Tunnelling Microscopy. *Nat. Nanotechnol.* **2015**, *10* (11), 949–953.

(84) Akinwande, D.; Huyghebaert, C.; Wang, C.-H.; Serna, M. I.; Goossens, S.; Li, L.-J.; Wong, H.-S. P.; Koppens, F. H. L. Graphene and Two-dimensional Materials for Silicon Technology. *Nature* **2019**, *573* (7775), 507–518.

(85) International Roadmap for Devices and Systems, 2020 Edition https://irds.ieee.org/editions/2020.

(86) Pang, C.-S.; Wu, P.; Appenzeller, J.; Chen, Z. Sub-1nm EOT WS_2-FET with I_{DS} > 600μA/μm at V_{DS}=1V and SS < 70mV/dec at L_G=40nm. In *2020 IEEE International Electron Devices Meeting (IEDM)*; 2020; pp 3.4.1–3.4.4.

(87) Smets, Q.; Arutchelvan, G.; Jussot, J.; Verreck, D.; Asselberghs, I.; Mehta, A. N.; Gaur, A.; Lin, D.; Kazzi, S. E.; Groven, B.; Caymax, M.; Radu, I. Ultra-scaled MOCVD MoS_2 MOSFETs with 42nm Contact Pitch and 250μA/μm Drain Current. In *2019 IEEE International Electron Devices Meeting (IEDM)*; 2019; pp 23.2.1–23.2.4.

(88) Chau, R. Process and Packaging Innovations for Moore's Law Continuation and Beyond. In *2019 IEEE International Electron Devices Meeting (IEDM)*; 2019; pp 1.1.1–1.1.6.

(89) Samavedam, S. B.; Ryckaert, J.; Beyne, E.; Ronse, K.; Horiguchi, N.; Tokei, Z.; Radu, I.; Bardon, M. G.; Na, M. H.; Spessot, A.; Biesemans, S. Future Logic Scaling: Towards Atomic Channels and Deconstructed Chips. In *2020 IEEE International Electron Devices Meeting (IEDM)*; 2020; pp 1.1.1–1.1.10.

(90) Zhan, N.; Olmedo, M.; Wang, G.; Liu, J. Graphene Based Nickel Nanocrystal Flash Memory. *Appl. Phys. Lett.* **2011**, *99*, 113112.

(91) Imam, S. A.; Deshpande, T.; Guermoune, A.; Siaj, M.; Szkopek, T. Charge Transfer Hysteresis in Graphene Dual-dielectric Memory Cell Structures. *Appl. Phys. Lett.* **2011**, *99* (8), 082109.

(92) Hong, A. J.; Song, E. B.; Yu, H. S.; Allen, M. J.; Kim, J.; Fowler, J. D.; Wassei, J. K.; Park, Y.; Wang, Y.; Zou, J.; Kaner, R. B.; Weiller, B. H.; Wang, K. L. Graphene Flash Memory. *ACS Nano* **2011**, *5* (10), 7812–7817.

(93) Tian, H.; Deng, B.; Chin, M. L.; Yan, X.; Jiang, H.; Han, S.-J.; Sun, V.; Xia, Q.; Dubey, M.; Xia, F.; Wang, H. A Dynamically Reconfigurable Ambipolar Black Phosphorus Memory Device. *ACS Nano* **2016**, *10* (11), 10428–10435.

(94) Wang, J.; Zou, X.; Xiao, X.; Xu, L.; Wang, C.; Jiang, C.; Ho, J. C.; Wang, T.; Li, J.; Liao, L. Floating Gate Memory-based Monolayer MoS_2 Transistor with Metal Nanocrystals Embedded in the Gate Dielectrics. *Small* **2015**, *11* (2), 208–213.

(95) Hou, X.; Yan, X.; Liu, C.; Ding, S.; Zhang, D. W.; Zhou, P. Operation Mode Switchable Charge-trap Memory Based on Few-layer MoS_2. *Semicond. Sci. Technol.* **2018**, *33*, 034001.

(96) Hou, X.; Zhang, H.; Liu, C.; Ding, S.; Bao, W.; Zhang, D. W.; Zhou, P. Charge-trap Memory Based on Hybrid 0D Quantum Dot–2D WSe_2 Structure. *Small* **2018**, *14*, 1800319.

(97) Bertolazzi, S.; Krasnozhon, D.; Kis, A. Nonvolatile Memory Cells Based on MoS_2/Graphene Heterostructures. *ACS Nano* **2013**, *7* (4), 3246–3252.

(98) Wu, E.; Xie, Y.; Wang, S.; Wu, C.; Zhang, D.; Hu, X.; Liu, J. Tunable and Nonvolatile Multibit Data Storage Memory Based on $MoTe_2$/Boron Nitride/Graphene Heterostructures through Contact Engineering. *Nanotechnology* **2020**, *31* (48), 485205.

(99) Cheng, R.; Wang, F.; Yin, L.; Xu, K.; Ahmed Shifa, T.; Wen, Y.; Zhan, X.; Li, J.; Jiang, C.; Wang, Z.; He, J. Multifunctional Tunneling Devices Based on Graphene/h-BN/$MoSe_2$ van der Waals Heterostructures. *Appl. Phys. Lett.* **2017**, *110*, 173507.

(100) Yi, S.-G.; Park, M. U.; Kim, S. H.; Lee, C. J.; Kwon, J.; Lee, G.-H.; Yoo, K.-H. Artificial Synaptic Emulators Based on MoS_2 Flash Memory Devices with Double Floating Gates. *ACS Appl. Mater. Interfaces* **2018**, *10* (37), 31480–31487.

(101) He, C.; Tang, J.; Shang, D.-S.; Tang, J.; Xi, Y.; Wang, S.; Li, N.; Zhang, Q.; Lu, J.-K.; Wei, Z.; Wang, Q.; Shen, C.; Li, J.; Shen, S.; Shen, J.; Yang, R.; Shi, D.; Wu, H.; Wang, S.; Zhang, G. Artificial Synapse Based on van der Waals Heterostructures with Tunable Synaptic Functions for Neuromorphic Computing. *ACS Appl. Mater. Interfaces* **2020**, *12* (10), 11945–11954.

(102) Chen, Y.; Yu, J.; Zhuge, F.; He, Y.; Zhang, Q.; Yu, S.; Liu, K.; Li, L.; Ma, Y.; Zhai, T. An Asymmetric Hot Carrier Tunneling van der Waals Heterostructure for Multibit Optoelectronic Memory. *Mater. Horizons* **2020**, *7* (5), 1331–1340.

(103) Tran, M. D.; Kim, H.; Kim, J. S.; Doan, M. H.; Chau, T. K.; Vu, Q. A.; Kim, J.-H.; Lee, Y. H. Two-terminal Multibit Optical Memory via van der Waals Heterostructure. *Adv. Mater.* **2019**, *31*, 1807075.

(104) Li, D.; Wang, X.; Zhang, Q.; Zou, L.; Xu, X.; Zhang, Z. Nonvolatile Floating-gate Memories Based on Stacked Black Phosphorus-Boron Nitride-MoS_2 Heterostructures. *Adv. Funct. Mater.* **2015**, *25* (47), 7360–7365.

(105) Rodder, M. A.; Vasishta, S.; Dodabalapur, A. Double-gate MoS_2 Field-Effect Transistor with a Multilayer Graphene Floating Gate: A Versatile Device for Logic, Memory, and Synaptic Applications. *ACS Appl. Mater. Interfaces* **2020**, *12* (30), 33926–33933.

(106) Wu, E.; Xie, Y.; Wang, S.; Zhang, D.; Hu, X.; Liu, J. Multi-level Flash Memory Device Based on Stacked Anisotropic ReS_2-Boron Nitride-Graphene Heterostructures. *Nanoscale* **2020**, *12* (36), 18800–18806.

(107) Ahmed, T.; Islam, S.; Paul, T.; Hariharan, N.; Elizabeth, S.; Ghosh, A. A Generic Method to Control Hysteresis and Memory Effect in van der Waals Hybrids. *Mater. Res. Express* **2020**, *7* (1), 014004.

(108) Paul, T.; Ahmed, T.; Tiwari, K. K.; Thakur, C. S.; Ghosh, A. A High-performance

MoS_2 Synaptic Device with Floating Gate Engineering for Neuromorphic Computing. *2D Mater.* **2019**, *6*, 045008.

(109) Wang, S.; He, C.; Tang, J.; Lu, X.; Shen, C.; Yu, H.; Du, L.; Li, J.; Yang, R.; Shi, D.; Zhang, G. New Floating Gate Memory with Excellent Retention Characteristics. *Adv. Electron. Mater.* **2019**, *5*, 180726.

(110) Li, D.; Chen, M.; Zong, Q.; Zhang, Z. Floating-gate Manipulated Graphene-Black Phosphorus Heterojunction for Nonvolatile Ambipolar Schottky Junction Memories, Memory Inverter Circuits, and Logic Rectifiers. *Nano Lett.* **2017**, *17* (10), 6353–6359.

(111) Zhang, P.; Li, D.; Chen, M.; Zong, Q.; Shen, J.; Wan, D.; Zhu, J.; Zhang, Z. Floating-gate Controlled Programmable Non-volatile Black Phosphorus PNP Junction Memory. *Nanoscale* **2018**, *10* (7), 3148–3152.

(112) Zhang, E.; Wang, W.; Zhang, C.; Jin, Y.; Zhu, G.; Sun, Q.; Zhang, D. W.; Zhou, P.; Xiu, F. Tunable Charge-trap Memory Based on Few-layer MoS_2. *ACS Nano* **2015**, *9* (1), 612–619.

(113) Chen, H.; Liu, C.; Wu, Z.; He, Y.; Wang, Z.; Zhang, H.; Wan, Q.; Hu, W.; Zhang, D. W.; Liu, M.; Liu, Q.; Zhou, P. Time-tailoring van der Waals Heterostructures for Human Memory System Programming. *Adv. Sci.* **2019**, *6* (20), 1901072.

(114) Ferrari, A. C.; Meyer, J. C.; Scardaci, V.; Casiraghi, C.; Lazzeri, M.; Mauri, F.; Piscanec, S.; Jiang, D.; Novoselov, K. S.; Roth, S.; Geim, A. K. Raman Spectrum of Graphene and Graphene Layers. *Phys. Rev. Lett.* **2006**, *97* (18), 187401.

(115) Reich, S.; Ferrari, A. C.; Arenel, R.; Loiseau, A.; Bello, I.; Robertson, J. Resonant Raman Scattering in Cubic and Hexagonal Boron Nitride. *Phys. Rev. B* **2005**, *71*, 205201.

(116) Li, H.; Zhang, Q.; Yap, C. C. R.; Tay, B. K.; Edwin, T. H. T.; Olivier, A.; Baillargeat, D. From Bulk to Monolayer MoS_2: Evolution of Raman Scattering. *Adv. Funct. Mater.* **2012**, *22*, 1385–1390.

(117) Tonndorf, P.; Schmidt, R.; Böttger, P.; Zhang, X.; Börner, J.; Liebig, A.; Albrecht,

M.; Kloc, C.; Gordan, O.; Zahn, D. R. T.; Vasconcellos, S. M. De; Bratschitsch, R. Photoluminescence Emission and Raman. *Opt. Express* **2013**, *21* (4), 4908–4916.

(118) Yamamoto, M.; Wang, S. T.; Ni, M.; Lin, Y.; Li, S.; Aikawa, S.; Jian, W.-B.; Ueno, K.; Wakabayashi, K.; Tsukagoshi, K. Strong Enhancement of Raman Scattering from a Bulk-inactive Vibrational Mode in Few-layer $MoTe_2$. *ACS Nano* **2014**, *8* (4), 3895–3903.

(119) Amet, F.; Williams, J. R.; Garcia, A. G. F.; Yankowitz, M.; Watanabe, K.; Taniguchi, T.; Goldhaber-Gordon, D. Tunneling Spectroscopy of Graphene-Boron-Nitride Heterostructures. *Phys. Rev. B* **2012**, *85* (7), 073405.

(120) Hattori, Y.; Taniguchi, T.; Watanabe, K.; Nagashio, K. Determination of Carrier Polarity in Fowler-Nordheim Tunneling and Evidence of Fermi Level Pinning at the Hexagonal Boron Nitride/Metal Interface. *ACS Appl. Mater. Interfaces* **2018**, *10* (14), 11732–11738.

(121) Lee, J.-D.; Choi, J.-H.; Park, D.; Kim, K. Effects of Interface Trap Generation and Annihilation on the Data Retention Characteristics of Flash Memory Cells. *IEEE Trans. Device Mater. Reliab.* **2004**, *4* (1), 110–117.

(122) Aritome, S.; Satoh, S.; Maruyama, T.; Watanabe, H.; Shuto, S.; Hemink, G. J.; Shirota, R.; Watanabe, S.; Masuoka, F. A 0.67μm^2 Self-aligned Shallow Trench Isolation Cell (SA-STI Cell) for 3V-Only 256Mbit NAND EEPROMs. In *Proceedings of 1994 IEEE International Electron Devices Meeting*; 1994; pp 61–64.

(123) Liu, Y.; Matsukawa, T.; Endo, K.; O'uchi, S.; Tsukada, J.; Yamauchi, H.; Ishikawa, Y.; Mizubayashi, W.; Morita, Y.; Migita, S.; Ota, H.; Masahara, M. Experimental Study of Three-dimensional Fin-channel Charge Trapping Flash Memories with Titanium Nitride and Polycrystalline Silicon Gates. *Jpn. J. Appl. Phys.* **2014**, *53* (4S), 04ED16.

(124) Wang, H.; Wu, Y.; Cong, C.; Shang, J.; Yu, T. Hysteresis of Electronic Transport in Graphene Transistors. *ACS Nano* **2010**, *4* (12), 7221–7228.

(125) Bez, R.; Camerlenghi, E.; Modelli, A.; Visconti, A. Introduction to Flash Memory. *Proc. IEEE* **2003**, *91* (4), 489–502.

(126) Lee, G.-H.; Yu, Y.-J.; Lee, C.; Dean, C.; Shepard, K. L.; Kim, P.; Hone, J. Electron Tunneling through Atomically Flat and Ultrathin Hexagonal Boron Nitride. *Appl. Phys. Lett.* **2011**, *99*, 243114.

(127) Fomenko, V. S. *Handbook of Thermionic Properties*; Plenum Press: New York, 1966.

(128) Xu, Y.-N.; Ching, W. Y. Calculation of Ground-state and Optical Properties of Boron Nitrides in the Hexagonal, Cubic, and Wurtzite Structures. *Phys. Rev. B* **1991**, *44* (15), 7787–7798.

(129) *Keysight B1500A Semiconductor Parameter Analyzer Datasheet*.

(130) Liu, Y.; Guo, J.; Zhu, E.; Liao, L.; Lee, S.-J.; Ding, M.; Shakir, I.; Gambin, V.; Huang, Y.; Duan, X. Approaching the Schottky-Mott Limit in van der Waals Metal-Semiconductor Junctions. *Nature* **2018**, *557* (7707), 696–700.

(131) Wu, L.; Wang, A.; Shi, J.; Yan, J.; Zhou, Z.; Bian, C.; Ma, J.; Ma, R.; Liu, H.; Chen, J.; Huang, Y.; Zhou, W.; Bao, L.; Ouyang, M.; Pennycook, S. J.; Pantelides, S. T.; Gao, H.-J. Atomically Sharp Interface Enabled Ultrahigh-speed Non-volatile Memory Devices. *Nat. Nanotechnol.* **2021**, *16* (8), 882–887.

(132) Liu, L.; Liu, C.; Jiang, L.; Li, J.; Ding, Y.; Wang, S.; Jiang, Y.-G.; Sun, Y.-B.; Wang, J.; Chen, S.; Zhang, D. W.; Zhou, P. Ultrafast Non-volatile Flash Memory Based on van der Waals Heterostructures. *Nat. Nanotechnol.* **2021**, *16* (8), 874–881.

(133) McPherson, J. W.; Kim, J.; Shanware, A.; Mogul, H.; Rodriguez, J. Trends in the Ultimate Breakdown Strength of High Dielectric-constant Materials. *IEEE Trans. Electron Devices* **2003**, *50* (8), 1771–1778.

(134) Wu, J.; Juliano, P.; Rosenbaum, E. Breakdown and Latent Damage of Ultra-thin Gate Oxides under ESD Stress Conditions. In *Electrical Overstress/Electrostatic Discharge Symposium Proceedings 2000 (IEEE Cat. No.00TH8476)*; 2000; pp 287–295.

(135) Wu, J.; Rosenbaum, E. Gate Oxide Reliability under ESD-like Pulse Stress. *IEEE Trans. Electron Devices* **2004**, *51* (9), 1528–1532.

(136) McPherson, J. W. Time Dependent Dielectric Breakdown Physics – Models Revisited. *Microelectron. Reliab.* **2012**, *52* (9), 1753–1760.

(137) Hattori, Y.; Taniguchi, T.; Watanabe, K.; Nagashio, K. Layer-by-layer Dielectric Breakdown of Hexagonal Boron Nitride. *ACS Nano* **2015**, *9* (1), 916–921.

(138) Stathis, J. H. Percolation Models for Gate Oxide Breakdown. *J. Appl. Phys.* **1999**, *86* (10), 5757–5766.

(139) Fujita, S.; Noguchi, H.; Ikegami, K.; Takeda, S.; Nomura, K.; Abe, K. Technology Trends and Near-future Applications of Embedded STT-MRAM. In *2015 IEEE International Memory Workshop (IMW)*; 2015; pp 1–5.

(140) Siao, M. D.; Shen, W. C.; Chen, R. S.; Chang, Z. W.; Shih, M. C.; Chiu, Y. P.; Cheng, C.-M. Two-dimensional Electronic Transport and Surface Electron Accumulation in MoS_2. *Nat. Commun.* **2018**, *9* (1), 1442.

(141) Hattori, Y.; Taniguchi, T.; Watanabe, K.; Nagashio, K. Impact Ionization and Transport Properties of Hexagonal Boron Nitride in a Constant-voltage Measurement. *Phys. Rev. B* **2018**, *97* (4), 45425.

(142) Liao, W.; Zhao, S.; Li, F.; Wang, C.; Ge, Y.; Wang, H.; Wang, S.; Zhang, H. Interface Engineering of Two-dimensional Transition Metal Dichalcogenides towards Next-generation Electronic Devices: Recent Advances and Challenges. *Nanoscale Horizons* **2020**, *5* (5), 787–807.

(143) Liu, Y.; Huang, Y.; Duan, X. Van der Waals Integration before and beyond Two-dimensional Materials. *Nature* **2019**, *567* (7748), 323–333.

(144) Gehring, A. Simulation of Tunneling in Semiconductor Devices, Technische Universität Wien, 2003.

(145) Muto, K. http://www.th.phys.titech.ac.jp/~muto/. (View 2021/06/01)

Publications

A. Original Papers for the Journals with Peer Review

(1) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Understanding the Memory Window Overestimation of 2D Materials Based Floating Gate Type Memory Devices by Measuring Floating Gate Voltage. *Small* **2020**, *16* (47), 2004907.

(2) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Material and Device Structure Designs for 2D Memory Devices Based on the Floating Gate Voltage Trajectory. *ACS Nano* **2021**, *15* (4), 6658–6668.

B. Proceedings for the Conferences with Peer Review

(1) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. 50 ns Ultrafast P/E Operation in 2D Heterostructured Non-volatile Memory Device, *2021 International Conference on Solid State Devices and Materials (SSDM)*, H-6-04, Online, September 2021. (Oral, Scheduled)

(2) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Understanding the Tunneling Behavior in 2D Based Floating Gate Type Memory Device by Measuring Floating Gate Voltage, *2020 International Conference on Solid State Devices and Materials (SSDM)*, pp. 579–580, H-9-04, Online, September 2020. (Oral)

(3) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Understanding the Device Operation of Ambipolar Channel Based 2D Memory Devices by Trajectory of Floating Gate Voltage, *78th Device Research Conference*, Online, June 2020. (Oral)

(4) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nagashio, K. High Temperature Retention Study of MoS_2/*h*-BN/MoS_2 Hetero-stack Based Non-

volatile Memory, *Recent Progress in Graphene & 2D Materials Research (RPGR) 2019*, 9p-9, Matsue, Japan, October 2019. (Poster)

(5) **Sasaki, T.**; Taniguchi, T.; Watanabe, K.; Nagashio, K. 388 K High Temperature Retention Study of 2D Hetero-stack Based Non-volatile Memory, *2019 International Conference on Solid State Devices and Materials (SSDM)*, pp. 199–200, D-6-02, Nagoya, Japan, September 2019. (Oral)

C. Proceedings for the Conferences without Peer Review

(1) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Strong Breakdown Strength of *h*-BN under Fast Voltage Pulse Stress Results in the Ultra-fast Operation of 2D Heterostructured Memory Devices, *The 82nd Japan Society of Applied Physics Autumn Meeting*, 13a-N307-1, Nagoya, Japan & Online, September 2021. (Oral, Scheduled)

(2) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Sub 1μs Programming Operation in 2D Heterostructured Memory Devices, *The 68th Japan Society of Applied Physics Spring Meeting*, p. 15-111, 19a-Z31-11, Online, March 2021. (Oral, **Highlighted**)

(3) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. V_{FG} Trajectory Based Understanding of 2D Memory Device Operation, *The 81st Japan Society of Applied Physics Autumn Meeting*, p. 15-115, 11p-Z29-11, Online, September 2020. (Oral)

(4) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Understanding the Memory Window Overestimation of 2D Memory Devices by Measuring Floating Gate Voltage, *The 81st Japan Society of Applied Physics Autumn Meeting*, p. 15-114, 11p-Z29-10, Online, September 2020. (Oral)

(5) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Understanding the Operation of 2D Memory Device by Measuring Floating Gate

Voltage, *The 67th Japan Society of Applied Physics Spring Meeting*, p. 14-081, 12p-A401-4, Tokyo, Japan, march 2020. (Oral)

(6) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. High Temperature Retention Study of 2D Hetero-stack Based Non-volatile Memory, *UTokyo-NTU joint conference*, Tokyo, Japan, December 2019. (Oral)

(7) **Sasaki, T.**; Ueno, K.; Taniguchi, T.; Watanabe, K.; Nishimura, T.; Nagashio, K. Operation Mechanism Analysis of MoS_2/*h*-BN/Graphite Stack Based Non-volatile Memory Device, *The 80th Japan Society of Applied Physics Autumn Meeting*, p. 15-081, 20a-E201-1, Sapporo, Japan, September 2019. (Oral)

The following works in Tohoku University are not directly related to this thesis:

A. Original Papers for the Journals with Peer Review

(1) **Sasaki, T**; Endoh, T. Cross-point-type Spin-Transfer-Torque Magnetoresistive Random Access Memory Cell with Multi-pillar Vertical Body Channel MOSFET, *Japanese Journal of Applied Physics.* **2018**, 57 (4S), 04FN09.

B. Proceedings for the Conferences with Peer Review

(1) **Sasaki, T**; Endoh, T. Cross Point Type 1T-1MTJ STT-MRAM Cell with 60 nm Multi-pillar Vertical Body Channel MOSFET under 55 nm p-MTJ and Its Beyond for High Density STT-MRAM, *2017 International Conference on Solid State Devices and Materials (SSDM)*, pp. 31-32, A-5-03, Sendai, Japan, September 2017. (Oral)

(2) **Sasaki, T**; Endoh, T. A Study on High Density and Low Power 1-nMOS and 1-bottom-pin-MTJ Type STT-MRAM Cell with Vertical Body Channel MOSFET, *Asia-Pacific Workshop on Fundamentals and Applications of Advanced Semiconductor Devices (AWAD) 2017*, 5B-3, Gyeongju, Korea, July 2017. (Oral)

(3) **Sasaki, T.**; Muraguchi, M.; Shinada, T.; Endoh, T. A Study of Strain Profile in Channel Region of Vertical MOSFET for Improving Drivability, *Asia-Pacific Workshop on Fundamentals and Applications of Advanced Semiconductor Devices (AWAD) 2015*, 1A-2, Jeju Island, Korea, June–July 2015. (Oral)

C. Proceedings for the Conference without Peer Review

(1) **Sasaki, T.**; Endoh, T. A Study on High Density and Low Power 1-nMOS and 1-bottom-pin-MTJ Type STT-MRAM Cell with Vertical Body Channel MOSFET, *Lorraine-Mainz-Tohoku joint Seminar: Prospect of Future Spintronics 2017*, SP-9, Sendai, Japan, October–November 2017. (Oral)

Award

(1) 53rd SSDM Young Researcher Award, International Conference on Solid State Devices and Materials (SSDM), 2020.