Circuit design of NAND flash memory for high-speed programming

NAND 型フラッシュメモリの書き込み高速化に向けた回路設計に関する研究

Ken Takeuchi

竹内 健

January 2006

2006 年 1 月

Abstract

This paper presents circuit design techniques to increase the program throughput of NAND flash memories. First, a new theoretical framework to increase the program throughput is developed. In the theory, the programming is modeled and the program throughput is expressed explicitly in terms of the key parameters, such as the simultaneously written states, the program pulse width, the verify read time, the circuit noise and the page size.

In addition, based on this new model, the guidelines for the high speed programming are developed for three kinds of NAND flash technologies, respectively, i.e. 3V MLC (3 volt multi-level cell), LV MLC (Low voltage multi-level cell) and 3V SLC (3 volt single-level cell).

Furthermore, four novel circuit techniques are presented to accelerate the programming of the 3V multi-level cell, the low voltage multi-level cell and the 3V single-level cell. The program speed target speed is 10MB/sec, 5MB/sec and 30MB/sec for the 3V multi-level cell, the low voltage multi-level cell and the 3V single-level cell, respectively.

Among four key circuit techniques, the multipage cell technology is developed to increase the program speed of the 3V multi-level cell and the low voltage multi-level cell. In this new scheme, the two bits in one memory cell is assigned to two row addresses, whereas in the conventional multi-level cell the two bits in one memory cell is assigned to two column addressed. As a result, the three programmed states are programmed in two operations, while they are programmed in one operation in the case of the conventional multi-level cell. The simultaneously written states, Nstate is decreased from three to one or two and therefore the program speed is accelerated by 130%. Moreover, the random read speed is doubled and the reliability is drastically improved by the order of four. Because of the multipage cell technologies, the multi-level NAND flash memory is successfully commercialized first in the world in 2001. This technology is used in all multi-level NAND flash memory products.

Second, the low load capacitance technology is developed to improve the program performance of the low voltage multi-level cell. In this technology, the cell load capacitance of the charge pump is drastically reduced. As a result, the bit-line ramp up time is decreased and the program speed is accelerated by 70%. In addition to the program speed improvement, the energy consumption is saved by 53%, the chip size is decreased by 5% and moreover the production cost is reduced by 5%. This novel technology will make it possible to realize below 1.8 volt operation NAND flash memories.

Third, the low noise technology is developed for the 3V multi-level cell and the low voltage multi-level cell. Two schemes are proposed. For the 3V multi-level cell, the new array architecture eliminating the source-line is developed. By using this new architecture, the circuit noise originated from the high resistance of the source-line as well as the capacitive coupling between bit-lines, are eliminated. For the low voltage multi-level cell, a novel Vcc-bit-line sensing scheme is proposed to eliminate the circuit noise. By using these

technologies, the program speed of the 3V multi-level cell and the low voltage multi-level cell are improved by 77% and 57%, respectively. These new technologies will accelerate the program speed of future multi-level NAND flash memories and will realize a high-speed 10MB/sec 3V multi-level cell and 5MB/sec low voltage multi-level cell.

Fourth, the parallel write technology is proposed for the 3V single-level cell. By using a bit-line as a dynamic latch circuit, two memory cells can be programmed at the same time with just one latch circuit. Consequently, the page size is doubled without any circuit area overhead and the program speed is improved by 73%. A very fast, 30MB/sec 3V multi-level cell will be realized by using this technology.

At the end of this thesis, the future program speed trend of the 3V multi-level cell, the low voltage multi-level cell and the 3V single-level cell are discussed. The program speed targets of the 3V multi-level cell, low-voltage multi-level cell and 3V single-level cell discussed above are fulfilled with this research. The program speed of the 3V multi-level cell has been 130% improved by using the multipage cell technology. Then, the performance of the 3V multi-level cell has been 77% improved by the low noise technology. In total, the program speed of the 3V multi-level cell is improved by 307%. As a result, the world's first 1MB/sec was realized and the world's first multi-level NAND flash memory was commercialized in 2001. Moreover, 10MB/sec 3V multi-level cell is realized, which makes it possible to store HDTV movies or to store a music or movie data as fast as FTTH. The low voltage multi-level cell will be commercialized by using the multipage cell technology and the low load capacitance technology. The programming of the low voltage multi-level cell will be accelerated further by using the low noise technology. Totally, the program speed is improved by 514% and 5MB/sec low voltage multi-level cell will be realized, which enables us to take or download DVD movies with mobile phones. Finally, as for the 3V single-level cell, the program speed is improved by 73% with the parallel write technology. A very fast programming faster than 30MB/sec will be realized, which makes it possible for digital still cameras to continuously take very high-resolved pictures with more than 10M pixels.

Acknowledgment

I would like to express my special thanks to Professor Takayasu Sakurai, the University of Tokyo for invaluable guidance and continuous encouragement. I am also grateful to Professor Tadashi Shibata, Professor Shin-ichi Takagi, Professor Toshiro Hiramoto, Associate Professor Minoru Fujishima and Associate Professor Makoto Ikeda, all of the University of Tokyo, for their advice and support as the qualifying examination committee members.

I would like to thank Dr. Tohru Furuyama, Mr. Tomoharu Tanaka, Mr. Kenichi Imamiya, Mr. Hiroto Nakai, Mr. Kazuhisa Kanazawa, Dr. Fujio Masuoka, Mr. Kazunori Ohuchi, Dr. Shigeyoshi Watanabe, Mr. Kiyoshi Kobayashi, Dr. Riichiro Shirota, Mr. Shigeo Ohshima, Dr. Koji Sakui, Dr. Junichi Miyamoto, Mr. Masaki Momodomi, Mr. Yukihito Oowaki, Mr. Shigeru Atsumi, Dr. Hisashi Hara, Dr. Akimichi Hojo, Dr. Yoichi Unno, and Dr. Kenji Maeguchi for their warm consideration and continuous encouragement ever since I joined the ULSI Research Laboratories of Toshiba Corporation.

I am grateful to colleagues of mine at Toshiba Corporation for their contributions. Without their help, this work could not have been successful. I especially thank Dr. Toru Tanzawa, Mr. Hiroshi Nakamura, Mr. Shinji Satoh, and Dr. Kazuhiro Shimizu. I appreciate Dr. G.J. Hemink, Mr. Yoshihisa Sugiura, Mr. Toshihiko Himeno, Dr. Tamio Ikehashi, Mr. Kazushige Kanda, Mr. Koji Hosono, Mr. Noboru Shibata, Mr. Seiichi Aritome, Mr. Kazuo Hatakeyama, Dr. Tohru Maruyama, Mr. Kenji Tsuchiya, Mr. Tatsuya Tanaka for their fruit full discussion.

Finally, I would like to express my heart-felt thanks to my wife, Tomoko Takeuchi and my sons, Wataru Takeuchi and Nozomu Andrew Takeuchi who have continuously supported me with their love and patience throughout this work.

CONTENTS

CHAPTER 1
INTRODUCTION 6
1.1 Background 6
1.2 Research objectives 10
1.3 Chapter organization and overview 12

CHAPTER 2
PRINCEPLES OF HIGH-SPEED DESIGN 22
2.1 Analysis of program operation 22
2.1.1 Bit by bit program verify scheme 22
2.1.2 Incremental program voltage scheme 23
2.1.3 Expression of program throughput 24
2.2 Guidelines for high speed design 27
2.3 Summary 32

CHAPTER 3
MULTIPAGE CELL TECHNOLOGY 36
Acceleration by reducing simultaneously written states, Nstate
3.1 Introduction 36
3.2 Multipage cell architecture 37
3.3 New NAND flash system 40
3.4 Circuit implementations 44
3.5 Reliability consideration 46
3.6 Experimental results 53
3.7 Summary 58

CHAPTER 4
LOW LOAD CAPACITANCE TECHNOLOGY 63
Acceleration by decreasing program pulse width, Tpulse
4.1 Introduction 63
4.2 Design issues at low Vcc 65
4.3 Source-line programming scheme 72
4.4 Circuit implementations 77
4.5 Experimental results 80
4.6 Design considerations 81
4.6.1 Power consumption 81
4.6.2 Area/Manufacturing cost penalty 81
4.7 Summary 84

CHAPTER 5
LOW NOISE TECHNOLOGY 88
Acceleration by reducing program cycles, Npulse
5.1 Introduction 88
5.2 Circuit noise issues deteriorating Vth control 89
5.2.1 Inter bit-line capacitive coupling noise 90
5.2.2 Source-line noise 93
5.3 Double-level Vth select gate array architecture 97
5.3.1 Circuit implementations 97
5.3.2 Experimental results 100
5.4 Vcc bit-line shield sensing scheme 101
5.4.1 Circuit implementation 102
5.4.2 Experimental results 106
5.4.3 Erase verify opreration 106
5.5 Summary 110

CHAPTER 6
PARALLEL WRITE TECHNOLOGY 114
Acceleration by increasing the page size
6.1 Introduction 114
6.2 Dual page programming scheme 115
6.3 New NAND flash protocol 119
6.4 Circuit implementations 122
6.5 Experimental results 125
6.6 Design considerations 127
6.6.1 Charge leakage issue 127
6.6.2 Capacitive coupling noise 127
6.6.3 Power consumption 128
6.7 Summary 128

CHAPTER 7
DISCUSSION AND CONCLUSION 134
7.1 Discussion on future program speed trend 134
7.2 Conclusion 136
7.3 Perspective 139

List of research achievement

CHAPTER 1
INTRODUCTION

1.1 Background

The increasing demand for hand-held digital equipment, such as digital cameras and digital audio players has created a need for a low-cost, high-density, high-speed programming file memory. For such requirements, a NAND flash memory [1-16] is preferable, as it inherently has many advantages over other flash memories such as AND [17-21], DINOR [22-23], Split-gate NOR [24-26], NROM [27-29], FN-tunneling NOR [30-31], and Hot-electron injection NOR [32-44] flash memories. One is a smaller cell size, as its memory cells are arranged in series [1-3]. Another is a high reliability. The bi-polarity FN-programming/erase improves the charge to breakdown (Qbd) characteristics and decreases the stress-induced leakage current [45-50]. Besides these, the low power consumption is also obtained by using Fowler-Nordheim (FN) tunneling for both program and erase operations [51-55].

The FN-based programming also allows a page-based program operation and drastically increases the program throughput. Already, a 4.4M-byte/sec 256-Mbit NAND flash memory [8] has been reported, whose program throughput is thirty times larger than that of the NOR flash memory [32-44]. However, even this programming speed is not fast enough to store high-resolution moving pictures in digital movie cameras.

In addition, recently, a lot of attention has been paid to multi-level cell technology [4, 6, 10,12-15,18,20,36,38,51,52,56-62] because they drastically reduce the cost per bit. For example, the memory cell density can be doubled without a die size increase if the four levels of data can be stored in one memory cell. However, in the multi-level cell, there is a critical problem of the performance degradation. In the case of single-level cell, i.e. two-level cell, “1” is the erased state and “0” is the erased state as shown in Fig.1.1. On the other hand, in the multi-level cell, “1”, “2” and “3” are the programmed states.

To realize multi-level flash memories, the programmed threshold voltage, Vth, of a memory cell needs to be controlled precisely [51, 52, 59, 60]. This is because as shown in Fig.1.1 the maximum Vth of the multi-level cell is higher than that of the single-level cell, which adversely affects the data retention characteristics due to the larger electric field across the tunnel oxide during storage [60] [64]. The Vth can be controlled precisely by

injecting electrons little by little [59]. As a result, the Vth distribution will be narrowed and the maximum Vth will be decreased. Consequently, the precise Vth control can solve the reliability degradation problem but it will drastically degrade the program performance because electron injection is slowed down.

On the other hand, in terms of the supply voltage, the increasing demand for low voltage/low power portable equipment in the consumer marketplace has created a need for a low voltage/low power flash memory. In particular, sub-1.8V operation is essential for systems requiring extended battery lifetime such as cellular phones, personal information devices, handheld global positioning systems and portable instrumentation [54]. Already 1.8V operations of a NOR flash memory [39] and a DINOR flash memory [23] [40] have been reported.

For these low voltage/low power applications, a NAND flash memory inherently has many advantages over other flash memory because of 1) the smaller cell size, 2) the lower power consumption, 3) the fast programming, and 4) the high reliability. However, in case of the NAND flash memory, the boosting of the huge bit-line capacitance leads to both larger power dissipation and a longer programming time at low Vcc [53] [63].

From a market point of view, there are three market segments, 1) mass market, 2) mobile market and 3) high-end market as shown in Table 1.1. First, in the mass market, the target application is the storage for the picture, movie, audio, PDA, STB (set top box), car navigation system, game, data storage such as USB memory and Silicon-disk. The program speed is fast (10MB/sec), the capacity is large (16Gbit) and the supply voltage is 3V. Considering these requirements, the most appropriate technology is 3V MLC (3V multi-level cell) technology. Multi-level cell is preferable over single-level cell because of the cost advantage.

Second, in the mobile market, the main application is the mobile phone. The program speed is moderate (5MB/sec), the capacity is rather small (4Gbit) and the supply voltage should be below 1.8V. For there requirements, the low voltage MLC (low voltage multi-level cell) technology should be adopted. Again, the multi-level cell is preferable over the single-level cell because of the cost advantage.

Third, in the high-end market, the application is high resolution picture with continuous shooting or high resolution movie. The programming should be very fast (30MB/sec) because high capacity data should be transferred. The capacity is large (8Gbit) and the supply voltage is 3V. For these market requirements, the 3V SLC (3V single-level-cell) technology is the promising candidate. Although the cost of the single-level cell is almost twice as large as the multi-level cell, in the high-end market the multi-level cell cannot satisfy the speed requirement and as a result the single-level cell is preferred.

Fig.1.2 summarizes the comparison of the NAND flash memory and HDD. The cost of HDD is 20-50% of that of the NAND flash memory. As for the size, the NAND flash memory can be reduced below 1cm^2, while HDD cannot be reduced below 1.5cm^2. In

terms of the thickness, the NAND flash memory can be decreased below 1mm. On the other hand, thickness of HDD cannot be reduced below 5mm. The program and the read performance is almost the same. In terms of the power consumption and the supply voltage, the NAND flash memory has a big advantage over HDD. Also, HDD is not reliable when shocked. In summary, the NAND flash memory has many advantages such as small size, thin thickness, low power consumption, low supply voltage operation and high reliability compared with HDD in the mobile equipment market. As a result, although HDD has an advantage in terms of cost, much more mobile electronic devices are using a NAND flash memory instead of HDD.

Fig.1.3 shows the program speed trend of 3V SLC (3V single-level cell) [9, 12, 13, 16], 3V MLC (3V multi-level cell) [9, 10, 12-15] and LV MLC (low voltage multi-level cell). According as 1) the number of pixels of the image processor increases, 2) the clock frequency of the microprocessor increases and finally 3) the bandwidth of wire-line and wireless communication increases, a higher speed programming, 50% faster per year, is essential in all markets.

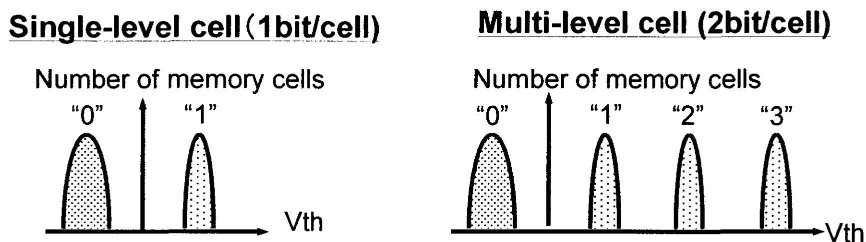


Fig.1.1 Vth distribution of single-level cell and multi-level cell

NAND Technology	3V multi-level cell	Low voltage multi-level cell	3V single-level-cell
Market	Mass market	Mobile market	High-end market
Application	Picture, Movie, Audio, PDA, STB, Car Navigation, Game, Data storage (USB memory, Si-disk)	Mobile phone	High resolution picture w. continuous shooting High resolution movie
Program speed	One third of 3V MLC (10MB/sec)	One sixth of 3V SLC (5MB/sec)	Fastest (30MB/sec)
Capacity	Largest (16Gbit)	One fourth of 3V MLC (4Gbit)	One half of 3V MLC (8Gbit)
Vcc	3V	<1.8V	3V

Table1.1 NAND flash memory market

	NAND flash	1inch HDD
Relative cost	-	20-50%
Size	<1cm^2	1.5cm^2
Thickness	<1mm	5mm
Program speed	10MB/sec	10MB/sec
Read speed	30MB/sec	30MB/sec
Operation current	30mA	300mA
Power supply	1.6V - 3.6V	2.7V - 3.6V
Reliability	-	Need extra sensor protection

Fig.1.2 Comparison of NAND flash memory and HDD

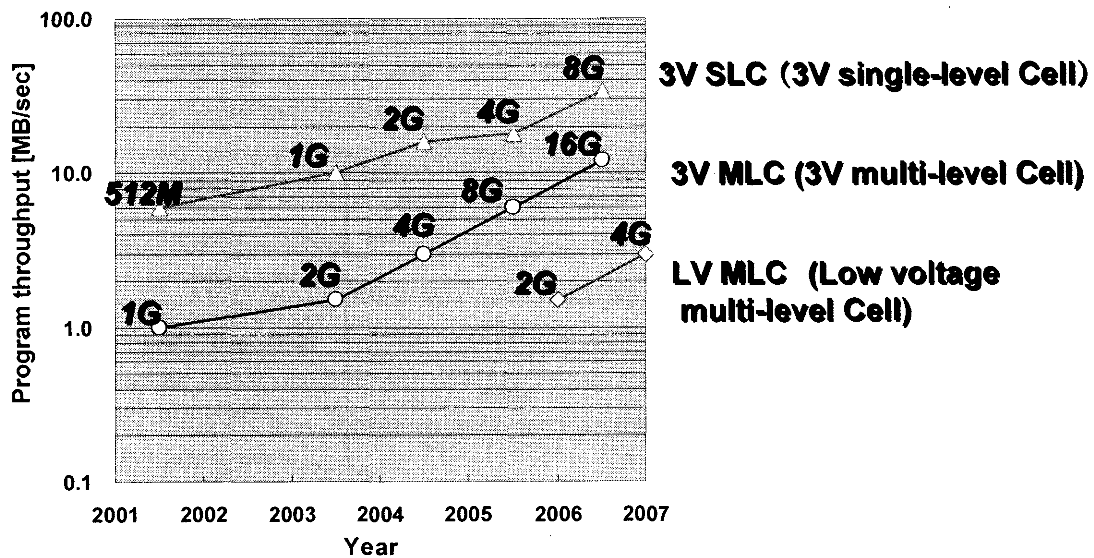


Fig.1.3 Program speed trend of NAND flash memories

1.2 Research objectives

The objective of this research is to study the circuit design technologies for high-speed programming in three categories of NAND flash memories, that is, 3V MLC (3V multi-level cell), LV MLC (low-voltage multi-level cell) and 3V SLC (3V single-level cell). Investigation focuses on the fast-programming circuit design since among many performance attributes the program speed is the most important one in high capacity NAND flash memories.

Fig.1.4 summarizes the program speed target of this research. The target speed is 10MB/sec, 5MB/sec and 30MB/sec for the 3V multi-level cell, the low voltage multi-level cell and the 3V single-level cell, respectively. In case of the 3V multi-level cell, at first 1MB/sec is required to commercialize a multi-level NAND flash memory and provide a low cost mobile storage to the society. Ultimately, in case of the 3V multi-level cell, 10MB/sec is required to makes it possible to store HDTV movies or to store a music or movie data as fast as FTTH. In case of the low voltage multi-level cell, 5MB/sec is required to take or download DVD movies with mobile phones. Finally, as for the 3V single-level cell, the target speed is 30MB/sec, which makes it possible for digital still cameras to continuously take very high-resolved pictures with more than 10M pixels.

In the case of 3V multi-level cell, the most critical issue is to develop new circuit technologies which overcome the speed degradation of the multi-level cell. In the multi-level cell, in order to achieve a high reliability the precise Vth control is essential, which drastically makes worse the performance.

In case of the low-voltage multi-level cell, in addition to the Vth control problem discussed above the program time increase due to the long ramp up time of the bit-line should be

Product	Market	Program speed target	Speed improvement needed to achieve the target	Benefit for the society
3V MLC	Mass market	1MB/s	100%	Commercialize multi-level cell and create low-cost mobile storage
		10MB/s	300%	Take HDTV movie/Download movie or music data with FTTH level speed
LV MLC	Mobile market	5MB/s	500%	Create mobile phone with DVD-movie shooting or downloading
3V SLC	High-end market	30MB/s	70%	Create digital still camera with continuous shooting of 10M pixel high resolution pictures

Fig.1.4 Program speed target of this research

solved.

In case of the single-level cell, the parallel operation is most effective in accelerating the programming. However, the area increase related with the page size increase should be overcome.

Although there is a strong market demand for the fast-programming NAND flash memory, there is no theory or framework to analyze the program speed.

Corresponding to these motivations, this research has the following concrete objectives:

1) Develop a theoretical framework to numerically analyze the program speed.
2) Based on the new theory, make clear the problems in accelerating the speed in 3V multi-level cell, low-voltage multi-level cell and 3V single-level cell, respectively.
3) Develop the guidelines for fast-programming which overcome the problems above.
4) Develop the circuit technologies to achieve a fast programming in 3V multi-level cell, low-voltage multi-level cell and 3V single-level cell, respectively based on the new guidelines.
5) Analyze the reliability degradation at lower Vcc and evaluate its effect on the speed.
6) Analyze the cell array noise and clarify its influence on the speed.
7) Investigate circuit design techniques to minimize the simultaneously written states, Nstate to accelerate the programming of the multi-level cell
8) Develop circuit design techniques to reduce the cell load capacitance to overcome the speed degradation problem at low Vcc.
9) Investigate circuit design techniques to eliminate noise and accelerate the programming of the multi-level cell.
10) Investigate circuit design techniques to increase the page size and maximize the programming speed of the single-level cell.
11) Define the future program speed trend.

In general, high-speed circuit techniques such as pipelining or wide bandwidth adversely affect the chip size or production costs. On the other hand, this research focus on creating new circuit technologies which improve the program performance without any chip size or production cost overhead or preferably can improve other attributes such as the energy consumption, the chip size and the production costs. This viewpoint is very important because the most important attribute of NAND flash memory is the cost advantage over any other storage devices and therefore fast-programming technologies which sacrifice the production costs make no sense.

In the conventional circuit design, the simultaneously written state in the multi-level cell and the page size are parameters given to circuit designers as boundary conditions and unchangeable parameters. There has no research work investigating the performance of the NAND flash memories at low Vcc. In addition, no research has been done regarding the circuit noise and its effect on the performance.

The significance of this research in terms of engineering resides in the following points:

1) A theory to analyze the program operation is established whereas there was no model or theory to analyze the program throughput so far.
2) Based on the theory, guidelines for high-speed programming are proposed.
3) The circuit performance is optimized for each of different NAND flash memories, the 3V multi-level cell, the low-voltage multi-level cell and the 3V single-level cell.
4) Vcc dependence on the reliability and its influence on the performance are analyzed, which has previously not been studied. According as the analysis, new circuit techniques optimizing the cell load capacitance are developed to resolve the speed problem at low Vcc.
5) Cell array noise and its effect on the speed are investigated, whereas there has been no literature to study the noise so far. Based on the analysis, new circuit techniques are proposed to remove the noise.
6) The simultaneously written states, Nstate is treated as objectives of design optimization whereas it has been given condition so far. As a result, a better program performance of MLC is achieved.
7) The page size is expanded by circuit techniques whereas it has been fixed as the number of page buffers thus far. Consequently, the program performance of SLC is improved.
8) Most technologies are used in commercial NAND flash memories. The world's first multi-level NAND flash memory was realized in 2001 with technologies developed through this research. Furthermore, the research results have been used in all multi-level NAND flash memories and have become de-fact standard technologies.
9) This research has improved the program speed of the 3V multi-level cell, low-voltage multi-level cell and the 3V single-level cell by 307%, 514% and 77%, respectively. As a result, the target speed of the 3V multi-level cell (10MB/sec), low-voltage multi-level cell (5MB/sec) and the 3V single-level cell (30MB/sec) are realized.

1.3 Chapter organization and overview

This paper is organized as shown in Fig.1.5. As shown in Fig.1.6, the technologies developed in this research focus on the memory core circuits.

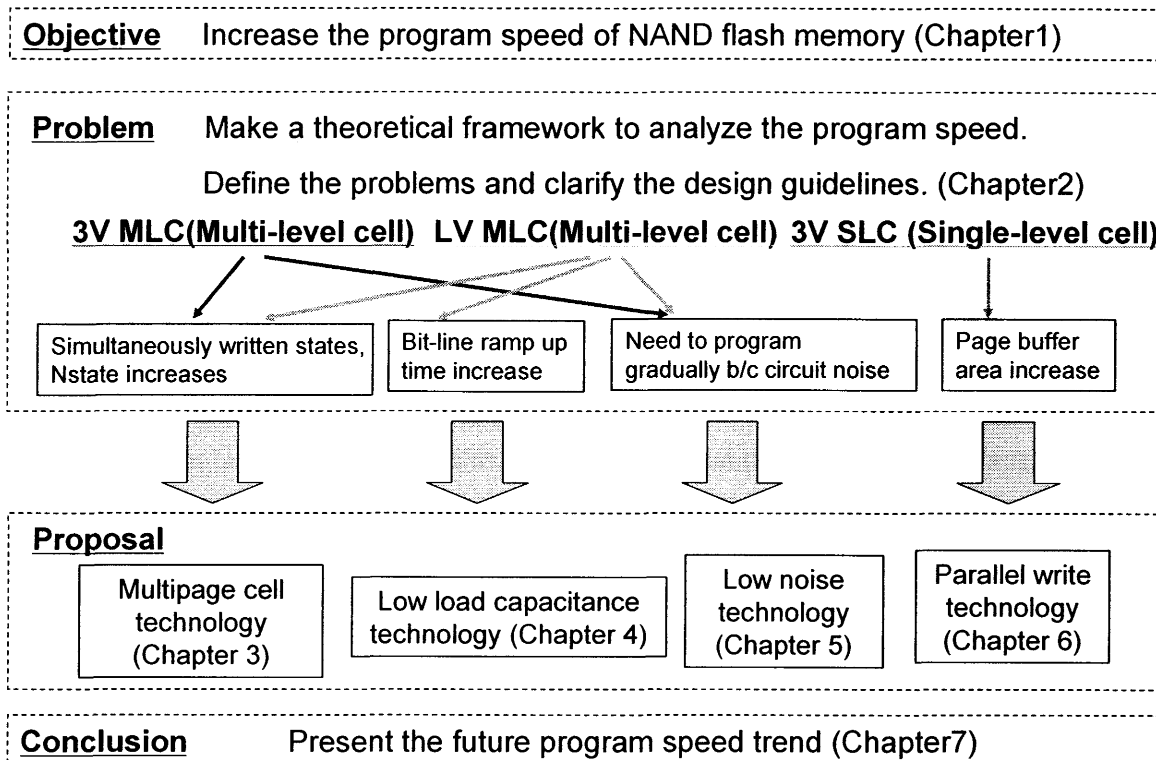


Fig.1.5 Structure of this thesis

Chapter 2 presents a new theoretical framework on the program speed. The programming is modeled and the program throughput is expressed explicitly in terms of key parameters, such as the simultaneously written states, the program pulse width, the verify read time, the circuit noise and the page size. Also, based on this new model, the problems in accelerating the speed are clarified. Finally, the guidelines for the high speed programming are developed for three NAND flash technologies, respectively.

In Chapter 3 a multipage cell technology for the 3V multi-level cell and the low-voltage multi-level cell is introduce and discussed. High speed multi-level cell is realized by decreasing the simultaneously written states, Nstate. The circuit implementations as well as the NAND flash system suitable for the multipage cell technology are presented. Moreover, the reliability improvement of the propose scheme is discussed. Finally, experimental results are reported.

Chapter 4 describes the low load capacitance technology for the low voltage multi-level cell. In this technology, the cell load capacitance of the charge pump is drastically reduced. Circuit implementations are explained and experimental results for a 0.25um 256Mbit NAND flash memory are reported. In addition, design considerations are addressed and the

future implementation plan on the low-voltage multi-level cell are discussed.

The low noise technology for the 3V multi-level cell and the low voltage multi-level cell is introduced in Chapter 5. The circuit noise and its effect on the performance are first analyzed. Then, two key technologies eliminating are presented. For the 3V multi-level cell, the new array architecture eliminating the source-line is developed. For the low voltage multi-level cell, a novel Vcc-bit-line sensing scheme is proposed to eliminate the circuit noise. Circuit implementations and experimental results for a 256Mbit NAND flash memory are shown. Also, future possibilities are discussed.

Chapter 6 introduces the parallel write technology for the 3V single-level cell. By using a bit-line as a dynamic latch circuit, two memory cells can be programmed at the same time with just one latch circuit. Consequently, the page size is doubled without any circuit area overhead. Circuit implementations and the new NAND flash protocol are presented. Design consideration especially about the noise to the dynamic latch is discussed. Finally, experimental results and the future implementation plan into products are shown.

Conclusions base on this research and discussions about the future program speed trend are described in Chapter7. In Chapter 7, perspectives on the future high-speed technologies are also discussed.

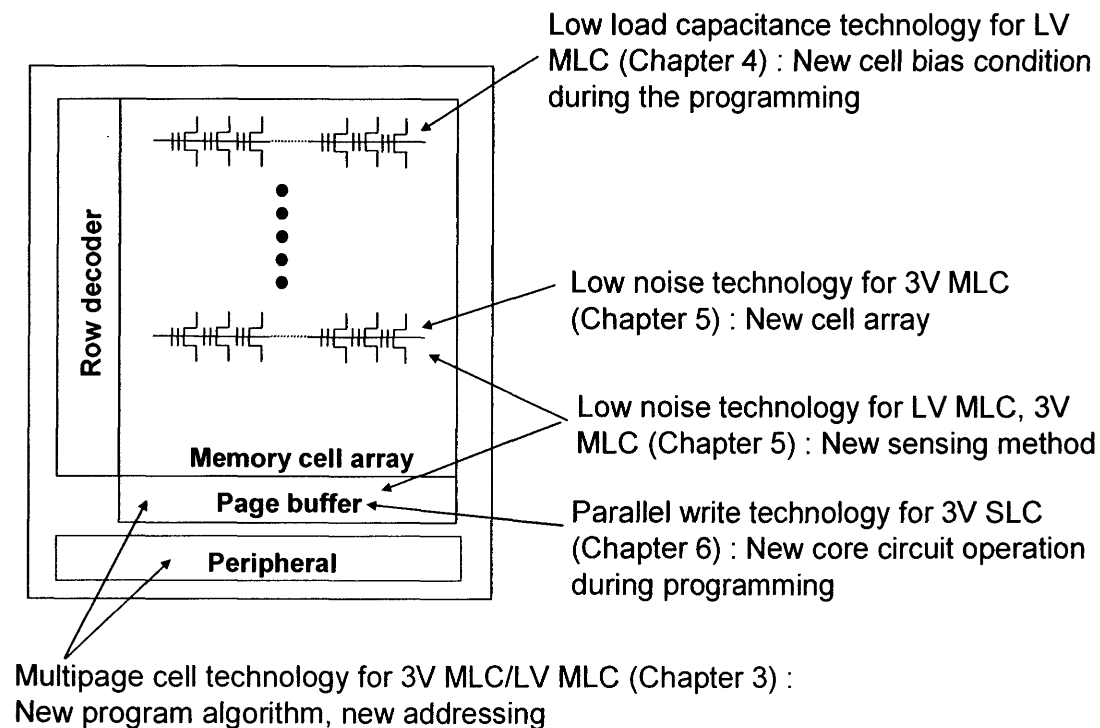


Fig.1.6 New technologies in this research

References

[1] F. Masuoka, M. Momodomi, Y. Iwata and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structured cell," in *IEDM Tech. Dig.,* 1987, pp.552-555.

[2] K. Imamiya, Y. Iwata, Y. Sugiura, H. Nakamura, H. Oodaira, M. Momodomi, Y. Ito, T. Watanabe, H. Araki, K. Narita, K. Masuda and J. Miyamoto, "A 35 ns-cycle-time 3.3 V-only 32 Mb NAND flash EEPROM", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.32-33.

[3] Kang-Deog Suh, Byung-Hoon Suh, Young-Ho Um, Jin-Ki Kim, Young-Joon Choi, Yong-Nam Koh, Sung-Soo Lee, Suk-Chon Kwon, Byung-Soon Choi, Jin-Sun Yum, Jung-Hyuk Choi, Jang-Rae Kim and Hyung-Kyu Lim, "A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.128-129.

[4] Tae-Sung Jung, Young-Joon Choi, Kang-Deog Suh, Byung-Hoon Suh, Jin-Ki Kim, Young-Ho Lim, Yong-Nam Koh, Jong-Wook Park, Ki-Jong Lee, Jung-Hoon Park, Kee-Tae Park, Jang-Rae Kim, Jeong-Hyong Lee and Hyung-Kyu Lim, "A 3.3 V 128 Mb multi-level NAND flash memory for mass storage application", in *ISSCC Digest of Technical Papers*, Feb., 1996, pp.32-33.

[5] J.K. Kim, K. Sakui, S.S. Lee, J. Itoh, S.C. Kwon, K. Kanazawa, J.J. Lee, H. Nakamura, K.Y. Kim, T. Himeno, J.R. Kim, K. Kanda, T.S. Jung, Y. Oshima, K.D. Suh, K. Hashimoto, S.T. Ahn, and J. Miyamoto, "A 120 mm^2 64 Mb NAND flash memory achieving 180 ns/byte effective program speed," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1996, pp.168-169.

[6] H. Nobukata, S. Takagi, K. Hiraga, T. Ohgishi, M. Miyashita, K. Kamimura, S. Hiramatsu, K. Sakai, T. Ishida, H. Arakawa, M. Itoh, I. Naiki and M. Noda, "A 144 Mb 8-level NAND flash memory with optimized pulse width programming," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1999, pp.39-40.

[7] Tae-Sung Jung, Do-Chan Choi, Sung-Hee Cho, Myong-Jae Kim, Seung-Keun Lee, Byung-Soon Choi, Jin-Sun Yum, San-Hong Kim, Dong-Gi Lee, Jong-Chang Son, Myung-Sik Yong, Heung-Kwun Oh, Sung-Bu Jun, Woung-Moo Lee, Ejaz Haq, Kang-Deog Suh, Syed Ali and Hyung-Kyu Lim, "A 3.3 V 16 Mb nonvolatile virtual DRAM using a NAND flash memory technology", in *ISSCC Digest of Technical Papers*, Feb., 1997, pp.398-399.

[8] K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, K. Takeuchi, T. Ikehashi, K. Kanda, K. Hosono, R. Shirota, S. Aritome, K. Shimizu, K. Hatakeyama and K. Sakui " A 130mm^2 256Mb NAND Flash with Shallow Trench Isolation Technology", *ISSCC Digest of Technical Papers*, pp.112-113, 1999.

[9] H.Nakamura, K.Imamiya, T.Himeno, T.Yamamura, T.Ikehashi, K.Takeuchi, K.Kanda, K.Hosono, T.Futatsuyama, K.Kawai, R.Shirota, N.Arai, Fumitaka Arai, K.Hatakeyama,

H.Hazama, Masanobu Saito, H.Meguro, K.Conley, K. Quader and J.Chen, "A 125mm^2 1Gb NAND Flash Memory with 10MB/s Program Throughput," *ISSCC Digest of Technical Papers*, 2002, pp.106-107.

[10] Taehee Cho, Young-Taek Lee, Euncheol Kim, Jinwook Lee, Sunmi Choi, Seungjae Lee, Dong-Hwan Kim, Wook-Kee Han, Young-Ho Lim, Jae-Duk Lee, Jung-Dal Choi, and Kang-Deog Suh, "A 3.3 V 1 Gb multi-level NAND flash memory with non-uniform threshold voltage distribution", in *ISSCC Digest of Technical Papers*, 2001, pp.28-29.

[11] June Lee, Heung-Soo Im, Dae-Seok Byeon, Kyeong-Han Lee, Dong-Hyuk Chae, Kyong-Hwa Lee, Young-Ho Lim, Jung-Dal Choi, Young-Il Seo, Jong-Sik Lee, and Kang-Deog Suh, "A 1.8V 1Gb NAND flash memory with 0.12um STI process technology", in *ISSCC Digest of Technical Papers*, Feb., 2002, pp.80-81.

[12] S.Aritome, "Advanced flash memory technology and trends for file storage application", in *IEDM Tech. Dig.*, 2000, pp.763-766.

[13] K. Kobayashi, "Toshiba's NAND flash strategy", in *CEATEC 2004 Dig. Tech. Papers*, Oct. 2004.

[14] S. Lee, Y.T. Lee, W.K. Han, D.H. Kim, M.S. Kim, S.H. Moon, H. Cho, J.W. Lee, D.S.Byeon, Y.H. Lim, H.S. Kim, S.H. Hur, and K.D. Suh, "A 3.3V 4Gb four-level NAND flash memory with 90nm CMOS technology", in *ISSCC Digest of Technical Papers*, Feb., 2004, pp.52-53.

[15] T.Hara, K.Fukuda, K.Kanazawa, N.Shibata, K.Hosono, H.Maejima, M.Nakagawa, T.Abe, M.Kojima, M.Fujiu, Y.Takeuchi, K.Amemiya, M.Morooka, T.Kamai, H.Nasu, K.Kawano, C.M.Wang, K.Sakurai, N.Tokiwa, H.Waki, T.Maruyama, S.Yoshikawa, M.Higashitani, T.D.Pham and T.Watanabe, "A 146mm^2 8Gb NAND flash memory with 70nm COMS technology", in *ISSCC Digest of Technical Papers*, 2005, pp.44-45.

[16] Yong-Sik Yim, Kwang-Shik Shin, Sung-Hoi Hur, Jae-Duk Lee, Ihn-Gee Balk, Hong-Soo Kim, Soo-Jin Chai, Eun-Young Choi, Min-Cheol Park, Dong-Seok Eun, Sung-Bok Lee, Hye-Jin Lim, Sun-Pil Youn, Sung-Hun Lee, Tae-Jung Kim, Han-Soo Kim, Kyu-Charn Park and Ki-Nam Kim, "70nm NAND flash technology with 0.025 um^2 cell size for 4Gb flash memory", in *IEDM Tech. Dig.,* 2003, pp.819-822.

[17] A. Nozoe, T. Yamazaki, H. Sato, H. Kotani, S. Kubono, K.; Manita, T. Tanaka, T. Kawahara, M. Kato, K. Kimura, H. Kume, R. Hori, T. Nishimoto, S. Shukuri, A. Ohba, Y. Kouro, O. Sakamoto, A. Fukumoto and M. Nakajima, "A 3.3 V high-density AND flash memory with 1 ms/512B erase and program time", in *ISSCC Digest of Technical Papers*, 1995, pp.126-127.

[18] A. Nozoe, H. Kotani, T. Tsujikawa, K. Yoshida, K. Furusawa, M. Kato, T. Nishimoto, H. Kume, H. Kurata, N. Miyamoio, S. Kubono, I. Kanamitsu, K. Koda, T. Nakayama, Y. Kouro, A. Hosogane, N. Ajika, and K. Kobayashi, "A 256 Mb multilevel flash memory with 2 MB/s program rate for mass storage applications", in *ISSCC Digest of Technical Papers*, Feb., 1999, pp.110-111.

[19] T. Ishii, K. Oshima, H. Sato, S. Noda, J. Kishimoto, H. Kotani, A. Nozoe, K. Furusawa, T. Yoshitake, M. Kato, M. Takaheshi, A. Sato, S. Kubano, K. Manita, K. Koda, T. Nakayama and A. Hosogane, "A 126.6 mm^2 AND-type 512 Mb flash memory with 1.8 V power supply", in *ISSCC Digest of Technical Papers*, Feb., 2001, pp.30-31.

[20] K. Yoshida, O. Tsuchiya, Y. Yamaguchi, J. Kishimoto, Y. Ikeda, S. Narumi, Y. Takase, K. Furusawa, K. Izawa, T. Yoshitake, T. Kobayashi, H.Kurata and M. Kanemitsu, "A 1 Gb multilevel AG-AND-type flash memory with 10 MB/s programming throughput for mass storage application", in *ISSCC Digest of Technical Papers*, Feb., 2003, pp.288-289.

[21] H. Miwa, T. Tanaka, K.; Oshima, Y. Nakamura, T. Ishii, A. Ohba, Y. Kouro, T. Furukawa, Y. Ikeda, O. Tsuchiya, R. Hori and K. Miyazawa, "A 140 mm^2 64 Mb AND flash memory with a 0.4um technology", in *ISSCC Digest of Technical Papers*, Feb., 1996, pp.34-35.

[22] S. Kobayashi, M. Mihara, Y. Miyawaki, M. Ishii, T. Futatsuya, A. Hosogane, A. Ohba, Y. Terada, N. Ajika, Y. Kunori, K. Yuzuriha, M. Hatanaka, H. Miyoshi, T. Yoshihara, Y. Uji, A. Matsuo, Y. Taniguchi and Y. Kiguchi, "A 3.3 V-only 16 Mb DINOR flash memory", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.122-123.

[23] M. Mihara, Y. Miyawaki, O. Ishizaki, T. Hayasaka, K. Kobayashi, T. Omae, H. Kimura, S. Shimizu, H. Makimoto, Y. Kawajiri, M. Wada, H. Sonoyama and J. Etoh, "A 29 mm^2 1.8V-only 16Mb DINOR Flash Memory with Gate-Protected Poly-Diode (GPPD) Charge Pump," in *ISSCC Digest of Technical Papers*, Feb., 1999, pp.114-116.

[24] R. Cernea, G. Samachisa, C. Su, H. Tsai, Y. Kao, C. Wang, Y.; Chen, A. Renninger, T. Wong, J. Brennan, and J. Haines, "A 1 Mb flash EEPROM" in *ISSCC Digest of Technical Papers*, Feb., 1989, pp.138-139.

[25] R. Cernea, D.J. Lee, M. Mofidi, E.Y. Chang, Wy-Yi Chien, L. Goh, Yupin Fong, J.H. Yuan, G. Samachisa, D.C. Guterman, S. Mehrotra, K. Sato, H. Onishi, K. Ueda, F. Noro, K. Mijamoto, M. Morita, K. Umeda, and K. Kubo, "A 34 Mb 3.3 V serial flash EEPROM for solid-state disk applications", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.126-127.

[26] S. Mehrotra, J.H. Yuan, R.A. Cernea, W.Y. Chien, D.C. Guterman, G. Samachisa, R.D. Norman, M. Mofidi, W. Lee, Y. Fong, A. Mihnea, E. Harari, R.W. Gregor, E.P. Eberhardt, J.R. Radosevich, K.R. Stiles, R.A. Kohler, C.W. Leung and T.J. Mulrooney, "Serial 9Mb flash EEPROM for solid state disk applications", in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1992, pp.24-25.

[27] Y. Sofer, M. Edan, Y. Betser, M. Grossgold, E. Maayan and B. Eitan, "A 55 mm^2 256 Mb NROM flash memory with embedded microcontroller using an NROM-based program file", in *ISSCC Digest of Technical Papers*, Feb., 2004, pp.48-49.

[28] E. Maayan, R. Dvir, J. Shor, Y. Polansky, Y. Sofer, I. Bloom, D. Avni, B. Eitan, Z. Cohen, M. Meyassed, Y. Alpern, H. Palm, E.S. Kamienski, P. Haibach, D. Caspary, S. Riedel, and R. Knofler, "A 512 Mb NROM flash data storage memory with 8 MB/s data rate", in *ISSCC Digest of Technical Papers*, Feb., 2004, pp.100-101.

[29] J. Willer, C. Ludwig, J. Deppe, C. Kleint, S. Riedel, J. Sachse, M. Krause, R. Mikalo, E.S. Kamienski, S. Parascandola, T. Mikolajick, J. Fischer, M. Isler, K. Kuesters, I. Bloom, A. Shapir, E. Lusky, and B. Eitan, "110nm NROM technology for code and data flash products", in *Symp. VLSI Technology Dig. Tech. Papers*, June 2004, pp.76-77.

[30] T. Takeshima, H. Sugawara, H. Takada, Y. Hisamune, K. Kanamori, T. Okazawa, T. Murotani and I. Sasaki, "A 3.3 V single-power-supply 64 Mb flash memory with dynamic bit-line latch (DBL) programming scheme", in *ISSCC Digest of Technical Papers*, Feb., 1994, pp.148-149.

[31] M. Ohkawa, H.; Sugawara, N. Sudo, M. Tsukiji, K. Nakagawa, M. Kawata, K. Oyama, T. Takeshima and S. Ohya, "A 98 mm^2 3.3 V 64 Mb flash memory with FN-NOR type 4-level cell", in *ISSCC Digest of Technical Papers*, Feb., 1996, pp.376-379.

[32] A. Baker, R. Alexis, S. Bell, V. Dalvi, R. Durante, E. Baer, M. Fandrich, O. Jungroth, J. Kreifels, M. Landgraf, K. Lee, H. Pon, M. Rashid, R. Rozman, J. Tsang, K. Underwood, C. Yarlagadda, "A 3.3 V 16 Mb flash memory with advanced write automation", in *ISSCC Digest of Technical Papers*, Feb., 1994, pp.146-147.

[33] T. Jinbo, H. Nakata, K. Hashimoto, T. Watanabe, K. Ninomiya, T. Urai, M. Koike, T. Sato, N. Kodama, K. Oyamaand T. Okazawa, "A 5 V-only 16 Mb flash memory with sector-erase mode", in *ISSCC Digest of Technical Papers*, Feb., 1992, pp.154-155.

[34] K. Seki, H. Kume, Y. Ohji, T. Tanaka, T. Adachi, M. Ushiyama, K. Shimohigashi, T. Wada, K. Komori, T. Nishimoto, K. Izawa, T. Hagiwara, Y. Kubota, K. Shohji, N. Miyamoto, S. Saeki and N. Ogawa, "An 80 ns 1 Mb flash memory with on-chip erase/erase-verify controller", in *ISSCC Digest of Technical Papers*, Feb., 1990, pp.60-61.

[35] D. Mills, M. Bauer, A. Bashir, R. Fackenthal, K. Frary, T. Gullard, C. Haid, J. Javanifard, P. Kwong, D. Leak, S. Pudar, M. Rashid, R. Rozman, S. Sambandan, S.; Sweha and J. Tsang, "A 3.3V 50MHz synchronous 16Mb flash memory", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.120-121.

[36] M. Bauer, R. Alexis, G. Atwood, B. Baltar, A. Fazio, K. Frary, M. Hensel, M. Ishac, J. Javanifard, M. Landgraf, D. Leak, K. Loe, D. Mills, P. Ruby, R. Rozman, S. Sweha, S. Talreja and K. Wojciechowski, "A multilevel-cell 32 Mb flash memory", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.132-133.

[37] S. Atsumi, A. Umezawa, M. Kuriyama, N. Banba, N. Ohtsuka, N. Tomita, Y. Iyama, T. Miyaba, R. Sudoh, E. Kamiya, M. Tanimoto, Y. Hiura, Y. Araki, E. Sakagami, N. Arai, S. Mori, "A 3.3 V-only 16 Mb flash memory with row-decoding scheme", in *ISSCC Digest of Technical Papers*, Feb., 1996, pp.42-43.

[38] G. Campardo, R. Micheloni, S. Commodaro, E. Yero, M. Zammabio, S. Mognoni, A. Sacco, M. Picca, A. Manstleita, M. Scotti, I. Motta, C. Golla, A. Pierin, A. Ohba, T. Fulatsuya, R.; Makabe, S. Kawai, Y. Kai, S. Shimizu, T. Ohnakado, I. Sugihara, R. Bez, A. Grossi, A. Modelli, P. Khouri, and G. Torelli, "A 40 mm^2 3 V 50 MHz 64 Mb 4-level cell

NOR-type flash memory", in *ISSCC Digest of Technical Papers*, Feb., 2000, pp.274-275.

[39] S. Atsumi, A. Umezawa, T. Tanzawa, T. Taura, H.; Shiga, Y. Takano, T. Miyaba, M. Matsui, H. Watanabe, K. Isobe, S. Kitamura, S. Yamada, M. Saito, S. Mori, and T. Watanabe, "A channel-erasing 1.8 V-only 32 Mb NOR flash EEPROM with a bit-line direct-sensing scheme", in *ISSCC Digest of Technical Papers*, Feb., 2000, pp.276-277.

[40] B. Pathak, A. Cabrera, G. Christensen, A. Darwish, M. Goldman, R. Haque, J. Jorgensen, R. Kajley, T. Ly, F. Marvin, S. Monasa, Q. Nguyen, D. Pierce, A. Sendrowski, I. Sharif, H. Shimoyoshi, A. Smidt, R. Sundaram, M. Taub, W. Tran, R. Trivedi, P. Walimbe and E. Yu, "A 1.8 V 64 Mb 100 MHz flexible read while write flash memory", in *ISSCC Digest of Technical Papers*, Feb., 2001, pp.32-33.

[41] T.Tanzawa, A. Umezawa, T. Taura, H. Shiga, T. Hara, Y. Takano, T. Miyaba, N. Tokiwa, K. Watanabe, H. Watanabe, K. Masuda, K. Naruke, H. Kato, S. Atsumi, "A 44 mm^2 4-bank 8-word page read 64 Mb flash memory with flexible block redundancy and fast accurate word-line voltage controller", in *ISSCC Digest of Technical Papers*, Feb., 2002, pp.102-103.

[42] H. Stiegler, B. Ashmore, R. Bussey, M. Gill, S. Lin, M. McConnell, D. McElroy, J. Schreck, P. Shah, P. Truong, A.L. Esquivel, J. Paterson and B. Riemenschneider, "A 4Mb 5V-only flash EEPROM with sector erase", in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1990, pp.103-104.

[43] S.C. Tsao, J.E. Frayer, C.S. Pang, Y. Ma, K. Kwon Y. Choi, D. Kim, J. Kim, and J. Park, "A 5V-only 16M flash memory using a contactless array of source-side injection cells", in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1995, pp.77-78.

[44] J. C. Chen, T. H. Kuo, L. E. Cleveland, C. K. Chung, N. Leong, Y. K. Kim, T. Akaogi and Y. Kasa, "A 2.7V only 8Mbx16 NOR Flash Memory", in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1996, pp.172-173.

[45] S. Aritome, R. Shirota, R. Kirisawa, T. Endoh, R. Nakayama, K. Sakui, and F. Masuoka, "A reliable bi-polarity write/erase technology in flash EEPROMs", in *IEDM Tech. Dig.,* 1990, pp.111-114.

[46] S. Satoh, T. Nakamura, K. Shimizu, K. Takeuchi, H. Iizuka, S. Aritome, and R. Shirota, "A novel gate-offset NAND cell (GOC-NAND) technology suitable for high-density and low-voltage-operation flash memories", in *IEDM Tech. Dig.,* 1999, pp.271-274.

[47] A. Goda, W. Moriyama, H. Hazama, H. Iizuka, K. Shimizu, S. Aritome, and R. Shirota, "A novel surface-oxidized barrier-SiN cell technology to improve endurance," in *IEDM Tech. Dig.,* 2000, pp.771-774.

[48] R. Kirisawa, S. Aritome, R. Nakayama, T. Endoh, R. Shirota and F. Masuoka, "A NAND structured cell with a new programming technology for highly reliable 5 V-only flash EEPROM," in *Symp. VLSI Technology Dig. Tech. Papers*, June 1990, pp.129-130.

[49] S. Shuto, S. Yamada, S. Aritome, T. Watanabe and K. Hashimoto, "Read disturb degradation mechanism for source erase flash memories," in *Symp. VLSI Technology*

Dig. Tech. Papers, June 1996, pp.242-243.

[50] E.C. Yang, L. Cheng-Jye, C. Tien-Sheng, L. Ming-Chi, and C.C. Hsu, "Novel Bi-directional tunneling program/erase NOR (BiNOR) type flash EEPROM," in *Symp. VLSI Technology Dig. Tech. Papers*, June 1999, pp.207-210.

[51] K.Takeuchi, T.Tanaka and H.Nakamura, "A double-level-Vth select gate array architecture for multi-level NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1995, pp.69-70.

[52] K.Takeuchi, T.Tanaka and T.Tazawa, "A multi-page cell architecture for high-speed programming multi-level NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1997, pp.67-68.

[53] K. Takeuchi, S. Satoh, T. Tanaka, K. Imamiya, and K. Sakui, " a negative Vth cell architecture for highly scalable, excellently noise immune and highly reliable NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1998, pp.234-235.

[54] K. Takeuchi, S. Satoh, K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, T. Ikehashi, K. Kanda, K. Hosono and K. Sakui, "a source-line programming scheme for low voltage operation NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1999, pp.37-38.

[55] K. Takeuchi and T. Tanaka, "A dual page programming scheme for high-speed multi-Gb-scale NAND flash memories," in Symp. *VLSI Circuits Dig. Tech. Papers*, June 2000, pp.156-157.

[56] T. Kobayashi, Y. Sasago, H. Kurata, S. Saeki, Y. Goto, T. Arigane, Y. Okuyama, H. Kume, and K. Kimura, " A giga-scale assist-gate (AG)-AND-type flash memory cell with 20-MB/s programming throughput for content-downloading applications," in *IEDM Tech. Dig.,* 2001, pp.29-32.

[57] Y. Sasago, T. Arigane, H. Kurata, S. Saeki, Y. Goto, S. Kamohara, T. Kobayashi, and H. Kume, " 10-MB/s multi-level programming of Gb-scale flash memory enabled by new AG-AND cell technology", in *IEDM Tech. Dig.,* 2002, pp.952-954.

[58] Y. Sasago, H. Kurata, T. Arigane, K. Otsuga, T. Kobayashi, Y. Ikeda, T. Fukumura, S. Narumi, A. Sato, T. Terauchi, M. Shimizu, S. Noda, K. Kozakai, O. Tsuchiya and K. Furusawa," 90-nm-node multi-level AG-AND type flash memory with cell size of true $2F^2$/bit and programming throughput of 10 MB/s", in *IEDM Tech. Dig.,* 2003, pp.824.826.

[59] T.Tanaka, T.Tanazawa and K.Takeuchi, "A 3.4-Mbyte/sec Programming 3-Level NAND Flash Memory Saving 40% Die Size Per Bit", in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1997, pp.65-66.

[60] G.J.Hemink, T.Tanaka, T.Endoh, S.Aritome and R.Shirota, "Fast and Accurate Programming Method for Multi-level NAND Flash EEPROMs," *Symp. VLSI Technology Dig. Tech. Papers*, June 1995, pp.129-130.

[61] K.Takeuchi, T.Tanaka and H.Nakamura, "a double-level-Vth select gate array architecture for multi-level NAND flash memories", *IEEE Journal of Solid-State Circuits*,

vol.31, no.4, April 1996, pp.602-609.

[62] K.Takeuchi, T.Tanaka and T.Tazawa, “A multipage cell architecture for high-speed programming multi-level NAND flash memories”, *IEEE Journal of Solid-State Circuits*, vol.33, no.8, August 1998, pp.1228-1238.

[63] K. Takeuchi, S. Satoh, K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, T. Ikehashi, K. Kanda, K. Hosono and K. Sakui, “A source-line programming scheme for low voltage operation NAND flash memories,” *IEEE Journal of Solid-State Circuits*, vol.35 no.5, May 2000, pp.672-681.

[64] K. Takeuchi, S. Satoh, T. Tanaka, K. Imamiya, and K. Sakui, " A negative Vth cell architecture for highly scalable, excellently noise immune and highly reliable NAND flash memories," *IEEE Journal of Solid-State Circuits*, vol.34, no.5, May 1999, pp.675-684.

[65] K. Takeuchi and T. Tanaka, “A dual page programming scheme for high-speed multi-Gb-scale NAND flash memories,” *IEEE Journal of Solid-State Circuits*, vol.36 no.5, May 2001, pp.744-751.

CHAPTER 2
PRINCIPLES OF HIGH-SPEED DESIGN

2.1 Analysis of program operation

2.1.1 Bit by bit program verify scheme

Fig.2.1 shows a memory block in a conventional NAND-type memory cell array [1-3]. The memory block is composed of a number of NAND cell units, for example 4k units which share 16 word-lines (CG1~CG16), two select gates (SG1,SG2) and a common diffused source-line. Each unit is connected to a bit-line control circuit via a bit-line (BL0~BL2). All memory cells, which share the same control gates are simultaneously read and programmed by using the page buffer as shown in Fig.2.2. As shown in Fig.2.3, all memory cells connected to the same word-lines compose a page and are programmed at the same time.

The threshold voltages of the memory cells are determined in the following manner [4], [5]. After the bit-lines are precharged to a high level, for example 1.3V, a reference voltage, e.g. 0.5V shown in Fig.2.4 is applied to the selected word-line, while the select gates are raised to Vread and the un-selected control gates are raised to a read voltage, Vread as shown in

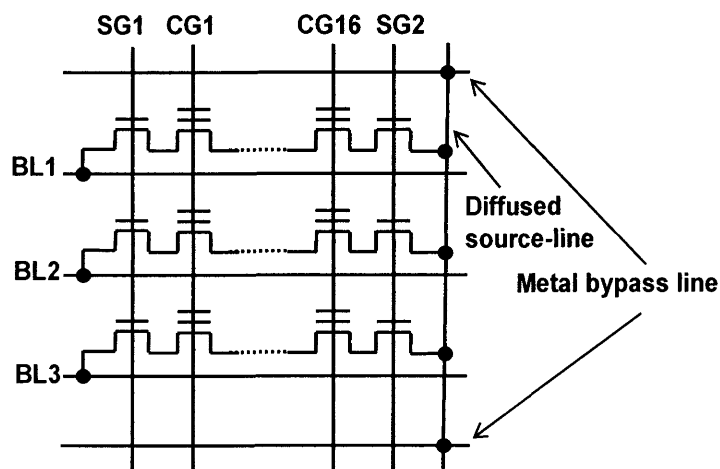


Fig.2.1 NAND flash memory cell array

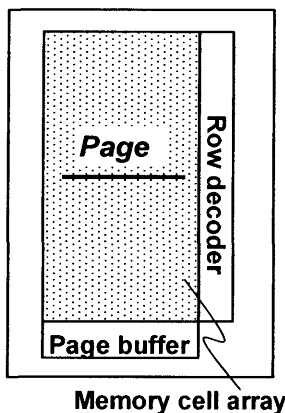


Fig.2.2 Page programming

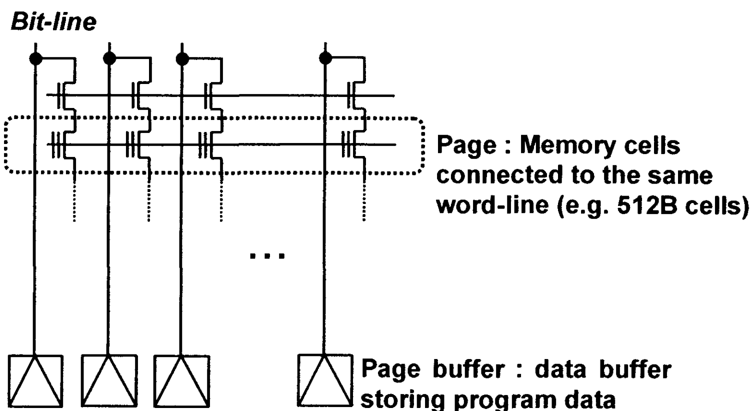


Fig.2.3 Memory cell array and core circuits

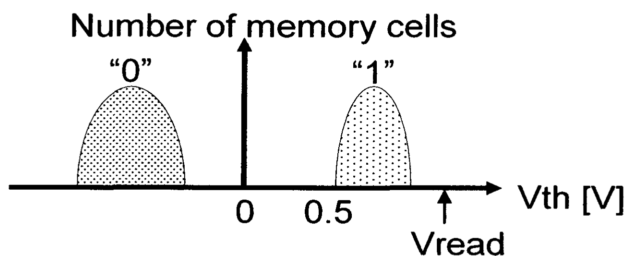


Fig.2.4 Vth distribution

threshold voltage of the corresponding memory cell is determined to be lower than the reference voltage, 0.5V. On the other hand, if the bit-line voltage stays above 1V, the threshold voltage of the memory cell is determined to be higher than the reference voltage, 0.5V.

In a bit-by-bit program operation [6], [7], program verify, which monitors the threshold voltage of the memory cells is carried out after each programming pulse. If the threshold voltage reaches the desired level, additional programming of the corresponding memory cells is inhibited. Programming and program verify are repeated until all selected memory cells are sufficiently programmed. Therefore, by using this program scheme, in principle, the threshold voltage of the memory cells can be controlled precisely.

2.1.2 Incremental program voltage scheme

As shown in Fig. 2.7, during the operation of the program, Vpgm such as 18V is applied to the word-line, while the channel of the cell is grounded. The staircase program pulse is shown in Fig.2.8 [8], [9] where the verify steps are carried out after each pulse.

2.1.3 Expression of program throughput

In principle, the Vth distribution, ⊿Vth shown in Fig.2.9, can be narrowed down to ⊿Vpgm [9]. In fact, an array noise, such as a source line noise [10] or an inter bit-line coupling noise [11], increases ⊿Vth by ⊿Vnoise. In addition, the capacitive coupling between floating gates increases the Vth distribution by ⊿Vfg [12]. As a result, ⊿Vth increases to ⊿Vpgm+⊿Vnoise+⊿Vfg [13], [14]. ⊿Vth is decided by the reliability characteristics of the memory cell. Suppose that Vth of "3" state in Fig.2.9 should be less than 3.1V because if the Vth is higher than 3.1V, the electric field across the tunnel oxide becomes higher and as a result the data retention characteristics becomes unacceptably worse. Considering that each level should be separated by 0.6V, ⊿Vth should be 0.65V. As a result, ⊿Vpgm should be ⊿Vth-⊿Vnoise-⊿Vfg, that is, 0.65-⊿Vnoise-⊿Vfg.

Fig.2.10 allows estimation of the number of program pulses, Npulse. As the number of

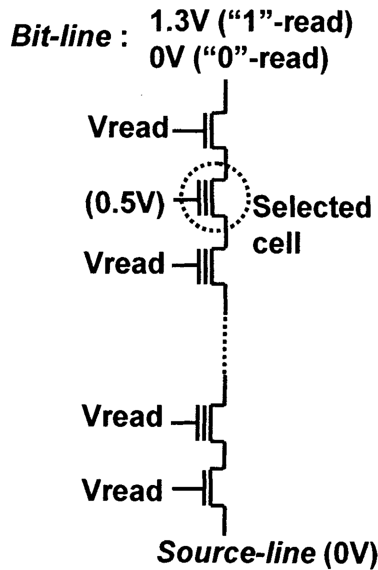


Fig.2.5 Read operation

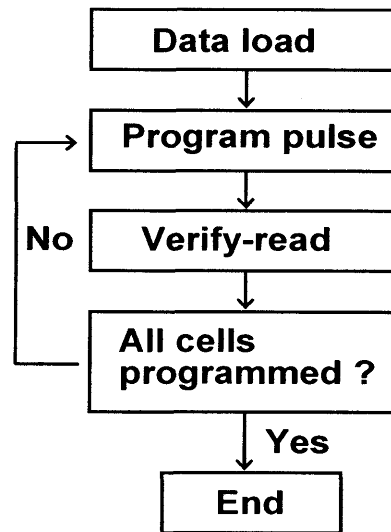


Fig.2.6 Program algorithm

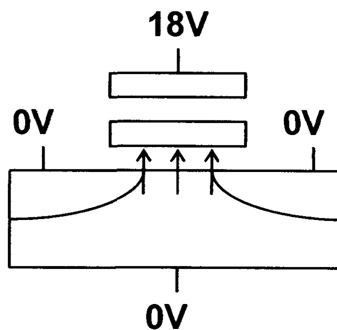


Fig.2.7 FN tunnel injection program

program pulses is decided by the program characteristics of the slowest cell, the characteristics of the slowest cell is described below. At the first program pulse, the fastest cell is sufficiently programmed and the Vth of the slowest cell is ⊿Vth0 lower than the verify voltage (0.5V) [13], [14]. Then, the Vth change of the slowest cell caused by each program pulse is constant with a value of ⊿Vpgm. Thus, ⊿Vth0/ ⊿Vpgm cycles are necessary to program the slowest cell [7], [8]. Consequently, the number of program pulses, Npulse is expressed as follows.

$$\Delta Vpgm = \Delta Vth - \Delta Vnoise - \Delta Vfg$$

$$Npulse = \frac{\Delta Vth0}{\Delta Vpgm}$$

Then, the number of pulses of the single-level cell is shown above. In case of four-level cell, three states, that is, "1"-,"2"- and "3"-states should be programmed. Therefore, the total program pulses is three times larger than that of the single-level cell [13-17]. By using the simultaneously written states, Nstate, the total program time, Tprog is expressed as follows [13], [14].

$$Tprog = (Tpulse + Tvfy) \times Npulse \times Nstate$$

$$= (Tpulse + Tvfy) \times \frac{\Delta Vth0}{\Delta Vpgm} \times Nstate$$

$$= (Tpulse + Tvfy) \times \frac{\Delta Vth0}{\Delta Vth - \Delta Vnoise - \Delta Vfg} \times Nstate$$

Nstate = 1 for 2-level cell, 3 for 4-level cell

Tvfy is the verify read time and Tpulse is the program pulse width as shown in Fig.2.8. Finally, the program throughput is expressed as a function of Tprog and the page size are as follows.

$$\text{Pr}\,ogram\ \ througput = \frac{Page\ \ size}{Tprog}$$

$$Tprog = (Tpulse + Tvfy) \times \frac{\Delta Vth0}{\Delta Vth - \Delta Vnoise - \Delta Vfg} \times Nstate$$

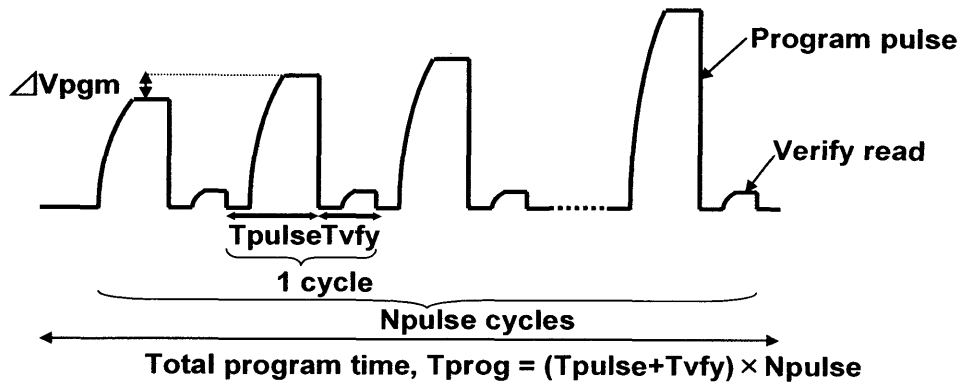


Fig.2.8 Word-line waveform during programming

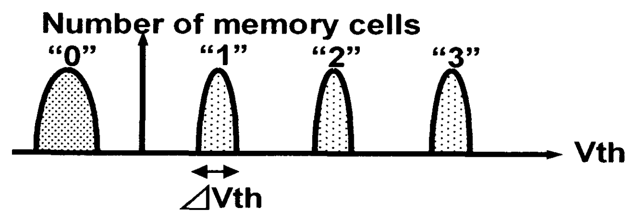


Fig.2.9 Vth distribution of the multi-level cell

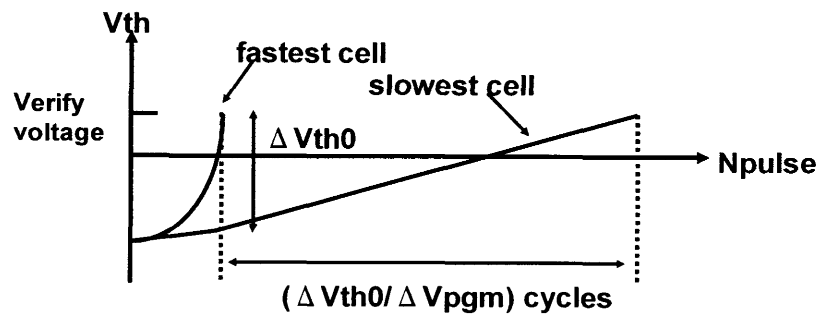


Fig.2.10 Program characteristics

2.2 Guidelines for high-speed design

Guidelines for high-speed program can be derived from the expression of the program throughput. Table 2.1 summarizes the high-speed techniques and the methodology. Following five methods are effective to increase the program throughput.

1) Reduce simultaneously written states, Nstate
2) Decrease program pulse width, Tpuse
3) Decrease the verify read time, Tvfy
4) Reduce program cycles, Npulse
5) Increase page size

Which method to adopt depends on the markets, that is, 3V multi-level cell, low-voltage multi-level cell and 3V single-level cell. Table 2.2 summarizes the problems, guidelines and solutions for high-speed programming in the 3V multi-level cell, the low-voltage multi-level cell and the 3V single-level cell.

A. 3V multi-level cell

The problems are:

1) Simultaneously written states, Nstate is large [13], [17]
2) Circuit noise, ⊿Vnoise drastically increases the program time [10,11, 18, 19]

As shown in Fig.2.11, in the multi-level cell, the target Vth distribution, ⊿Vth is much

<table>
<tr><th colspan="3">High speed techniques</th><th>Methodology</th></tr>
<tr><td colspan="3">Reduce simultaneously written states, Nstate</td><td>Multipage cell technology (Chapter 3)</td></tr>
<tr><td rowspan="3">Decrease program pulse width, Tpulse</td><td colspan="2">Speed up bit-line/source-line ramp up</td><td>Low load capacitance technology (Chapter 4)</td></tr>
<tr><td colspan="2" rowspan="2">Speed up word-line ramp up
(Decrease RC delay time of word-line)</td><td>New low-R material for word-line
New low-k material for ILD</td></tr>
<tr><td>Hierarchical row decoder</td></tr>
<tr><td rowspan="3">Decrease verify read time, Tvfy</td><td colspan="2" rowspan="2">Speed up word-line ramp up
(Decrease RC delay time of word-line)</td><td>New low-R material for word-line
New low-k material for ILD</td></tr>
<tr><td>Hierarchical row decoder</td></tr>
<tr><td colspan="2">Decrease bit-line capacitance</td><td>New low-R material for bit-line
New low-k material for ILD</td></tr>
<tr><td rowspan="5">Reduce program cycles, Npulse</td><td rowspan="2">Increase ⊿Vpgm</td><td>Decrease noise, ⊿Vnoise</td><td>Low noise technology (Chapter 5)</td></tr>
<tr><td>Decrease FG-FG coupling, ⊿Vfg</td><td>New cell structure such as thin FG</td></tr>
<tr><td colspan="2">Decrease Vth distribution width w/o verify, ⊿Vth0</td><td>Reduce variation of cell coupling ratio</td></tr>
<tr><td colspan="2" rowspan="2">Increase Vth distribution width, ⊿Vth, which makes worse the cell reliability</td><td>Improve reliability by fixing quality of tunnel oxide</td></tr>
<tr><td>Improve reliability by introducing more efficient ECC</td></tr>
<tr><td colspan="3">Increase page size</td><td>Parallel write technology (Chapter 6)</td></tr>
</table>

Table 2.1 High speed technologies

narrower than that of the single-level cell. Because of the noise ⊿Vnoise, ⊿Vpgm drastically decreases, which leads to increase the program cycles, Npulse and as a result increases the programming time, Tprog. Therefore, the guidelines for the high-speed program in the 3V multi-level cell are as follows.

1) Reduce simultaneously written states, Nstate [13], [14].
2) Reduce program cycles, Npulse by decreasing the circuit noise, ⊿Vnoise [10, 11, 18, 19].

Based on the guidelines above, following two circuit techniques are proposed as a solution.

1) The multipage cell architecture reduces the simultaneously written states, Nstate (Chapter 3).
2) The low noise technology decreases the circuit noise (Chapter 5).

B. Low-voltage multi-level cell

As for the problems, in addition to two problems mentioned in the 3V multi-level cell, the program time drastically increases as shown in Fig.2.12 because of the long bit-line ramp up [20], [21]. Therefore, the problems of the low-voltage multi-level cell are as follows.

1) Simultaneously written states, Nstate is large [13], [14].
2) Circuit noise, ⊿Vnoise drastically increases the programming time [10, 11, 18, 19].
3) Tpulse is large because of the long bit-line ramp up time [20], [21].

The guidelines for high-speed program of the low-voltage multi-level cell are as follows.

1) Reduce simultaneously written states, Nstate [13], [14].
2) Reduce program cycles, Npulse by decreasing the circuit noise, ⊿Vnoise [10, 11, 18, 19].
3) Reduce the program pulse width by decreasing the core load capacitance [20, 21].

As a solution, three circuit technologies are proposed as follows.

1) The multipage cell architecture reduces the simultaneously written states, Nstate (Chapter 3).
2) The low noise technology decreases the circuit noise (Chapter 5).
3) The low load capacitance technology decreases the program pulse width (Chapter 4).

C. 3V single-level cell

As for the 3V single-level cell, the simultaneously written states, Nstate is one and cannot be decreased. Moreover, as shown in Fig. 2.11, reducing the circuit noise is not as effective as the multi-level cell. Therefore, so as to increase the program throughput, the page size should be increased. However, as shown in Fig. 2.13, the additional page buffer increases the chip size by 5% [22], [23]. Hence, the problem of single-level cell is as follows.

1) Additional page buffers increase the chip area [22], [23].

The guidelines for high-speed program in the 3Vsingle-level cell are as follows.

1) Reduce the area overhead caused by the page size increase [22], [23].

Following circuit technology is proposed as a solution.

1) The parallel write technology increase the page size without circuit area overhead (Chapter 6).

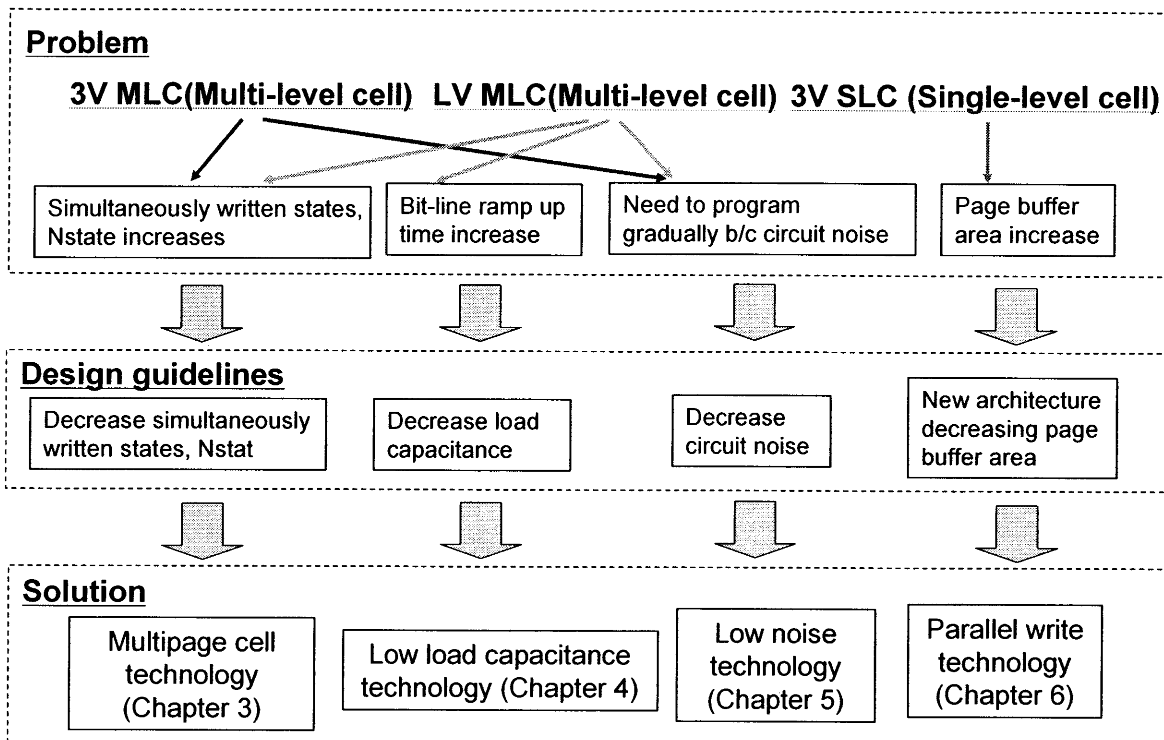


Table 2.2 Summary of problems, guidelines and solutions for the high-speed design.

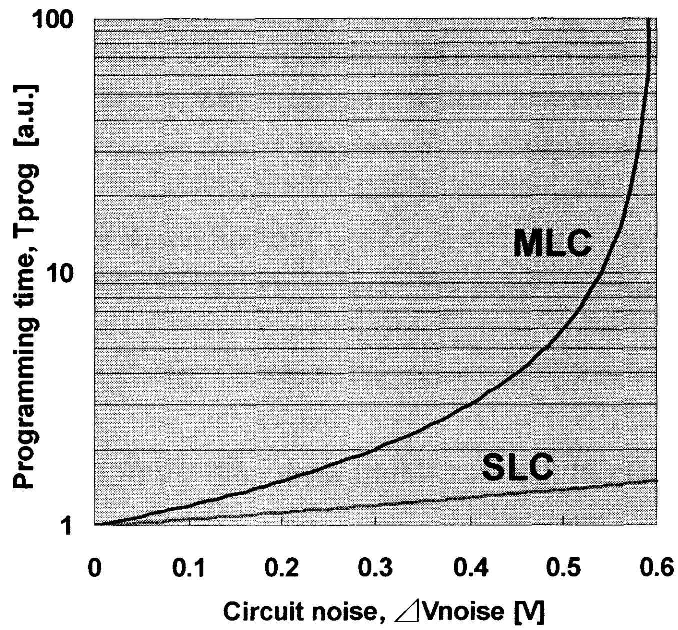


Fig.2.11 Sensitivity analysis of the circuit noise on the programming time

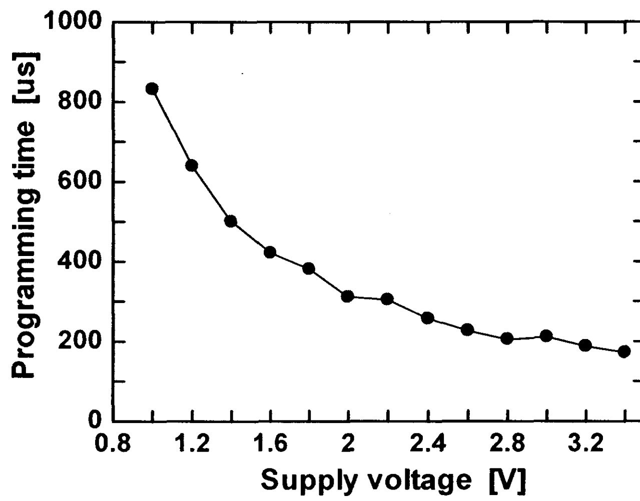
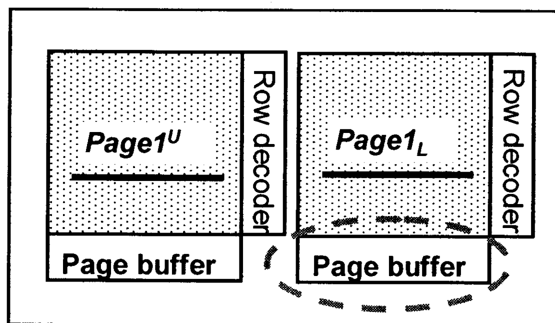
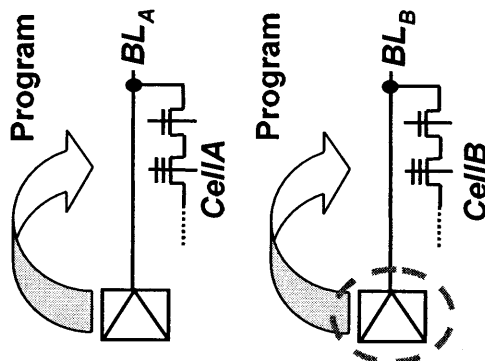


Fig.2.12 Program speed degradation at low Vcc





Additional page buffer

Fig.2.13 Chip size increase in case that the page size is doubled.

2.3 Summary

Basic principles of the program operation are described. Furthermore, an analytical framework of the program throughput is presented. The program speed is expressed as a function of key parameters, such as the simultaneously written states, the program pulse width, the verify read time, the circuit noise and the page size. By using the framework, the problems, the guidelines and the solutions for the high-speed design are summarized in Table 2.2. In the case of 3V multi-level cell, the key guidelines are 1) to reduce simultaneously written states, Nstate as well as 2) to reduce program cycles, Npulse by decreasing the circuit noise, ⊿Vnoise. As for the low-voltage multi-level cell, guidelines are 1) to reduce simultaneously written states, Nstate, 2) to reduce program cycles, Npulse by decreasing the circuit noise, ⊿Vnoise and 3) to reduce the program pulse width by decreasing the core load capacitance. Finally, the guideline of the 3V single-level cell is to 1) to reduce the area overhead caused by the page size increase.

References

[1] F. Masuoka, M. Momodomi, Y. Iwata and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structured cell," in *IEDM Tech. Dig.,* 1987, pp.552-555.

[2] K. Imamiya, Y. Iwata, Y. Sugiura, H. Nakamura, H. Oodaira, M. Momodomi, Y. Ito, T. Watanabe, H. Araki, K. Narita, K. Masuda and J. Miyamoto, "A 35 ns-cycle-time 3.3 V-only 32 Mb NAND flash EEPROM", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.32-33.

[3] Kang-Deog Suh, Byung-Hoon Suh, Young-Ho Um, Jin-Ki Kim, Young-Joon Choi, Yong-Nam Koh, Sung-Soo Lee, Suk-Chon Kwon, Byung-Soon Choi, Jin-Sun Yum, Jung-Hyuk Choi, Jang-Rae Kim and Hyung-Kyu Lim, "A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.128-129.

[4] K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, K. Takeuchi, T. Ikehashi, K. Kanda, K. Hosono, R. Shirota, S. Aritome, K. Shimizu, K. Hatakeyama and K. Sakui " A 130mm^2 256Mb NAND Flash with Shallow Trench Isolation Technology", *ISSCC Digest of Technical Papers*, pp.112-113, 1999.

[5] H.Nakamura, K.Imamiya, T.Himeno, T.Yamamura, T.Ikehashi, K.Takeuchi, K.Kanda, K.Hosono, T.Futatsuyama, K.Kawai, R.Shirota, N.Arai, Fumitaka Arai, K.Hatakeyama, H.Hazama, Masanobu Saito, H.Meguro, K.Conley, K. Quader and J.Chen, "A 125mm^2 1Gb NAND Flash Memory with 10MB/s Program Throughput," *ISSCC Digest of Technical Papers*, 2002, pp.106-107.

[6] T.Tanaka, M.Momodomi, Y.Iwata, Y.Tanaka, H.Oodaira, Y.Itoh, R.Shirota, K.Ohuchi and F.Masuoka,"A 4-Mbit NAND-EEPROM with Tight Programmed Vt Distribution," *Symp. VLSI Circuits Dig. Tech. Papers*, June 1990, pp.105-106.

[7] T.Tanaka, Y.Tanaka, H.Nakamura, H.Oodaira, S.Aritome, R.Shirota and F.Masuoka, "A quick intelligent program architecture for 3V-only NAND EEPROMs," *Symp. VLSI Circuits Dig. Tech. Papers*, June 1992, pp.20-21.

[8] K.D.Suh, B.H.Suh, Y.H.Lim, Y.J.Choi, Y.N.Koh, S.S.Lee, S.C.Kwon, B.S.Choi, J.S.Yum, J.H.Choi, J.R.Kim and H.K.Lim, "A 3.3V 32Mb NAND flash memory with incremental step pulse programming scheme," in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.128-129.

[9] G.J.Hemink, T.Tanaka, T.Endoh, S.Aritome and R.Shirota, "Fast and accurate programming method for multi-level NAND flash EEPROM's", in Symp. *VLSI Technology Dig. Tech. Papers*, June 1995, pp.129-130.

[10] K.Takeuchi, T.Tanaka and H.Nakamura, "A double-level-Vth select gate array architecture for multi-level NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1995, pp.69-70.

[11] K.Takeuchi, T.Tanaka and H.Nakamura, "A double-level-Vth select gate array

architecture for multi-level NAND flash memories", *IEEE Journal of Solid-State Circuits*, vol.31, no.4, April 1996, pp.602-609.

[12] J.D.Lee, S.H.Hur, Y.S.Yim, S.S.Cho, H.S.Kim, K.J.Joo, T.Cho, Y.T.Lee, J.D.Choi and M.Y.Lee, "Effects of parasitic capacitance on NAND flash memory cell operation," in *NVSMW Tech. Dig.*, Aug. 2001, pp.90-92.

[13] K.Takeuchi, T.Tanaka and T.Tazawa, "A multi-page cell architecture for high-speed programming multi-level NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1997, pp.67-68.

[14] K.Takeuchi, T.Tanaka and T.Tazawa, "A multipage cell architecture for high-speed programming multi-level NAND flash memories", *IEEE Journal of Solid-State Circuits*, vol.33, no.8, August 1998, pp.1228-1238.

[15] H. Nobukata, S. Takagi, K. Hiraga, T. Ohgishi, M. Miyashita, K. Kamimura, S. Hiramatsu, K. Sakai, T. Ishida, H. Arakawa, M. Itoh, I. Naiki and M. Noda, "A 144 Mb 8-level NAND flash memory with optimized pulse width programming," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1999, pp.39-40.

[16] Tae-Sung Jung, Young-Joon Choi, Kang-Deog Suh, Byung-Hoon Suh, Jin-Ki Kim, Young-Ho Lim, Yong-Nam Koh, Jong-Wook Park, Ki-Jong Lee, Jung-Hoon Park, Kee-Tae Park, Jang-Rae Kim, Jeong-Hyong Lee and Hyung-Kyu Lim, "A 3.3 V 128 Mb multi-level NAND flash memory for mass storage application", in *ISSCC Digest of Technical Papers*, Feb., 1996, pp.32-33.

[17] T.Tanaka, T.Tanazawa and K.Takeuchi, "A 3.4-Mbyte/sec Programming 3-Level NAND Flash Memory Saving 40% Die Size Per Bit", in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1997, pp.65-66.

[18] K. Takeuchi, S. Satoh, T. Tanaka, K. Imamiya, and K. Sakui, " A negative Vth cell architecture for highly scalable, excellently noise immune and highly reliable NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1998, pp.234-235.

[19] K. Takeuchi, S. Satoh, T. Tanaka, K. Imamiya, and K. Sakui, " A negative Vth cell architecture for highly scalable, excellently noise immune and highly reliable NAND flash memories," *IEEE Journal of Solid-State Circuits*, vol.34, no.5, May 1999, pp.675-684.

[20] K. Takeuchi, S. Satoh, K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, T. Ikehashi, K. Kanda, K. Hosono and K. Sakui, "A source-line programming scheme for low voltage operation NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1999, pp.37-38.

[21] K. Takeuchi, S. Satoh, K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, T. Ikehashi, K. Kanda, K. Hosono and K. Sakui, "A source-line programming scheme for low voltage operation NAND flash memories," *IEEE Journal of Solid-State Circuits*, vol.35 no.5, May 2000, pp.672-681.

[22] K. Takeuchi and T. Tanaka, "A dual page programming scheme for high-speed multi-Gb-scale NAND flash memories," in Symp. *VLSI Circuits Dig. Tech. Papers*, June

2000, pp.156-157.

[23] K. Takeuchi and T. Tanaka, "A dual page programming scheme for high-speed multi-Gb-scale NAND flash memories," *IEEE Journal of Solid-State Circuits*, vol.36 no.5, May 2001, pp.744-751.

[24] Jin-Gyun Kim, Jae-Young Ahn, Hong-Suk Kim, Ju-Wan Lim, Chae-Ho Kim, Shu, H., Hasebe, K., Sung-Hoi Hur, Jong-Ho Park, Hee-Seok Kim, Yu-Gyun Shin, U-In Chung and Joo-Tae Moon, "A highly manufacturable low-k ALD-SiBN process for 60nmNAND flash devices and beyond", in *IEDM Tech. Dig.,* 2004, pp.1063-1066.

[25] Yong-Sik Yim, Kwang-Shik Shin, Sung-Hoi Hur, Jae-Duk Lee, Ihn-Gee Balk, Hong-Soo Kim, Soo-Jin Chai, Eun-Young Choi, Min-Cheol Park, Dong-Seok Eun, Sung-Bok Lee, Hye-Jin Lim, Sun-Pil Youn, Sung-Hun Lee, Tae-Jung Kim, Han-Soo Kim, Kyu-Charn Park and Ki-Nam Kim, "nm NAND flash technology with 0.025um^2/cell size for 4Gb flash memory", in *IEDM Tech. Dig.,* 2004, pp.827-830.

[26] Jung-Dal Choi, Seong-Soon Cho, Yong-Sik Yim, Jae-Duk Lee, Hong-Soo Kim, Kyung-Joong Joo, Sung-Hoi Hur, Heung-Soo Im, Joon Kim, Jeong-Woo Lee, Kang-Ill Seo, Man-Sug Kang, Kyung-Hyun Kim, Jeong-Lim Nam, Kyu-Charn Park and Moon-Yong Lee, "Highly manufacturable 1 Gb NAND flash using 0.12um process technology" in *IEDM Tech. Dig.,* 2001, pp.25-28.

[27] K. Kobayashi, "Toshiba's NAND flash strategy", in *CEATEC 2004 Dig. Tech. Papers*, Oct. 2004.

[28] F. Arai, N. Arai, S. Satoh, T. Yaegashi, E. Kamiya, Y. Matsunaga, Y. Takeuchi, H. Kamata, A. Shimizu, N. Ohtami, N. Kai, S. Takahashi, W. Moriyama, K. Kugimiya, S. Miyazaki, T. Hirose, H. Meguro, K. Hatakeyama, K. Shimizu and R. Shirota, "High-density (4.4F^2) NAND flash technology using Super-Shallow Channel Profile (SSCP) engineering", in *IEDM Tech. Dig.,* 2000, pp.775-778.

CHAPTER 3
MULTIPAGE CELL TECHNOLOGY
Acceleration by reducing simultaneously written states, Nstate

3.1 Introduction

The major design problem in the multi-level cell is that the simultaneously written states, Nstate is three times larger than that of the single-level cell. Consequently, the program cycle is three times larger and therefore the total programming time is three times larger than that of the single-level cell

As a solution to this problem, a simultaneous multi-level program has been developed [1]. In this method, the threshold voltage during programming is controlled by the drain voltage under a constant control gate condition. Therefore, the memory cells on a word-line can be programmed to three different threshold voltages during the same programming period by changing the bit-line voltage. Then, the three programmed levels could be verified cell by cell. However, in this chapter it is shown that when the simultaneous multi-level program is used in NAND flash memories, the program speed will be drastically degraded. One reason is that the number of verify sequences is increased to three in a four-level cell system in comparison with just one sequence in a two-level cell system. The other is that the bit-line voltage during programming is limited in case of the self boosted program inhibit voltage method [2], and as a result, three states actually cannot be programmed at the same speed.

To overcome this problem, a new architecture, a multipage cell architecture is proposed [3], [4]. Two bits store in one memory cell is assigned to different page addresses while in the conventional scheme they are assigned to different column addresses. As a result, the simultaneously written states, Nstate, decreases to one or two and hence the program time is 57% shorter than the conventional multi-level cell. By using this new technology, the random read access also becomes half of the conventional scheme because one or two cycles are required in the proposed read, while three cycles are needed in the conventional read.

Since the target application of a multi-level flash memory is mass storage, the core circuit must be small so as to have a small die size. Especially, in NAND flash memories, the page buffer has to be minimized because each bit-line contains a page buffer circuit for

page program and page read operations [5], [6]. To meet this requirement, a compact intelligent 3-level page buffer has been proposed [5], capable of successfully realizing precise Vth control, fast programming and small circuit area.

Although several page buffer circuits for a four-level cell have already been proposed [1], [6], [7], none of them can meet all the above requirements. Some [1], [7] use multi-level sensing reference circuits and can precisely control the threshold voltage of memory cells using the cell by cell verify scheme, with the drawback of an area penalty. The other [6] minimizes the circuit area by eliminating reference circuits but needs a longer program time because each state is programmed separately. In this paper, so as to achieve precise Vth control, fast programming and small circuit area, a compact four-level page buffer circuit was newly developed.

The data retention problem in the multi-level flash memory is much harder than the single bit case, since the threshold voltage of a memory cell is higher [8-13] and the electric field across the tunnel oxide is higher, and what is worse, the voltage differences between each level becomes smaller. To improve the data retention characteristics of a multi-level memory cell, as a new NAND flash system best suitable for this technology, a preferential page select method has also been proposed, where each memory cell is programmed preferentially to the lower Vth level. With the preferential page select method, the IC error rate decreases by as much as 43% and operation at a highly reliable level can be realized.

Furthermore, it is shown that the reliability improves by the order of four since even if one memory cell fails, only one bit fails in one page, while two bits fain in the conventional scheme.

In Section 3.2, the concept of the multi-page cell architecture and the chip architecture of a proposed four-level cell are described. The preferential page select method is given in Section 3.3. In Section 3.4, as circuit implementations, the compact four-level page buffer is introduced and key device operations are described. The reliability improvement is discussed in Section 3.5. The experimental results are reported in Section 3.6.

3.2 Multipage cell architecture

A. Concept

Fig.3.1 shows the conventional and the proposed four-level cell. Two bit data stored in a conventional cell correspond to two Y-addresses (Y1,Y2). These two bit data belong to the same page. Also, three levels (the "1"-, "2"- and "3"-level) are programmed during the same operation [1], [6]. On the other hand, the proposed cell contains two "pages". Here, a "page" means the group of memory cells that are programmed simultaneously. In other words, the two-bit data of each cell correspond to two X-addresses, X1, X2 and the

programming of X1 and that of X2 are performed during different operations.

As shown in Fig.3.2, at the first page program, the memory cell is programmed to the "1"-state. After that, the "0"- or "1"-cell are respectively programmed to the "2"- or "3"-state during the second page program. Fig.3.3 shows the Vth distribution. In the conventional method, the ⊿Vth of each programmed state is the same [1], [6], [7], as shown in Fig.3.3(a). Conversely, "1"-program is controlled more precisely than "2"-or "3"-program in the proposed method, as described below (Fig.3.3(b)).

B. Chip architecture

In NAND Flash memories, the program operations are performed in page units to increase program throughput [14-16]. A schematic chip overview of the conventional [6] and the proposed 128Mb flash memory is shown in Fig.3.4. Table 3.1 explains the core

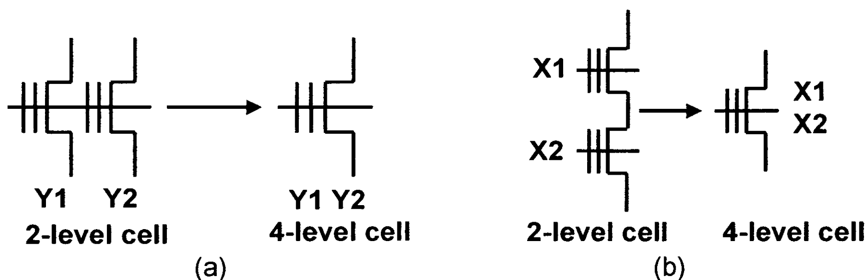


Fig.3.1 (a) Conventional four-level cell.
(b) Propose four-level cell

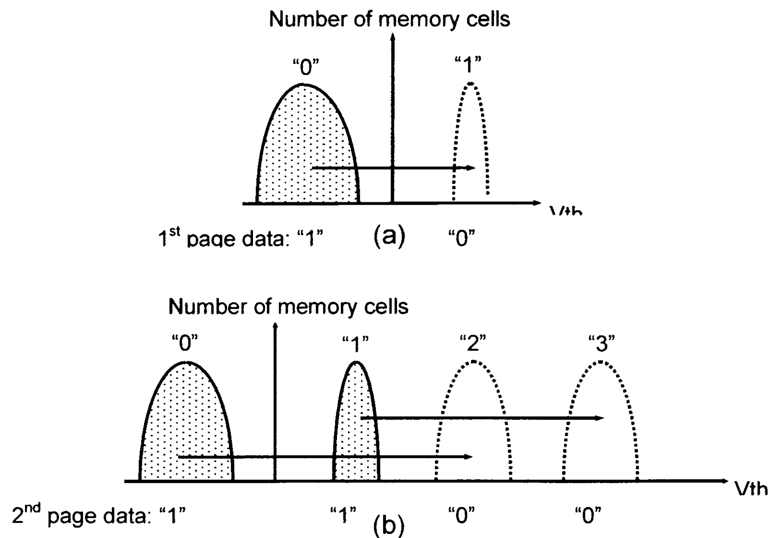


Fig.3.2 Program operation of the proposed architecture.
(a) 1st page program . (b) 2nd page program

circuit configuration. In both the conventional [6] and the proposed four-level cell array, each page buffer includes two latch circuits to store two bit data read out from or programmed into a four-level cell. Two adjacent bit-lines share one page buffer so as to reduce the page buffer area. In both conventional and proposed four-level cell array, two bit-lines sharing a page buffer belong to different pages. Program or read is performed in quite different operations because two latch circuits are needed to program or read one memory cell connecting to one of the neighboring bit-lines.

In the conventional 128Mb cell array [6], 4k cells share the same control gate and 2k page buffers are activated during a read or a program operation. Although each page buffer needs two latch circuits, the number of page buffers is halved. Under the same page size condition, the total number of latch circuits in the conventional 128Mb four-level cell array is the same as the conventional 64M two-level cell array and the circuit area does not increase.

On the other hand, in the proposed four-level cell array, 8k cells share the same control gate so as to make the page size 512 bytes and 4k page buffers are included. Compared with the conventional 128Mb four-level cell array and the 64Mb two-level cell array under the same page size condition, there is no circuit area penalty because the number of X-decoders is halved, whereas the circuit area of page buffer is doubled, as seen in Table 3.1.

In NAND Flash memories, the cell read current is as small as 1uA. Therefore the read access time is determined by the bit-line capacitance and the delay time of the word-line boosting has little effect on the access time. In the proposed array, as the bit-line capacitance is halved, the (verify) read is accelerated.

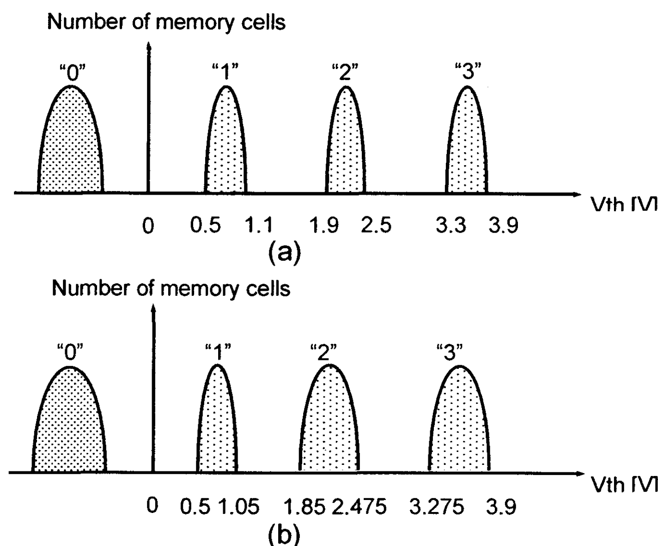


Fig.3.3 Vth distribution of the multi-level cell.
(a) Conventional. (b) Proposed.

	64Mb 2-level cell	Conventional 128Mb 4-level cell [6] (Fig.3.4(a))	Proposed 128Mb 4-level cell (Fig.3.4(b))
Page size	4kbit	4kbit	4kbit
Page buffer circuit : Na	4k	2k	4k
Latch circuit per one page buffer : Nb	1	2	2
Total latch circuit : Na×Nb	4k	4k	8k
X-address decoder (Block selector)	1k	1k	512

Table 3.1 Core circuit configuration

3.3 New NAND flash system

In a conventional NAND Flash memory, the program cycle starts form the source-line side cell to the bit-line side cell so as to prevent any interference between selected and unselected cells. Random page selection is not recommended during program operation. Such a restriction is not permitted in NOR flash memory where the size of program data is small and random page access must be adopted. However, in NAND flash memory, the restriction of page selection does not slow down the program performance of the memory system, because in mass storage applications each program data has a large quantity and a lot of pages can be successively programmed. For example, in case of a digital camera with a 400k-pixel CCD image sensor, the compressed image data is about 40 Kbytes and as many as 80 pages can be successively programmed.

In the proposed multi-level NAND flash system, a new page selection rule has been adopted. Each memory cell has a first page and second page in it, as described above. During the program operation, the first page of each word-line must be selected first. In other words, the second page program before the first page program is inhibited. Such a restriction is acceptable in mass storage applications, as previously discussed.

Fig.3.5 shows the X-addresses of the proposed 128 Mb four-level multi-page cell array. The first half of all X-addresses, that is, from "A0" to "A8191" correspond to the first pages of the word-lines, from "WL0" to "WL8191". In addition, the latter half of all X-addresses, that is, from "A8192" to "A16383" correspond to the second pages of the word-lines, from "WL0" to "WL8191".

For example, the first X-address, "A0" corresponds to the first page of the word-line, "WL0" and the next X-address, "A1" corresponds to the first page of the word-line, "WL1". At first, all the first pages, from "WL0" to "WL8191", are programmed. Next, the second page program is carried out to "WL0". In case the input data with a file size of 64 Mb is stored in a 128 Mb proposed cell array, all the first pages are selected and all the 64 M memory cells are programmed to the "0" or "1"-state, as shown in Fig.3.6(b). On the other hand, in the conventional method, 32 M memory cells are programmed to the "0","1","2" or"3"-state and the rest are still in the erased states, as shown in Fig.3.5(a). In the worst case, 32 M memory cells are in the "3"-states in the conventional method.

By using the proposed preferential page select method, the number of "3"-state cells decreases and data retention characteristics improves, as explained below. During shutdown (retention), the floating gate has a potential due to its stored charges. If the electric field across the tunnel oxide is high enough, electrons are ejected from the floating gate and Vth decreases. The leakage current during a retention is approximately described by the FN equation:

$$I_{leak} = A \cdot E_{ox}^{2} \cdot \exp(-\frac{B}{E_{ox}})$$

where A and B are characteristic constants and Eox is the electric field across the tunnel oxide. Eox during shutdown (retention) is expressed as

$$E_{ox} = \frac{C_{ono}}{C_{ox} + C_{ono}} \cdot \frac{Vth - Vthi}{T_{ox}}$$

where Cono, Cox, and Tox are ONO capacitance, tunnel oxide capacitance and tunnel oxide thickness [17-19]. Vthi is a thermal equilibrium threshold voltage which is achieved by irradiating an ultraviolet light on a cell. In a NAND-type cell, Vthi is around 0V to suppress a read disturb. Therefore, as Vth is higher, the electric field across the tunnel oxide, Eox is higher and the leakage current, Ileak increases. In case of a four-level cell, Vth of the "3"-state is 2.7V higher than that of the "1"-state. A 2.7V difference of Vth increases the leakage current by more than three-orders of magnitude. As a result, the IC error rate during shutdown is determined by the failure rate of the "3"-state to the "2"-state. If the number of the "3"-state cells is decreased, the IC error rate also decreases.

Fig.3.7 shows the IC error rate. In this estimation, it is assumed that the failure rate of the "1"-state to the "0"-state is negligibly smaller than that of the "3"-state to the "2"-state. In the proposed method, the number of the "3"-state cells is drastically decreased. As a result, the data retention characteristics are definitely improved. In a solid-state file, only 70% of the total cells are counted as part of the capacity and the rest is used as a work area that provides space for garbage collection [20]. In this case, the IC error rate decreases by as much as 43% and a highly reliable operation can be realized.

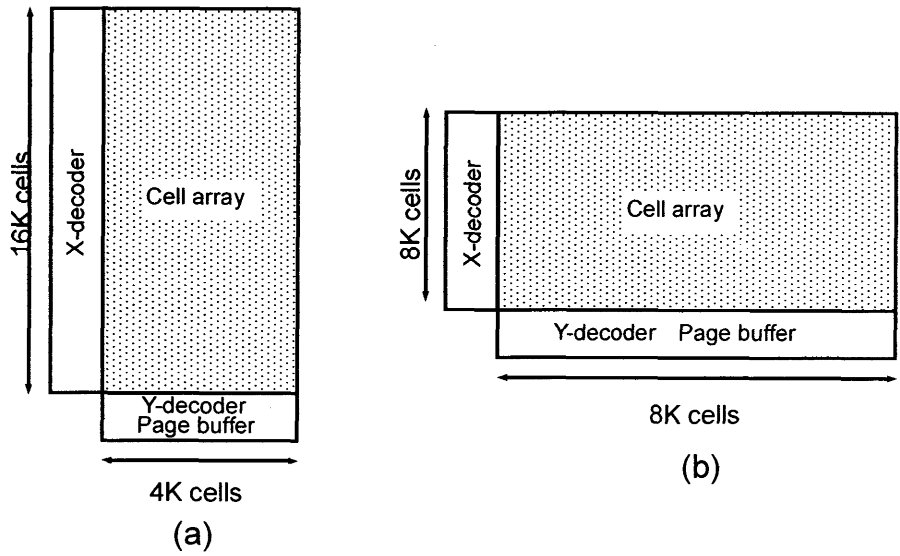


Fig.3.4 Cell array organization of a 128M-bit NAND flash memory. (a) Conventional [6]. (b) Proposed.

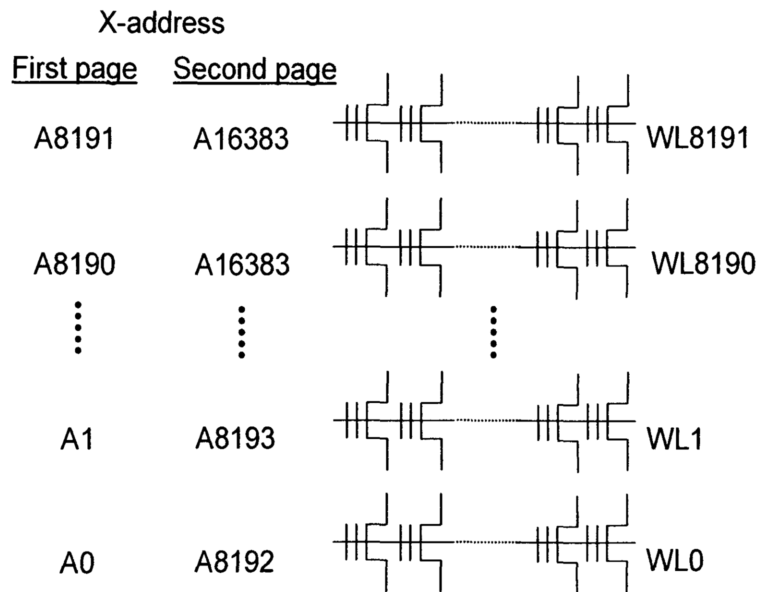


Fig.3.5 Allocation of X addresses to memory cells in a proposed NAND flash system

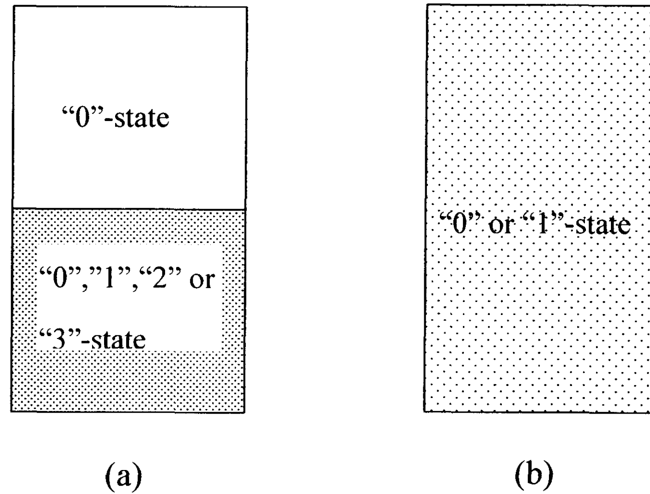


Fig.3.6 128M-bit four-level memory cell array storing 64Mbits of data. (a) In a conventional cell array, 32-Mbit memory cell are programmed to the "0", "1", "2" or "3" state and the remaining 32M memory cells are still in the erased state. (b) In a proposed cell array, all of the 64M memory cells are in the "0" or "1" state.

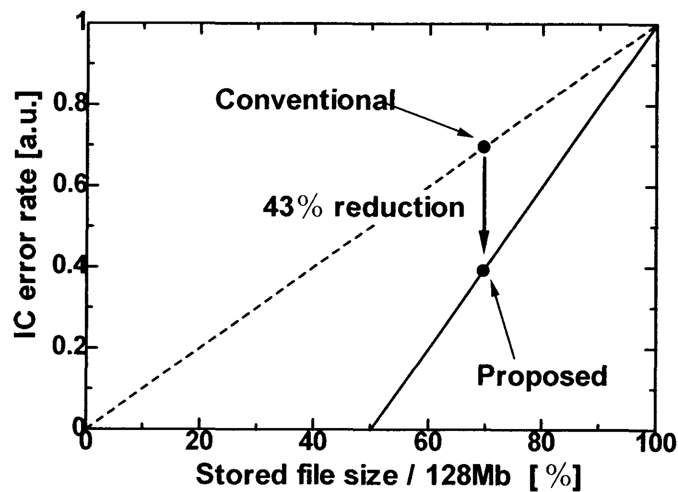


Fig.3.7 Reliability improvement of the new flash system.

3.4 Circuit implementations

A. Compact page buffer

The simplified circuit diagram of the proposed four-level compact page buffer for a multi-page cell is shown in Fig.3.8. To obtain a small die size, the reference circuit is eliminated and the number of transistors is minimized. Two bit-lines share one page buffer and are alternately selected during the program and read operation [21]. The data to be read out from or programmed into a memory cell is temporarily stored in a pair of flip-flop circuits, "Latch1"and "Latch2".

B. First page program operation

Fig.3.9(c) shows a flow chart of the program operation. The first page program is performed just like a two-level cell [13]. The timing chart for the first page program

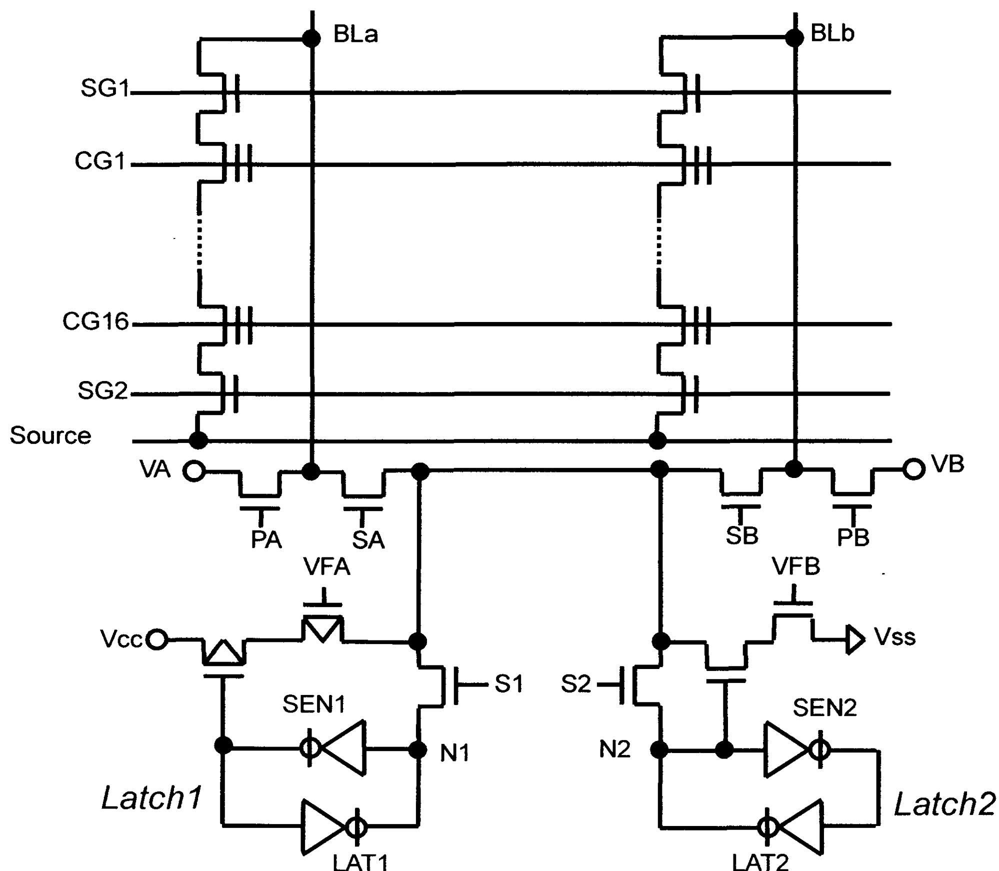


Fig.3.8 Compact page buffer.

operation can be seen in Fig.3.10(a). The bit-line, BLa is selected and the bit-line, BLb is unselected. The program data is input to "Latch1" as shown in Table 3.2. Program pulses are applied to the selected control gate (CG). If the first page data is "1", the bit-line voltage, Vch becomes Vcc. In this case, the select gate transistor connected to SG1 is turned off, because SG1 is set to Vcc. As a result, the whole channel of the selected memory cell is coupled to word-line signals and the memory cell remains in the erased state [2]. In the case in which the first page data is "0", Vch is 0V and the cell is programmed to the "1"-state. Program inhibit operation on an unselected bit-line, BLb, is performed by supplying Vcc using the PB signal in Fig.3.8.

After the first page program operation, the first page verify read operation is carried out. Fig.3.10(b) shows the timing chart of the first page verify read operation. The stored data in "Latch1" is modified, such that programming is executed only when the memory cell is not sufficiently programmed. The program pulses are gradually raised from 18.3 V by ΔVpgm. As shown in Fig.3.3(b), the Vth distribution, ΔVth of the "1"-state is reduced down to 0.55 V by decreasing ΔVpgm to 0.25 V.

C. Second page program operation

During the second page program, both "Latch1" and "Latch2" are used. The second page program data is input to "Latch1" through I/O, as can be seen in Fig.3.11(b). During the data load, the inverse polarity of the first page data stored in the memory cell is also read out to "Latch2". The additional first page read does not slow down the program speed because "1"-read is faster than the data load. Corresponding to the two-bit data latched in the page buffer, channel voltage Vch is changed, as shown in Table 3.3.

Fig.10(c) shows the timing diagrams of the second page program operation. In the case in which the second page data is "1", Vch is Vcc and the programming is inhibited. Thus, the memory cell stays at "0" or "1". If the page buffer data, N1 and N2, are respectively "0" and "1", Vch is 0V and the "1"-cell is programmed to the "3"-state. If both N1 and N2 are "0", Vch is 1.425V and the "0"-cell is programmed to the "2"- state. By raising the bit-line voltage of the "2"-program, the program speed of the "2"-program is adjusted to be the same as the "3"-program. After the second page program operation, the second page verify read operation is performed.

The timing chart of the second page verify read operation is shown in Fig.3.10(d). The second page verify read operation is composed of the "3"- and "2"-verify. During the "3"- and "2"-verify, the selected control gate is applied respectively to 3.275V and 1.85V. In case of a two-level cell [9] or a three-level cell [5], after the memory cell data is read out to the bit-line, the bit-line voltage is modified according to the one-bit data stored in one flip-flop circuit.

To realize the four-level cell-by-cell verify read operation, the key operation is that after

the data of the selected memory cell is read to the selected bit-line, the bit-line voltage is changed according to the two-bit data in both “Latch1” and “Latch2” by pulsing VFA and VFB. As a result, when a memory cell is programmed to the target Vth of its respective state, the associated “Latch1” is automatically reset to inhibit further programming by providing Vcc to the selected bit-line.

D. Read operation

The random read of the first and second page is performed during the same operation. One of four different states is identified by sequentially changing the word-line voltage [6], [7]. Therefore, the multiple reference circuit [1], [7] is eliminated and a much smaller circuit area can be obtained. The timing diagram is shown in Fig.3.10(e). The read operation is composed of three phases and the word-line is applied to three different levels corresponding to three programmed states. During the first phase, the selected control gate is biased to 2.875 V and the bit-line data is latched in “Latch1”.

At the time of the second phase, the selected control gate is set to 1.45 V and then the bit-line data is latched in “Latch2”. Finally, at the third phase, the selected control gate is applied to 0V. After the memory cell data is read out to the selected bit-line, BLa, the bit-line voltage is modified according to the stored data in both “Latch1” and “Latch2” by activating the VFB and VFA signals. Then, the bit-line data is latched to “Latch1”. The final data is shown in Table 3.4. The first and the second page data are latched respectively in “Latch1” and “Latch2”.

The read access is further accelerated by adopting a new configuration shown in Fig.3.12. In the case of the new configuration, in the second page program, “0”- and “1”- states are programmed to “3”- and “2”-states, respectively. The programming time is the same as the configuration of Fig.3.2. As shown in Fig.3.13, in the new configuration, one (2nd page read) or two (1st page read) cycles are required, while three cycles are needed in the conventional scheme. Therefore, in average, the read access time can be half of the conventional scheme.

3.5 Reliability consideration

Fig.3.14 shows the NAND flash file system. Between the host and the NAND flash memory, the flash controller is inserted to manage the file system such as the bad block management, write/erase cycle control and ECC (error correcting code). As shown in Fig.3.14, the error correction is operated in the unit of the sector. The sector size is 512Byte. For example, in the commercial multi-level NAND flash memories, 4bit correction is operated per one sector, that is, 512Byte.

In the case of the conventional multi-level cell, if one memory cell fails, two bits among one sector, 512Bytes fail. On the other hand, in the proposed architecture, even if one memory cell fails, only one bit among one sector fails and therefore a higher reliability can be obtained.

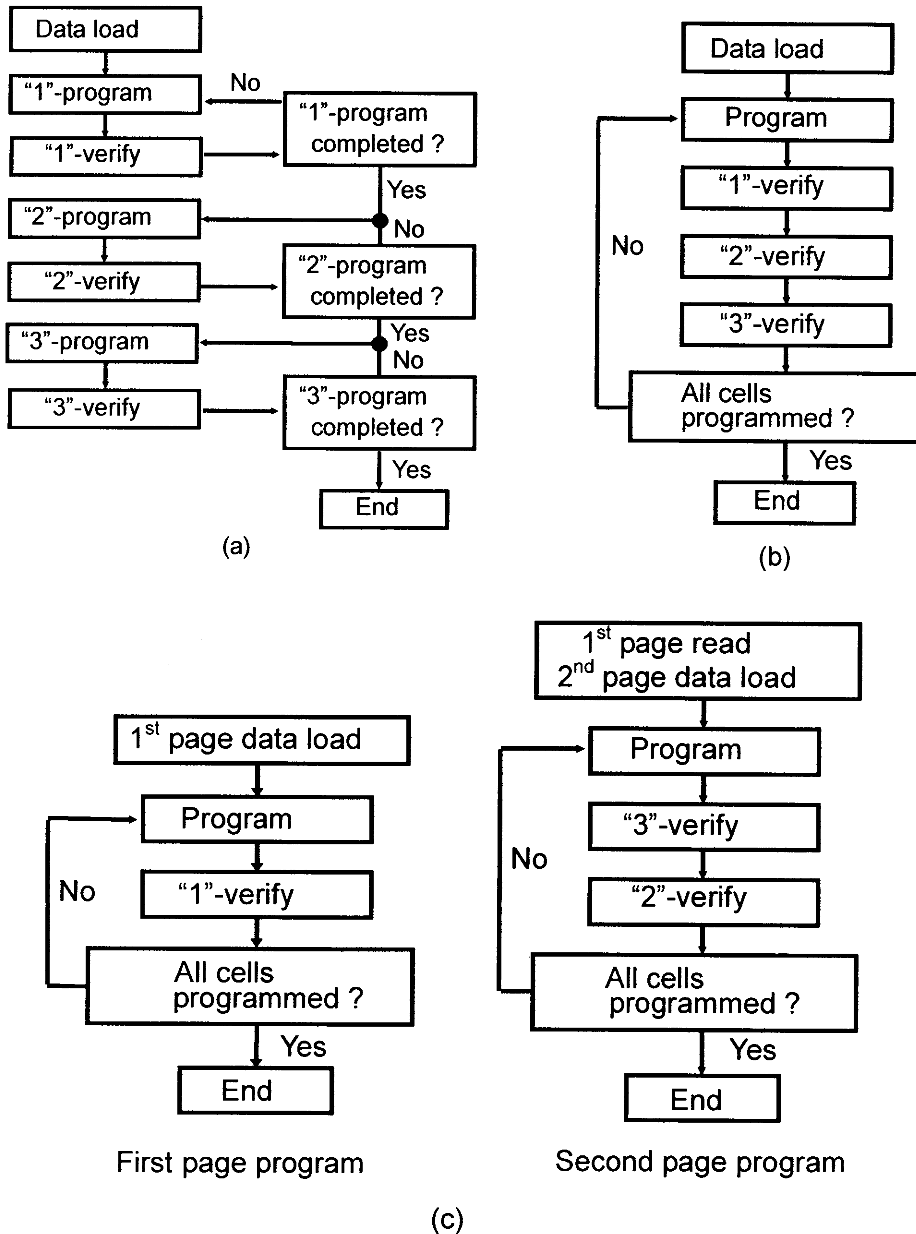
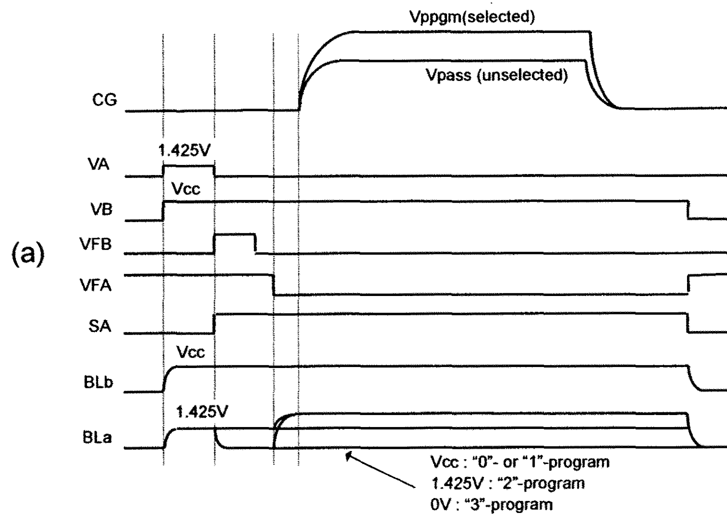
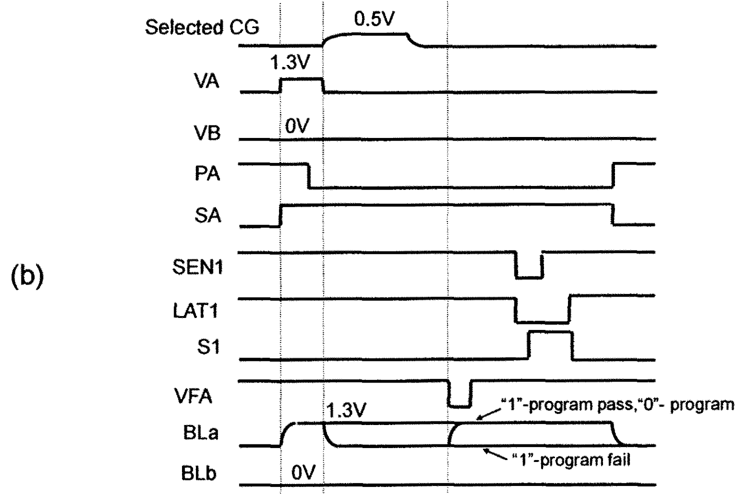


Fig.3.9 Program algorithm of (a) the conventional without simultaneous multi-level program [6], (b) the conventional with simultaneous multi-level program and (c) the proposed.

(a)



(b)



(c)

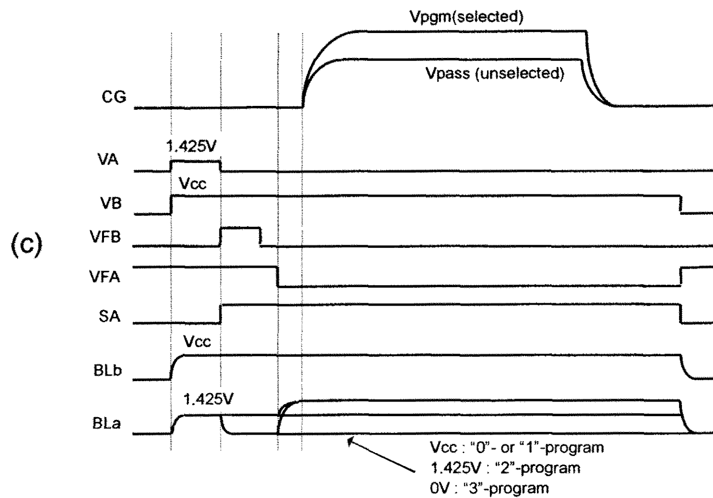
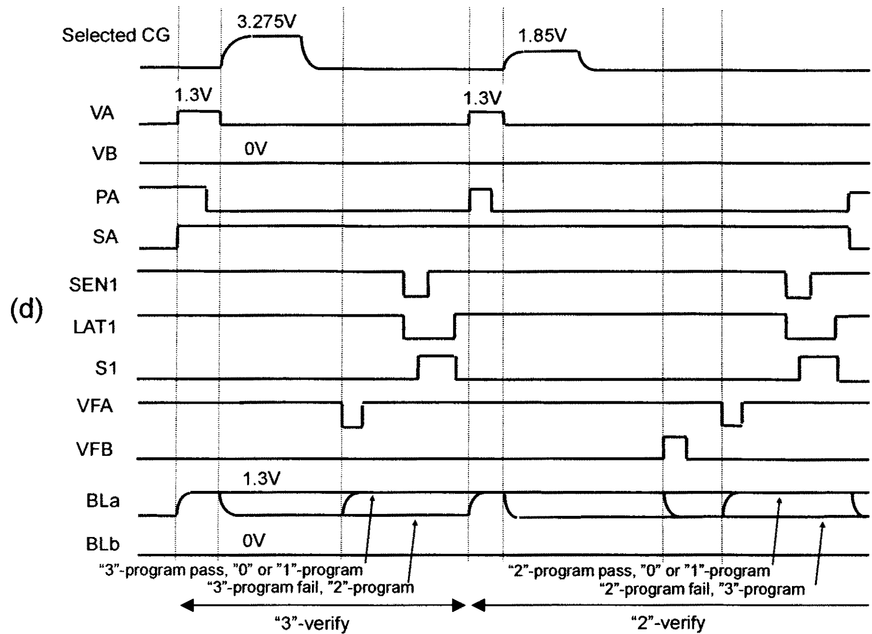


Fig.3.10 Signal timing diagram. (a) First page program. (b) First page verify read. (c) Second page program.



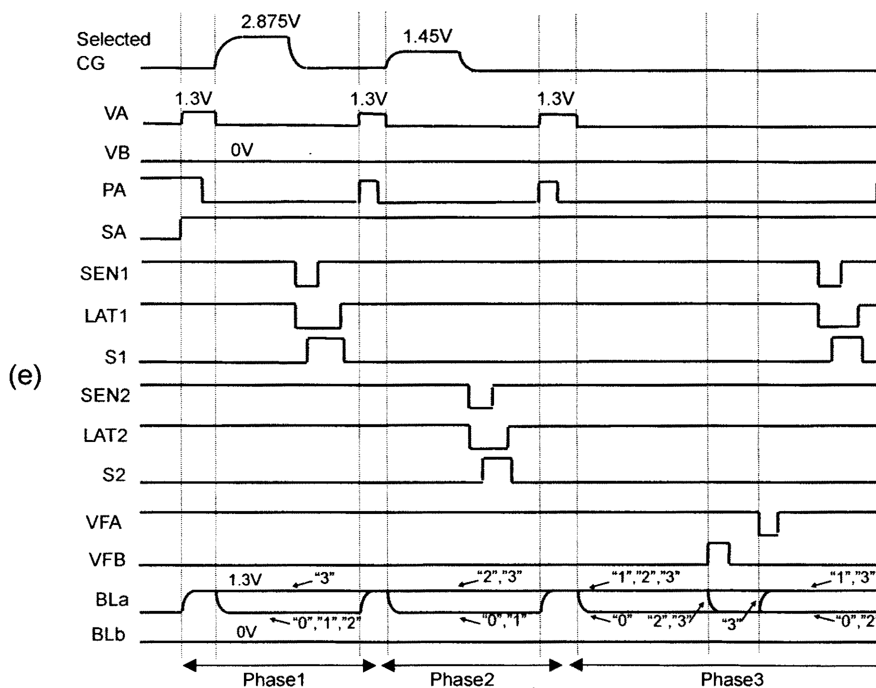


Fig.3.10 Signal timing diagram. (d) Second page verify-read. (e) Read.

	Memory cell state "0"-state	 "1"-state
1st page data, N1	1	0
Bit-line voltage, Vch	Vcc	0V

Table 3.2 Relation between program data stored in a page buffer and the bit-line voltage, Vch during the first page program.

	Memory cell state "0"	 "1"	 "2"	 "3"
2nd page data, N1	1	1	0	0
1st page data, N2	0	1	0	1
Bit-line voltage, Vch	Vcc		1.425V	0V

Table 3.3 Relation between program data stored in a page buffer and the bit-line voltage, Vch during the second page program.

	Memory cell state "0"	 "1"	 "2"	 "3"
1st page data, N1	0	1	0	1
2nd page data, N2	0	0	1	1

Table 3.4 Read-out data latched in a page buffer.

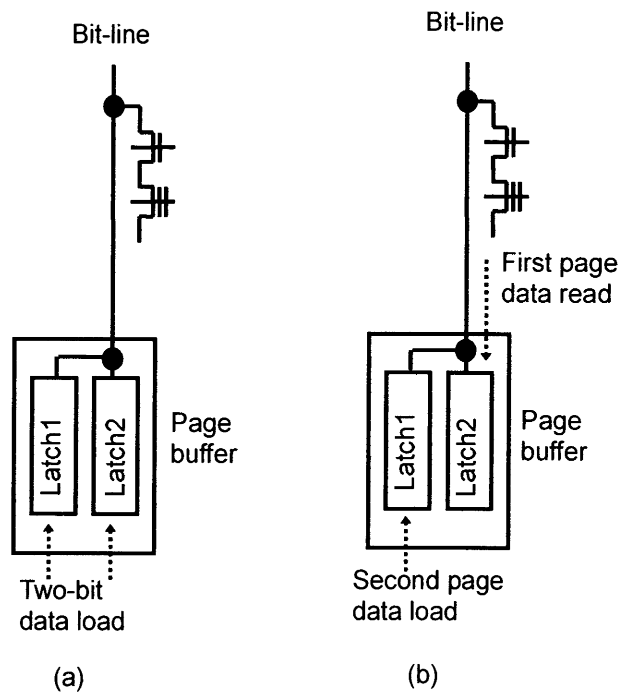


Fig.3.11 Data load operation of (a) the conventional method and (b) the second page program of the proposed method.

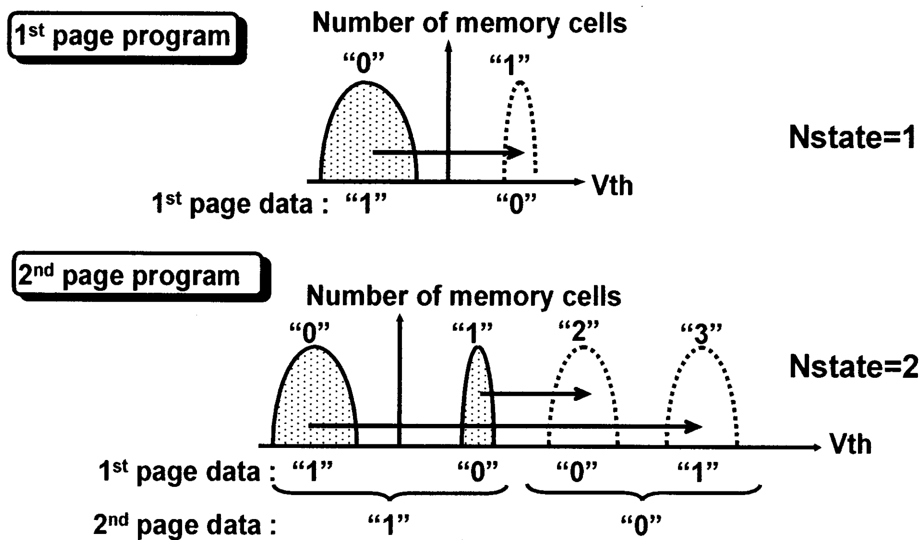


Fig.3.12 First and second page program of the new configuration.

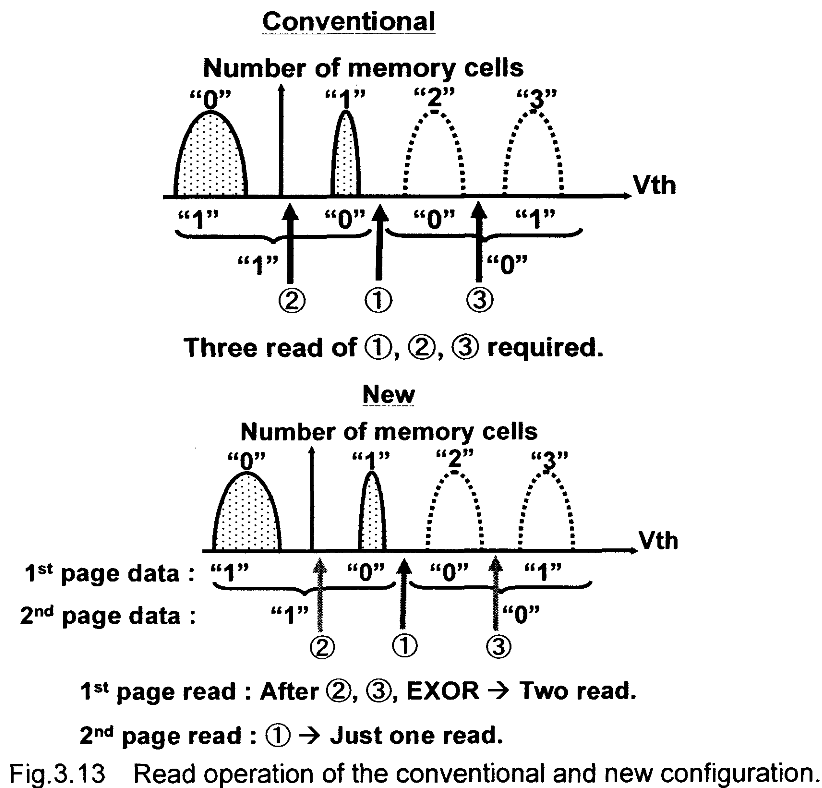


Fig.3.13 Read operation of the conventional and new configuration.

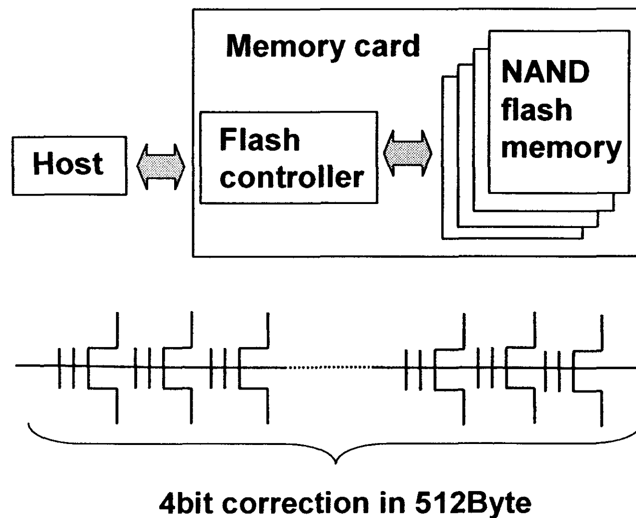


Fig.3.14 NAND flash system and the ECC (Error correcting code). The ECC engine in the flash controller chip detects and corrects the error.

3.6 Experimental results

A. Programming time

A-1. Conventional method without simultaneous multi-level program [6]

Fig.3.9(a) shows the program operation of the conventional method without a simultaneous multi-level program [6]. The program operation is composed of three program cycles. During the data load, two-bit program data are input to the page buffer, as seen in Fig.3.11(a). At the first program cycle, programming and program-verify of the "1"-state are repeated until the "1"-program is completed. Next at the second program cycle, the "2"-program and "2"-verify read are operated. Finally at the third program cycle, the "3"-program and "3"-verify are performed.

Each program operation is the same as the two-level cell, that is, the selected bit-line is grounded and the program inhibit bit-line is biased to Vcc. Program voltage increment, ⊿Vpgm is 0.3 V to make ⊿Vth 0.6 V. Therefore, the number of program pulses, Npulse during each program cycle is expressed as follows.

$$Npulse = 1 + \frac{\Delta Vth0 + \delta Vpgm}{\Delta Vpgm} = 10$$

δVpgm expresses the variation of ⊿Vpgm. As a result, the total program time, Tprog which is the sum of the first, second, and third program cycle is as follows.

$$Tprog = Tload + 3 \times (Tpulse + Tvfy) \times Npulse \ = 695us$$

Tload means the data loading time. The verify read time, Tvfy, is prolonged to 7.5 us because the maximum Vth is higher than the two-level cell and the decreased voltage difference between the control gate and Vth reduces the cell read current. As seen in Fig.3.15, in case the simultaneous multi-level program is not used, the program characteristics are seriously degraded because each state is programmed separately.

A-2. Conventional method with simultaneous multi-level program

Fig.3.9(b) shows the program operation of the conventional method with simultaneous multi-level program. During the data load, two-bit program data are input to the page buffer, as seen in Fig.3.11(a). During the program, the channel voltage of the memory cell, Vch, is changed according to the stored data in the page buffer. Then, three verify read sequences ("1"-,"2"-,"3"-verify) are subsequently carried out, where the control gate is applied to verify voltages, 0.5, 1.9 and 3.3 V, respectively. Ideally, Vch is 0, 1.4 and 2.8 V for the "3"-,"2"- and "1"- program where 1.4 V (2.8V) corresponds to the Vth difference between the "3" and "2" ("1").

The simulated write characteristics of the slowest cell are shown in Fig.3.15(a). ⊿Vpgm

is 0.3 V to make ⊿Vth 0.6 V, as can be seen in Fig.3.3(a). The three levels are programmed at the same speed. Including the Vch variation, δVch the number of program pulses, Npulse and the program time per page, Tprog are as follows.

$$Npulse = 1 + \frac{\Delta Vth0 + \delta Vpgm + \delta Vch}{\Delta Vpgm} = 10$$

$$Tprog = Tload + (Tpulse + 3 \times Tvfy) \times Npulse \ = 395us$$

δVch is 0.1 V. The verify read time, Tvfy, is 7.5 us. Compared with the two-level cell, the increased verify sequences, that occupy as much as 57% of the program time, seriously slow down the program speed, as shown in Fig.3.16.

Actually in the case of the "0"-program, the bit-line and the select gate connected to the bit-line are applied to Vcc and the select gate transistor is turned off [2]. If the minimum Vcc is 2.8 V, Vch must be lower than 1.5 V so as to turn on the select gate transistor during the "1"-,"2"-or "3"-program. Thus, Vch is 0 V for the "3"-program, 1.4 V for the "2"-program and 1.5 V for the "1"-program.

The initial value for Vpgm, Vpgm0, must be δVbl lower than the ideal case, where δVbl is the Vch difference between the actual and ideal case during the "1"-program. If this is not done, then the fastest cell would be programmed higher than the "1"-state. As a result, the "2"- and "3"-program become slower than the ideal case, as shown in Fig.3.15(b) and the number of program pulses, Npulse definitely increases.

$$Npulse = 1 + \frac{\Delta Vth0 + \delta Vpgm + \delta Vch + \delta Vbl}{\Delta Vpgm} = 14$$

$$Tprog = Tload + (Tpulse + 3 \times Tvfy) \times Npulse \ = 545us$$

δVbl is 1.3 V. The programming is entirely too slow, because the number of program pulses, Npulse increases due to the limitation of Vch, as seen in Fig.3.16.

A-3. Proposed method

The first page program is performed just like a two-level cell. ⊿Vth of the "1"-state is reduced down to 0.55 V by decreasing ⊿Vpgm to 0.25 V, as described above. As a result, the number of program pulses, Npulse1, and the program time per page, Tprog1, are expressed as follows.

$$Npulse1 = 1 + \frac{\Delta Vth0 + \delta Vpgm}{\Delta Vpgm} = 11$$

$$Tprog1 = Tload + (Tpulse + Tvfy) \times Npulse1 \ = 234.5us$$

In case of the "1"-program the bit-line is grounded, so there is no bit-line voltage fluctuation, that is, δVch is 0 V. The verify read time, Tvfy decreases to 4.5 us, due to the reduced bit-line capacitance. Compared with a conventional four-level cell, the program time decreases, because there is no Vch variation and only one verify sequence is required.

At the second page program, the channel voltage of the selected memory cell is 0 V for

the “3”-program and 1.425 V for the “2”-program, while the control gate is applied to Vpgm. Fig.3.15(c) shows the simulated write characteristics of the slowest cell. The “3”- and “2”-program are performed at the same speed. As a result, the number of program pulses, Npulse2, decreases. To make the maximum Vth the same as the conventional method, ⊿Vth of the “2”- and the “3”-state is increased to 0.625 V, as shown in Fig.3.3(b). Thus, ⊿Vpgm can be increased to 0.325 V, which further decreases Npulse2 and drastically improves program performance.

$$Npulse2 = 1 + \frac{\Delta Vth0 + \delta Vpgm + \delta Vch}{\Delta Vpgm} = 9$$

The verify sequences decrease to two sequences (“2”-and “3”-verify), compared with the three sequences in the conventional method. This also decreases the programming time per page, Tprog2.

$$Tprog2 = Tload + (Tpulse + Tvfy \times 2) \times Npulse2 \quad = 236us$$

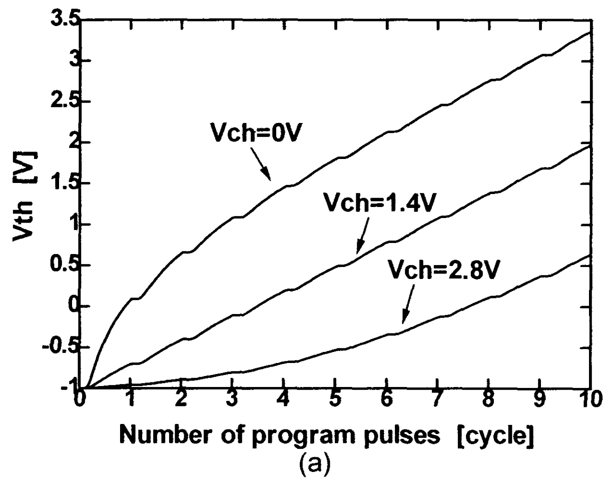
According to the preferential page select method described in Section 3.3, the first page has already been programmed before the second page program. Therefore, the program time of the second page is Tprog2, not Tprog1+Tprog2. Consequently, whichever page is selected, any page can be programmed within 236 us. Due to the reduced program pulses and verify sequences, a high programming speed of 236us/512byte or 2.2Mbyte/sec can be obtained, which is as much as 2.3 times faster than the conventional method, as shown in Fig.3.16.

B. Read access time

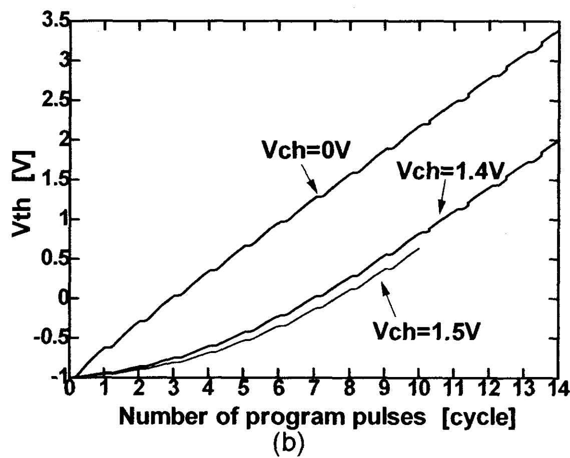
Fig.3.17 shows the random read access time. In the conventional scheme, three read cycle, “1”-, “2”- and “3”- read are required as shown in Fig.3.13. However, in the proposed scheme, in average one and half cycles are needed. Therefore, in average, the random read access time is 47% shorter than that of the conventional scheme.

C. Reliability

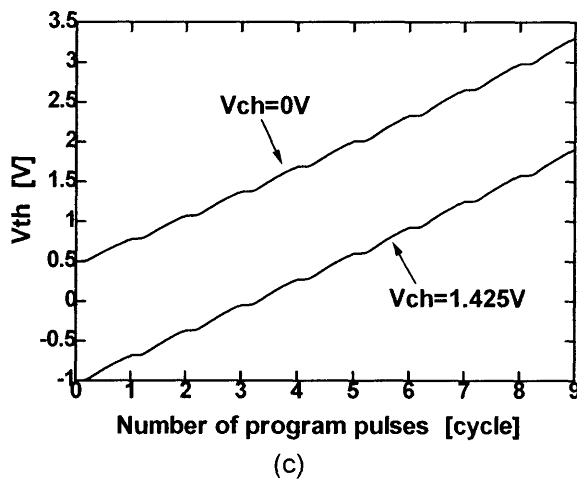
Fig.3.18 shows the reliability comparison between the conventional and the proposed scheme. In the proposed scheme even if one memory cell fails, error split into two pages. Hence, the reliability of the proposed scheme is four orders of magnitudes higher than that of the conventional scheme by using ECC. By utilizing the preferential page select scheme discussed in Section 3.3, the highly reliable multi-level NAND flash memory can be realized.



(a)



(b)



(c)

Fig.3.15 Program characteristics of the slowest cell. (a) Ideal case of the conventional method. ⊿Vpgm is 0.3V. (b) Actual case of the conventional method. ⊿Vpgm is 0.3V. (c) Proposed method. ⊿Vpgm is 0.325V. The coupling factor of the cell was 0.57 and the tunnel oxide thickness is 10nm.

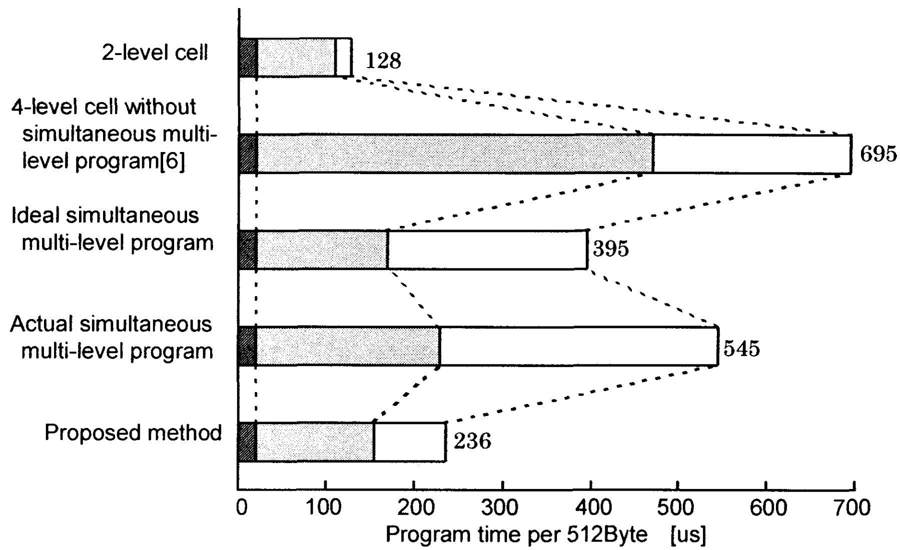


Fig.3.16 Programming time, Tprog. Dashed area is the data load. Dotted area is the program pulse and the open area is the verify read.

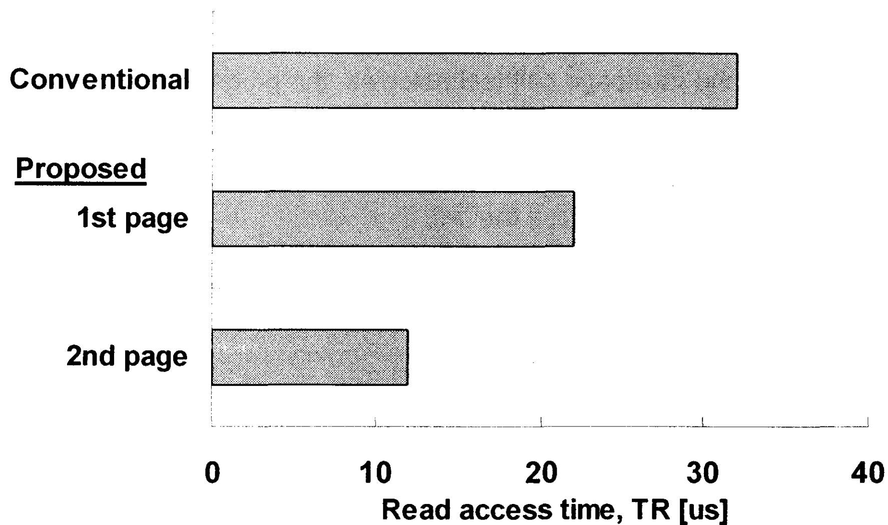


Fig.3.17 Read access time.

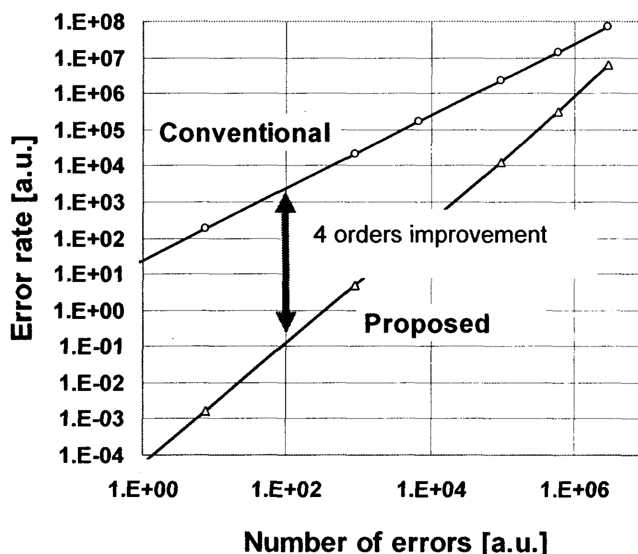


Fig.3.18 Reliability comparison with ECC.

3.7 Summary

A multi-page cell technology is proposed, allowing for both the precise control of the threshold voltage of a memory cell and fast programming without area penalty. The program speed is accelerated by 130%. A fast programming of more than 1MB/sec is achieved with this new technology.

The program speed trend of 3V multi-level cell and low-voltage multi-level cell is shown in Fig.3.19. Because of the multipage cell technologies, the program speed is doubled in both the 3V multi-level cell and the low-voltage multi-level cell. Due to this novel technology, the program speed of the multi-level cell becomes higher than 1MB/sec and consequently enabled us to commercialize the first multi-level 1Gbit products in 2001.

A microphotograph of the commercial 1Gbit multi-level NAND flash memory with 0.16um CMOS double metal technology is shown in Fig.3.20(a) [22]. Moreover, a microphotograph of the commercial 2Gbit multi-level NAND flash memory with 0.13um CMOS double metal technology is shown in Fig.3.20(b) [23]. The multipage cell technology is used in all multi-level cell NAND flash memory products such as 0.16um 1Gbit, 0.13um 2Gbit, 90nm 4Gbit [24], [25] and 70nm 8Gbit [26] products.

A small die size can be achieved with a newly developed compact four-level page buffer circuit. The random read access time becomes half because only one or two cycles are required during the read. A higher reliability with four orders of magnitudes less error rate is secured with this new architecture. In addition, a new NAND flash system, the preferential page select method is also proposed to improve the data retention

characteristics. The IC error rate decreases by as much as 43% and a highly reliable operation can be realized.

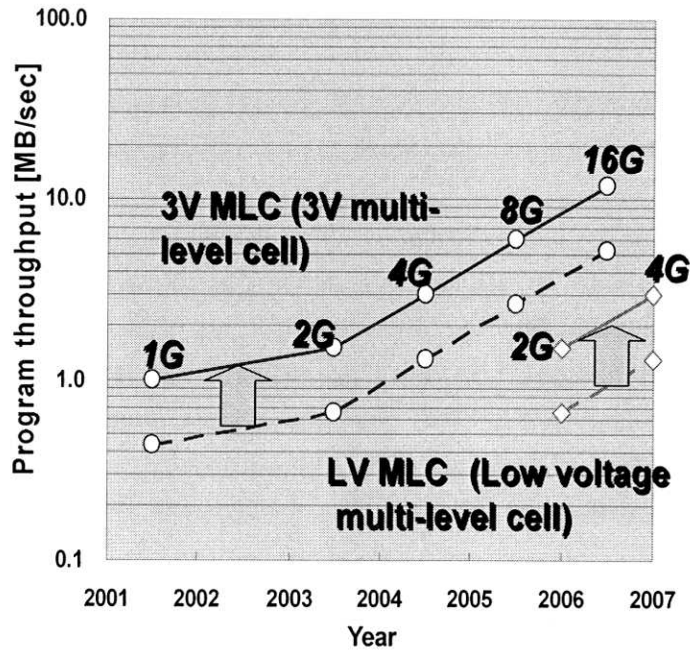


Fig.3.19 Program speed improvement with multipage cell technology.
(Circle : 3V MLC, Lozenge : LV MLC)

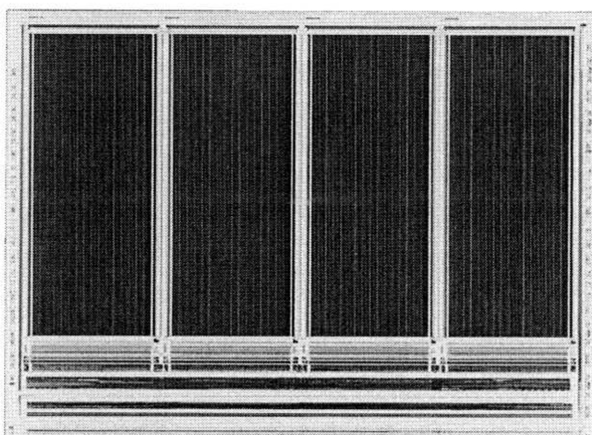


Fig.3.20 (a) Microphotograph of the 0.16um 1Gbit multi-level NAND flash memory [10].

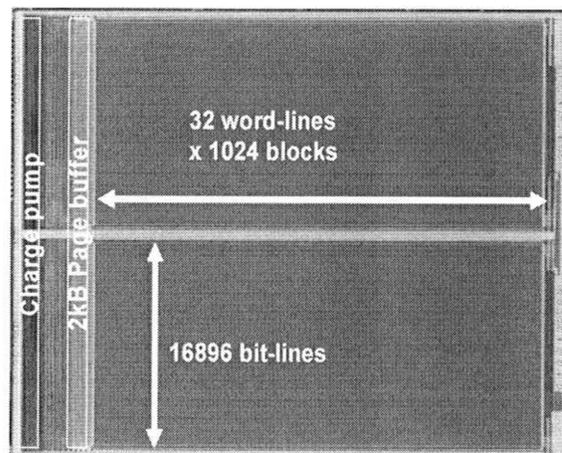


Fig.3.20 (b) Microphotograph of the 0.13um 2Gbit multi-level NAND flash memory [11].

References

[1] M.Ohkawa, H.Sugawara, N.Sudo, M.Tsukiji, K.Nakagawa, M.Kawata, K.Oyama, T.Takeshima and S.Ohya, "A 98mm^2 3.3V 64Mb Flash Memory with FN-NOR Type four-level Cell", in *ISSCC Digest of Technical Papers*, Feb. 1996, pp.36-37.

[2] K.D.Suh, B.H.Suh, Y.H.Lim, Y.J.Choi, Y.N.Koh, S.S.Lee, S.C.Kwon, B.S.Choi, J.S.Yum, J.H.Choi, J.R.Kim and H.K.Lim, "A 3.3V 32Mb NAND flash memory with incremental step pulse programming scheme," in *ISSCC Digest of Technical Papers*, Feb. 1995, pp.128-129.

[3] K.Takeuchi, T.Tanaka and T.Tazawa, "A multi-page cell architecture for high-speed programming multi-level NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1997, pp.67-68.

[4] K.Takeuchi, T.Tanaka and T.Tazawa, "A multipage cell architecture for high-speed programming multi-level NAND flash memories", *IEEE Journal of Solid-State Circuits*, vol.33, no.8, August 1998, pp.1228-1238.

[5] T.Tanaka, T.Tanazawa and K.Takeuchi, "A 3.4-Mbyte/sec Programming 3-Level NAND Flash Memory Saving 40% Die Size Per Bit", in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1997, pp.65-66.

[6] T.S.Jung, Y.J.Choi, K.D.Suh, B.H.Suh, J.K.Kim, Y.H.Lim, Y.N.Koh, J.W.Park, K.J.Lee, J.H.Park, K.N.Koh, J.W.Park, K.J.Lee, J.H.Park, K.T.Park, J.R.Kim, J.H.Lee and H.K.Lim, "A 3.3V 128Mb multi-level NAND flash memory for mass storage applications", in *ISSCC Digest of Technical Papers*, 1996, pp.32-33.

[7] M.Bauer, R.Alex, G.Atwood, B.Baltar, A.Fazio, K.Frary, M.Hensel, M.Ishac, J.Javanifard, M.Landgraf, D.Leak, K.Loe, D.Mills, P.Ruby, R.Rozman, S.Sweha, S.Talreja and K.Wojciechowski, "A multi-level-cell 32 Mb flash memory," in *ISSCC Dig. Tech. Papers*, Feb. 1995, pp.132-133.

[8] K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, K. Takeuchi, T. Ikehashi, K. Kanda, K. Hosono, R. Shirota, S. Aritome, K. Shimizu, K. Hatakeyama and K. Sakui " A 130mm^2 256Mb NAND Flash with Shallow Trench Isolation Technology", *ISSCC Digest of Technical Papers*, pp.112-113, 1999.

[9] H.Nakamura, K.Imamiya, T.Himeno, T.Yamamura, T.Ikehashi, K.Takeuchi, K.Kanda, K.Hosono, T.Futatsuyama, K.Kawai, R.Shirota, N.Arai, Fumitaka Arai, K.Hatakeyama, H.Hazama, Masanobu Saito, H.Meguro, K.Conley, K. Quader and J.Chen, "A 125mm^2 1Gb NAND Flash Memory with 10MB/s Program Throughput," *ISSCC Digest of Technical Papers*, 2002, pp.106-107.

[10] Taehee Cho, Young-Taek Lee, Euncheol Kim, Jinwook Lee, Sunmi Choi, Seungjae Lee, Dong-Hwan Kim, Wook-Kee Han, Young-Ho Lim, Jae-Duk Lee, Jung-Dal Choi, and Kang-Deog Suh, "A 3.3 V 1 Gb multi-level NAND flash memory with non-uniform threshold voltage distribution", in *ISSCC Digest of Technical Papers*, 2001, pp.28-29.

[11] June Lee, Heung-Soo Im, Dae-Seok Byeon, Kyeong-Han Lee, Dong-Hyuk Chae, Kyong-Hwa Lee, Young-Ho Lim, Jung-Dal Choi, Young-Il Seo, Jong-Sik Lee, and Kang-Deog Suh, "A 1.8V 1Gb NAND flash memory with 0.12um STI process technology", in *ISSCC Digest of Technical Papers*, Feb., 2002, pp.80-81.

[12] S.Aritome, "Advanced flash memory technology and trends for file storage application", in *IEDM Tech. Dig.*, 2000, pp.763-766.

[13] K. Kobayashi, "Toshiba's NAND flash strategy", in *CEATEC 2004 Dig. Tech. Papers*, Oct. 2004.

[14] F. Masuoka, M. Momodomi, Y. Iwata and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structured cell," in *IEDM Tech. Dig.*, 1987, pp.552-555.

[15] K. Imamiya, Y. Iwata, Y. Sugiura, H. Nakamura, H. Oodaira, M. Momodomi, Y. Ito, T. Watanabe, H. Araki, K. Narita, K. Masuda and J. Miyamoto, "A 35 ns-cycle-time 3.3 V-only 32 Mb NAND flash EEPROM", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.32-33.

[16] Kang-Deog Suh, Byung-Hoon Suh, Young-Ho Um, Jin-Ki Kim, Young-Joon Choi, Yong-Nam Koh, Sung-Soo Lee, Suk-Chon Kwon, Byung-Soon Choi, Jin-Sun Yum, Jung-Hyuk Choi, Jang-Rae Kim and Hyung-Kyu Lim, "A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.128-129.

[17] G.J.Hemink, T.Tanaka, T.Endoh, S.Aritome and R.Shirota, "Fast and Accurate Programming Method for Multi-level NAND Flash EEPROMs," *Symp. VLSI Technology Dig. Tech. Papers*, June 1995, pp.129-130.

[18] S. Aritome, R. Shirota, R. Kirisawa, T. Endoh, R. Nakayama, K. Sakui, and F. Masuoka, "A reliable bi-polarity write/erase technology in flash EEPROMs", in *IEDM Tech. Dig.*, 1990, pp.111-114.

[19] F. Arai, N. Arai, S. Satoh, T. Yaegashi, E. Kamiya, Y. Matsunaga, Y. Takeuchi, H. Kamata, A. Shimizu, N. Ohtami, N. Kai, S. Takahashi, W. Moriyama, K. Kugimiya, S. Miyazaki, T. Hirose, H. Meguro, K. Hatakeyama, K. Shimizu and R. Shirota, "High-density ($4.4F^2$) NAND flash technology using Super-Shallow Channel Profile (SSCP) engineering", in *IEDM Tech. Dig.*, 2000, pp.775-778.

[20] H.Niijima, "Design of a solid-state file using flash EEPROM," *IBM J. Res. Develop.*, vol.39 no.5, 1995, pp.531-545.

[21] T.Tanaka, Y.Tanaka, H.Nakamura, H.Oodaira, S.Aritome, R.Shirota and F.Masuoka, "A Quick Intelligent Programming Architecture and a Shielded Bit-line Sensing Method for 3V-only NAND Flash Memory," *IEEE J. Solid-State Circuits*, vol.29, November 1994, pp.1366-1373.

[22] S.Aritome, "Advanced flash memory technology and trends for file storage application", in *IEEE Tech. Dig. IEDM 2000*, pp.763-766.

[23] H.Nakamura, K.Imamiya, T.Himeno, T.Yamamura, T.Ikehashi, K.Takeuchi, K.Kanda,

K.Hosono, T.Futatsuyama, K.Kawai, R.Shirota, N.Arai, Fumitaka Arai, K.Hatakeyama, H.Hazama, Masanobu Saito, H.Meguro, K.Conley, K. Quader and J.Chen, "A 125mm^2 1Gb NAND Flash Memory with 10MB/s Program Throughput," *ISSCC Digest of Technical Papers*, 2002, pp.106-107.

[24] K. Kobayashi, "Toshiba's NAND flash strategy", in *CEATEC 2004 Dig. Tech. Papers*, Oct. 2004.

[25] D.S. Byeon, S.S.Lee, Y.H.Lim, J.S.Park, W.K.Han, P.S.Kwak, D.H.Kim, D.H.Chae, S.H.Moon, S.J.Lee, H.C.Cho, J.W.Lee, M.S.Kim, J.S.Yang, Y.W.Park, D.W.Bae, J.D.Choi, S.H.Hur and K.D.Suh, "An 8Gb multi-level NAND flash memory with 63nm STI CMOS process technology", in *ISSCC Digest of Technical Papers*, 2005, pp.46-47.

[26] T.Hara, K.Fukuda, K.Kanazawa, N.Shibata, K.Hosono, H.Maejima, M.Nakagawa, T.Abe, M.Kojima, M.Fujiu, Y.Takeuchi, K.Amemiya, M.Morooka, T.Kamai, H.Nasu, K.Kawano, C.M.Wang, K.Sakurai, N.Tokiwa, H.Waki, T.Maruyama, S.Yoshikawa, M.Higashitani, T.D.Pham and T.Watanabe, "A 146mm^2 8Gb NAND flash memory with 70nm COMS technology", in *ISSCC Digest of Technical Papers*, 2005, pp.44-45.

CHAPTER 4
LOAD CAPACITANCE REDUCTION TECHNOLOGY

Acceleration by decreasing program pulse width, Tpulse

4.1 Introduction

The increasing demand for low voltage/low power portable equipment in the consumer marketplace has created a need for a low voltage/low power flash memory. In particular, sub-1.8V operation is essential for systems requiring extended battery lifetime such as cellular phones, personal information devices, handheld global positioning systems and portable instrumentation [1-3].

For such a low voltage/low power market, there is a strong demand for low-voltage multi-level cell because of the low cost. In the case of the NAND flash memory, at a low supply voltage the tunnel oxide thickness of the memory cell cannot be scaled down to secure the reliability such as data retention [4-8]. Since the Tox of the cell is not scaled, the program voltage applied to the cell is not lowered at the low a low supply voltage. Consequently, at a low supply voltage the ramp up time to charge the huge bit-line capacitance, e.g. 20nF by a charge pump circuit drastically becomes dramatically long and seriously degrades the program performance.

In this chapter, the crucial issue in realizing a low voltage/low power operation NAND flash memory is explained first. In the conventional Vcc-bit-line programming scheme [9], the program disturbance becomes harsher as Vcc decreases. The minimum Vcc of the conventional Vcc-bit-line programming scheme is 2.0V and sub-1.8V operation cannot be realized.

To improve the program disturbance characteristics, the boosted-bit-line programming scheme has been proposed [10]. In this scheme, the program inhibit bit-line is raised to 4.5V by a charge pump circuit to decrease the program disturbance. However, the boosting of the huge bit-line capacitance leads to both a larger power dissipation and a longer programming time. Besides, the conventional boosted-bit-line programming scheme is subject to circuit area and manufacturing cost overheads. Therefore, to realize a low voltage/low power NAND flash memory, the key issue is how to improve the program

	Conv. Vcc-bit-line scheme	***Conv. boosted-Bit-line scheme***	***Proposed scheme***
Vcc (minimum)	× **(2.0V)**	○ **(None)**	○ **(None)**
Programming time	○ **(178us/page)**	× **(500us/page)**	○ **(192us/page)**
Chip size overhead	○ **(None)**	× **(5%)**	○ **(None)**
Additional process steps	○ **(None)**	× **(High-voltage PMOS req.)**	○ **(None)**
Energy dissipation	○ **(0.49uJ)**	× **(1.13uJ)**	○ **(0.53uJ)**

Table 4.1 Performance comparison (Vcc=1.4V)

disturbance characteristics without sacrificing the performance, the chip area or the manufacturing cost.

To overcome this problem, a new architecture, "a source-line programming scheme" [11], [12] is introduced. In order to improve the program disturbance characteristics, a high program inhibit voltage is applied to the channel from the low capacitance source-line as opposed to from the large capacitance bit-line of the conventional scheme. The bit-line swing is decreased from Vcc to 0.5V to lower the power consumption. In this way, by decreasing a core load capacitance, a high speed programming as well as a low energy dissipation are achieved.

Table 4.1 summarizes the performance comparison of the proposed and the conventional programming schemes. In the conventional boosted-bit-line programming scheme and the proposed scheme, the program disturbance is dramatically relaxed while it sets a limit to lower the supply voltage in the conventional Vcc-bit-line programming scheme. In the conventional boosted-bit-line programming scheme, the bit-line boosting increases the power dissipation, the programming time, the chip area and the manufacturing cost. In contrast, the proposed scheme drastically reduces the program disturbance without circuit area, manufacturing cost, program speed or power consumption overhead. As a result, a highly reliable program operation can be realized irrespective of the supply voltage. A very fast programming of 192us/page and a very low power operation of 22mW at 1.4V can be realized in the proposed scheme.

In Section 4.2, it is explained that the conventional Vcc-bit-line programming scheme cannot operate below 2.0V due to the program disturbance issue. Moreover, the problems of the conventional boosted-bit-line programming scheme are explained. In Section 4.3,

the concept and the chip architecture are described. The circuit implementations are discussed in Section 4.4 and the experimental results of are presented in Section 4.5. In Section 4.6, the design considerations such as the power consumption issue and cost issues are presented. Finally, the summary is given in Section 4.7.

4.2 Design issues at low Vcc

A. Problems of the conventional Vcc bit-line scheme.

The conventional Vcc bit-line programming scheme [13-15] is shown in Fig.4.1(a). The program disturbance is shown in Fig.4.2. During a programming, the selected control gate is Vpgm, for example 18V. The unselected control gates are Vpass that is about 10V. In the case of the "0"-program, the selected bit-line is set to 0V and the channel of the selected memory cell is grounded. The Vpass disturb occurs in the unselected memory cells in series with the "0"-program cell.

On the other hand, the program inhibit bit-line is applied to Vcc [9]. In this case, the select gate transistor connected to the bit-line is turned off, because the select gate is biased to Vcc. As a result, the whole channel of the program inhibit memory cells is raised by the capacitive coupling with control gates and the unselected memory cells remain in the erased state. The unselected cell depicted in Fig.4.2 is subject to the Vpgm disturb. To reduce the Vpgm disturb, the key issue is to achieve a higher channel voltage of the program inhibit cell, Vch.

Here, the expression of Vch will be presented in detail. Fig.4.3 shows the circuit model of the series-connected sixteen memory cells during the Vpgm disturb [11], [12]. In this figure, Cono and Cox are respectively the ONO capacitance and the tunnel oxide capacitance. Ctotal is the total channel capacitance of sixteen memory cells which is the sum of the junction capacitance between the inversion layer and P-well and that between the source/drain diffusion regions and P-well.

In the conventional Vcc-bit-line programming scheme, "0"-cells of the unselected control gates are not turned on until the control gates are raised to Vth0+Vchi. Vth0 is the Vth of the "0"-state. Vchi is the transferred bit-line voltage via the select transistor connected to the program-inhibit bit-line and can be expressed as Vcc-Vthsg where Vthsg is the threshold voltage of the select transistor. After "0"-cells are turned on, the channel is raised by the capacitive coupling with control gates. This means that the voltage swing of control gates that contributes to the raising of the channel is Vpass-Vth0-Vchi. Therefore, the channel voltage of the conventional Vcc-bit-line programming scheme, Vch, is expressed as follows [16].

$$Vch_conv.1 = Vchi + 15\gamma(Vpass - Vth0 - Vchi) + \gamma \cdot Vpgm$$
$$= (1 - 15\gamma) \cdot (Vcc - Vthsg) + 15\gamma(Vpass - Vth0) + \gamma Vpgm$$
$$\gamma = \frac{Cono \cdot Cox}{16Ctotal \cdot (Cono + Cox)} = 0.04$$

To reduce the Vpgm disturb, a higher Vch is needed [17]. In the conventional Vcc-bit-line programming scheme, as Vcc decreases, Vchi decreases. As a result, the channel voltage, Vch also decreases. This increases the electric field across the tunnel oxide and the Vpgm disturb becomes worse.

Fig.4.4 shows the measured Vth shift after the Vpgm and the Vpass disturb as a function of Vpass. To decrease the Vpass disturb, Vpass must be decreased [16]. On the other hand, so as to improve the Vpgm disturb characteristics, Vpass has to be large enough. The intersection of the Vpass disturb and the Vpgm disturb is where the program disturb can be minimized. As shown in Fig.4.4, the conventional Vcc-bit-line programming scheme causes a large Vth shift. Fig.4.5 shows the allowable program disturb time so as not to cause an incorrect operation. The program disturb time is defined as when the Vth shifts by 1.5V. The program disturb time decreases as Vcc decreases. The minimum Vcc of the conventional Vcc-bit-line programming scheme is 2.0V.

B. Problems of the conventional boosted bit-line scheme.

In order to reduce the program disturbance, the boosted-bit-line programming scheme has been proposed [2]. Fig.4.1(b) shows the boosted-bit-line programming scheme. What's different from the conventional Vcc-bit-line programming scheme is that the program inhibit bit-line and SGD are raised to 4.5V by a charge pump circuit to reduce the program disturbance. In this case, Vchi is expressed as 4.5-Vthsg, and the channel voltage, Vch, is expressed as follows.

$$Vch_conv.2 = Vchi + 15\gamma(Vpass - Vth0 - Vchi) + \gamma \cdot Vpgm$$
$$= (1 - 15\gamma) \cdot (4.5 - Vthsg) + 15\gamma(Vpass - Vth0) + \gamma Vpgm$$
$$= Vch_conv.1 + (1 - 15\gamma) \cdot (4.5 - Vcc)$$

The value of the parameter "γ" used in this equation is the same as that used in the equation of Vch_conv.1. Much larger Vch reduces the electric field across the tunnel oxide and decreases the Vpgm disturb. The Vth shift caused by the program disturbance is strongly suppressed, as shown in Fig.4.4. As shown in Fig.4.5, a highly reliable program operation can be realized irrespective of Vcc.

Despite the excellent program disturbance characteristics, the bit-line boosting causes many problems, such as a slow programming, a large power consumption, an enlarged chip size and a manufacturing cost overhead at low Vcc. These drawbacks will be discussed below.

B-1. Degraded program speed

In NAND flash memories, four thousand memory cells are programmed at the same time to increase the program throughput. Therefore, in the conventional boosted-bit-line programming scheme, the huge capacitance, as much as 17nF, of the four thousand bit-lines is boosted to 4.5V. As the supply voltage decreases, the rise time of the bit-line boosting increases. Consequently, the enlarged rising time of the bit-lines results in a longer programming time as shown in Fig.4.6. The programming time is prolonged to 500us/page at 1.4V and the programming is unacceptably slow.

B-2. Large power dissipation

Fig.4.7 shows the energy dissipation per programming pulse. In the conventional boosted-bit-line programming scheme, the bit-line charging current increases the energy dissipation and this increase is particularly marked at low Vcc.

B-3. Chip size and manufacturing cost overheads

In addition to the prolonged programming time and the enlarged power consumption, the conventional boosted-bit-line programming scheme is subject to chip size and manufacturing cost overheads.

In the conventional Vcc-bit-line programming scheme, a page buffer connecting to a bit-line is composed of low-voltage transistors, because the maximum voltage applied to the bit-line is Vcc [11], [12]. The high-voltage switch transferring Vpgm or Vpass to the word-lines is composed of high-voltage NMOS. Therefore, a high-voltage PMOS is not used to lower the manufacturing cost. However, in the conventional boosted-bit-line programming scheme, each page buffer must be made of high-voltage NMOS and high-voltage PMOS to apply 4.5V to the bit-line. The circuit area of the high-voltage transistor is much larger than that of the low-voltage transistor. As the NAND flash memory has as many as four thousand page buffer circuits to realize the page-based fast parallel programming, the chip size overhead is as much as five percent. Moreover, the manufacturing cost also increases because a high-voltage PMOS is required.

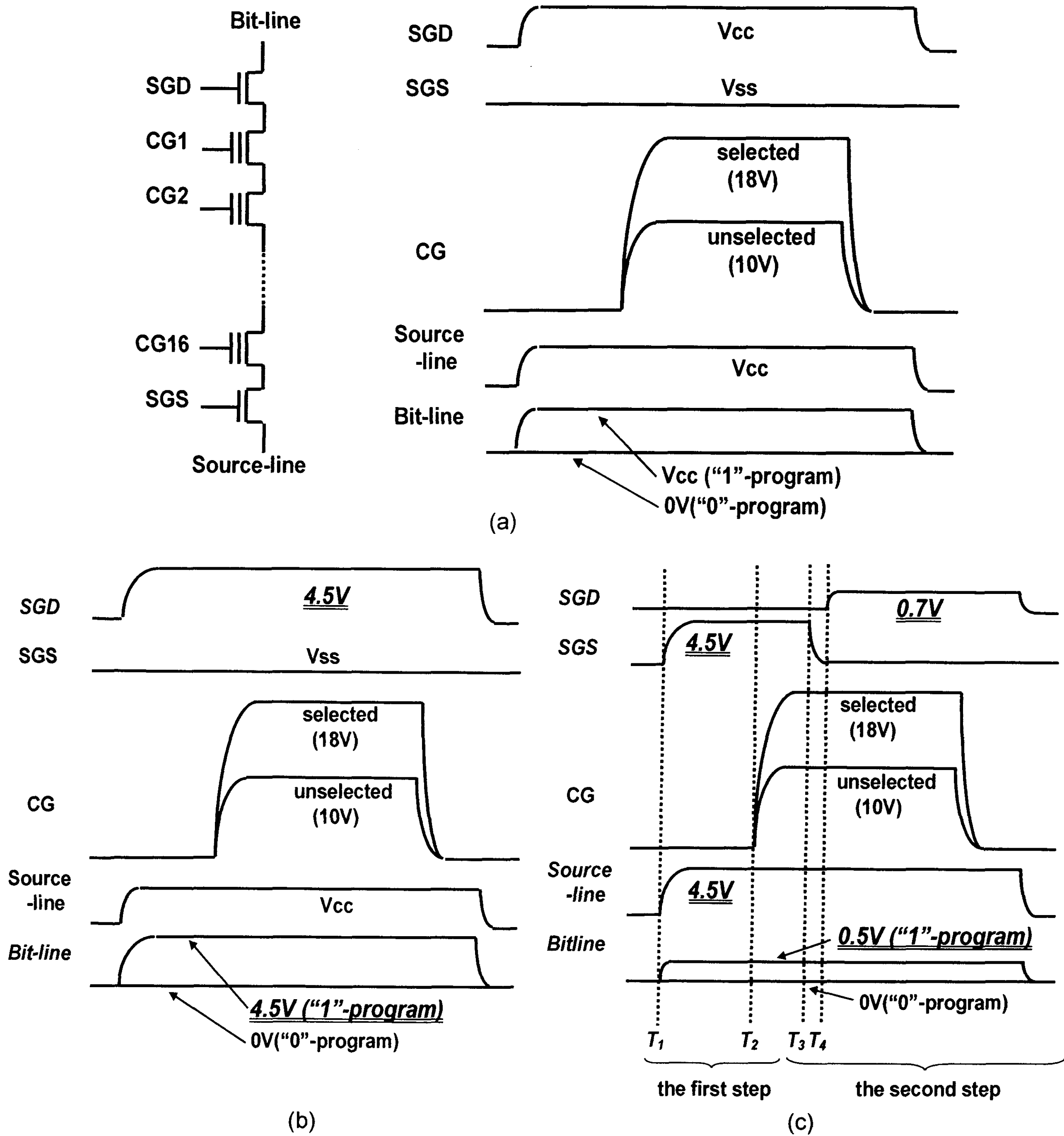


Fig.4.1 Signal timing diagram. (a) Conventional Vcc-bit-line programming scheme. (b) Conventional boosted bit-line programming scheme. (c) Proposed source-line programming scheme.

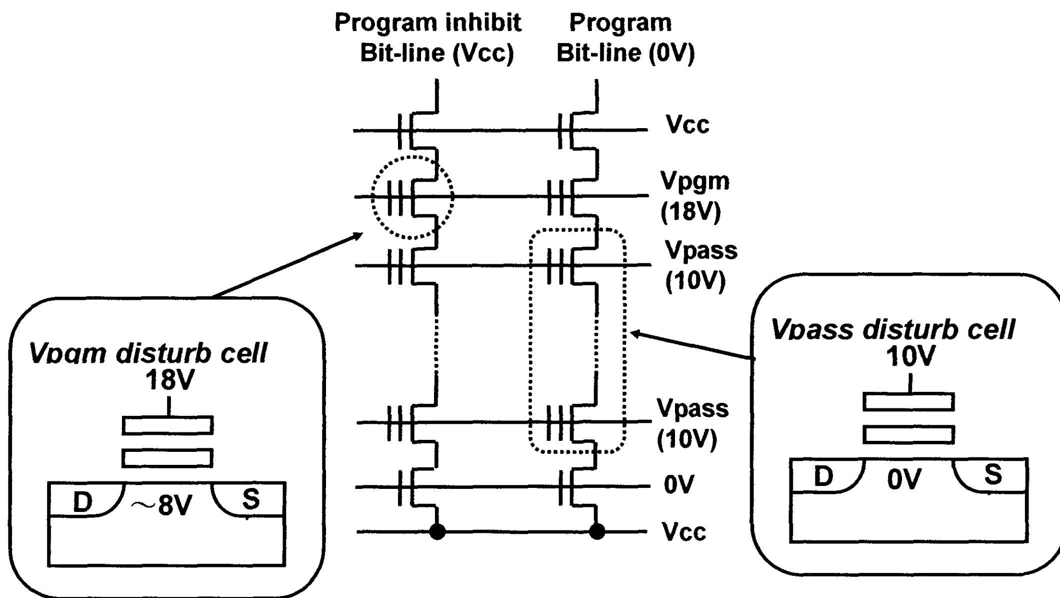


Fig.4.2 Program disturbance of the conventional Vcc bit-line programming scheme.

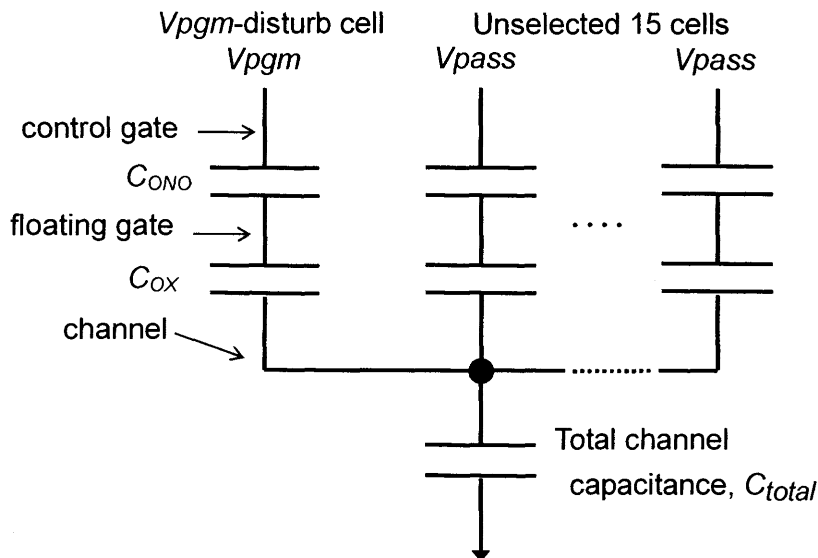


Fig.4.3 Circuit model of the sixteen memory cells in series during the Vpgm disturb. Cono and Cox are the ONO capacitance and the tunnel oxide capacitance. Ctotal is the total channel capacitance which is the sum of the junction capacitance between the inversion layer and P-well and that between the source/drain diffusion regions and P-well.

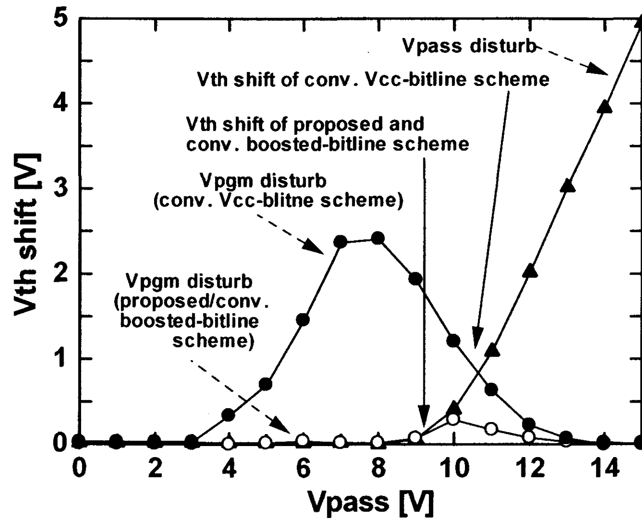


Fig.4.4 Threshold voltage shift after the Vpgm and the Vpass disturb measured at 85℃. Vcc=1.8V. To decrease the Vpass disturb, Vpass must be decreased. The Vpgm disturb becomes smaller as Vpass increases. The intersection of the Vpass disturb and the Vpgm disturb is where the program disturb is minimized.

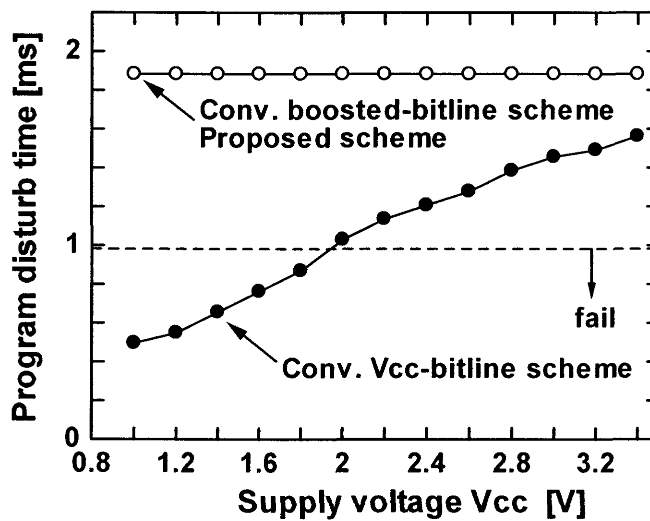


Fig.4.5 Allowable program disturb time measured at 85℃. The program disturb time is defined as when the Vth shifts by 1.5V. The minimum value of Vcc is 2.0V in the conventional Vcc-bit-line programming scheme.

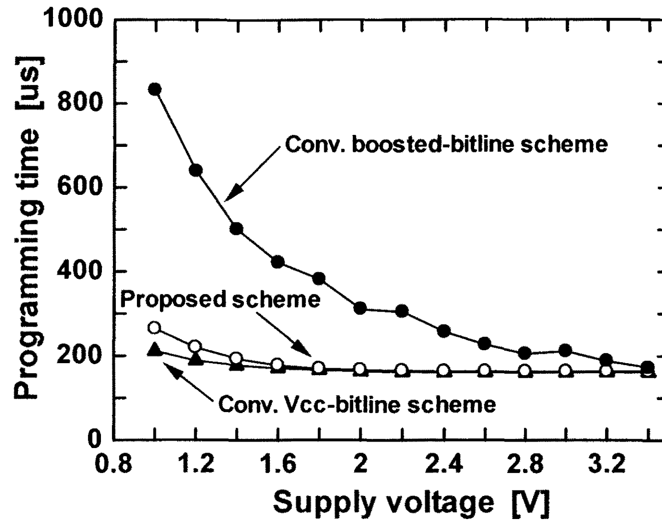


Fig.4.6 Programming time per page. In the conventional boosted-bit-line programming scheme, the bit-line boosting extremely increases the total programming time at low Vcc.

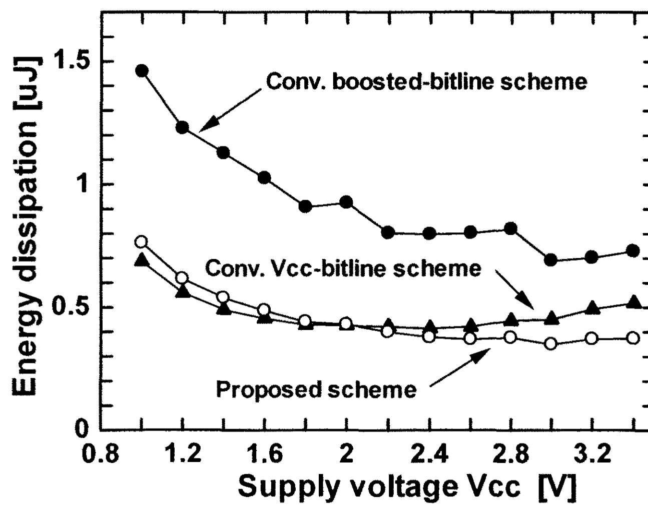


Fig.4.7 Energy dissipation per programming pulse. In the conventional boosted-bit-line programming scheme, the bit-line charging current increases the energy dissipation and this increase is particularly marked at low Vcc.

4.3 Source-line programming scheme

4.3.1 Principle

The concept of the new programming scheme is shown in Fig.4.1(c) [11], [12]. There are two important points. First, in order to reduce the program disturbance, a higher program inhibit voltage, for example 4.5V, is applied from the source-line to the channel, whereas the program inhibit voltage is biased from the bit-line in the conventional schemes (Fig.4.1(a),(b)). The other point is that the bit-line swing is decreased from Vcc to 0.5V to reduce the power dissipation. This small bit-line swing can be used because a high program inhibit voltage is applied from the source-line to the channel and a high bit-line voltage is unnecessary.

Referring to Fig.4.1(c) and Fig.4.8, a detailed description of the proposed program operation is given below. In Fig.4.8, the upper figures show the "0"-program that means the program selected. The lower figures of Fig.4.8 correspond to the "1"-program, that is, the program inhibit operation. The proposed program operation is composed of two steps.

A. The first step

The first step is depicted in Fig.4.8(a). In the first step, the channels of both the "0"- and "1"-program cells are boosted to Vch. First, the source-line and SGS are biased to 4.5V at T1 in Fig.4.1(c) and a high voltage is transferred to the channel of both the "0"-program and the "1"-program cells. SGD is 0V. At T2, Vpgm and Vpass are applied to the selected and the unselected control gates, respectively. By the capacitive coupling with the control gates, the channels of all memory cells, that is, both the "0"- and "1"-program cells, are raised to Vch, 8V, as shown in Fig.4.8(a).

B. The second step

Next, we proceed to the second step shown in Fig.4.8 (b). In the second step, the channel voltage is changed according to the programming data and an electron injection is performed. At first, SGS is discharged to 0V at T3. After that, SGD is applied to 0.7V at T4. At this time, the channel voltage is changed or kept according to the program data, as shown in Fig.4.8(b). In the case of the "0"-program, the bit-line is grounded and the channel is discharged from Vch to 0V. As a result, electrons are injected from the channel to the floating gate. Here, SGD must be higher than 0.7V to turn on the select transistor connected to the grounded bit-line in the case of the "0"-program. On the other hand, in the case of the "1"-program, the bit-line should be higher than 0.5V and SGD is turned off. Therefore, the channel of the "1"-program cell is kept at Vch and the electron injection is

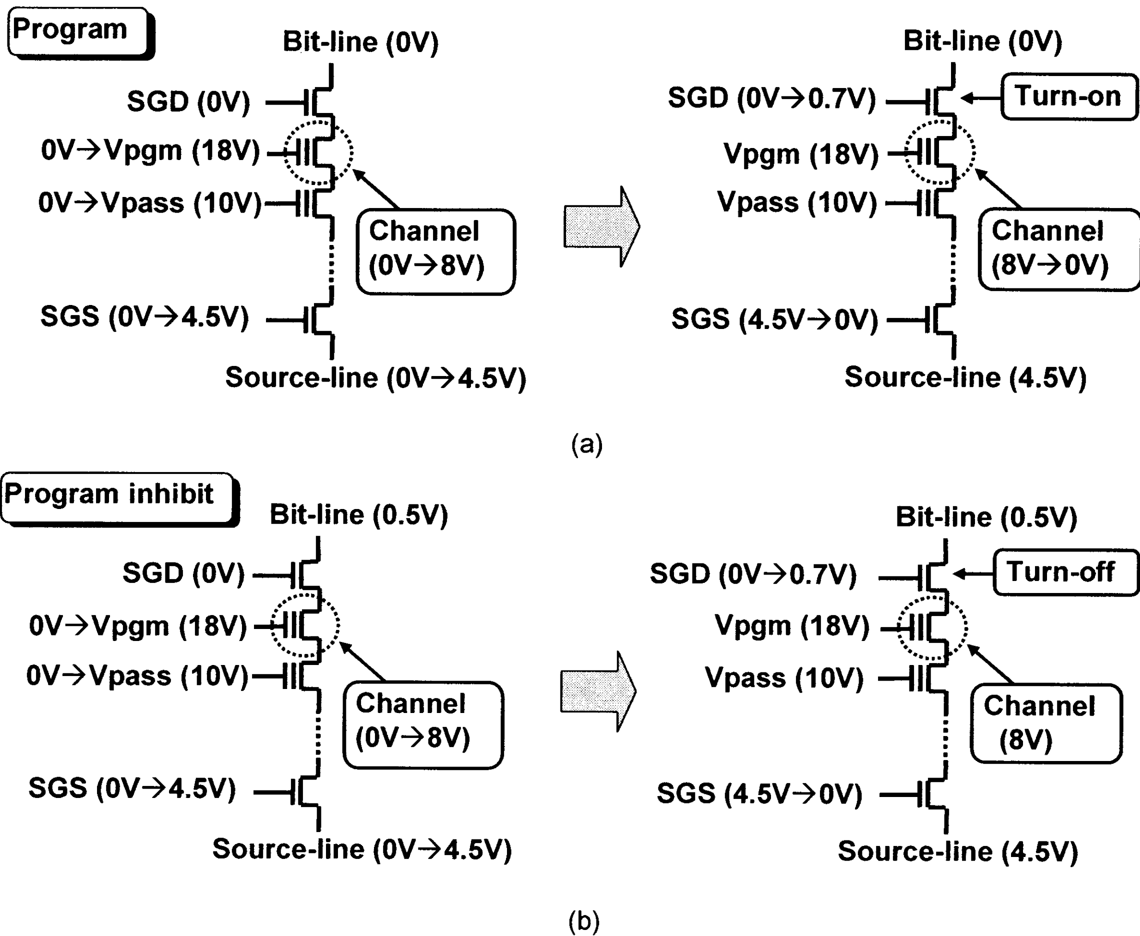


Fig.4.8 (a) The first and (b) the second steps of the proposed program operation. The upper figures show the "0"-program that means the program selected. The lower figures correspond to the "1"-program, that is, the program inhibit operation.

prohibited.

C. Load capacitance reduction

Here, it is explained why a large source-line swing / small bit-line swing adopted in the proposed scheme is preferable. The first requirement is that a higher channel voltage is needed to reduce the program disturbance. Therefore, either the bit-line or the source-line must be raised higher than Vcc. In view of the power consumption, the capacitance to be boosted should be small. Also, as for the program speed, the load capacitance of the charge pump circuit should be small, because the long rising time degrades the program

speed. Therefore, between the bit-line and the source-line, one having smaller capacitance should be boosted higher than Vcc.

Then, which is larger, the bit-line capacitance or the source-line capacitance? Fig.4.9(a) shows an array configuration of a NAND flash memory and Fig.4.9(b) shows a top view of memory cells [18], [19]. The cross point of an active area line (horizontal line) and a polycide word-line (vertical line) forms a memory cell. 16 memory cells are arranged

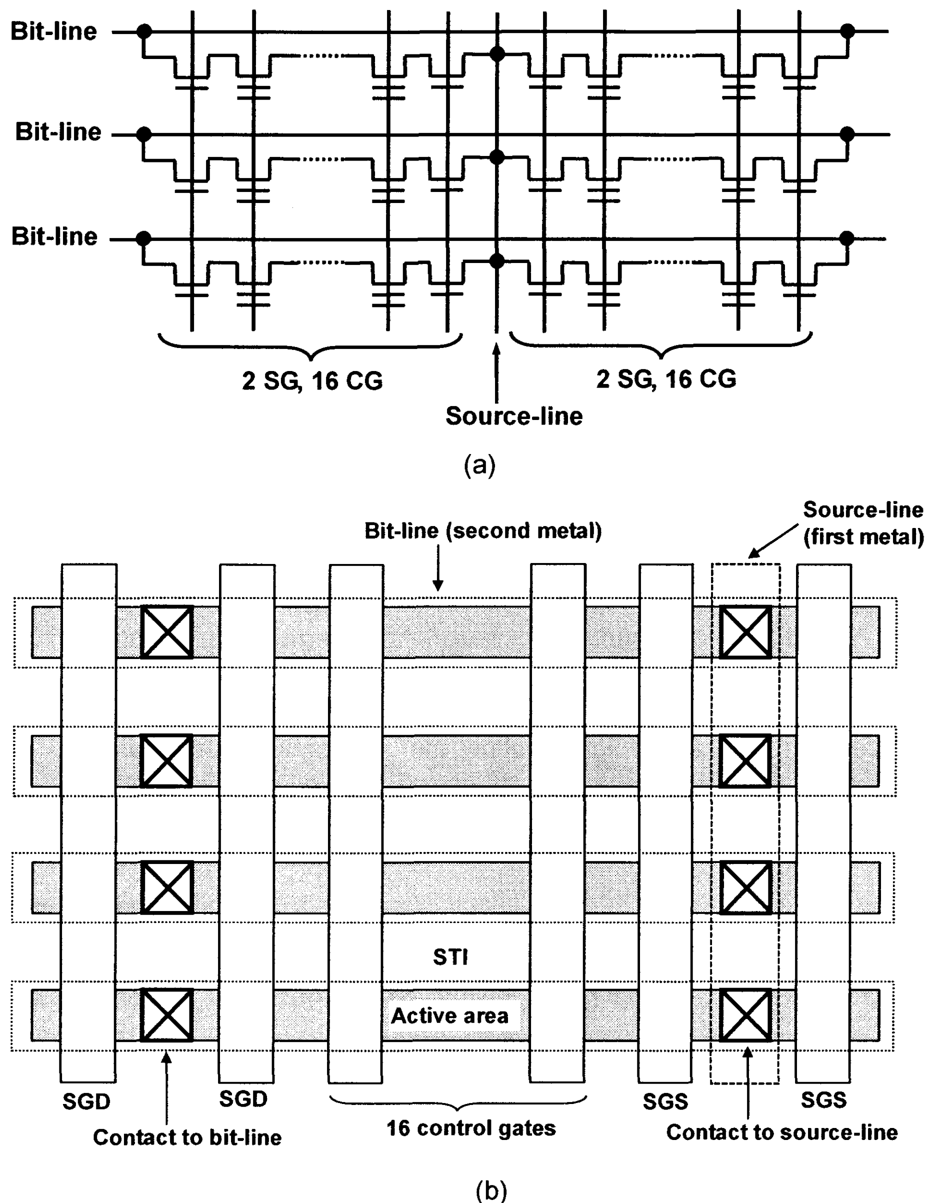


Fig.4.9 (a) A NAND-type cell array. (a) An equivalent circuit and (b) a layout of memory cells [8, 9]. A source-line is arranged every thirty-two control gates and four select gates.

between two select transistors and contacts to the bit-line or the source-line are arranged every eighteen vertical lines. The source diffusion layer is connected to the first metal (W) parallel to the word-line. The drain diffusion layer is connected to the bit-line which consists of the second metal layer parallel to the active area line.

The total bit-line capacitance in a chip, *Cbit-line*, and the total source-line capacitance in a chip, *Csource-line*, are expressed as follows:

$$C_{bit-line} = C_{bit-line_wire} + C_{junction}$$

$$C_{source-line} = C_{source-line_wire} + C_{junction}$$

where *Cbit-line_wire*, *Csource-line_wire*, and *Cjunction* are the bit-line wire capacitance, the source-line wire capacitance and the junction capacitance, respectively. The junction capacitance of the bit-line is the same as that of the source-line because the number of contacts to the memory cells is the same.

What is unique about the NAND flash memory is that the total bit-line wire-capacitance in a chip, *Cbit-line_wire*, is much larger than the total source-line wire capacitance in a chip, *Csource-line_wire*, or the junction capacitance, *Cjunction*. As shown in Fig.4.9(b), the bit-lines cover all memory cells. On the other hand, one source-line is arranged every thirty-two control gates and four select gates. Therefore, the number of the source-line is less than one tenth o f the bit-line. Thus, the total source-line wire capacitance in a chip, *Csource-line_wire*, is much smaller than the total bit-line wire capacitance in a chip, *Cbit-line_wire*. Also, as for the junction capacitance, *Cjunction*, one contact is shared by thirty-two memory cells and four select transistors.

Thus, the junction capacitance, *Cjunction*, is less than one-tenth of the bit-line wire capacitance, *Cbit-line_wire*. Due to this unique characteristic, the source-line capacitance is much smaller than the bit-line capacitance. The total bit-line capacitance is as much as 17nF. The source-line capacitance is 2nF and about one-tenth of the bit-line capacitance. Therefore, in order to achieve excellent program disturbance characteristics, low power consumption and a high programming speed, the source-line boosting / the small bit-line swing is the best choice.

E. Chip architecture

The block diagram of the proposed NAND flash memory is shown in Fig.4.10. The read charge pump circuit (Vread pump) generates an unselected control gate voltage during a read. During a programming, the source-line and SGS are boosted by this read charge pump circuit and no additional charge pump circuit is required. The output voltage of the read charge pump circuit is transferred to the source-line via the high voltage NMOS switch. Therefore, a high-voltage PMOS is not needed. Compared with the conventional

Vcc-bit-line programming scheme, the overhead is just this peripheral high voltage switch and its area increase is just 0.01 percent and negligibly small.

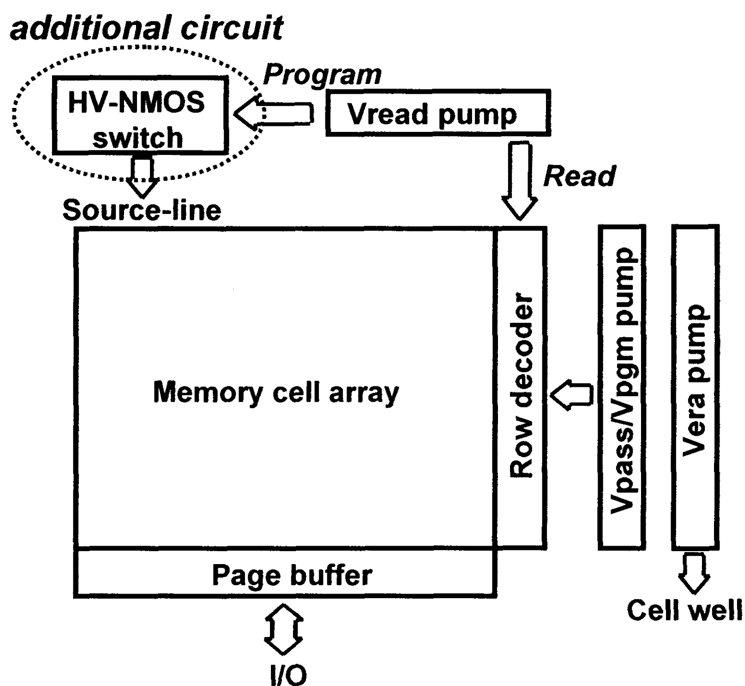


Fig.4.10 A block diagram of the proposed NAND flash memory. In a program operation, a source-line is boosted by the read charge pump circuit (Vread pump) which generates an unselected word-line voltage during a read.

4.4 Circuit implementations

In this section, circuit techniques to realize the proposed program operation are described. The simplified circuit diagram of the page buffer is shown in Fig.4.11. The program data or the data read out from a memory cell is temporarily stored in the page buffer. In order to suppress the inter bit-line capacitive coupling noise, a shielded bit-line sensing method is used [20]. Two bit-lines share one page buffer and are alternately selected during a program and a read operation. During a read operation, an unselected bit-line is grounded and forms a shield between the neighboring selected bit-lines.

A. Program operation

In Fig.4.11, BLa is selected. Selected BLa is biased to 0V or 0.5V, according to the program data stored in the page buffer. Unselected BLb is applied to 0.5V to prohibit the

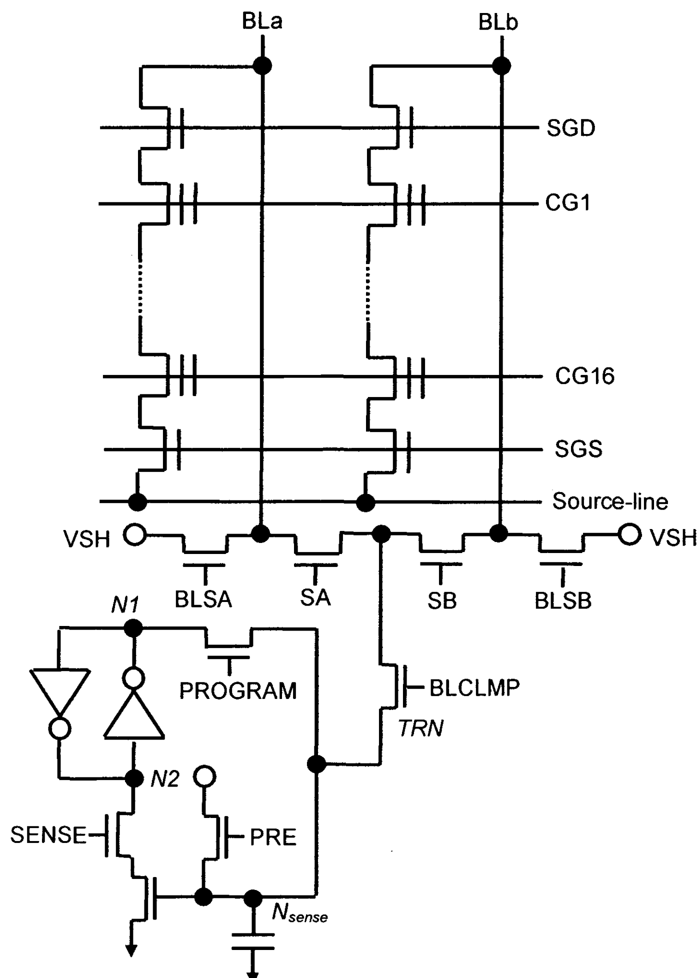


Fig.4.11 A simplified schematic diagram of a page buffer.

electron injection into the floating gate. The SA-signal is biased to Vcc and the BLSA-signal is Vss. The BLCLMP-signal is biased to an intermediate voltage, *Vmid*. *Vmid* is expressed as follows:

$$Vmid = 0.5 + Vthn$$

where Vthn is the threshold voltage of the NMOS, TRN in Fig.4.11. The PROGRAM-signal is applied to Vcc. In the case of the "0"-program, the page buffer node, N1 in Fig.4.11 is Vss and BLa is biased to 0V. As a result, electrons are injected from the grounded channel to the floating gate. If the program data is "1", N1 is Vcc. As the BLCLMP-signal is biased to *Vmid*, BLa is applied to 0.5V.

The voltage regulator to generate *Vmid* is shown in Fig.4.12. *Vref* is 0.5V generated by a band-gap reference circuit. In Fig.4.12, VA is controlled to be 0.5V. As a result, *Vmid* becomes *0.5+Vthn*. By using this voltage regulator, even if the threshold voltage of the NMOS, TRN in Fig.4.11, fluctuates due to temperature or process variation, the bit-line voltage can be unchanged.

As for the unselected bit-line, BLb, the SB-signal is grounded and the BLSB-signal is applied to *Vmid*. The VSH-signal is Vcc. As a result, BLb is applied to the program inhibit voltage, 0.5V and the electron injection into the memory cells connected to BLb is prohibited.

B. Verify read operation

After the program pulse is applied, the verify read operation is performed. The timing chart for the verify read operation is shown in Fig.4.13. Throughout the verify read operation, the PROGRM-signal is Vss. At Tr1, the BLCLMP-signal is biased to 2V using the voltage regulator shown in Fig.4.12. As a result, the selected bit-line, BLa is

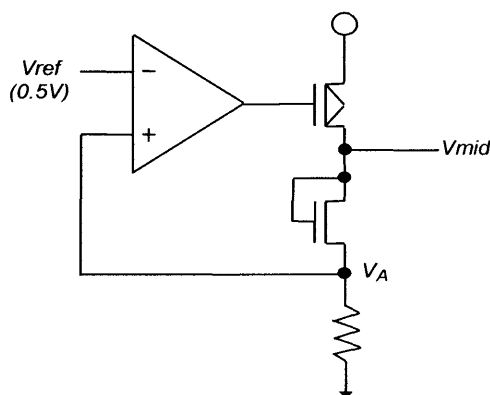


Fig.4.12 An intermediate voltage generator.

precharged to 1V. Then, the selected control gate is applied to 0.5V at Tr2. The unselected control gates and the select gates are Vread (3.5V). If the memory cell is successfully "0"-programmed, the cell read current does not flow and BLa is kept at the precharged voltage. Otherwise, the cell read current flows from BLa to the source-line and the voltage of BLa decreases. At Tr3, the BLCLMP-signal is biased to 1.7V. In case that the memory cell is "0"-programmed, the NMOS, TRN is turned off and the sense node, Nsense is kept at Vcc. In case of the "1"-program or the "0"-program fail, the NMOS, TRN is turned on and the charge at the sense node, Nsense is discharged to BLa. As a result, the voltage of Nsense decreases below 0.5V. Then, the SENSE-signal is activated at Tr4 and the reprogram data is transferred to the page buffer. The reprogram data is modified such that the re-programming is executed only on memory cells that have not been successfully programmed [22].

C. Read operation.

The read operation is the same as the verify read operation except that N1 and N2 in Fig.4.11 are reset to Vss and Vcc, respectively, at Tr1 in Fig.4.13.

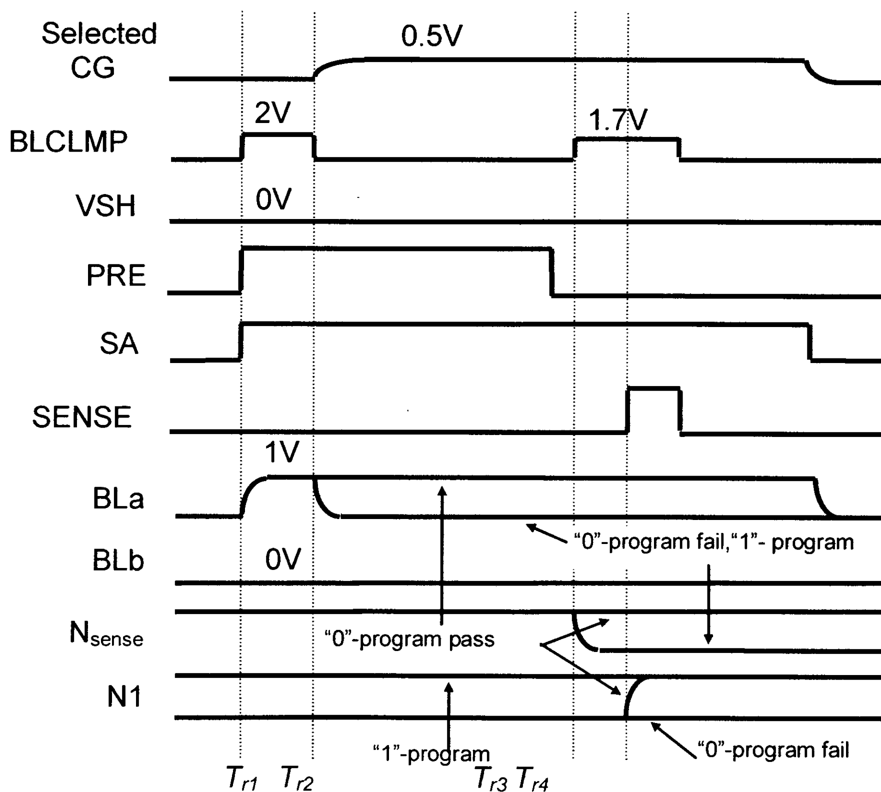


Fig.4.13. Signal timing diagram of the verify read operation.

4.5 Experimental results

A. Improved program disturbance characteristics

In the proposed scheme, a high program inhibit voltage, for example 4.5V, is applied from the source-line to the channel. In this case, Vchi is 4.5-Vthsg and the channel voltage, Vch, is expressed as follows.

$$Vch_proposed. = Vchi + 15\gamma(Vpass - Vth0 - Vchi) + \gamma \cdot Vpgm$$
$$= Vch_conv.1 + (1 - 15\gamma)\cdot(4.5 - Vcc)$$

The value of the parameter "γ" used in this equation is the same as that used in the equation of Vch_conv.1. Much larger Vch reduces the electric field across the tunnel oxide and decreases the Vpgm disturb. As can be seen in Fig.4.4, the Vth shift caused by the program disturbance is greatly decreased. As a result, the allowable program disturb time is long enough, as shown in Fig.4.5 and a highly reliable program operation can be realized irrespective of Vcc.

B. High-speed programming

As discussed in Section 4.3.1, the source-line capacitance is much smaller than the bit-line capacitance. Therefore, the source-line boosting in the proposed program

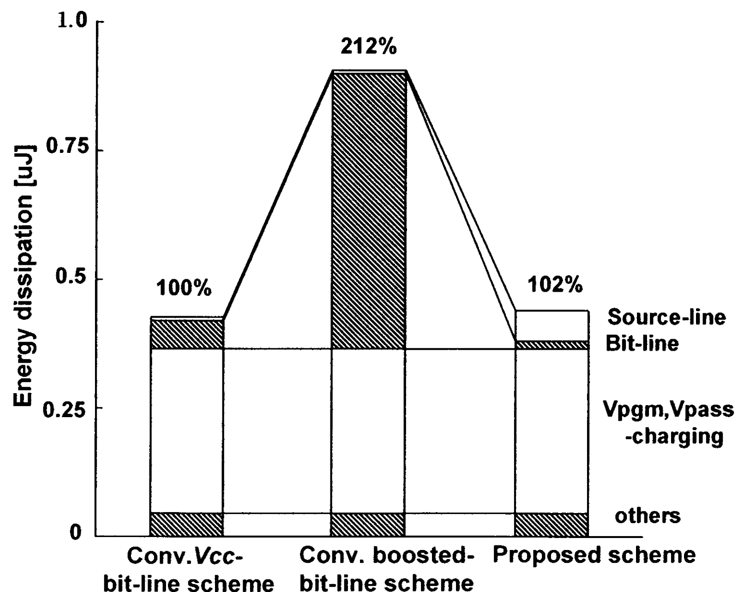


Fig.4.14 Comparison of energy dissipation per programming pulse.

operation can be performed during the setup of the bit-line that corresponds to the interval from T1 to T2 in Fig.4.1(c) and does not degrade the program speed. An additional timing from T2 to T4 increases the programming time by only two percent. Fig.4.6 shows the programming time per page. At each Vcc, the number of stages and the capacitance of the read charge pump circuit are optimized so that the word-lines rise within a predetermined period during a read. In the proposed scheme, a fast programming of 192us/page at 1.4V can be achieved.

4.6 Design considerations

4.6.1 Power consumption

In the proposed program operation, the bit-line voltage can be reduced from Vcc to 0.5V. This drastically decreases the bit-line charging current, as shown in Fig.4.14. Therefore, despite the source-line boosting, the power consumption of the proposed scheme is almost the same as that of the conventional Vcc-bit-line scheme and much energy dissipation at lower Vcc is caused by the Vpgm and Vpass boosting.

4.6.2 Area/Manufacturing cost penalty

As discussed in Section 4.3.1, the overhead is only the peripheral high-voltage switch whose circuit area is just 0.01 percent of the chip area. Then, a high-voltage PMOS is unnecessary. Therefore, the proposed scheme can be realized without any circuit area or manufacturing cost overhead.

The proposed scheme has been realized in a commercial 256Mbit NAND flash memory. The chip has been fabricated using a 0.25um CMOS-STI technology [19], [20] resulting in a die size of 130mm^2 and an effective cell size of 0.29um^2. Key technology parameters are summarized in Table 4.2. Fig.4.15 shows the micrograph of the proposed 256Mbit NAND flash memory chip. The measured waveforms of the proposed program operation are shown in Fig.4.16. There is actually no area increase in comparison with the conventional 256Mbit NAND flash memory [21].

Technology	**: 0.25um p-sub double-well CMOS**
	1-poly, 1-polycide, 3-metal
	Self-aligned shallow trench isolation
Cell size	**: 0.29um^2**
Chip size	**: 130mm^2**
Tunnel oxide	**: 9nm**
Inter-poly dielectric	**: 14nm (effective)**
Gate oxide	**: 40nm (High voltage)**
	9nm (Low voltage)

Table 4.2 Process parameters

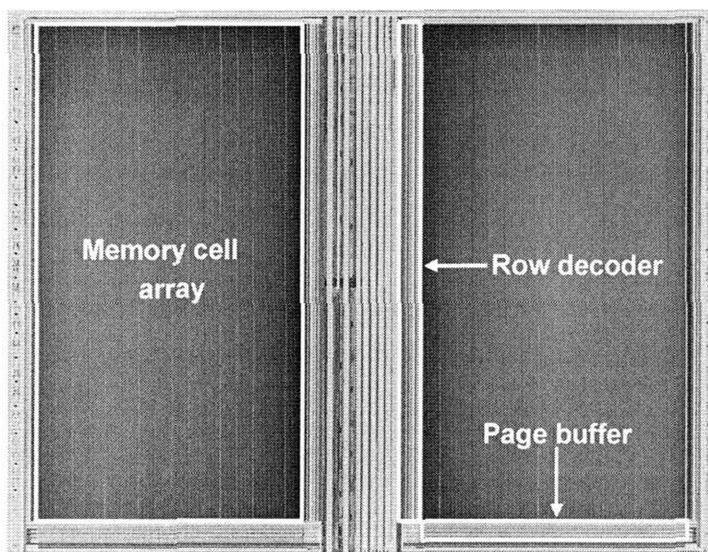


Fig.4.15 A micrograph of the proposed 256Mbit NAND flash memory chip.

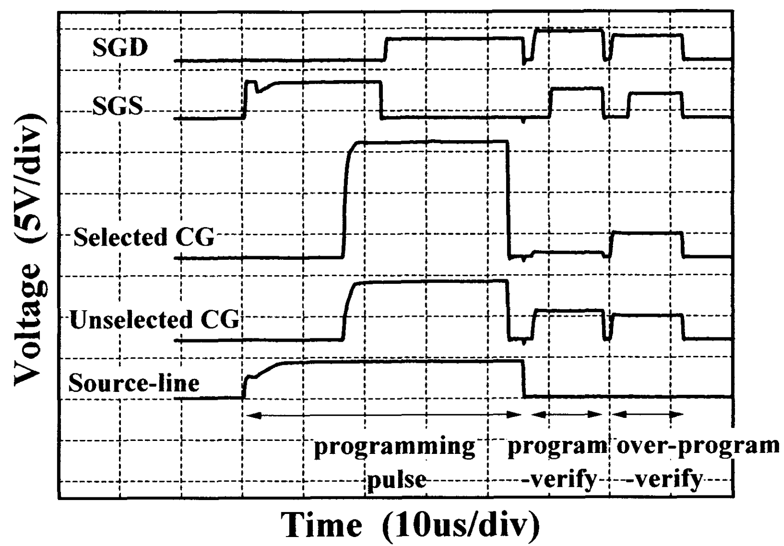


Fig.4.16 Measured waveforms of the proposed program operation. After a programming pulse, a verify read operation and an over-program verify read operation are performed.

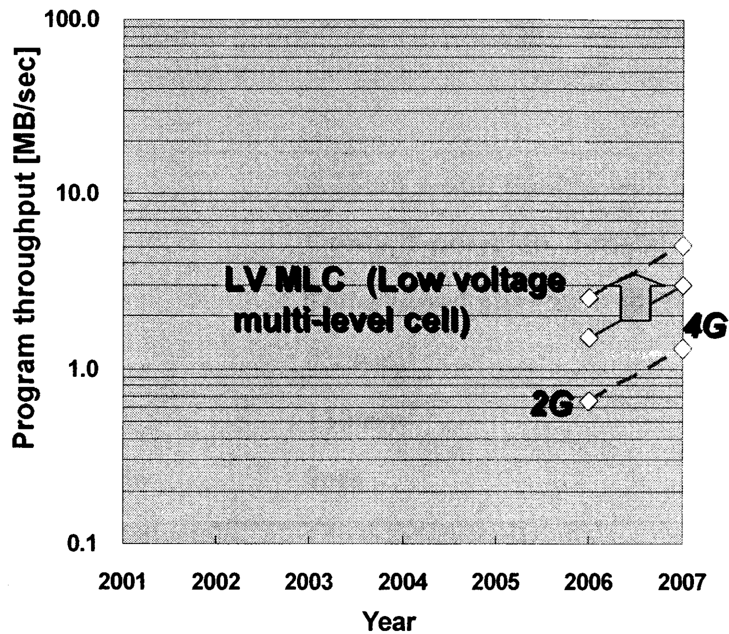


Fig.4.17 Program speed improvement with the low load capacitance technology.

4.7 Summary

A low load capacitance technology is proposed. The proposed architecture shows excellent program disturbance characteristics irrespective of the supply voltage. The program speed is improved by 70%. In addition, small energy dissipation and low cost are achieved at the same time. Fig.4.17 shows the program speed trend of the low-voltage multi-level cell. By combining the multi-page cell technologies described in Chapter 3 and the low load capacitance technology, the 2.5MB/sec low-voltage multi-level NAND flash memory can be realized.

The proposed scheme is verified with commercial 0.25um 256Mbit NAND flash memory. The new architecture will make it possible to realize a commercial low-voltage multi-level NAND flash memory.

References

[1] S. Atsumi, A. Umezawa, T. Tanzawa, T. Taura, H.; Shiga, Y. Takano, T. Miyaba, M. Matsui, H. Watanabe, K. Isobe, S. Kitamura, S. Yamada, M. Saito, S. Mori, and T. Watanabe, "A channel-erasing 1.8 V-only 32 Mb NOR flash EEPROM with a bit-line direct-sensing scheme", in *ISSCC Digest of Technical Papers*, Feb., 2000, pp.276-277.

[2] M. Mihara, Y. Miyawaki, O. Ishizaki, T. Hayasaka, K. Kobayashi, T. Omae, H. Kimura, S. Shimizu, H. Makimoto, Y. Kawajiri, M. Wada, H. Sonoyama and J. Etoh, "A 29 mm^2 1.8V-only 16Mb DINOR Flash Memory with Gate-Protected Poly-Diode (GPPD) Charge Pump," in *ISSCC Digest of Technical Papers*, Feb., 1999, pp.114-116.

[3] B. Pathak, A. Cabrera, G. Christensen, A. Darwish, M. Goldman, R. Haque, J. Jorgensen, R. Kajley, T. Ly, F. Marvin, S. Monasa, Q. Nguyen, D. Pierce, A. Sendrowski, I. Sharif, H. Shimoyoshi, A. Smidt, R. Sundaram, M. Taub, W. Tran, R. Trivedi, P. Walimbe and E. Yu, "A 1.8 V 64 Mb 100 MHz flexible read while write flash memory", in *ISSCC Digest of Technical Papers*, Feb., 2001, pp.32-33.

[4] S. Aritome, R. Shirota, R. Kirisawa, T. Endoh, R. Nakayama, K. Sakui, and F. Masuoka, "A reliable bi-polarity write/erase technology in flash EEPROMs", in *IEDM Tech. Dig.,* 1990, pp.111-114.

[5] S. Satoh, T. Nakamura, K. Shimizu, K. Takeuchi, H. Iizuka, S. Aritome, and R. Shirota, "A novel gate-offset NAND cell (GOC-NAND) technology suitable for high-density and low-voltage-operation flash memories", in *IEDM Tech. Dig.,* 1999, pp.271-274.

[6] A. Goda, W. Moriyama, H. Hazama, H. Iizuka, K. Shimizu, S. Aritome, and R. Shirota, "A novel surface-oxidized barrier-SiN cell technology to improve endurance," in *IEDM Tech. Dig.,* 2000, pp.771-774.

[7] R. Kirisawa, S. Aritome, R. Nakayama, T. Endoh, R. Shirota and F. Masuoka, "A NAND structured cell with a new programming technology for highly reliable 5 V-only flash EEPROM," in *Symp. VLSI Technology Dig. Tech. Papers*, June 1990, pp.129-130.

[8] S. Shuto, S. Yamada, S. Aritome, T. Watanabe and K. Hashimoto, "Read disturb degradation mechanism for source erase flash memories," in *Symp. VLSI Technology Dig. Tech. Papers*, June 1996, pp.242-243.

[9] K. D. Suh, B. H. Suh, Y. H. Lim, J. K. Kim, Y. J. Choi, Y. N. Koh, S. S. Lee, S. C. Kwon, B. S. Choi, J. S. Yum, J. H. Choi, J. R. Kim, and H. K. Lim, "A 3.3V 32Mb NAND flash memory with incremental step pulse programming scheme," in *ISSCC Digest of Technical Papers*, Feb. 1995, pp.128-129.

[10] T. S. Jung, D. C. Choi, S. H. Cho, M. J. Kim, S. K. Lee, B. S. Choi, J. S. Yum, S. H. Kim, D. G. Lee, J. C. Son, M. S. Young, H. K. Oh, S. B. Jun, W. M. Lee, E. Haq, K. D. Suh, S. Ali and H. K. Lim, "A 3.3V 16Mb Non-Volatile Virtual DRAM using a NAND Flash Memory Technology", *ISSCC Digest of Technical Papers*, 1997, pp.398-399.

[11] K. Takeuchi, S. Satoh, K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, T. Ikehashi, K.

Kanda, K. Hosono and K. Sakui, "A source-line programming scheme for low voltage operation NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1999, pp.37-38.

[12] K. Takeuchi, S. Satoh, K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, T. Ikehashi, K. Kanda, K. Hosono and K. Sakui, "A source-line programming scheme for low voltage operation NAND flash memories," *IEEE Journal of Solid-State Circuits*, vol.35 no.5, May 2000, pp.672-681.

[13] K. Imamiya, Y. Iwata, Y. Sugiura, H. Nakamura, H. Oodaira, M. Momodomi, Y. Ito, T. Watanabe, H. Araki, K. Narita, K. Masuda and J. Miyamoto, "A 35 ns-cycle-time 3.3 V-only 32 Mb NAND flash EEPROM", in *ISSCC Digest of Technical Papers*, Feb., 1995, pp.32-33.

[14] Tae-Sung Jung, Young-Joon Choi, Kang-Deog Suh, Byung-Hoon Suh, Jin-Ki Kim, Young-Ho Lim, Yong-Nam Koh, Jong-Wook Park, Ki-Jong Lee, Jung-Hoon Park, Kee-Tae Park, Jang-Rae Kim, Jeong-Hyong Lee and Hyung-Kyu Lim, "A 3.3 V 128 Mb multi-level NAND flash memory for mass storage application", in *ISSCC Digest of Technical Papers*, Feb., 1996, pp.32-33.

[15] J.K. Kim, K. Sakui, S.S. Lee, J. Itoh, S.C. Kwon, K. Kanazawa, J.J. Lee, H. Nakamura, K.Y. Kim, T. Himeno, J.R. Kim, K. Kanda, T.S. Jung, Y. Oshima, K.D. Suh, K. Hashimoto, S.T. Ahn, and J. Miyamoto, "A 120 mm^2 64 Mb NAND flash memory achieving 180 ns/byte effective program speed," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1996, pp.168-169.

[16] S. Satoh, H. Hagiwara, T. Tanzawa, K. Takeuchi, and R. Shirota, "A novel isolation-scaling technology for NAND EEPROMs with the minimized program disturbance," in *IEDM Tech. Dig.*, 1997, pp.291-294.

[17] K. Takeuchi, S. Satoh, T. Tanaka, K. Imamiya, and K. Sakui, " A negative Vth cell architecture for highly scalable, excellently noise immune and highly reliable NAND flash memories," *IEEE J. Solid-State Circuits*, vol.34, no.5, May 1999, pp.675-684.

[18] K. Shimizu, K. Narita, H. Watanabe, E. Kamiya, Y. Takeuchi, T. Yaegashi, S. Aritome, and T. Watanabe, " A novel high-density $5F^2$ NAND STI cell technology suitable for 256Mbit and 1Gbit flash memories", *IEDM Dig. Tech. Papers*, 1997, pp.271-274.

[19] Y. Takeuchi, K. Shimizu, K. Narita, E. Kamiya, T. Yaegashi, K. Amemiya and S. Aritome, "A Self-Aligned STI Process Integration for Low Cost and Highly Reliable 1Gbit Flash Memories", in *Symp. VLSI Technology Dig. Tech. Papers*, June 1998, pp.102-103.

[20] T. Tanaka, Y. Tanaka, H. Nakamura, K. Sakui, H. Oodaira, R. Shirota, K. Ohuchi, F. Masuoka, and H. Hara, "A quick intelligent programming architecture and a shielded bit-line sensing method for 3V-only NAND flash memory," *IEEE J. Solid-State Circuits*, vol.29, November 1994, pp.1366-1373.

[21] K. Imamiya, Y. Sugiura, H. Nakamura, T. Himeno, K. Takeuchi, T. Ikehashi, K. Kanda, K. Hosono, R. Shirota, S. Aritome, K. Shimizu, K. Hatakeyama and K. Sakui " A 130mm^2

256Mb NAND Flash with Shallow Trench Isolation Technology", *ISSCC Digest of Technical Papers*, 1999, pp.112-113.