Copyright © 1988, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

MULTIPORT MEMORY DESIGN CONSIDERATIONS FOR PARALLEL EXECUTION CPU ARCHITECTURES

by

Y. Nakagome, G.A. Uvieghara, and D.A. Hodges

Memorandum No. UCB/ERL M88/83

22 December 1988

MULTIPORT MEMORY DESIGN CONSIDERATIONS FOR PARALLEL EXECUTION CPU ARCHITECTURES

by

Y. Nakagome*, G.A. Uvieghara, and D.A. Hodges

*Permanent Address: Hitachi Central Research Laboratory
P.O. Box 2, Kokubunji
Tokyo, 185, Japan

Memorandum No. UCB/ERL M88/83

22 December 1988

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TITLE PAGE

ABSTRACT

Parallel execution architecture is one of the efficient ways to increase CPU performance. High bandwidth communications between multiple execution units and on-chip memories is vital to implementing parallel execution architectures on a single chip. Multiport memory is one of the feasible methods to increase the communication bandwidth. This report treats the design considerations for multiport memories based on a multiport memory cell having more than two ports.

The key design issue is to realize a highly stable and high speed memory with a limited cost (area). Analytic expressions of noise margins for possible multiport memory cell configurations are presented as a measure of stability. Analyses have been verified through SPICE simulations. Optimal design of multiport memory cell has become possible through these analyses. Active area of multiport memories has also been studied. It has been found that the pseudo-static single-ended access cell has better performance compared to other configurations.

It has been revealed that a sensitive and high-speed sense amplifier is needed for the efficient design of multiport memories. Class-A push-pull current sense amplifier shows better performance in the case where bitlines are highly capacitive.

These studies show that the pseudo-static single-ended multiport memory cell is applicable to an on-chip multiport memory having more than two ports.

ACKNOWLEDGEMENT

We would like to thank Deog K. Jeong and Daebum Lee for their helpful advice and for providing MOSFET model parameters.

This work was supported by grants from Hitachi Central Research Laboratory. Thanks go to Dr. Yasutsugu Takeda, Dr. Shojiro Asai, Dr. Toshiaki Masuhara, and Mr. Yuji Kuwahara for supporting this research work.

Table of Contents

CHAPTER 1 INTRODUCTION 2
1.1 Background 2
1.2 Research Objective 3
1.3 Organization of The Report 3

CHAPTER 2 MULTIPORT MEMORY ARCHITECTURES 4
2.1 Multiple Function Units System 4
2.2 Multiport Memory 5

CHAPTER 3 DESIGN OF MULTIPORT MEMORY CELL 7
3.1 Possible Multiport Memory Cell Configurations 7
3.2 Stability Analysis 9
3.2.1 Static Noise Margin 9
3.2.2 Derivation of Static Noise Margin 10
3.2.2.1 Differentially Accessed Cell 10
3.2.2.2 Single-ended Access Cell 12
3.2.2.3 Modified Single-ended Access Cell 15
3.2.3 Verification of Analysis 15
3.2.3.1 SPICE Simulation 15
3.2.3.2 Test Circuit for Noise Margin Measurement 17
3.3 Optimal Design of Multiport Memories 18
3.3.1 Differentially Accessed Cell 18
3.3.2 Single-ended Access Cell 19
3.3.3 Modified Single-ended Access Cell 20
3.3.4 Comparison of Noise Margins 21
3.4 Active Area Estimation 21
3.4.1 Scale Up Factor 22
3.4.1.1 Differentially Accessed Cell 23
3.4.1.2 Single-ended Access Cell 24
3.4.2 Normalized Active Area 26
3.4.2.1 Differentially Accessed Cell 26
3.4.2.2 Single-ended Access Cell 27
3.4.2.3 The Impact of Dummy Cell on Active Area 28

3.4.3 Discussion 30
3.5 Choice of Multiport Memory Cell 31

CHAPTER 4 SENSE AMPLIFIER DESIGN 33
4.1 Possible Sense Circuit Configurations 33
4.1.1 Voltage Sense Amplifiers 33
4.1.2 Current Sense Amplifiers 34
4.2 Analysis and Design of Sense Amplifiers 35
4.2.1 Cross-coupled Dynamic Sense Amplifier 35
4.2.2 Transresistance Sense Amplifier 38
4.2.3 Class-A Push-Pull Current Sense Amplifier 40
4.3 Discussion 43

CHAPTER 5 CONCLUSION 45

APPENDIX 1 DERIVATION OF STATIC NOISE MARGINS FOR DIFFERENTIALLY ACCESSED CELL 47

APPENDIX 2 DERIVATION OF STATIC NOISE MARGINS FOR SINGLE-ENDED ACCESS CELL 50

REFERENCES 56

CHAPTER 1 INTRODUCTION

1.1 Background

A very efficient way to increase CPU performance is by exploiting parallelism using multiple function units. A fast and dense register file, which enables high bandwidth on-chip communications, is indispensable to implementing parallel-execution CPU architectures on a single chip. High bandwidth communications can be achieved through (1) employing time multiplexed read and write operations to perform multiple operations within a single clock cycle, or (2) employing a multiport memory to support concurrent read and/or write operations. Combining these two methods is a well-known technique in RISC (Reduced Instruction Set Computer) CPU chips [1]. The referenced example performs two-port read and single-port write in a time-multiplexed manner by using a dual-port register file.

Multiport memories with more than two ports are increasingly important in parallel-execution CPU architectures, since the time multiplexed approach may face speed constraints posed by some difficulties in reducing access time with a given process technology.

A few works to implement a multiport memory on a single chip have been reported. One example is the parallel image processor using a four-port (two ports for reading and other two ports for writing) register file[2]. A multiport memory will be needed for implementing multi-instruction set computers such as VLIW (Very Long Instruction Word) architectures[3]. Another demand is coming from data-flow computer architectures such as HPSm (High Performance Substrate micro-architecture) which was proposed and developed at the Computer Science Division of University of California at Berkeley[4][5].

HPSm needs a special memory which performs 6-READs, 3-WRITEs and 6-ASSOCIATIVE WRITEs in a single clock cycle[6][7]. A research project to implement this architecture on a single chip using a smart triple-port register file is under progress. Other design difficulties for smart memory come from supporting out-of-order execution and handling of branch prediction and exception. There will be separate reports regarding these issues.

1.2 Research Objective

The purpose of this report is to clarify the design issues of multiport memory which have not been touched on by past works. Of special concern is the feasibility of multiport memory with more than two ports. The primary task of this report is:

(1) to give quantitative measures of the performance of the possible multiport memories for design optimization,

(2) to choose the promising multiport memory cell configuration based on cost comparisons, and

(3) to present the promising sense amplifier circuits suitable to the multiport memory chosen.

1.3 Organization of The Report

A macroscopic view of multiport memory architectures is presented in Chapter 2. In Chapter 3, stability analyses of multiport memories are presented, and optimal design of multiport memory is described. In Chapter 4 designs of fast and low-power sense amplifiers for multiport memories are presented. The conclusions are presented in Chapter 5.

CHAPTER 2 MULTIPORT MEMORY ARCHITECTURES

2.1 Multiple Function Units System

There are two possible memory architectures to support multiple function units (MFU). Figure 2.1(a) shows an approach using Multiple Single-Port Register Files (MSRF). Another approach is to use a Multiport Register File (MRF) as shown in Figure 2.1(b). The latter approach calls for a new multiport memory cell for the core of the memory.

The MSRF approach has the following advantages :

(1) Design of the register file is easy since it can make use of the conventional register file design.

(2) It operates fast enough because each register file has small area by the use of a compact cell structure.

On the other hand, it suffers from the following disadvantages:

(1) It consumes large area (area is proportional to the number of the ports :N).

(2) Power dissipation is proportional to the number of ports.

(3) It needs special controls to maintain data consistency for writing data back to the register files.

The MRF approach has the following advantages due to having a single cell array :

(1) It has a potential to reduce both area and power dissipation.

(2) It does not need special attention to data consistency, and with proper safeguards it is capable of multiport writing.

The disadvantages of MRF are as follows :

(1) It requires a new multiport memory cell and assurance that it has enough operating

margins in all possible operating conditions.

(2) It needs special efforts in layout design style to surmount its irregularity.

(3) Access time increases unless we employ special circuit design techniques, since the memory core occupies large area and signal lines are highly capacitive.

Multiport register file has a potential for achieving good performance per cost. But the design complexity problems mentioned above are not well understood. In order to form a fair judgement we have to know the performance/cost tradeoffs of the MRF quantitatively. Thus, in this report we focus on the MRF and study its performance and cost quantitatively.

2.2 Multiport Memory

Figure 2.2 shows the basic configuration of an N-port Multiport Register File. A memory cell consists of a storage element and N switches. N word lines control these switches to establish electrical connection between a storage element and N bitlines. This configuration allows concurrent read operations from N registers and concurrent write operations into N registers (and even their combinations). This is a simple and understandable configuration, but it has some drawbacks as follows:

(1) Read operation may be slow since a limited cell current must be shared among N bitlines in the case of N-port reading.

(2) It needs careful design to assure the stability of the storage element because electrical connection between storage element and bitlines is strong in the case of N-port reading.

If we can surmount these design difficulties, this configuration appears to be the most promising.

Figure 2.3 shows one of the other possible configurations to improve stability and speed for multiport reading. The feature of this configuration is to use two sets of N switches each for reading and writing. The two sets of switches are controlled by two sets of N word lines (read word lines and write word lines). The storage element controls N read switches together with N read word lines for the read operation, and is directly connected to bitlines through one of the write switches for the write operation. This configuration allows a faster and more stable read operation than the simple configuration shown in Figure 2.2. The disadvantage of this cell is that it needs large area since the number of transistors and signal lines increase. The philosophy of this circuit comes from that of the three transistor DRAM (Dynamic Random Access Memory) cell.

There is another possible configuration that prevents multiport read operation. An example is shown in Figure 2.4. The storage element in this case is the same one used for the basic configuration shown in Figure 2.2. The feature of this technique is to inhibit multiple word line selection by the use of a match detection circuit which sense the occurrence of match in decoder outputs and generates transfer signals to control the data transfer circuits located next to the I/O buffers. The data transfer circuits copy back the data to the ports inhibited. The stability and speed of the storage element can be improved through this method. On the other hand, it needs complicated control for inhibiting word lines and generating transfer signals. Furthermore, special attention must be paid to the increase of delay caused by inserting the above circuits into the critical path.

This report treats the analysis and design of the basic configuration shown in Figure 2.2. The feasibility of the other two configurations has not yet been studied.

CHAPTER 3 DESIGN OF MULTIPORT MEMORY CELL

This chapter discusses, first, possible multiport memory cell configurations. Next in section 3.2, we derive the static noise margins of the possible cell configurations, which is a useful measure of cell stability. In section 3.3, the optimized design for each cell configuration is presented. In section 3.4, the active area for each cell configuration is calculated to evaluate the cost. Finally, we discuss the selection of multiport memory cell based on the above analysis.

3.1 Possible Multiport Memory Cell Configurations

As far as dual-port memories are concerned, several kinds of cell configurations have been proposed. These are classified into two configurations. One of these is the differentially accessed cell configuration using two pairs of bitlines for data transmission[8]. Another is the single-ended access cell configuration using two single bitlines for data transmission[1,9]. Using the analogy of dual-port memories, we address two possible cell configurations for multiport memories with more than two ports.

Figure 3.1 shows the differentially accessed multiport memory cell. The storage element consists of the conventional cross-coupled inverter pair. Two sets of N access transistors control the connection between N bitline pairs and two internal storage nodes. The major advantage of this cell is high signal/noise ratio since this configuration allows cancellation of the common mode coupling noise on the bitline pairs. Another advantage is high speed by utilizing a sensitive and high speed differential sense amplifier. On the contrary, it suffers from a relatively large area due to an increase of the number of signal lines (mainly

bitlines).

Figure 3.2 shows a multiport single-ended access cell. It needs the minimum number of access transistors and bitlines, and has a potential of a more compact cell structure compared to the other configurations. Furthermore, it is relatively stable compared with the differential sense amplifier since one of the inverters is unloaded and can maintain its nonlinearity during read operations. But, at the same time, it makes the write operation difficult. So, special circuit techniques such as those listed below are essential for the write operation:

(1) Write operation is carried out in a differential manner by using both access transistors for the write operation[1].

(2) Boosting a word line for the write operation[10].

The drawback of the first approach is that it can not perform multiport writing. The second approach suffers from complicated design since it needs a three level driver circuit. Another possible method is to use a clocked inverter for the storage element to make it operate in a pseudo-static fashion. Figure 3.2 (b) shows a pseudo-static single-ended cell with one-sided clocking. Because it is more difficult to write "1" than to write "0", this configuration seems to be reasonable. Figure 3.2 (c) shows a pseudo-static single-ended cell with two-sided clocking for a safer write operation.

Figure 3.3 (a) shows a modified version of the single-ended cell. One bitline is located at each side of the cell. This configuration can reduce the electrical connectivity between bitlines and the internal storage node, so the stability is expected to improve. Figure 3.3 (b) is a practical version of the modified single-ended cell, which improves the write operation by the use of a disabling clock. The complexity of this cell arises from the need to select

one of the two disabling clocks, depending on which bitline is selected.

3.2 Stability Analysis

In this section, derivations of static noise margins for the different cell configurations addressed above are presented. We also present the results of verification through SPICE simulation.

3.2.1 Static Noise Margin

Detailed analyses of static noise margin in SRAM (Static Random Access Memory) cells have been reported by E.Seevinck et.al.[11][12][13]. We make use of their definition of static noise margin and the mathematical expressions. Moreover, the concept of write noise margin is introduced for the complete analysis of cell operation. At first, we explain the definition of read and write noise margins in a differentially accessed cell and in a single-ended access cell.

Figure 3.4 (a) shows a circuit diagram of a differentially accessed cell for the read operation. For simplicity, we assume that the bitline voltage is clamped at the supply voltage, V_{dd}. Voltage sources V_n denotes the static noise sources. These static noise sources represent the variations of transistor parameters such as transconductance, β, or threshold voltage, V_T. Figure 3.4 (b) shows the transfer characteristics of the cross-coupled inverter pair (flip-flop) when V_n=0 Assume that the initial state of the flip-flop is $V_1<V_2$ as shown in the figure. Figure 3.4 (c) shows the transfer characteristics at the onset of being flipped by increasing the noise voltage, V_n. The amount of noise voltage corresponds to the length of the side of the maximum possible square. This means that if the noise voltage is smaller than the length of the side of the square in the Figure 3.4(c) , the state of the flip-flop is

stable. Here, we define the Read Noise Margin (*RNM*) as the length of the side of the maximum possible square.

The Write Noise Margin (*WNM*) in a differentially accessed cell can be defined by analogy. Figure 3.5 (a) shows a circuit diagram for the write operation. Assume that the initial state of the flip-flop is expressed as $V_1<V_2$. Figure 3.5 (b) shows the transfer characteristics when $V_n=0$. The write noise margin is defined as the length of the side of the minimum possible square. This means that if the amount of noise voltage is smaller than the length of the side of the square the state of flip-flop can be flipped.

The static noise margins in a single-ended access cell can be defined in the same manner. Figure 3.6 (a) shows the definition of Read Noise Margins (*RNM*s). Note that the *RNM* for the "1" read operation is different from that for the "0" read operation. RNM_H denotes read noise margin for reading "1" and RNM_L denotes read noise margin for reading "0", respectively.

Figure 3.6 (b) shows the definition of Write Noise Margins (*WNM*s) in the single-ended access cell. V_{dd} is applied to the bitline for writing "1" while V_{ss} is applied for writing "0". Note that the nonlinearity of Inverter 2 is not degraded in any operating condition since it is not loaded.

3.2.2 Derivation of Static Noise Margin

3.2.2.1 Differentially Accessed Cell

Mathematical expressions of static noise margin of SRAM cells are presented by E.Seevinck et.al.[11]. Based on their analysis, we derive a Read Noise Margin (RNM) for multiport (N-port) reading and Write Noise Margin (WNM) for single-port writing in a dif-

ferentially accessed cell.

Figure 3.7 (a) shows a circuit schematic for the N-port read operation. Assume here that level "0" is stored at the output of Inverter 1 (INV1) and level "1" is stored at the output of INV2. The two transistors Q1p and Q2a can be ignored since they are either cutting off or weakly conducting. Q1a represents N access transistors, so its transconductance becomes $N \times \beta_{na}$. For simplicity, we make the following assumptions:

(1) I-V characteristics of each MOSFET can be expressed by the basic MOS model equations.

(2) Threshold voltage, V_T, and transconductance, β, of each MOSFET is not modulated by drain-source voltage, V_{DS}, or substrate bias, V_{BB}.

(3) Threshold voltage of an N-channel MOSFET is equal to the absolute value of that of a P-channel MOSFET.

Read Noise Margin (RNM) for a differentially accessed cell is expressed as,

$$RNM = V_T - \frac{1}{k+1}\left[\frac{V_{dd} - \frac{2r+N}{r+N}V_T}{1 + \frac{r}{k(r+N)}} - \frac{V_{dd} - 2V_T}{1 + \frac{k}{p} + [\frac{1}{p}(1+2k+\frac{k^2}{p})]^{1/2}}\right] \qquad (3.1)$$

where,

$$r = \frac{\beta_{nd}}{\beta_{na}}, p = \frac{\beta_{pd}}{\beta_{nd}}$$

$$k = \frac{r}{r+N}\left[\sqrt{\frac{(r+N)V_r^2}{(r+N)V_r^2 - NV_s^2}} - 1\right]$$

$$V_s = V_{dd} - V_T$$

$$V_r = V_{dd} - \frac{2r+N}{r+N}V_T.$$

Figure 3.7 (b) shows a circuit schematic for the single-port write operation. Here, we ignore Q1p and Q2n since their contribution to the circuit operation is negligible. The

detailed derivation is given in APPENDIX 1. Write Noise Margin (WNM) can be expressed as follows,

$$WNM = \frac{1}{k+1}\left[\frac{\sqrt{1+r(2k+1)}-1}{r}(V_{dd}-V_T)-(k-1)V_T-\frac{(\frac{rp}{k}-1)V_{dd}+(2rp+1)V_T}{rp-1+\frac{rp}{k}}\right] \quad (3.2)$$

where, (i) if $rp = 1$:

$$k = \frac{V_{dd}(V_{dd}-2V_T)}{2(V_{dd}-3V_T)^2}$$

or, (ii) if $rp \neq 1$:

$$k = \frac{rp}{rp-1}\left[\frac{V_r+(rp-1)V_{dd}}{[V_r^2+rp(rp-1)V_p^2]^{1/2}}-1\right]$$

$$V_r = V_{dd}-(2rp+1)V_T$$

$$V_p = \sqrt{V_{dd}(V_{dd}-4V_T)}$$

3.2.2.2 Single-ended Access Cell

Figure 3.8 shows a circuit schematic for the N-port read operation. V_{bl} in this figure denotes the bitline voltage for the read operation. As was mentioned before, the read noise margins are different for the "0" and "1" read operations, so the following analysis is carried out for these two cases. Please see APPENDIX 2 for the detailed derivation.

We ignore the transistor Q1p for the "0" read operation as shown in Figure 3.8 (a). And it is assumed that the access transistor Q1a operates in the linear region, that is $V_{bl} < V_{dd} - V_T$. The read noise margin for the N-port "0" reading can be expressed as,

$$RNM_L = V_T - \frac{1}{k+1}\left[\frac{V_{dd}-\frac{2r+N}{r+N}V_T}{1+\frac{r}{k(r+N)}} - \frac{V_{dd}-2V_T}{1+\frac{k}{q}+[\frac{1}{q}(1+2k+\frac{k^2}{q})]^{1/2}}\right] \quad (3.3)$$

where,

$$r = \frac{\beta_{nd}}{\beta_{na}}, q = \frac{\beta_{pf}}{\beta_{nf}}$$

$$k = \frac{r}{r+N}\left[\sqrt{\frac{(r+N)V_r^2}{(r+N)V_r^2 - NV_s^2}} - 1\right]$$

$$V_s = \sqrt{V_{bl}(2V_{dd} - 2V_T - V_{bl})}$$

$$V_r = V_{dd} - \frac{2r+N}{r+N}V_T$$

In the case of reading "1", it is determined by the value of the bitline voltage, V_{bl}, and beta ratio, rp/N, whether Q1a operates in the saturation region or not. So, analysis is carried out for these two cases. The bitline voltage to give the boundary between the saturation and linear region can be expressed as follows,

$$V_{b0} = V_{dd} - V_T - \sqrt{\frac{rp}{N}V_T(2V_{dd} - 5V_T)} \tag{3.4}$$

where,

$$r = \frac{\beta_{nd}}{\beta_{na}}, p = \frac{\beta_{pd}}{\beta_{nd}}$$

If V_{bl} is smaller than V_{b0}, Q1a operates in the linear region. Hence, the read noise margin can be expressed as follows,

(1) if $V_{bl} \le V_{b0}$,

$$RNM_H = V_T - \frac{1}{k+1}\left[\frac{rpV_{dd} - (2rp+N)V_T}{rp - N + \frac{rp}{k}} - \frac{V_{dd} - 2V_T}{1 + qk + [q(qk^2+2k+1)]^{1/2}}\right] \tag{3.5}$$

where, (i) if $rp = N$,

$$k = \frac{V_{dd}(V_{dd} - 2V_T) - V_s^2}{2(V_{dd} - 3V_T)^2}$$

$$V_s = \sqrt{V_{bl}(2V_{dd} - 2V_T - V_{bl})}$$

or, (ii) if $rp \ne N$,

$$k = \frac{rp}{rp-N}\left[\frac{(rp-N)V_{dd} + NV_r}{[(rp-N)(NV_s^2 + rpV_p^2) + N^2V_r^2]^{1/2}} - 1\right]$$

$$V_s = \sqrt{V_{bl}(2V_{dd} - 2V_T - V_{bl})}$$
$$V_r = V_{dd} - (\frac{2rp}{N} + 1)V_T$$
$$V_p = \sqrt{V_{dd}(V_{dd} - 4V_T)}$$

(2) if $V_{bl} > V_{b0}$,

$$RNM_H = V_T - \frac{V_{dd} - 2V_T}{k+1}\left[\frac{1}{1+\frac{1}{k}} - \frac{1}{1 + qk + [q(qk^2 + 2k + 1)]^{1/2}}\right] \quad (3.6)$$

where,

$$k = \sqrt{\frac{rpV_r^2}{rpV_r^2 - NV_s^2}} - 1$$
$$V_s = V_{dd} - V_T - V_{bl}$$
$$V_r = V_{dd} - 2V_T$$

The expression of the "0" write noise margin can be given by substituting $V_{bl} = 0$ into the equation (3.5) and inverting the sign. Thus, it is expressed as,

$$WNM_L = -V_T + \frac{1}{k+1}\left[\frac{rpV_{dd} - (2rp+1)V_T}{rp - 1 + \frac{rp}{k}} - \frac{V_{dd} - 2V_T}{1 + qk + [q(qk^2 + 2k + 1)]^{1/2}}\right] \quad (3.7)$$

where, (i) if $rp = 1$,

$$k = \frac{V_{dd}(V_{dd} - 2V_T)}{2(V_{dd} - 3V_T)^2}$$

or, (ii) if $rp \neq 1$

$$k = \frac{rp}{rp-1}\left[\frac{(rp-1)V_{dd} + V_r}{[rp(rp-1)V_p^2 + V_r^2]^{1/2}} - 1\right]$$
$$V_r = V_{dd} - (2rp+1)V_T$$
$$V_p = \sqrt{V_{dd}(V_{dd} - 4V_T)}$$

We also get the expression for the "1" write noise margin as,

$$WNM_H = -V_T + \frac{1}{k+1}\left[\frac{V_{dd} - \frac{2r+1}{r+1}V_T}{1 + \frac{r}{k(r+1)}} - \frac{V_{dd} - 2V_T}{1 + \frac{k}{q} + [\frac{1}{q}(1+2k+\frac{k^2}{q})]^{1/2}}\right] \tag{3.8}$$

where,

$$k = \frac{r}{r+1}\left[\sqrt{\frac{(r+1)V_r^2}{(r+1)V_r^2 - V_s^2}} - 1\right]$$

$$V_s = V_{dd} - V_T$$

$$V_r = V_{dd} - \frac{2r+1}{r+1}V_T$$

3.2.2.3 Modified Single-ended Access Cell

For the write operation, we can make use of the same expression for the single-ended access cell. But for the read operation, it is difficult to get the analytical expressions. This is because the voltages of the internal storage nodes come close to the bitline voltage and all transistors in the cell can not be ignored. So, in this case, we estimate the noise margin from the transfer characteristics of the cell.

3.2.3 Verification of Analysis

In this section, we briefly describe the results of verification of analysis through SPICE simulations. We also touch on the the test circuit designed for measuring noise margins.

3.2.3.1 SPICE simulation

In Figure 3.9, the method to verify the analysis through SPICE simulation is shown.

For simulating a read noise margin, the timing sequence shown in Figure 3.9 (a) is used. That is:

(1) set V_1 and V_2 to the initial value by using NODESET command,

(2) apply V_{dd} to the word line so that the access transistor is on,

(3) increase V_n step by step until the cell flips,

then, the noise margin is equal to the V_n at the onset of flipping. A write noise margin can also determined in the same manner as shown in Figure 3.9 (b).

Figure 3.10 shows the noise margins in the differentially accessed cell simulated through SPICE together with the plots calculated from the analytic expressions presented above. The conditions for the calculations and simulations are: $V_{dd} = 5$ V, $V_T = 0.85$ V, and $p = 0.44$. Good agreement between analyses and SPICE simulations is observed.

Another concern is the secondary effect of transistor parameters such as body effect on the noise margins. The effect of bulk threshold parameter, gamma (γ), on noise margins was examined as shown in Figure 3.10. The read noise margin increase by increasing gamma while the write noise margin decreases. This is because the conductance of the access transistor decreases due to the body effect for higher gamma's.

Figure 3.11 shows the noise margins in the single-ended access cell simulated through SPICE together with the plots calculated from the analytic expressions presented above. The conditions for the calculations and simulations are: $V_{dd} = 5$ V, $V_T = 0.85$ V, $V_{bl} = 2.5$ V , $p = 0.47$. and $q = 1.08$. Also in this case, good agreement between analyses and SPICE simulations is observed. According to the results of SPICE simulation, gamma does not affect both the "0" read noise margin (RNM_L) and the "0" write noise margin (WNM_L). This is because the source voltage of the access transistor is around the V_{ss} level. In the case of reading "1", the noise margin increases for larger gammas. This is because the substrate bias ($V_{BS} = -2.5$ V) modulate the threshold voltage of the access transistors.

The above comparison has proved that the analytic expression can be useful for estimating the stability in multiport memories. Secondary effects of transistor parameters

have to be taken into account for a more accurate estimation of the noise margin.

3.2.3.2 Test Circuit for Noise Margin Measurement

We have designed a test circuit to verify the analyses. Figure 3.12 shows the circuit diagram for measuring read noise margins in the single-ended access cell. The noise voltage set circuit is used to set the noise voltage between an input of an inverter and an output of another inverter. The noise voltage is dynamically stored in two capacitors to simulate the floating voltage sources shown in Figure 3.6 (a). The amount of noise voltage can be controlled by changing the external biases, V_{NL} and V_{NH}. Also, stored information ("0" or "1") can be controlled through the external LOW input. Note that the capacitances should have large enough values so that the parasitic capacitances at the inverter inputs can be negligible.

Figure 3.12 (b) shows a timing diagram to measure a noise margin. Initial setting is done by forcing the state of the flip-flop to one of the two states and storing noise voltages in the capacitors while maintaining the word line voltage at the V_{ss} level. For examining the noise margin, apply V_{dd} to the word line(s) and observe the internal state of the flip-flop through the buffer. If the state of the inverter does not flip, the read noise margin has a value larger than the noise voltage set by the former initialization. If the state of the inverter flips, the read noise margin has a value smaller than the noise voltage set by the former initialization. Thus the maximum noise voltage to prevent the cell from flipping gives the read noise margin.

This test circuit has been embedded in the RAT (Register Alias Table)[7] test chip and has been sent for fabrication.

3.3 Optimal Design of Multiport Memories

This section describes the optimal design of multiport memories based on the noise margin analyses presented in section 3.2.

3.3.1 Differentially Accessed Cell

Figure 3.13 shows the plot of read noise margins (RNM) and write noise margins (WNM) calculated from equations (3.1) and (3.2). Plots are shown as a function of a beta ratio, $r = \beta_{nd}/\beta_{na}$, for different beta ratios, $p = \beta_{pd}/\beta_{nd}$. We see the following from this figure:

(1) Read noise margin (RNM) is a monotonically increasing function of r, while write noise margin (WNM) is a monotonically decreasing function.

(2) Read noise margin (RNM) is a monotonically increasing function of p, while write noise margin (WNM) is a monotonically decreasing function.

These results show that we should carefully choose beta ratios, r and p, for ensuring enough noise margins.

Figure 3.14 shows a scheme to determine the optimal beta ratio, r_{opt}, for a given p. The critical operations in an N-port multiport memory are a single-port write ($r = r_{opt}$) and an N-port read ($r = r_{opt}/N$). For simplicity, we define the optimal beta ratio r_{opt} as r which gives equal value of noise margin for a single-port write and an N-port read. Also we call this value of noise margin as the maximum noise margin,$NM_{\max}$. It is important to note that in the actual design we should take care about other effects such as dynamic noise during read/write operation, sensitivity of noise margins to supply voltage, and sensitivity of noise margins to process variations for determining optimal beta ratios.

Figure 3.15 shows the plots of optimal beta ratio (r_{opt}) as a function of the number of the ports, N. r_{opt} is a monotonically increasing function of N. Figure 3.16 shows the plot of maximum noise margin (NM_{max}) as a function of the number of the ports, N. These results show that the beta of P-channel MOSFET ,β_{pd} ,does not play an important role in determining the noise margin, but it should be appropriately large enough to avoid an increase of r_{opt}. An example of optimal design for triple-port memory is: $p = 0.5$ and $r_{opt} = 2.6$. It gives the noise margin of about 0.44 V.

3.3.2 Single-ended Access Cell

The optimization of the single-ended cell is rather complicated compared with that of differentially accessed cell, because we must consider four noise margins at a time and the variation of bitline voltage for read operation.

Figure 3.17 shows an example of plots for a static single-ended access cell. It is assumed that the bitline voltage for the read operation is 2.5 V. We clearly see in this figure that the "1" level writing is very difficult and we cannot ensure enough noise margins for all operations, as is described in Chapter 3.1. So, we must give up the static approach at this moment. In this section, we focus on a one-sided clocking version of the pseudo-static single-ended cell and ignore the "1" write operation. Furthermore, we fix the bitline voltage for the read operation at 2.5 V since it is practical from the viewpoint of the sense amplifier design.

The first task is to get the right balance between the "0" and "1" read operations by minimizing the difference between the "0" and "1" read noise margins. "0" read noise margin, RNM_L, and "1" read noise margin, RNM_H are plotted against beta ratio, $q = \beta_{pf}/\beta_{nf}$, for different p's in Figure 3.18. RNM_L is an increasing function of q while RNM_H is a

decreasing one. This means that there is an optimal q ($= q_{opt}$) to give equal noise margins for "0" and "1" read operations for each given p. In other words, optimization of the logic threshold of inverter 2 is needed to get the right balance.

In Figure 3.19, q_{opt} is plotted against beta ratio $p = \beta_{pd}/\beta_{nd}$ for different r's. To make design easy, we approximate the relation between q_{opt} and p as a linear function,

$$q_{opt} \approx 1.919 \times p^{0.738}. \tag{3.9}$$

In Figure 3.20 the noise margins of the pseudo-static single-ended cell are plotted against r for optimal beta ratio, q_{opt}. Three curves representing RNM_H, RNM_L and WNM_L cross at around the same point. Thus equation (3.9) gives a favorable balance of noise margins.

Through this relation we can determine the q_{opt} for given p, and can determine the r_{opt} accordingly. In Figure 3.21, r_{opt} is shown as a function of the number of the ports, N. Note that the value of r_{opt} is comparatively small compared with that of the differentially accessed cell. This is because the read operation is more stable than that of the differentially accessed cell due to the strong non-linearity of one of the inverters.

Figure 3.22 shows maximum noise margin, NM_{max}, as a function of the number of the ports, N. An example of optimal design for triple-port memory is: $p = 0.5$, $q = 1.15$ and $r_{opt} = 1.6$. It gives the noise margin of about 0.52 V.

3.3.3 Modified Single-ended Access Cell

The read noise margins for the modified single-ended cell are calculated through the transfer characteristics of the flip-flop. Figure 3.23 shows an example of the calculated transfer characteristics of this cell.

Figure 3.24 shows the read noise margin as a function of r. The analytic expression for the single-ended access cell is used for the plot of the write noise margin. Please note that the "1" write operation is also difficult in this cell structure, so the pseudo-static approach is assumed.

In Figure 3.25, r_{opt} is shown as a function of the number of the ports, N, and Figure 3.22 shows maximum noise margin, NM_{max}, as a function of the number of the ports, N.

3.3.4 Comparison of Noise Margins

Noise margins for optimally designed multiport memories are compared. For the single-ended cell and the modified single-ended cell only a pseudo-static version with one-sided clocking is considered for this comparison.

Figure 3.27 shows the noise margins plotted against the number of ports, N, for three cell configurations optimally designed for $p = 0.5$. The single-ended cell has the largest noise margin of the three configurations. We can expect a noise margin of 0.5 V for the triple-port single-ended memory cell. The modified single-ended cell has no advantage over the single-ended cell due to its low noise margin and its increased number of signal lines. Hence, we exclude the modified single-ended access cell as a candidate for the multiport memory. In the following section, we study other aspects such as area and speed.

3.4 Active Area Estimation

Another important design issue to achieve high performance per cost is the cell area and the speed. It is difficult to specify what design parameters affect the cell area, but it seems to be dependent on the following:

(1) Total area consumed by the active transistors.

(2) Number of signal lines.

(3) Regularity of the circuit layout

(4) Process technology (buried contact, interconnection etc.).

Among them, (1) and (2) appears to have a major effect on the cell area. As far as the number of the signal lines is concerned, the pseudo-static single-ended access cell has an advantage over the differentially accessed cell if the number of ports is greater than two (N>2). In this section, we focus on the active area for two cell configurations to confirm the feasibility of the single-ended cell.

It is a well known fact that area and speed are closely related. The smaller the cell area, the smaller the capacitance of the signal lines as well as that of other parasitics. At the same time, reducing the transistor size in the cell results in the reduction of the driving capability. To figure out the relation between area and signal line capacitance is a difficult task, because it is closely related to the fabrication technology. So, we take only the signal current into account as the criterion for estimating the area. In other words, we calculate the area so that we can get the standard signal current.

The important thing to note is that if the memory cell with the minimum size access transistor provides enough signal current, the single-ended cell has an area smaller than that of the differential cell. This means that the following analysis is true when the size of the access transistor is bigger than that of the minimum size.

3.4.1 Scale Up Factor

The amount of signal current is mainly determined by the size of the access transistor. So, it is important to know the dependence of the size of the access transistor(β_{na}) on the number of the ports(N). Accordingly, we define the scale up factor, F_s, as a matter of

convenience. Scale up factor, F_s, is defined as the ratio of β_{na} (transconductance of the access transistor) to β_0. Here, β_0 is a constant relating to the standard signal current.

Figure 3.28(a) shows the definition of the scale up factor of the differentially accessed cell. The transconductance β_0 is defined as the transistor parameter which gives the standard signal current under the bias condition of $V_{GS} = V_{DS} = 5\ V$, and W_0 denotes the channel width of this transistor. The transconductance of the access transistor, β_{na}, in the differentially accessed multiport memory cell becomes larger than β_0 in order to get the same amount of signal current. This is because the output voltage of the Inverter 1 is larger than V_{ss} and the V_{GS} of the access transistor becomes less than 5 V. The scale up factor, F_s is expressed as,

$$F_s = \frac{\beta_{na}}{\beta_0} = \frac{W_{na}}{W_0}. \tag{3.10}$$

Figure 3.28(b) shows the scale up factor in the single-ended accessed cell. In the single-ended access case, the signal current for the "1" read is smaller than that for the "0" read if the beta ratio p is smaller than or equal to 1. This is because the V_{GS} for the "1" read is 2.5 V while the V_{GS} for the "0" read is greater than 2.5 V. Thus, we only treat the "1" read case. From the viewpoint of the signal current, the single-ended cell has a disadvantage since the V_{GS} of the access transistor is one half of that of the differentially accessed cell.

At first, we would like to derive the scale up factor for these two cell configurations.

3.4.1.1 Differentially Accessed Cell

In Figure 3.28(a), the signal current i_s can be expressed as,

$$i_s = \frac{\beta_{na}}{2} (V_{dd} - V_1 - V_T)^2 \tag{3.11}$$

$$i_s = \frac{\beta_{nd}}{N} V_1 (V_{dd} - V_T - \frac{V_1}{2})^2. \tag{3.12}$$

where V_1 is the output voltage of the inverter 1. By eliminating i_s, we get,

$$V_1 = (1 - \sqrt{\frac{r}{r+N}})(V_{dd} - V_T). \tag{3.13}$$

Thus, the signal current becomes,

$$i_s = \frac{\beta_{na}}{2} \frac{r}{r+N} (V_{dd} - V_T)^2. \tag{3.14}$$

This signal current should be equal to the standard signal current. So we get,

$$i_s = \frac{\beta_0}{2} (V_{dd} - V_T)^2. \tag{3.15}$$

>From (3.14) and (3.15) we get the scale up factor as,

$$F_s = \frac{\beta_{na}}{\beta_0} = \frac{r+N}{r}, \tag{3.16}$$

where,

$$r = \frac{\beta_{nd}}{\beta_{na}}.$$

3.4.1.2 Single-ended Access Cell

In Figure 3.28(b), the access transistor operates in the saturation region under the following condition,

$$V_{bl} \geq V_{dd} - V_T - \sqrt{\frac{rp}{N} V_T (2V_{dd} - 5V_T)} \tag{3.17}$$

where,

$$r = \frac{\beta_{nd}}{\beta_{na}}, p = \frac{\beta_{pd}}{\beta_{nd}}.$$

Substituting $V_{bl} = \frac{V_{dd}}{2}$ into (3.17) yields,

$$\frac{rp}{N} \geq \frac{(V_{dd} - 2V_T)^2}{4V_T(2V_{dd} - 5V_T)} \tag{3.18}$$

We calculate the scale up factor for the following two cases.

(1) If the access transistor operates in the saturation region:

$$\left[\frac{rp}{N} \geq \frac{(V_{dd} - 2V_T)^2}{4V_T(2V_{dd} - 5V_T)} \right]$$

The signal current is expressed as,

$$i_s = \frac{\beta_{na}}{2} (\frac{V_{dd}}{2} - V_T)^2. \tag{3.19}$$

>From (3.15) and (3.19) we get the scale up factor as,

$$F_s = \frac{\beta_{na}}{\beta_0} = \frac{(V_{dd} - V_T)^2}{(\frac{V_{dd}}{2} - V_T)^2}. \tag{3.20}$$

(2) If the access transistor operates in the linear region:

$$\left[\frac{rp}{N} < \frac{(V_{dd} - 2V_T)^2}{4V_T(2V_{dd} - 5V_T)} \right]$$

The signal current is expressed as,

$$i_s = \beta_{na} (V_1 - \frac{V_{dd}}{2})(\frac{V_{dd}}{2} - V_T - \frac{V_1 - \frac{V_{dd}}{2}}{2}) \tag{3.21}$$

$$i_s = \frac{\beta_{pd}}{N} (V_{dd} - V_1)(V_{dd} - V_T - \frac{V_{dd} - V_1}{2}). \tag{3.22}$$

The scale up factor can be given by eliminating i_s from these equations and by solving for V_1 for two cases.

(a) If $rp = N$,

$$V_1 = \frac{V_{dd}(7V_{dd} - 12V_T)}{8(V_{dd} - 2V_T)}. \tag{3.23}$$

Substituting (3.23) into (3.21) yields,

$$i_s = \frac{\beta_{na}}{2} \frac{V_{dd}(V_{dd} - 4V_T)(3V_{dd} - 4V_T)(5V_{dd} - 8V_T)}{64(V_{dd} - 2V_T)^2}. \tag{3.24}$$

Thus, we get,

$$F_s = \frac{64(V_{dd} - V_T)^2(V_{dd} - 2V_T)^2}{V_{dd}(V_{dd} - 4V_T)(3V_{dd} - 4V_T)(5V_{dd} - 8V_T)}. \tag{3.25}$$

(a) If $rp \neq N$,

$$i_s = \frac{\beta_{na}}{2} \frac{[V_a - (rp + N)V_b][(3rp - N)V_b - V_a]}{4(rp - N)^2}. \tag{3.26}$$

where,

$$V_a = \sqrt{(rp + N)^2(V_{dd} - 2V_T)^2 + rp(rp - N)V_{dd}(3V_{dd} - 4V_T)} \tag{3.27}$$

$$V_b = V_{dd} - 2V_T. \tag{3.28}$$

Thus, we get,

$$F_s = \frac{4(rp - N)^2(V_{dd} - V_T)^2}{[V_a - (rp + N)V_b][(3rp - N)V_b - V_a]}. \tag{3.29}$$

3.4.2 Normalized Active Area

We define the normalized active area as,

$$A_N = \frac{W_{total}}{W_0}. \tag{3.30}$$

where W_{total} is the summation of the channel width in a cell, and W_0 is the channel width of the transistor having the transconductance of β_0. We compare the cell area using this measure.

3.4.2.1 Differentially Accessed Cell

The normalized area for the differentially accessed cell is expressed as,

$$A_N = 2(N + r + rp\eta)\, F_s. \tag{3.31}$$

where,

$$r = \frac{\beta_{nd}}{\beta_{na}}, p = \frac{\beta_{pd}}{\beta_{nd}},$$

and η is the beta ratio of N-channel MOSFET to P-channel MOSFET for the same channel width and expressed as,

$$\eta = \frac{\beta_{NMOS}}{\beta_{PMOS}}.$$

η is 2.5 ~ 3.0 in the conventional CMOS process. Substituting $r = r_{opt}$ and equation (3.16) into (3.31) yields,

$$A_N = 2[N + r_{opt}(1 + p\eta)] \frac{r_{opt} + N}{r_{opt}}. \tag{3.32}$$

F_s and A_N are plotted as a function of the number of the ports, N, in Figure 3.29 and Figure 3.30, respectively. Here, $\eta = 2.5$ is assumed. As is shown in the Figure 3.30, the smaller the beta ratio p the smaller the active area, especially for a large value of N. This means that the size of the access transistor determines the active area, since the size of the access transistor decreases by decreasing p as is shown in Figure 3.29. As is mentioned in section 3.3.1, decreasing p suffers from an increase of $r = \beta_{nd}/\beta_{na}$. So, balancing these two ratios is needed to get a compact and stable cell. For example, the normalized active area of a triple-port cell is 38 for $p = 0.5$. It is 2.4 times bigger than the conventional single-port (SRAM) cell.

3.4.2.2 Single-Ended Access Cell

The normalized active area for the pseudo-static single-ended cell with one-side clocking (one-side clocking version) is expressed as,

$$A_N = [N + r_{opt}(4 + p\eta)]\, F_s. \tag{3.33}$$

The normalized active area for the two-sided clocking version becomes,

$$A_N = [N + 4r_{opt}(1 + p\eta)]\, F_s. \tag{3.34}$$

In these expressions, we ignore the area occupied by Inverter 2, since it drives only the input of Inverter 1 and can be small.

F_s and A_N for the one-sided clocking version are plotted as a function of the number of the ports, N, in Figure 3.31 and Figure 3.32, respectively. Because the V_{GS} of the access transistor is small (2.5 V), the scale up factor is as much as 7 or more (1~4 for the differential cell). This shows that the single-ended cell consumes a large active area to get a reasonable signal current. The active area depends strongly on the beta ratio p. This is because the size of the P-channel MOSFET determines the amount of signal current. For example, the normalized active area of a triple-port cell is 81 for $p = 0.5$. It is 1.8 times bigger than the single-port cell.

F_s and A_N for the two-sided clocking version are plotted as a function of the number of the ports, N, in Figure 3.33 and Figure 3.34, respectively. The criterion of the noise margin (NM_{cri}) is 0.8 V in these calculations. In this condition, the beta ratio r_{opt} increases proportionally with the number of the ports, N. The result is that the scale up factor is constant for different Ns. By the same reason, the normalized active area is proportional to the number of the ports.

Thus far, to maintain a constant signal current and to ensure a certain amount of noise margin, we must increase the active area proportionally with the number of the ports. To avoid a significant increase of the area, a special circuit or an improved array structure to sense the small signal current is needed. In the next section, we present a technique using a dummy cell and discuss its effectiveness.

3.4.2.3 The Impact of Dummy Cell on Active Area

The problem of the single-ended cell is in resolving the difference between the signal currents when reading "0" and "1". It can be compensated by using a dummy cell.

The memory array configuration for using a dummy cell is shown in Figure 3.35(a). The memory array is divided into two halves and the sense amplifiers are located between these two half arrays. Each bitline has a dummy cell. When the memory cells in the left hand side are selected, the dummy cells at the right hand side are selected, and vice versa.

Figure 3.35(a) shows the circuit scheme of a dummy cell. The idea is to drive the dummy bitline by the intermediate value of the two signal currents ("0" and "1" read). By this means, equal signal voltage differences for "0" and "1" read operation are expected as shown in Figure 3.35(b). The drawbacks of this means are:

(1) It adds a certain amount of delay in the case of the "0" read. This is because the voltage of the dummy bitline drops before the sense amplifiers are triggered. So, it takes some time to reach the logic threshold of the next stage.

(2) Peak current increases because of an increased amplitude of the dummy bitline in the case of the "0" read operation.

(3) It increases the area.

(4) Additional logic control is needed to select one of the dummy cells.

The required scale up factor in this case is expressed as follows:

(1) If the access transistor operates in the saturation region:

$$\left[\frac{rp}{N} \geq \frac{(V_{dd} - 2V_T)^2}{4V_T(2V_{dd} - 5V_T)} \right]$$

$$F_s = \frac{2(V_{dd} - V_T)^2}{[\frac{r}{r+N}\frac{V_{dd}}{2}(\frac{3}{2}V_{dd} - 2V_T) + (\frac{V_{dd}}{2} - V_T)^2]}. \tag{3.35}$$

(2) If the access transistor operates in the linear region:

$$\left[\frac{rp}{N} < \frac{(V_{dd}-2V_T)^2}{4V_T(2V_{dd}-5V_T)}\right]$$

(a) if $rp = N$, then

$$F_s = \frac{2(V_{dd}-V_T)^2}{\left[\frac{r}{r+N}\frac{V_{dd}}{2}(\frac{3}{2}V_{dd}-2V_T)+\frac{V_{dd}(V_{dd}-4V_T)(3V_{dd}-4V_T)(5V_{dd}-8V_T)}{64(V_{dd}-2V_T)^2}\right]} \quad (3.36)$$

(a) if $rp \neq N$, then

$$F_s = \frac{2(V_{dd}-V_T)^2}{\left[\frac{r}{r+N}\frac{V_{dd}}{2}(\frac{3}{2}V_{dd}-2V_T)+\frac{[V_a-(rp+N)V_b][(3rp-N)V_b-V_a]}{4(rp-N)^2}\right]} \quad (3.37)$$

F_s and A_N for the one-sided clocking version with a dummy cell are plotted as a function of the number of the ports, N, in Figure 3.36 and Figure 3.37, respectively. By employing a dummy cell, A_N becomes insensitive to the beta ratio p. This is because the signal voltage is determined not only by the size of the P-channel MOSFET. For example, the normalized active area of a triple-port cell is 52 for $p = 0.5$. It is 64 % of the area without a dummy cell.

F_s and A_N for the two-sided clocking version with a dummy cell are plotted as a function of the number of the ports, N, in Figure 3.38 and Figure 3.39, respectively. The scaling factor decreases to around three to five. The normalized active area of a triple-port cell is 52 for $p = 0.5$, and is about 54 % of that without using a dummy cell.

3.4.3 Discussion

A comparison of the normalized active area is shown in Figure 3.40. The normalized active area of the single-ended cell with one-sided clocking is about two times as big as that

of the differentially accessed cell when $N = 3$. The two-sided clocking version consumes twice the area of the one-sided clocking version. So, the single-ended cell seems infeasible without special circuit techniques.

Figure 3.41 shows some ideas to solve this problem. The first one is to employ a split-bitline configuration as shown in Figure 3.41(b). In this configuration, the bitline capacitance (C_{BL2}) is expected to be one half of the total bitline capacitance (C_{BL1}). In this case, the signal current to get the same signal voltage can be a half of the standard one. In Figure 3.42, the normalized area for split-bitline based single-ended cells are compared with that of the differentially accessed cell. As shown in this figure, the normalized area for the one-sided clocking version of the single-ended cell is comparable to that for the differentially accessed cell.

The second one is to employ a dummy cell in addition to the split-bitline architecture as shown in Figure 3.41(c). In Figure 3.43, the normalized area for a dummy cell and split-bitline based single-ended cells are compared with that of the differentially accessed cell. In this case, the normalized area of the two-sided clocking version of the single-ended cell is comparable to that of the differentially accessed cell.

3.5 Choice of Multiport Memory Cell

As far as the noise margins are concerned, the pseudo-static single-ended cell is superior to other configurations.

>From the viewpoint of the active area, split-bitline based single-ended cell and the differentially accessed cell seem feasible.

The single-ended cell has the advantage of a smaller number of the signal lines over other configurations if the number of the ports is greater than three.

Based on these considerations, the pseudo-static single-ended access cell (one- or two-sided clocking) seems to be promising for the multiport memory cell.

Note that this selection must be accompanied with a sensitive and high speed sensing scheme. The split bitline scheme or a dummy cell is an effective approach by improving a memory array configuration.

To make this choice robust, the study of the sense amplifier is indispensable. Therefore, in the next chapter, the sense amplifier circuit fit for the single-ended cell is investigated.

CHAPTER 4 SENSE AMPLIFIER DESIGN

As was mentioned in the previous chapter, one of the keys to achieving a compact multiport memory cell is a sensing circuit that can sense the signal current provided by the minimum size memory cell. This chapter discusses, first, possible sense amplifier configurations for the single-ended multiport memory cell. Next, we analyze speed, power and peak current in these sense amplifiers. Finally, we compare the performance of the sense amplifiers.

4.1 Possible Sense Circuit Configurations

There are two possible configurations for the sensing circuit as shown below:

(1) Voltage sensing method.

(2) Current sensing method.

The first one is to amplify the small input voltage (0.2 ~ 0.5 V) to the large output voltage (~ 5 V), and is conventionally used for the sensing circuit of DRAMs and SRAMs. The second one is used for the sensing circuit of EPROMs (Erasable Programmable Read Only Memories) or EEPROMs (Electrically Erasable Programmable Read Only Memories). The difference is that the first one has a relatively large input impedance while the second one has a small one. Consequently, the voltage swing in the current sense amplifiers is smaller than that of the voltage sense amplifiers. Next, we show some examples for the single-ended memory cell.

4.1.1 Voltage Sense Amplifiers

One possible way of the voltage sensing scheme is to combine the split-bitline array configuration (Figure 3.41) and a dynamic cross-coupled sense amplifier (SA) which is conventionally used for DRAMs (or it is called an open-bitline structure). Figure 4.1 shows an example of the sensing circuit. The advantages of this scheme are:

(1) Sensing speed is relatively high due to a decreased bitline capacitance.

(2) It is easy to get $\frac{V_{dd}}{2}$ (the only thing to do is to short bitline pairs).

On the other hand, it has the following disadvantages:

(1) Timing control (from the rising edge of the word-line to the edge of the SA triggering pulse) is critical.

(2) Dynamic power is relatively large since every bitline is driven to V_{dd} or V_{ss} level for every read operation.

4.1.2 Current Sense Amplifiers

Figure 4.2 shows the transresistance sense amplifier[14], an example of the current sense amplifier. The current sense amplifier is basically an I-V (current to voltage) converter. The transresistance sense amplifier consists of an inverter (amplifier) and a feedback resistor. By this configuration, we can get the output voltage which is proportional to the input current. In this sensing circuit, the voltage swing of the input node is suppressed to a low level. So it has the following advantages:

(1) Sensing speed is high especially for heavy input capacitance.

(2) Dynamic power is decreased due to a reduced voltage swing.

On the other hand, it has the following disadvantages:

(1) It consumes DC power since it needs DC bias current.

(2) Precharging a bitline takes much time after a write operation (write recovery),

unless there is a special circuit.

(3) The peak current for precharging is high.

In figure 4.2(b), we use a reset transistor and a reset pulse to shunt the feedback resistor for improving the write recovery characteristic.

4.2 Analysis and Design of Sense Amplifiers

In this section, the performances of three kinds of sense amplifiers (one voltage SA and two current SA) are analyzed.

The first task is to set a target for the design of a sense amplifier. Figure 4.3 shows the calculated signal current for the triple-port cell with minimum size access transistors. The calculation is based on the transistor parameters for a 1.6μm CMOS process provided through the MOSIS service. Other transistor sizes are determined based on the optimal design presented in Chapter 3. The result shows that we can expect as small as 40μA per port in the case of the triple-port read operation. So, we choose 50μA per port as the target of the signal current for the following analyses.

We assume that the logic threshold voltage of the next stage to be 2.5 V. Then, V_{IL} and V_{IH} becomes 2.125 V and 2.875 V, respectively[15]. So, we assume that 2.5±0.5 V is needed for the sense amplifier output to drive the next stage correctly. This is used for calculating the sensing delay time.

4.2.1 Cross-coupled Dynamic Sense Amplifier

Figure 4.4 shows the circuit diagram for analysis. For simplicity, the following are assumed:

(1) The voltage difference between BL and the precharge level ($\frac{V_{dd}}{2}$) is same as the

difference between $\overline{\text{BL}}$ and the precharge level.

(2) Transconductances (β) of all MOSFETs are equal.

(3) Both cross-coupled MOSFETs operate in the saturation region. This means that the voltage difference between BL and $\overline{\text{BL}}$ is less than the threshold voltage (V_T). (It is enough to estimate a transmission delay.)

At first, we perform DC analysis. The currents for three N-channel MOSFETs in Figure 4.3 are expressed as,

$$i_{1n} = \frac{\beta}{2}\left(\frac{V_{dd}}{2} - x - y - V_T\right)^2 \tag{4.1}$$

$$i_{2n} = \frac{\beta}{2}\left(\frac{V_{dd}}{2} + x - y - V_T\right)^2 \tag{4.2}$$

$$i_n = \beta y \left(V_{dd} - \frac{y}{2} - V_T\right). \tag{4.3}$$

>From these equations we get,

$$y = \frac{V_p}{3} - \frac{\sqrt{2V_q{}^2 - 24x^2}}{6}. \tag{4.4}$$

where,

$$V_p = 2V_{dd} - 3V_T$$
$$V_q = 5V_{dd}{}^2 - 12V_{dd}V_T + 6V_T{}^2.$$

Substituting (4.4) into (4.3) yields,

$$i_n = \frac{\beta}{2}\left[\frac{2}{3}V_{dd} - V_T - \frac{\sqrt{2V_q{}^2 - 24x^2}}{6}\right]\left[\frac{\sqrt{2V_q{}^2 - 24x^2}}{6} + \frac{4}{3}V_{dd} - V_T\right]. \tag{4.5}$$

We get the amount of the peak current by substituting $x = 0$ into (4.5),

$$i_{peak} = \frac{\beta}{2}\left[\frac{2}{3}V_{dd} - V_T - \frac{\sqrt{2}}{6}V_q\right]\left[\frac{\sqrt{2}}{6}V_q + \frac{4}{3}V_{dd} - V_T\right]. \tag{4.6}$$

If we assume $V_{dd} = 5V$ and $V_T = 1V$,(4.6) becomes,

$$i_{peak} \approx 1.33 \times \beta. \tag{4.7}$$

Next, we calculate the delay time. We define two delay times, τ_{d1} and τ_{d2}, as follows. τ_{d1} is the time needed for establishing a voltage difference of 200 mV between BL and $\overline{BL}$, and τ_{d2} is the time from triggering the sense amplifiers to getting 1 V difference between BL and $\overline{BL}$. τ_{d1} is expressed as,

$$\tau_{d1} = 0.2 \times \frac{C_{bl}}{i_s}. \tag{4.8}$$

We get τ_{d2} by solving the following differential equation,

$$\begin{aligned} \frac{dx}{dt} &= \frac{i_{2N} - i_{1N}}{C_{bl}} \\ &= \frac{\beta}{C_{bl}}(V_{dd} - 2V_T - 2y)\,x. \end{aligned} \tag{4.9}$$

The variation of x by changing y from 0 to 0.5 is small (0.347~0.390), so we can approximate that y is constant. Then, (4.9) becomes,

$$\frac{dx}{dt} \approx \frac{\beta}{C_{bl}}(V_{dd} - 2V_T - 2y_0)\,x. \tag{4.10}$$

where,

$$y_0 = \frac{V_p}{3} - \frac{\sqrt{2}}{6}V_q.$$

By solving (4.10), we get,

$$x = x_0 \exp\left[\frac{\beta(V_{dd} - 2V_T - 2y_0)}{C_{bl}}\,t\right] \tag{4.11}$$

where, x_0 is the initial value of x. By substituting $x = 0.5$ and $x_0 = 0.1$ into (4.11), we get the expression of the delay time as,

$$\tau_{d2} = ln\,(5)\frac{C_{bl}}{\beta(V_{dd} - 2V_T - 2y_0)} \tag{4.12}$$

If we assume $V_{dd} = 5\text{V}$ and $V_T = 1\text{V}$,(4.12) becomes,

$$\tau_{d2} = 0.698 \times \frac{C_{bl}}{\beta}. \tag{4.13}$$

>From (4.8) and (4.13), the total sensing delay is expressed as,

$$\tau_d = \tau_{d1} + \tau_{d2}$$
$$= \left[\frac{0.2}{i_s} + \frac{0.698}{\beta} \right] C_{bl}. \qquad (4.14)$$

Figure 4.5 shows calculated delays and peak currents as functions of the transconductance of the transistors. Here, we assume that the number of the bitlines is 100 (assuming triple-port and 32-bit data). The circles in the figure are the delay times simulated with SPICE. Good agreement between the analysis and SPICE simulation is observed. This figure shows that the delay time improvement can not be expected and we suffer from high peak current by increasing transconductance over $1 \times 10^{-3} A/V^2$.

4.2.2 Transresistance Sense Amplifier

Figure 4.6 shows the circuit schematic of the transresistance sense amplifier used for the analysis. For simplicity, it is assumed that the transconductances of N-channel and P-channel MOSFETs are equal and both MOSFETs operate in the saturation region.

At first, we perform DC analysis. Current equations for MOSFETs are,

$$i_n = \frac{\beta}{2} (\frac{V_{dd}}{2} + x - V_T)^2 \qquad (4.15)$$

$$i_p = \frac{\beta}{2} (\frac{V_{dd}}{2} - x - V_T)^2. \qquad (4.16)$$

Thus, signal current is expressed as,

$$i_s = i_n - i_p$$
$$= \beta (V_{dd} - 2V_T)x. \qquad (4.17)$$

We get the input voltage change as,

$$x = \frac{i_s}{\beta (V_{dd} - 2V_T)}. \qquad (4.18)$$

The output voltage change is expressed as,

$$y = \left[R_F - \frac{1}{\beta (V_{dd} - 2V_T)} \right] i_s. \tag{4.19}$$

Then, the required value for the feedback resistor is,

$$R_F = \frac{y}{i_s} + \frac{1}{\beta (V_{dd} - 2V_T)}. \tag{4.20}$$

If we assume that $i_s = 50\mu\text{A}$ and $y = 1\text{V}$, (4.20) becomes,

$$R_F = 2{\times}10^4 + \frac{1}{\beta (V_{dd} - 2V_T)}. \tag{4.21}$$

The DC current for $x = 0$ is given by,

$$I_{dc} = \frac{\beta}{2} \left(\frac{V_{dd}}{2} - V_T\right)^2. \tag{4.22}$$

Next, we perform a transient analysis. In the following analysis, we ignore the output capacitance for simplicity. Then we get a differential equation,

$$\frac{dx}{dt} + Px = Q. \tag{4.23}$$

where,

$$P = \frac{\beta(V_{dd} - 2V_T)}{C_{bl}}$$

$$Q = \frac{i_s}{C_{bl}}.$$

By solving (4.23), we get,

$$x = \frac{i_s}{\beta(V_{dd} - 2V_T)} \left[1 - \exp\left[\frac{\beta(V_{dd} - 2V_T)}{C_{bl}} t \right] \right]. \tag{4.24}$$

If we assume that the output voltage change at the steady state is 1 V, the delay time is expressed as,

$$\tau = ln(2) \frac{C_{bl}}{\beta(V_{dd} - 2V_T)}. \tag{4.25}$$

We define the precharge time, τ_p, as the time needed to drive the input within 5 % of the steady state voltage. Assuming that the transconductance of the reset transistor is large

enough compared with that of the inverter transistor, we get the precharging time as,

$$\tau_p = \frac{C_{bl}}{\beta(V_{dd} - 2V_T)} \left[\frac{2V_T}{V_{dd} - V_T} + \ln\left[\frac{V_{dd}}{40} \right] \right]. \tag{4.26}$$

Figure 4.7 shows delay time, τ_d, precharging time , τ_p, and DC current, I_{DC}, as a function of the transconductance. This figure shows that the precharging time is about four times as much as the delay time. Thus, a special circuit is needed for the fast recovery after a write operation. The delay time is also simulated with SPICE, and the results are shown as circles in the figure. The reason why the analysis shows the smaller delay time is that we ignore the output capacitance for the analysis. The SPICE simulation results are shown in Figures 4.8 and 4.9, for different input (bitline) capacitances. These figures clearly show that the peak current during read operations can be suppressed by the reduction of the input voltage swing. Figure 4.8 also shows that the disadvantage of this amplifier is that it takes too much time for precharging an input.

4.2.3 Class-A Push-Pull Current Sense Amplifier

A problem of the transresistance sense amplifier is the slow recovery from a write operation. One possible way to solve this problem is to use an amplifier capable of driving high input capacitance. We present here an example of a current sense amplifier using a push-pull driver,

Figure 4.10(a) shows the schematic of the class-A push-pull current sense amplifier. The bias voltage V_{BN} and V_{BP} are applied to the gates of the push-pull driver so that the driver is in the class-A operation mode for the standard signal current (50). At the same time, the input current is mirrored to the output. Since the output capacitance is small enough compared with that of the input, a high slew rate is expected at the output even for a

small input current. RESET pulse is used for setting the output voltage at the logic threshold voltage of the next stage.

We analyze this circuit using a circuit shown in Figure 4.10(b). For simplicity, we assume that the upper half and the lower half of the circuits operate symmetrically, and that all transistors operate in the saturation region.

At first, we perform a DC analysis. The bias current is expressed as,

$$I_B = \frac{\beta_1}{2} V_B{}^2. \tag{4.27}$$

i_n and i_p are expressed as ,

$$I_n = \frac{\beta_1}{2} (V_B + x)^2 \tag{4.28}$$

$$I_n = \frac{\beta_1}{2} (V_B - x)^2. \tag{4.29}$$

Then, the signal current is expressed by using x as,

$$I_s = 2\beta_1 V_B x. \tag{4.30}$$

Thus the input voltage swing is expressed as,

$$x = \frac{i_s}{2\beta_1 V_B} = \frac{i_s}{2[2\beta_1 I_B]^{1/2}}. \tag{4.31}$$

The gate-source voltage of transistor Q_2 is expressed as,

$$V_{gs2} = \sqrt{\frac{\beta_1}{\beta_2}} \left[V_B + \frac{i_s}{2[2\beta_1 I_B]^{1/2}} \right] + V_T. \tag{4.32}$$

So, the condition that Q_1 operates in the saturation region is,

$$V_{gs2} < \frac{V_{dd}}{2} - V_B.$$

Substituting (4.32) yields,

$$\beta_2 > \frac{1}{8I_B} \left(\frac{i_s + 4I_B}{\frac{V_{dd}}{2} - V_T - [\frac{2I_B}{\beta_1}]^{1/2}} \right)^2. \tag{4.33}$$

Next, we perform a transient analysis. The differential equation is,

$$C_{bl}\frac{dx}{dt} + 2\beta_1 V_B x = i_s. \tag{4.34}$$

Solving this equation yields,

$$x = \frac{i_s}{2\beta_1 V_B}\left[1 - \exp\left(-\frac{2\beta_1 V_B}{C_{bl}}t\right)\right]. \tag{4.35}$$

If we neglect the parasitic capacitances, the output current is expressed as,

$$\begin{aligned} i_{out} &= \frac{\beta_3}{\beta_2}(i_n - i_p) \\ &= \frac{\beta_3}{\beta_2}i_s\left[1 - \exp\left(-\frac{2\beta_1 V_B}{C_{bl}}t\right)\right]. \end{aligned} \tag{4.36}$$

The output voltage is expressed as,

$$\frac{dy}{dt} = \frac{i_{out}}{C_0}. \tag{4.37}$$

>From (4.36) and (4.37), we get,

$$y = \frac{\beta_3}{\beta_2}\frac{i_s}{C_0}\left[t + t_a\left\{\exp\left(-\frac{t}{t_a}\right) - 1\right\}\right]. \tag{4.38}$$

where,

$$t_a = \frac{C_{bl}}{2\beta_1 V_B}.$$

A parabolic approximation of exponential function yields,

$$y \approx \frac{1}{2}\frac{\beta_3}{\beta_2}\frac{i_s}{C_0}\frac{t^2}{t_a}. \tag{4.39}$$

The delay time is expressed as,

$$\tau = \sqrt{\frac{\beta_2}{\beta_3}\frac{C_0}{i_s}t_a}.$$

This result suggests that the delay time is insensitive to the input capacitance, since it is proportional to the square root of the input capacitance. It is in contrast to the results of the previous sense amplifiers where the delay time is proportional to the input capacitance.

Figures 4.11 and 4.12 show the SPICE simulation results for two different input capacitances. The delay time does not increase much by doubling the input capacitance.

>From these results, this configuration allows high-speed sensing in the case that the input is highly capacitive. The only drawback of this sense amplifier is that it consumes a large area, since the transistor size should be large enough to operate in the saturation region.

4.3 Discussion

In this section, performances of the sense amplifiers are compared and discussed based on the previous results. Figure 4.13 shows the delay time calculated as a function of input capacitance. Transistor sizes are determined so that the the DC current becomes 10 mA for the current sense amplifier and the peak current becomes 100 mA for the dynamic cross-coupled sense amplifier. Note that a split bitline array configuration is assumed for the dynamic cross-coupled sense amplifier, so the capacitance used for the calculation is one half of that of the current sense amplifier. The delay time of the transresistance sense amplifier is almost equal to that of the dynamic cross-coupled sense amplifier. The class-A push-pull current sense amplifier shows better performance over the other two amplifiers when the input capacitance is larger than 2 pF (total BL capacitance).

Figure 4.14 shows the power dissipation calculated as a function of the input capacitance. The following are assumed for the calculation:

(1) cycle time = 100 ns,

(2) performs 2-READs and 2-WRITEs in a cycle, and

(3) has 100 bitlines ($\approx 3 \times 32$–*bit*).

We neglect the DC current of the dynamic cross-coupled sense amplifier during the sensing operation, and other dynamic power consumptions due to parasitics. This figure shows that if the DC bias current is less than 10 mA (total) and the input capacitance is larger than 1 pF, dynamic power dominates the total power dissipation.

Table 4.1 summarizes the design tradeoffs for three sense amplifiers. From these results, it seems feasible to sense the signal current of 50μA. For detailed discussions, further studies to estimate the areas and parasitics are needed.

CHAPTER 5 CONCLUSION

The feasibility of the multiport memory with more than two ports has been studied analytically. We conclude the following:

(1) The analytic expressions of the static noise margins of the possible multiport memory cells are derived. Optimal design of multiport memories becomes possible through this analysis.

(2) It is found that the one-side clocked pseudo-static single-ended cell is more stable than the static differentially accessed cell. Noise margin is expected to be as much as 0.5 V for a triple-port single-ended cell. For more stable triple-port memories or multiport memories with more than three ports, the two-side clocked single-ended access cell is feasible.

(3) The active area for the optimally designed multiport memory cell is estimated. The single-ended cell has an active area that is as small as that of the differentially accessed cell, by employing a split-bitline array configuration , a dummy cell, or using a high sensitive sense amplifier.

(4) Three possible sense amplifiers which are capable of sensing the signal current provided by the minimum size single-ended cell are presented. It is revealed that the class-A push-pull current sense amplifier has better performance for high bitline capacitances (>2 pF). DC power dissipation of the current sense amplifier is not a dominant factor if the circuit is properly designed. A special circuit technique is needed for fast recovery after a write operation.

Through these results, it is concluded that the triple-port memory is feasible through the use of the pseudo-static single-ended cell.

Several topics which future research may address are:

(1) efficient layout design style for multiport memories;

(2) study of the effect of process variation on the stability of the multiport memory;

(3) study of the effect of the internal and external noise on the stability of the multiport memory;

(4) other possible memory cell configurations to improve operating margins as presented in Chapter 2.2.

APPENDIX 1 DERIVATION OF STATIC NOISE MARGIN OF DIFFERENTIALLY ACCESSED CELL

The detailed analysis of the read noise margin of the differentially accessed cell (SRAM) was performed by E. Seevinck et.al.[11]. Here, we derive the write noise margin of the differentially accessed cell.

In Figure A1.1, equating currents of the MOSFETs for each inverter yields,

$$\frac{\beta_{na}}{2}(V_{dd} - x - V_N - V_T)^2 = \beta_{nd}(x + V_N)\left[y + V_N - V_T - \frac{x + V_N}{2} \right] \tag{A1.1}$$

$$\beta_{pd}(V_{dd} - y)\left[V_{dd} - x - V_T - \frac{V_{dd} - y}{2} \right] = \beta_{na} y \left[V_{dd} - V_T - \frac{y}{2} \right]. \tag{A1.2}$$

By using $V_s = V_{dd} - V_T$, $r = \frac{\beta_{nd}}{\beta_{na}}$, and $p = \frac{\beta_{pd}}{\beta_{nd}}$, (A1.1) and (A1.2) can be rewritten as,

$$(V_s - V_N - x)^2 = r(x + V_N)(2y - x + V_N - 2V_T) \tag{A1.3}$$

$$rp(V_{dd} - y)(V_s - V_T - 2x + y) = y(2V_s - y). \tag{A1.4}$$

(A1.4) is rewritten as,

$$(rp - 1)y^2 + 2[V_s - rp(V_T + x)]y - rpV_{dd}(V_s - V_T - 2x) = 0. \tag{A1.5}$$

We approximate (A1.5) as a linear function at $x \approx V_T$ as,

$$y = -kx + V_0, \tag{A1.6}$$

for following two cases.

(1) if $rp = 1$, then y becomes,

$$y = V_{dd} - \frac{V_{dd}(V_{dd} - 2V_T)}{2(V_{dd} - 2V_T - x)}. \tag{A1.7}$$

Thus, we get,

$$k = -\left[\frac{dy}{dt}\right]_{x = V_T} = \frac{V_{dd}(V_{dd} - 2V_T)}{2(V_{dd} - 3V_T)^2}. \tag{A1.8}$$

$$V_0 = (1 - k)V_{dd} + 4kV_T. \tag{A1.9}$$

(2) if $rp \neq 1$, then y becomes,

$$y = \frac{1}{rp - 1}\left[- V_s + rp\,(V_T + x) + \sqrt{[V_s - rp\,(V_T + x)]^2 + rp\,(rp - 1)V_{dd}(V_s - V_T - 2x)}\right] \tag{A1.10}$$

thus, we get,

$$k = \frac{rp}{rp - 1}\left[\frac{V_r + (rp - 1)V_{DD}}{[V_r^2 + rp\,(rp - 1)V_p^2]^{1/2}} - 1\right] \tag{A1.11}$$

$$V_0 = kV_T + \frac{\frac{rp}{k}V_{dd} - V_r}{rp - 1 + \frac{rp}{k}} \tag{A1.12}$$

where,

$$V_r = V_{DD} - (2rp + 1)V_T$$

$$V_p = \sqrt{V_{DD}(V_{DD} - 4V_T)}.$$

Substituting (A1.6) into (A1.3) yields,

$$[1 + r(2k + 1)]X^2 - 2(V_s + rA)X + V_s^2 = 0 \tag{A1.13}$$

where,

$$X = x + V_N$$

$$A = V_0 + (k + 1)V_N - V_T.$$

By using the condition of coinciding roots, we get,

$$V_s + rA = V_s\sqrt{1 + r(2k + 1)}. \tag{A1.14}$$

Substituting A into (A1.14) yields,

$$(k + 1)V_N = - V_0 + V_T + \frac{\sqrt{1 + r(2k + 1)} - 1}{r}V_s.$$

Thus, the write noise margin is expressed as,

$$WNM = V_N$$

$$= \frac{1}{k+1}\left[\frac{\sqrt{1+r(2k+1)}-1}{r}(V_{DD}-V_T)-(k-1)V_T-\frac{(\frac{rp}{k}-1)V_{DD}+(2rp+1)V_T}{rp-1+\frac{rp}{k}}\right] \quad \text{(A1.15)}$$

APPENDIX 2 DERIVATION OF STATIC NOISE MARGIN OF SINGLE-ENDED ACCESS CELL

We can use the same expressions of the read noise margins for the write noise margins by substituting, $V_{bl} = 0$ for "0" write, and $V_{bl} = V_{dd}$ for "1" write. Thus, we derive here the two read noise margins.

A2.1 "0" Read Noise Margin (RNM_L)

In Figure A2.1, only Q_{2n} operates in the saturation region. By equating currents of the MOSFETs for each inverter, we get,

$$(V_{bl} - y)\left[V_{dd} - V_T - y - \frac{V_{bl} - y}{2}\right] = \frac{r}{N} y \left[x - V_T - \frac{y}{2}\right] \qquad \text{(A2.1)}$$

$$(y + V_N - V_T)^2 = p\ (V_{dd} - V_N - x)\,(V_{dd} - V_N - 2V_T - 2y + x) \qquad \text{(A2.2)}$$

where, $r = \frac{\beta_{nd}}{\beta_{na}}$ and $p = \frac{\beta_{pd}}{\beta_{nd}}$. We approximate (A2.1) as a linear function at $x \approx V_{dd} - V_T$ as,

$$y = -kx + V_0.$$

(A2.1) can be rewritten as,

$$(r + N)y^2 - 2[N\ V_{dd} - (r+N)\ V_T + rN\ x]\ y + N\ V_s^2 = 0 \qquad \text{(A2.3)}$$

where,

$$V_s = \sqrt{V_{bl}(2V_{dd} - 2V_T - V_{bl})}.$$

Solving (A2.3) yields,

$$y = \frac{N\ V_{dd} - (r+N)\ V_T + rN\ x - \sqrt{[N\ V_{dd} - (r+N)\ V_T + rN\ x]^2 - (r + N)\ V_s^2}}{r + N} \qquad \text{(A2.4)}$$

Thus, we get,

$$k = -\left[\frac{dy}{dt}\right]_{x = V_{dd} - V_T}$$
$$= \frac{r}{r+N}\left[\sqrt{\frac{(r+N)V_r^2}{(r+N)V_r^2 - NV_s^2}} - 1\right] \tag{A2.5}$$

and,

$$V_0 = k\,(V_{dd} - V_T) + \frac{r+N}{r+N+\frac{r}{k}}\,V_T \tag{A2.6}$$

where,

$$V_r = V_{dd} - \frac{2r+N}{r+N}V_T.$$

The read noise margin can be given by substituting $y = -kx + V_0$ into (A2.2) as,

$$RNM_L = V_T - \frac{1}{k+1}\left[\frac{V_{dd} - \frac{2r+N}{r+N}V_T}{1 + \frac{r}{k(r+N)}} - \frac{V_{dd} - 2V_T}{1 + \frac{k}{q} + [\frac{1}{q}(1+2k+\frac{k^2}{q})]^{1/2}}\right]. \tag{A2.7}$$

A2.2 "1" Read Noise Margin (RNM_H)

At first, we calculate the bitline voltage , V_{b0} , which gives the bound whether Q_{1a} operates in the saturation region or in the linear region. Let $y = V_{dd} - V_T$ for $x = V_T$ for the inverter 1. It gives,

$$\frac{N\beta_{na}}{2}(V_{dd} - V_{b0} - V_T)^2 = \beta_{pd} V_T \left[V_{dd} - \frac{5}{2} V_T \right]. \quad \text{(A2.8)}$$

Thus we get,

$$V_{b0} = V_{dd} - V_T - \sqrt{\frac{rp}{N} V_T (2V_{dd} - 5V_T)}. \quad \text{(A2.9)}$$

We analyze the read noise margin for following two cases.

(1) if $V_{bl} \le V_{b0}$

In this case, Q_{1a} operates in the linear region. By equating currents of the MOSFETs for each inverter, we get,

$$(y - V_{bl}) \left[V_{dd} - V_T - V_{bl} - \frac{y - V_{bl}}{2} \right] = \frac{rp}{N}(V_{dd} - y) \left[V_{dd} - x - V_T - \frac{V_{dd} - y}{2} \right] \quad \text{(A2.10)}$$

$$(x - V_N) \left[y - V_N - V_T - \frac{x - V_N}{2} \right] = \frac{q}{2}(V_{dd} + V_N - V_T - y). \quad \text{(A2.11)}$$

We approximate (A2.10) as a linear function at $x \approx V_T$ as, $y = -kx + V_0$. (A2.10) can be rewritten as,

$$(rp - N)y^2 + 2[N\ V_{dd} - (rp+N)\ V_T - rpNx\,]\,y - N\ V_s^2 + rpV_{dd}(V_{dd} - 2V_T - 2x) = 0 \quad \text{(A2.12)}$$

where,

$$V_s = \sqrt{V_{bl}(2V_{dd} - 2V_T - V_{bl})}.$$

(a) If $rp = N$, then

$$y = V_{dd} + \frac{V_s^2 - V_{dd}(V_{dd} - 2V_T)}{2(V_{dd} - 2V_T - x)}. \tag{A2.13}$$

Thus, we get,

$$k = -\left(\frac{dy}{dt}\right)_{x = V_T} = \frac{V_{dd}(V_{dd} - 2V_T) - V_s^2}{2(V_{dd} - 3V_T)^2} \tag{A2.14}$$

and,

$$V_0 = (1 - k)V_{dd} + 4kV_T. \tag{A2.15}$$

(b) If $rp \neq N$, then,

$$y = \frac{-B + \sqrt{B^2 + (rp - N)[NV_s^2 + rpV_{dd}(V_{dd} - 2V_T - 2x)]}}{rp + N} \tag{A2.16}$$

where,

$$B = NV_{dd} - (rp+N)V_T - rpx.$$

Thus we get,

$$k = -\left(\frac{dy}{dt}\right)_{x = V_T}$$

$$= \frac{rp}{rp-N}\left[\frac{(rp-N)V_{dd} + NV_r}{[(rp-N)(NV_s^2 + rpV_p^2) + N^2V_r^2]^{1/2}} - 1\right] \tag{A2.17}$$

and,

$$V_0 = kV_T + \frac{\frac{rp}{Nk}V_{dd} - V_r}{\frac{rp}{N} - 1 + \frac{rp}{Nk}} \tag{A2.18}$$

where,

$$V_r = V_{dd} - (\frac{2rp}{N}+1)V_T$$

$$V_p = \sqrt{V_{dd}(V_{dd} - 4V_T)}.$$

Substituting $y = -kx + V_0$ into (A2.11) yields,

$$(qk^2 + 2k + 1)X^2 - 2(qkA - V_{dd} + 2V_T + A)X + qA^2 = 0 \qquad \text{(A2.19)}$$

where,

$$X = x - V_N$$
$$A = V_{dd} - V_0 - V_T + (k+1)V_N.$$

By using the condition of coinciding roots, we get,

$$(qk + 1)A - V_{dd} + 2V_T = -A\sqrt{q(qk^2 + 2k + 1)}. \qquad \text{(A2.20)}$$

Substituting A into (A2.20) yields,

$$(k+1)V_N = -V_{dd} + V_0 + V_T + \frac{V_{dd} - 2V_T}{1 + qk + [q(qk^2+2k+1)]^{1/2}}. \qquad \text{(A2.21)}$$

Thus, the read noise margin is expressed as,

$$RNM_H = V_T - \frac{1}{k+1}\left[\frac{rpV_{dd} - (2rp+N)V_T}{rp - N + \frac{rp}{k}} - \frac{V_{dd} - 2V_T}{1 + qk + [q(qk^2+2k+1)]^{1/2}}\right] \qquad \text{(A2.22)}$$

(2) If $V_{bl} > V_{b0}$

In this case, Q_{1a} operates in the saturation region. By equating currents of the MOSFETs for each inverter, we get,

$$\frac{1}{2}(V_{dd} - V_{bl} - V_T)^2 = \frac{rp}{N}(V_{dd} - y)\left[V_{dd} - x - V_T - \frac{V_{dd} - y}{2}\right] \qquad \text{(A2.23)}$$

$$(x - V_N)\left[y - V_N - V_T - \frac{x - V_N}{2}\right] = \frac{q}{2}(V_{dd} + V_N - V_T - y). \qquad \text{(A2.24)}$$

We approximate (A2.23) as a linear function at $x \approx V_T$ as, $y = -kx + V_0$. (A2.10) can be rewritten as,

$$y^2 - 2(V_T + x)y + \frac{N}{rp}V_s^2 - V_{dd}(V_{dd} - 2V_T - 2x) = 0 \qquad \text{(A2.25)}$$

where,

$$V_s = V_{dd} - V_T - V_{bl}.$$

Solving (A.25) yields,

$$y = V_T + x + \sqrt{(V_{dd} - V_T - x)^2 - \frac{N}{rp} V_s{}^2}. \qquad \text{(A2.26)}$$

Thus, we get,

$$k = -\left[\frac{dy}{dt}\right]_{x = V_T} = \sqrt{\frac{rpV_r^2}{rpV_r^2 - NV_s^2}} - 1 \qquad \text{(A2.27)}$$

and,

$$V_0 = (k + 2)V_T + \frac{V_r}{k + 1}. \qquad \text{(A2.28)}$$

The read noise margin can be given by substituting $y = -kx + V_0$ into (A2.24) as,

$$RNM_H = V_T - \frac{V_{dd} - 2V_T}{k+1}\left[\frac{1}{1 + \frac{1}{k}} - \frac{1}{1 + qk + [q(qk^2 + 2k + 1)]^{1/2}}\right]. \qquad \text{(A2.29)}$$

REFERENCES

[1] R.W.Sherburne,Jr., M.G.H.Katevenis, D.A.Patterson, and C.H.Sequin, "A 32-Bit NMOS Microprocessor with a Large Register File", *IEEE J. Solid-State Circuits,* vol.SC-19, no.5, pp682-689, Oct. 1984.

[2] J.P.Norsworthy, D.M.Pfeiffer, M.K.Corry, and J.A.Thompson, "A Parallel Image Processor Chip", *in ISSCC Dig. Tech. Papers,* pp.158-159, Feb. 1988.

[3] J.A.Fisher, "Very Long Instruction Word Architectures and The ELI-512", *Yale University Department of Computer Science, Technical Report 253,* 1982.

[4] Y.N.Patt, W.M.Hwu, and M.Shebanow, "HPS, A New Microarchitecture: Rationale and Introduction", *Proceedings of The 18th Annual Workshop on Microprogramming,* Asilomar, California, Dec. 1985.

[5] W.M.Hwu and Y.N.Patt, "HPSm, A High Performance Restricted Data Flow Architecture Having Minimal Functionality", *The 13th International Symposium on Computer Architecture Conference Proceedings,* Tokyo, Japan, Jun. 1987.

[6] W.M.Hwu, *Ph.D Dissertation,* Computer Science Division, EECS Dept., University of California, Berkeley, CA 94720, 1987.

[7] G.A.Uvieghara, "Circuit Design Considerations for the Register Alias Table (RAT) -A Smart Register File for the HPSm Dataflow Microprocessor", *M.S. Thesis,* EECS Dept., University of California, Berkeley, CA 94720, May 1987.

[8] H.Kadota, S.Ozawa, K.Kawakami, and E.Ichinohe, "A New Register File Structure for The High-Speed Microprocessor", *IEEE J. Solid-State Circuits,* vol.SC-17, no.5, pp892-897, Oct. 1982.

[9] K.J.O'Connor, "The Twin-Port Memory Cell", *IEEE J. Solid-State Circuits,* vol.SC-22, no.5, pp712-720, Oct. 1987.

[10] R.G.Stewart and A.G.F.Dingwall, "16k CMOS/SOS Asynchronous Static RAM", *in ISSCC Dig. Tech. Papers,* pp.104-105, Feb. 1979.

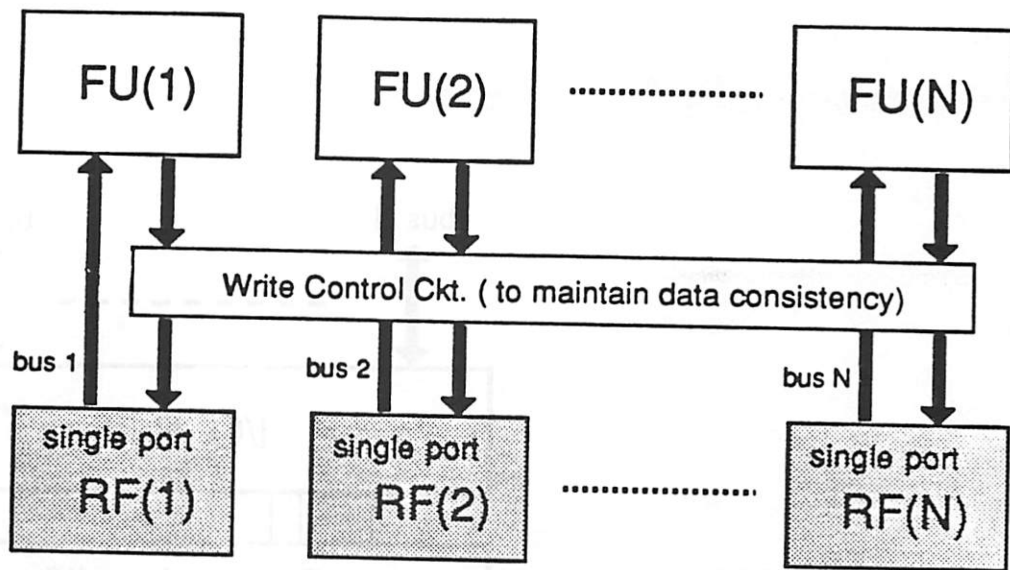
[11] E.Seevinck, F.J.List, and J.Lohstroh, "Static-Noise Margin Analysis of MOS SRAM Cells", *IEEE J. Solid-State Circuits,* vol.SC-22, no.5, pp748-754, Oct. 1987.

[12] J.Lohstroh, E.Seevinck, and J.D.Groot, "Worst-Case Static Noise Margin Criteria for Logic Circuits and Their Mathematical Equivalence", *IEEE J. Solid-State Circuits,* vol.SC-18, no.6, pp803-807, Dec. 1983.

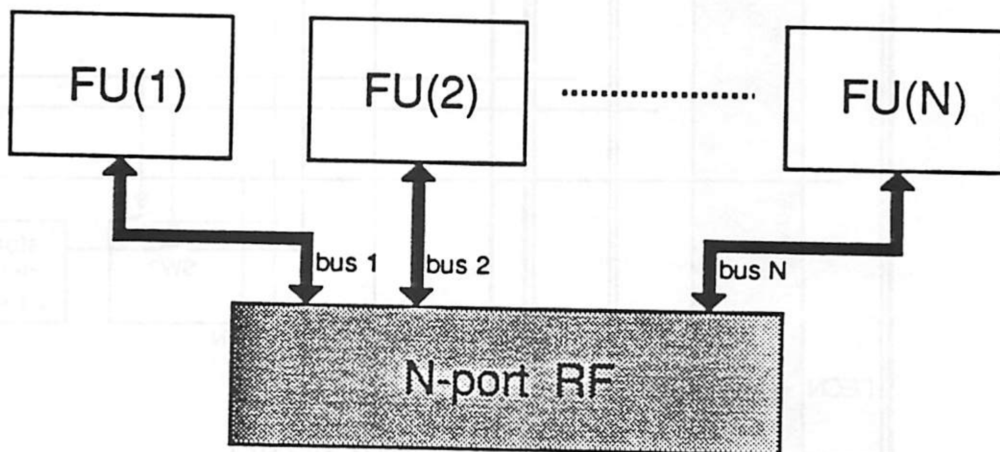
[13] J.Lohstroh, "Static and Dynamic Noise Margins of Logic Circuits", *IEEE J. Solid-State Circuits,* vol.SC-14, no.3, pp591-598, Jun. 1979.

[14] H.J.Shin, "Electronic Switching Techniques for Fiber Optic Networks", *U.C.Berkeley ERL, Memorandum No.UCB/ERL M88/15,* 1988.

[15] N.H.E. Weste and K.Eshraghian, *Principles of CMOS VLSI Design,* Addison-Wesley Publish Company, 1985.



(a)



(b)

Figure 2.1 Multiple fuction units systems with:(a) multiple singl-port register file and (b) multiport register file.

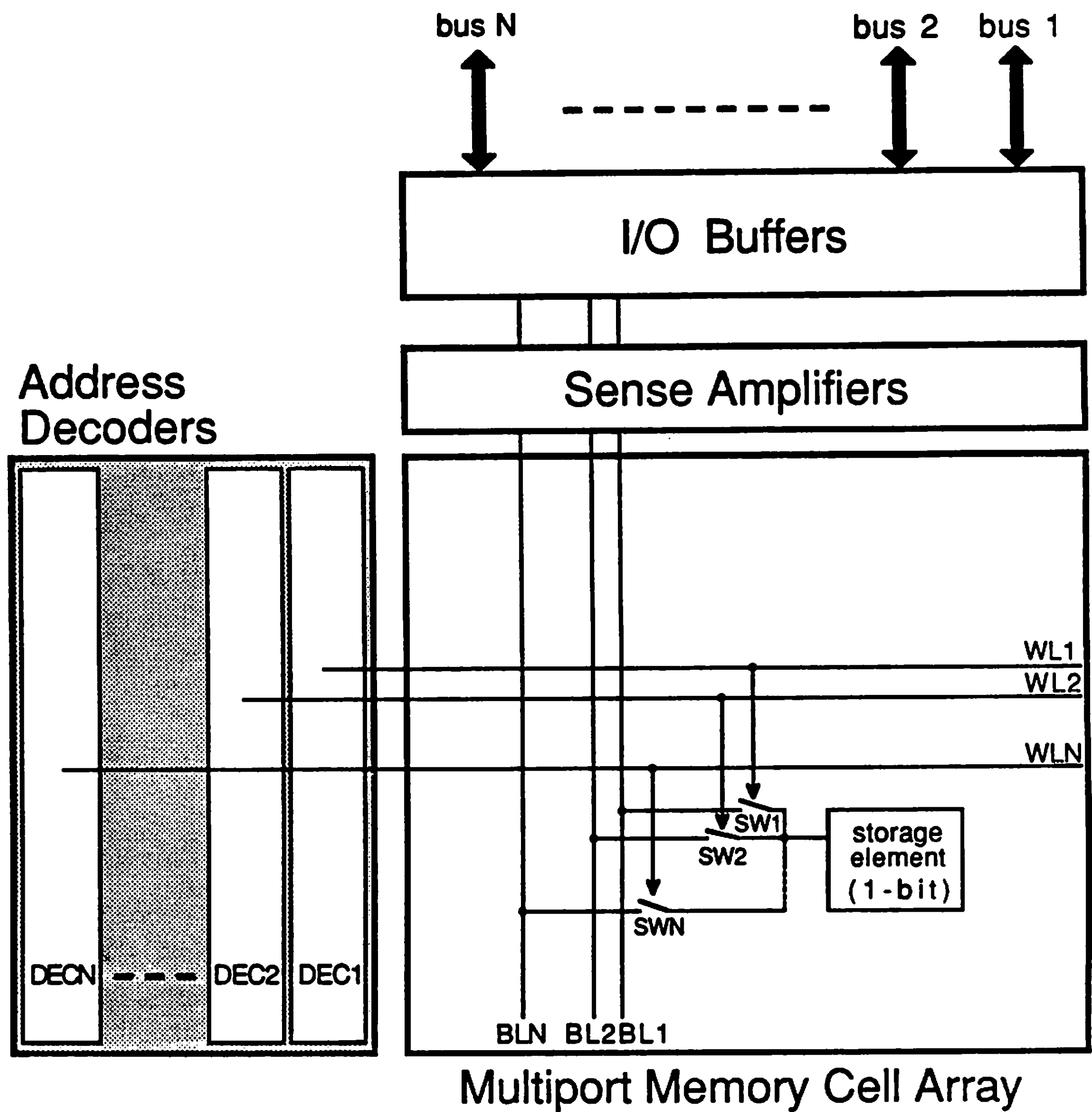


Figure 2.2 Multiport memory with common access-port memory cell.

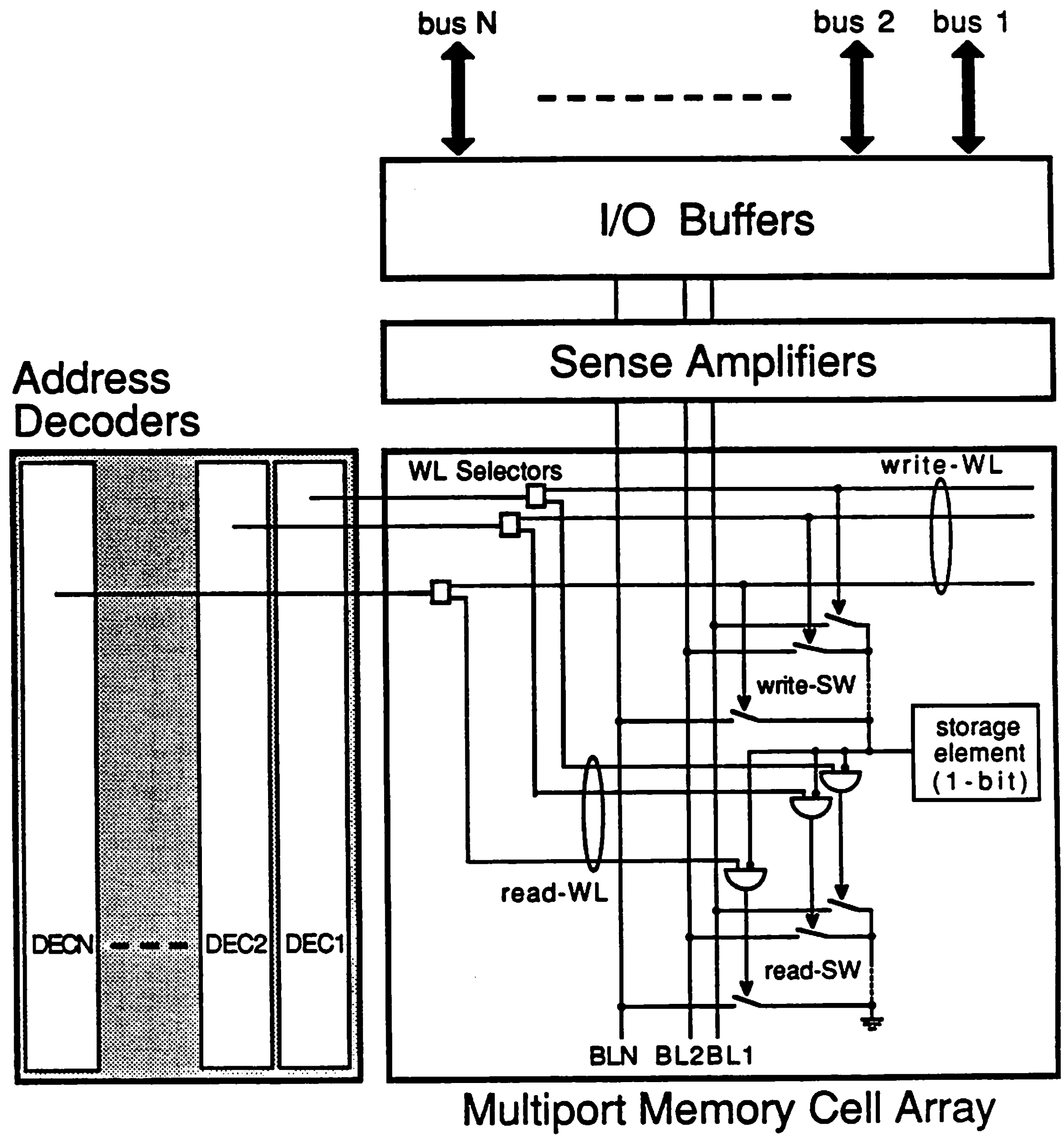


Figure 2.3 Multiport memory with seperate access-port memory cell for noise margin improvement.

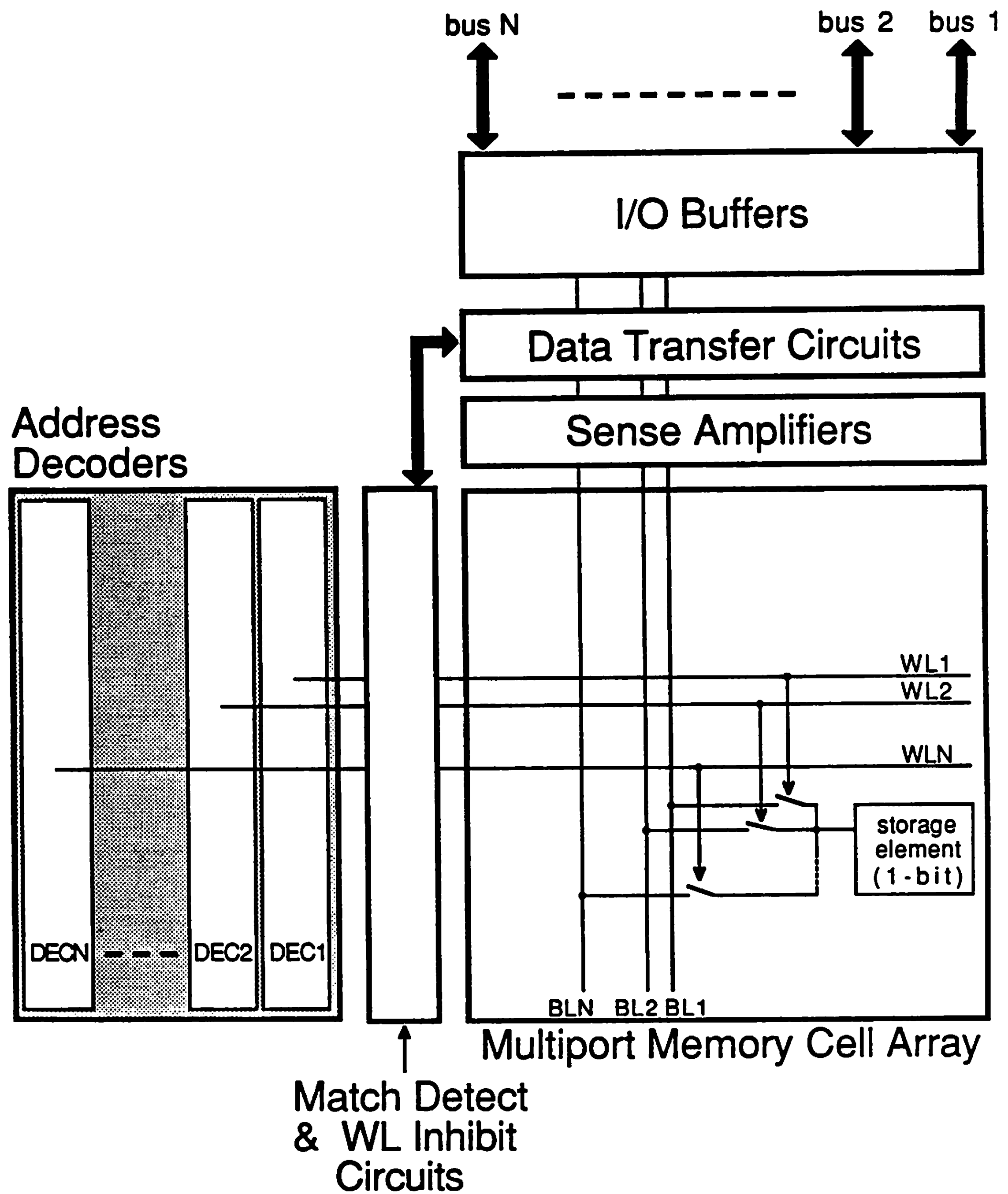


Figure 2.4 Multiport memory with common access-port memory cell and word line inhibit circuits for noise margin improvement.

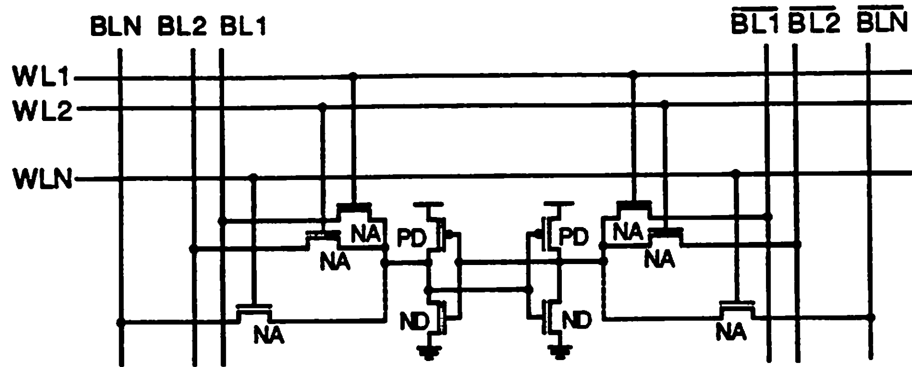
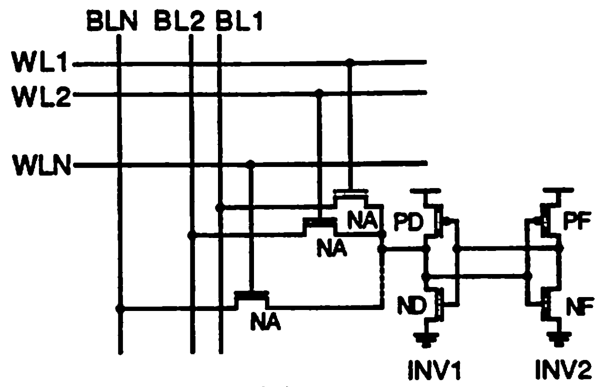
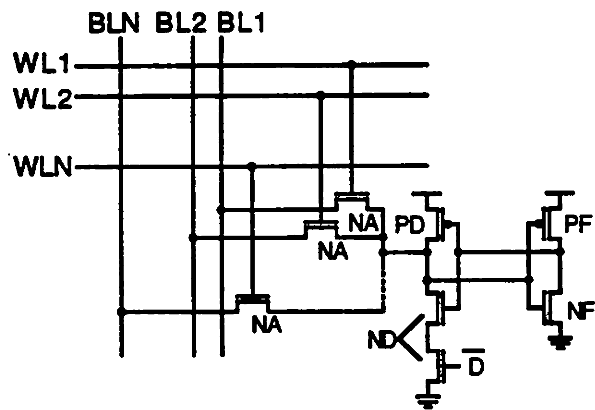


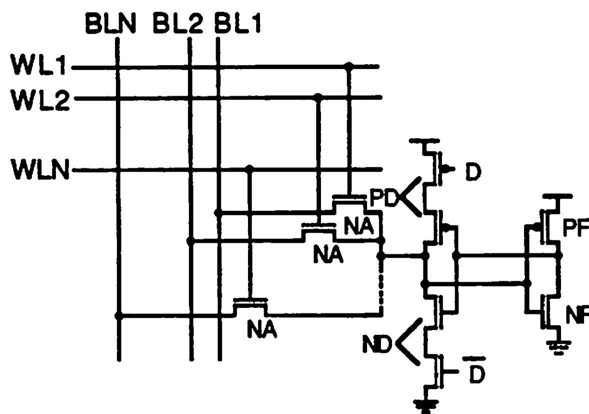
Figure 3.1 Differentially accessed multiport (N-port) memory cell.



(a)

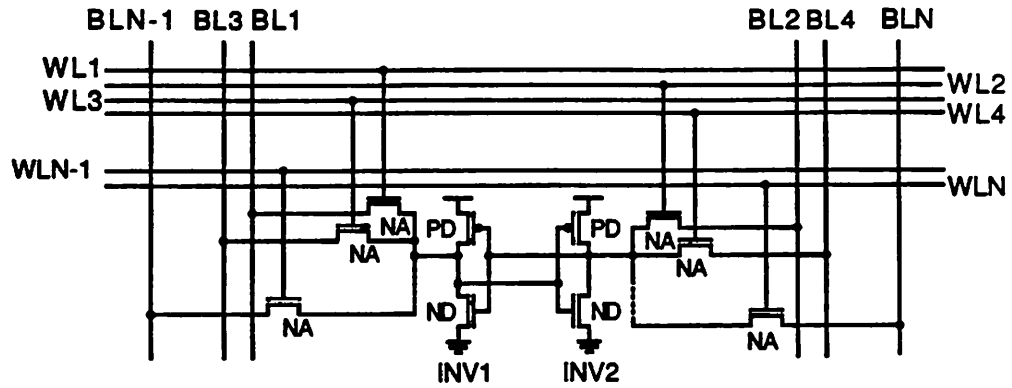


(b)

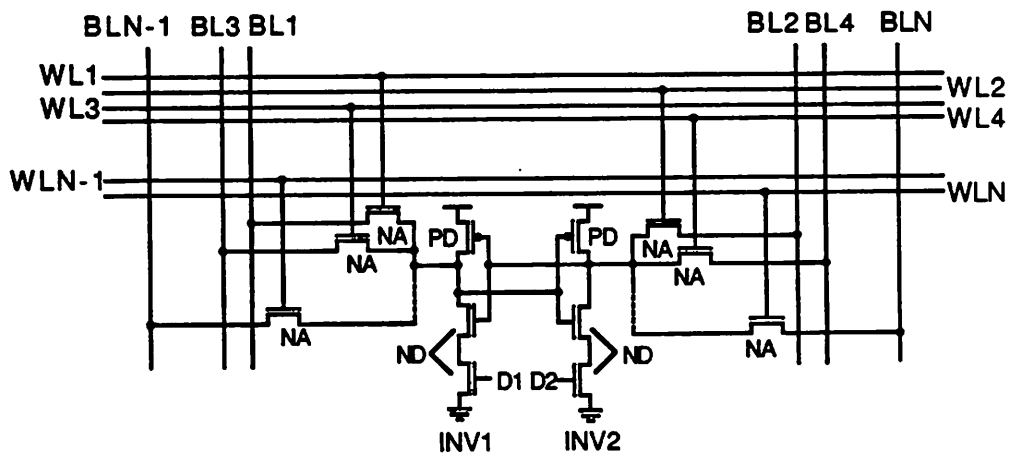


(c)

Figure 3.2 Single-ended access multiport(N-port) memory cells:(a) static version, (b) pseudo-static version with one-side clocking or (c) pseudo-static version with two-side clocking.



(a)



(b)

Figure 3.3 Modified single-ended access multiport(N-port) memory cells :(a) static version or (b) pseudo-static version with one-side clocking.

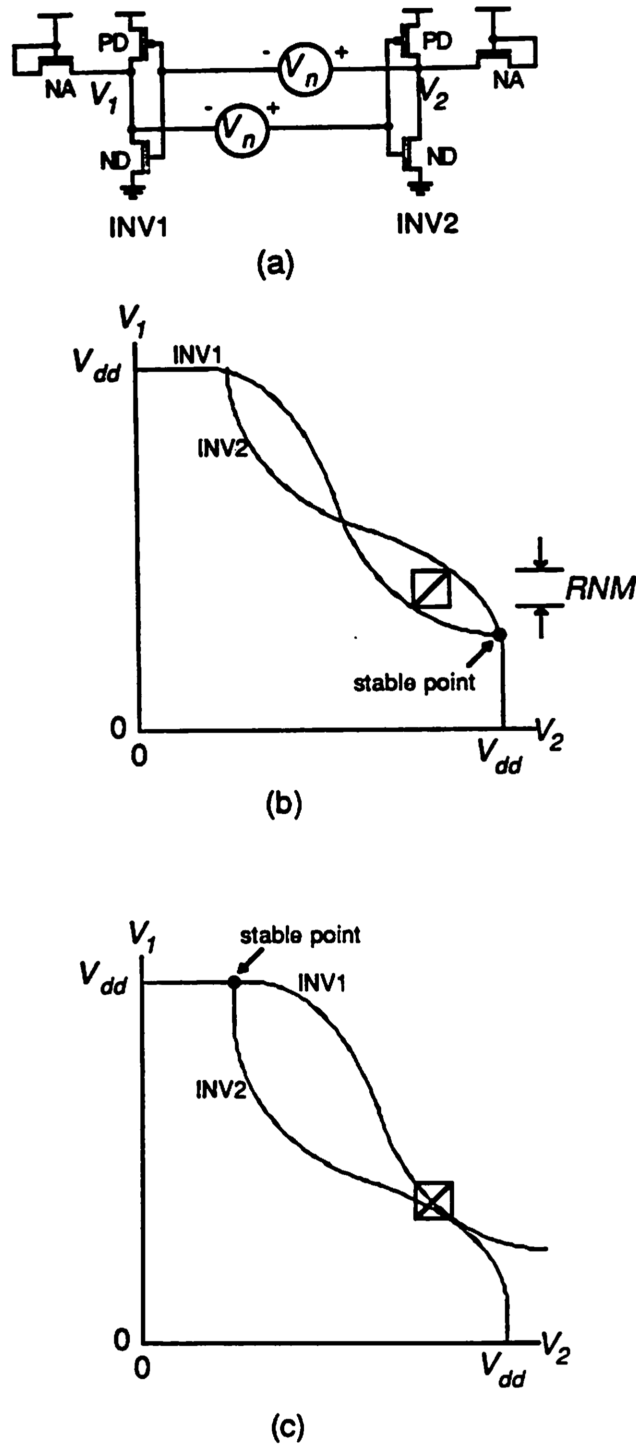
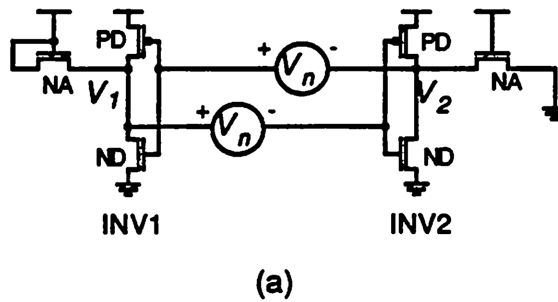


Figure 3.4 Definition of read noise margin for differentially accessed cell: (a) circuit schematic with static noise sources, and transfer characteristics (b) without noise voltage and (c) with critical noise voltage.



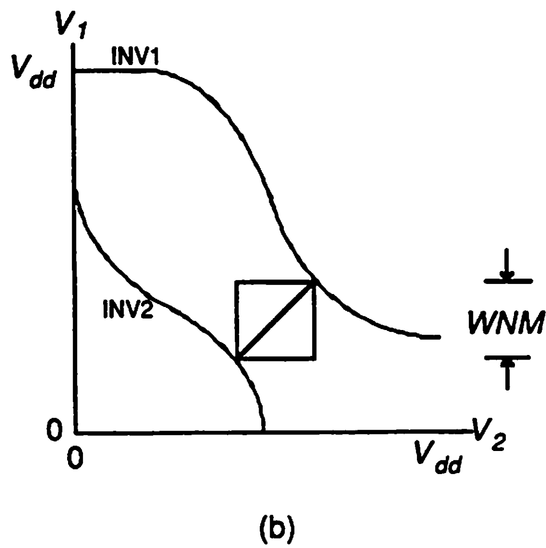


Figure 3.5 Definition of write noise margin for differentially accessed cell: (a) circuit schematic with static noise sources and (b) graphical representation of write noise margin.

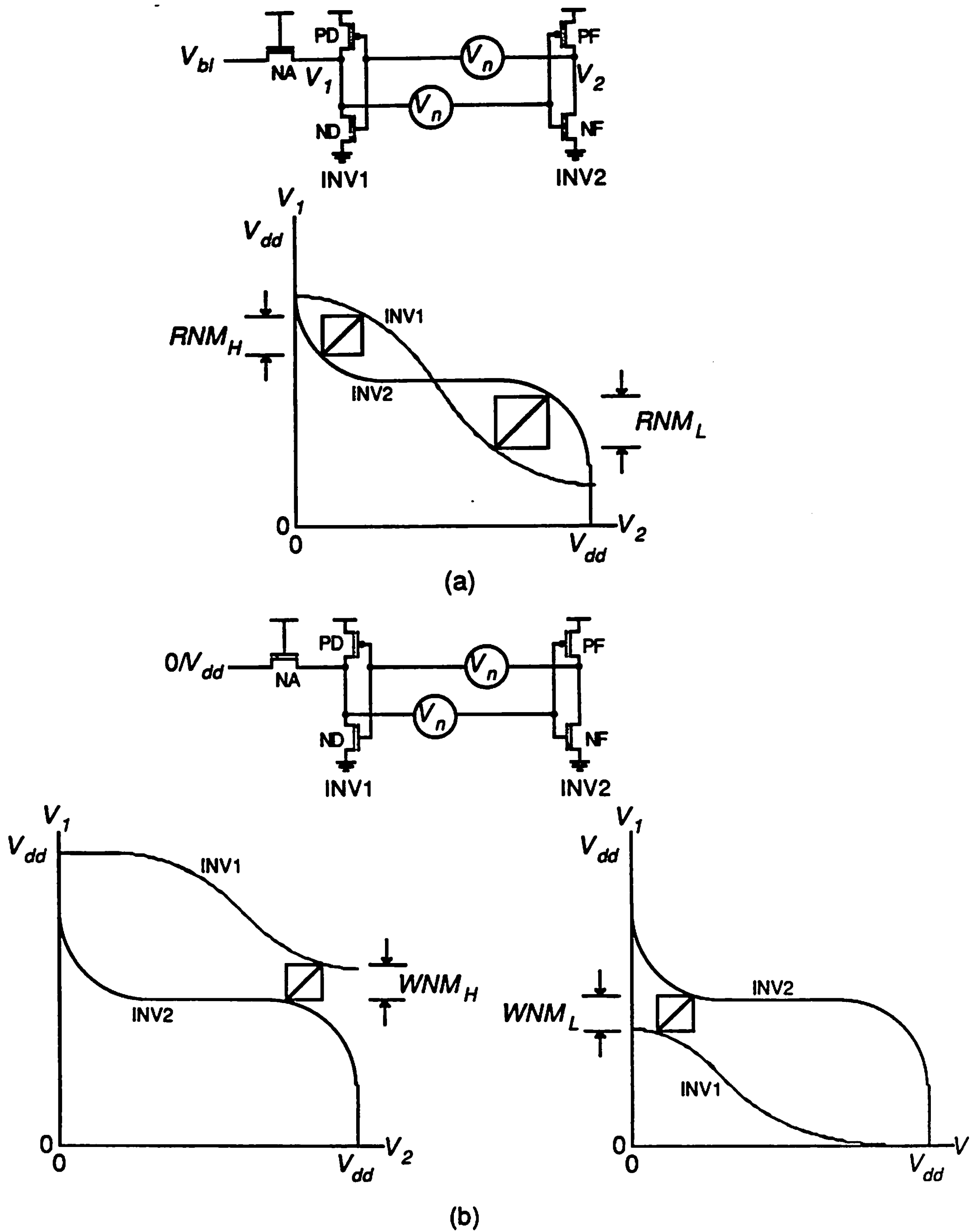
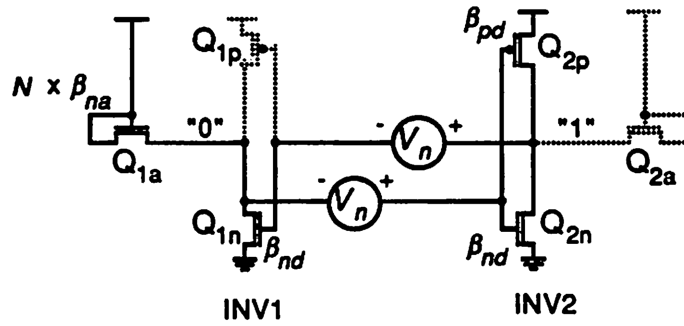
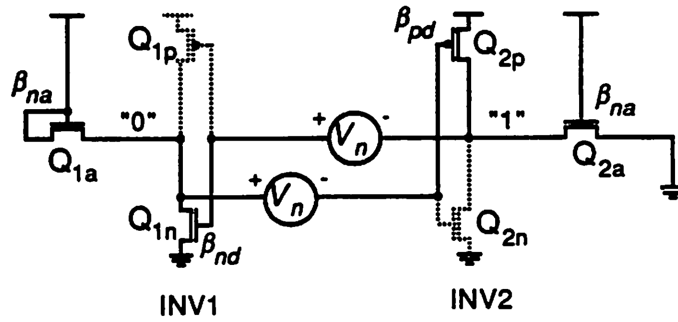


Figure 3.6 Definition of noise margins for single-ended access cell: (a) read noise margins and (b) write noise margins.



(a)



(b)

Figure 3.7 Circuit diagrams of differentially accessed cell for analysis: (a) read noise margin and (b) write noise margin.

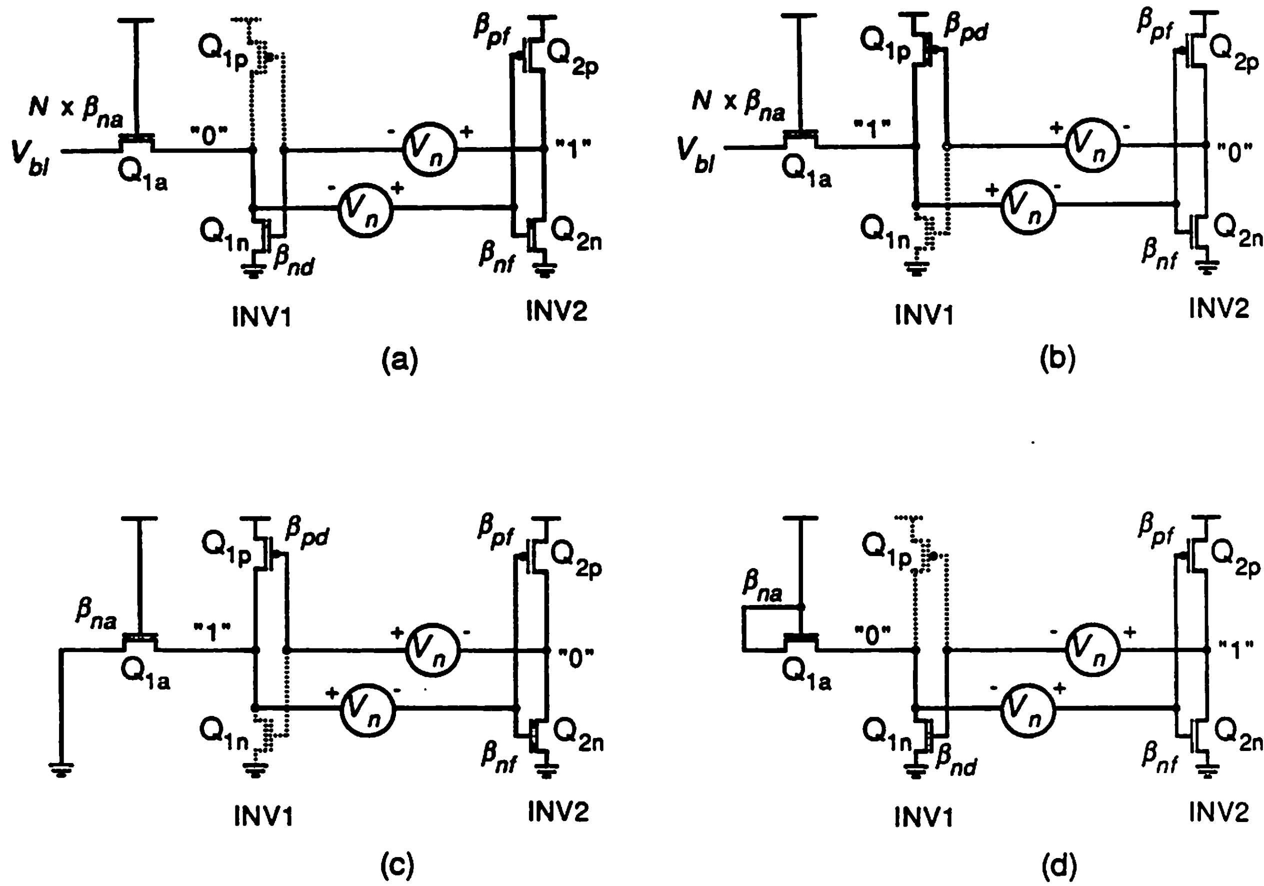
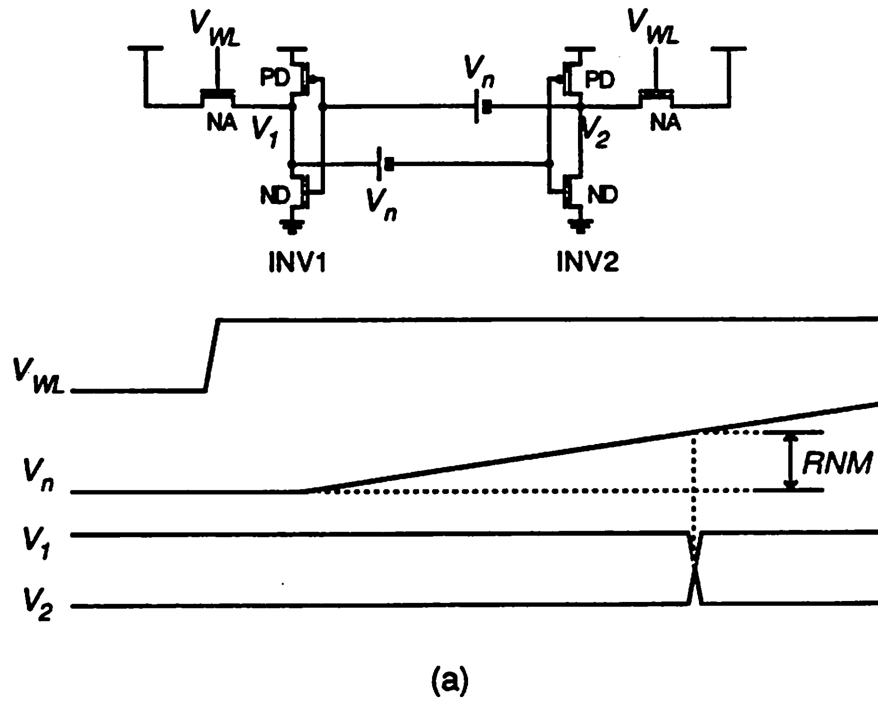
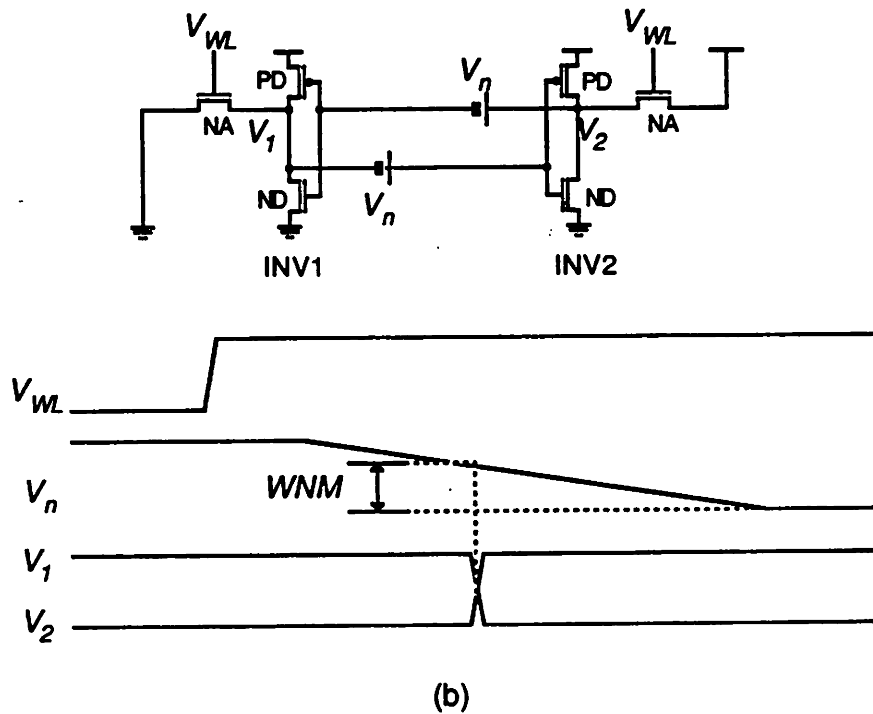


Figure 3.8 Circuit diagrams of single-ended access cell for analysis: (a) "0" read noise margin, (b) "1" read noise margin, (c) "0" write noise margin, and (b) "1" write noise margin.



(a)



(b)

Figure 3.9 SPICE simulation of noise margins for differentially accessed cell: (a) read noise margin and (b) write noise margin.

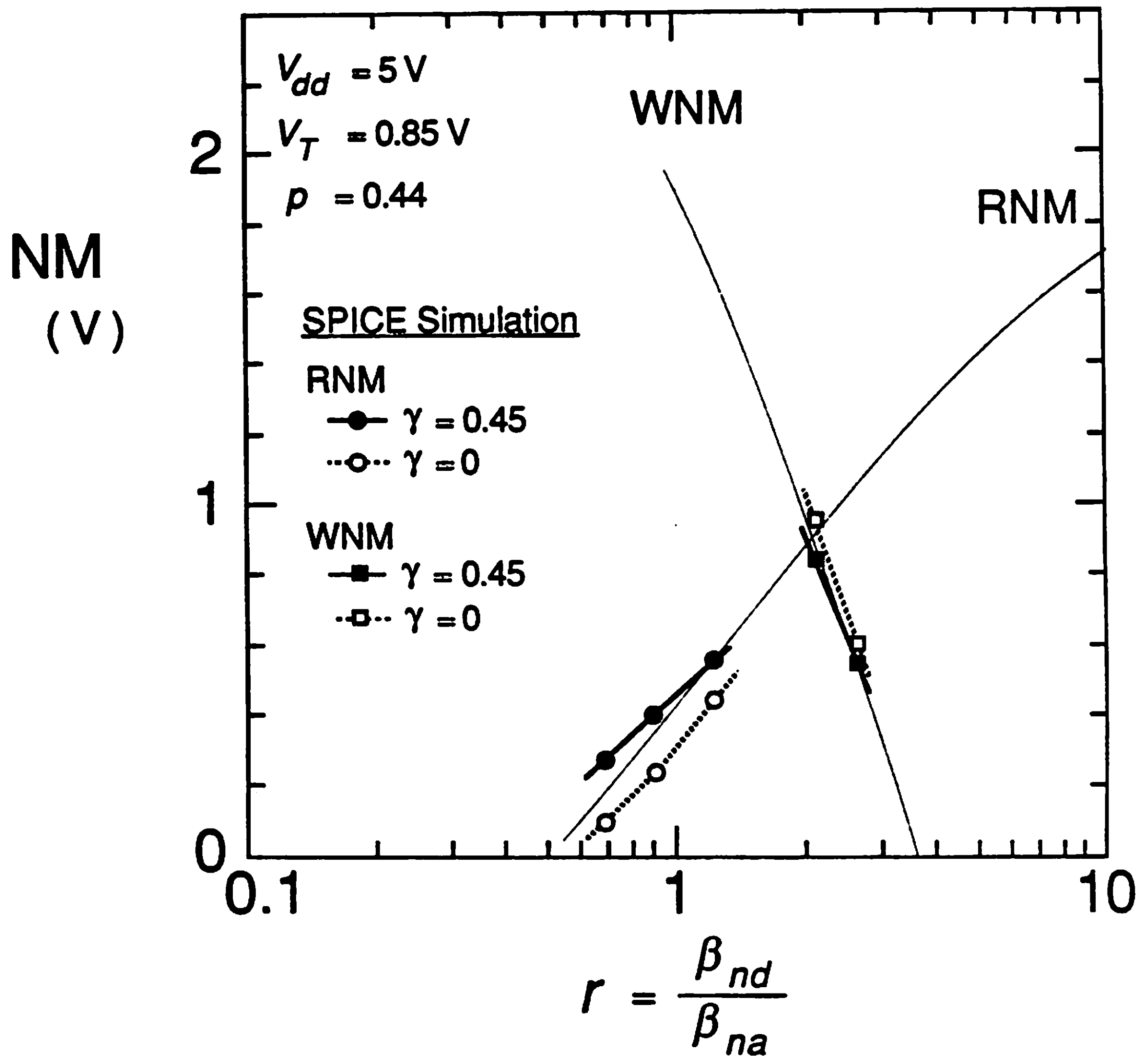


Figure 3.10 Calculated noise margins of differentially accessed cell together with SPICE simulation results.

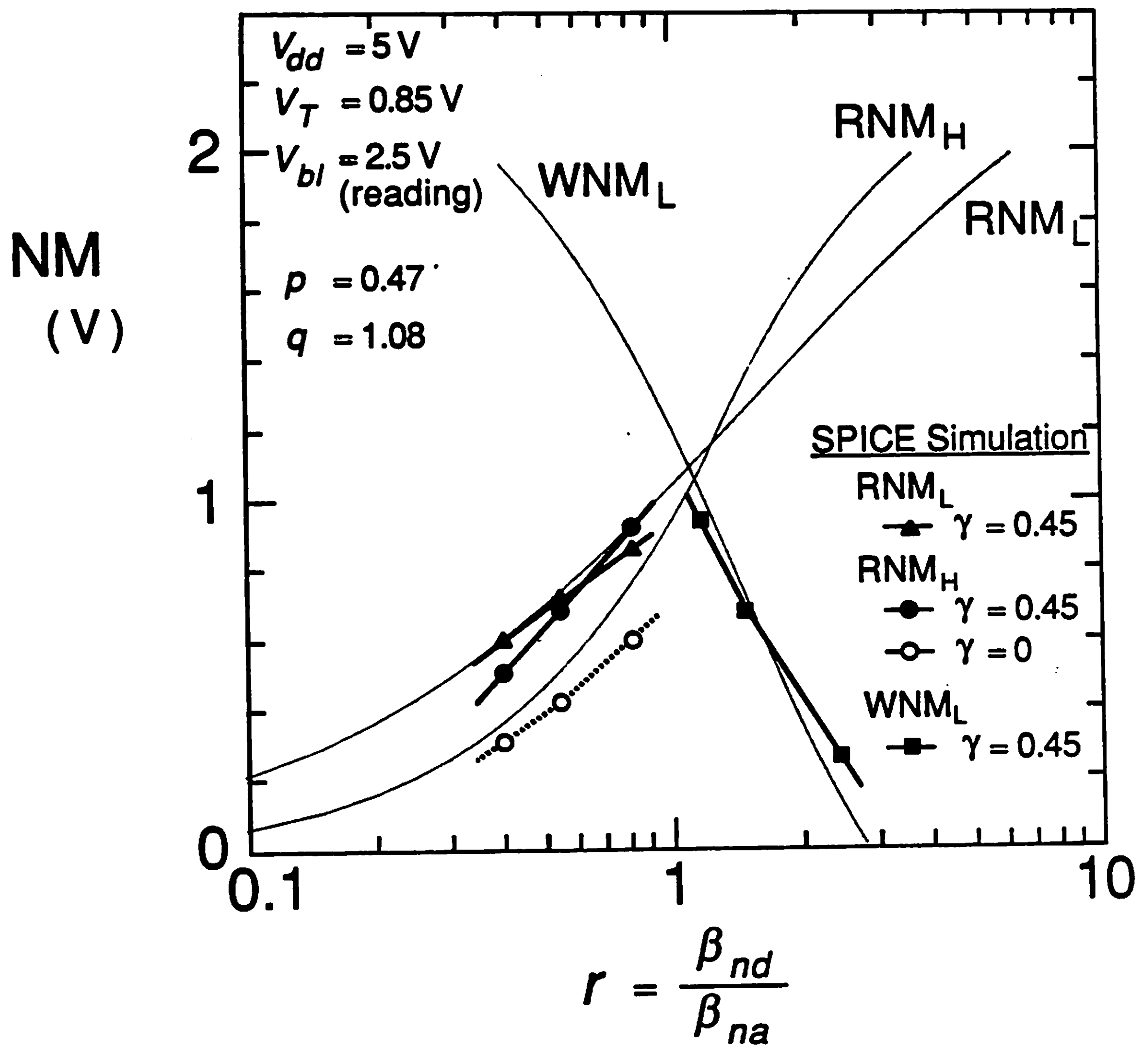


Figure 3.11 Calculated noise margins of single-ended access cell together with SPICE simulation results.

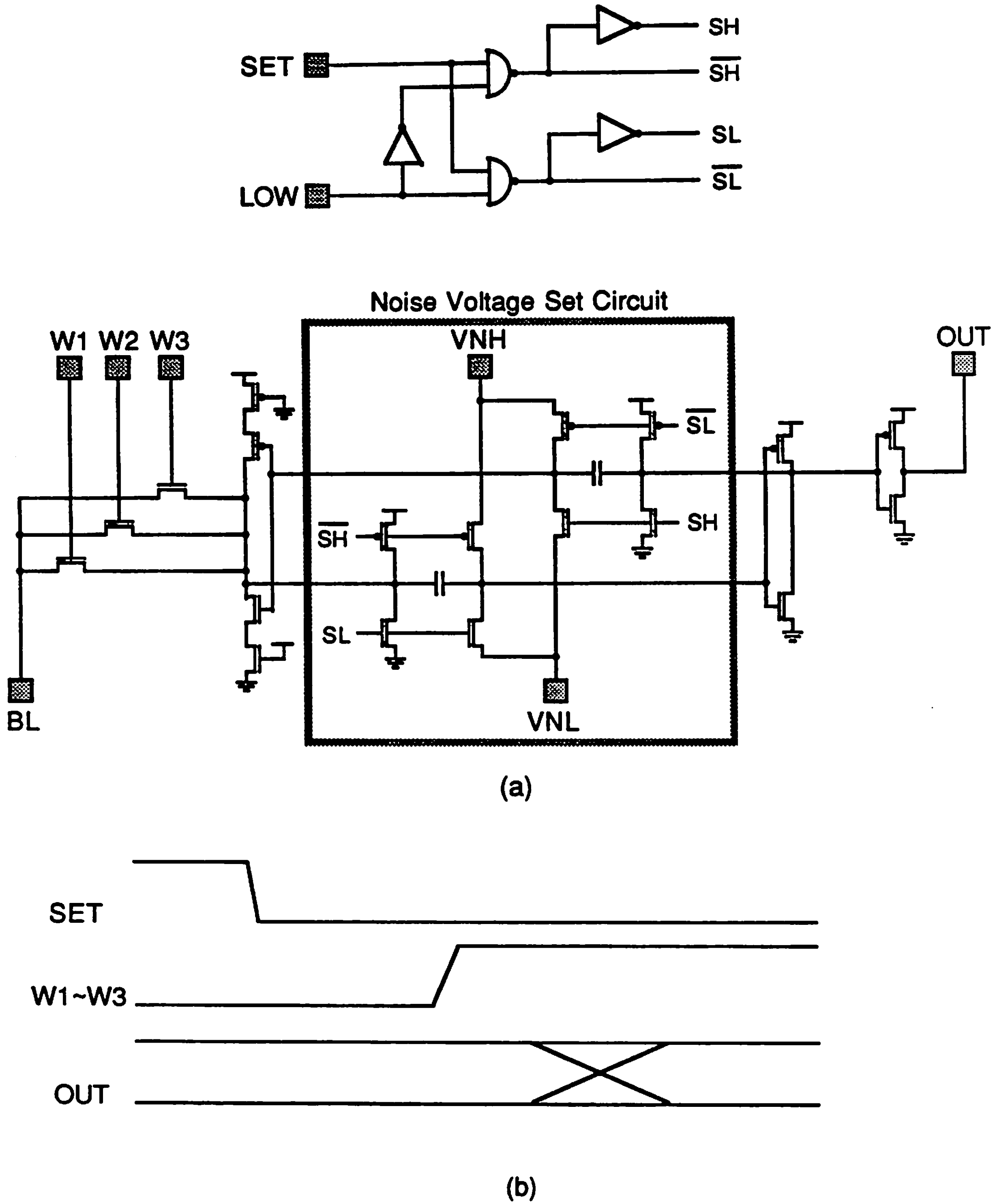


Figure 3.12 Test circuit designed for measuring noise margins of single-ended access cell.

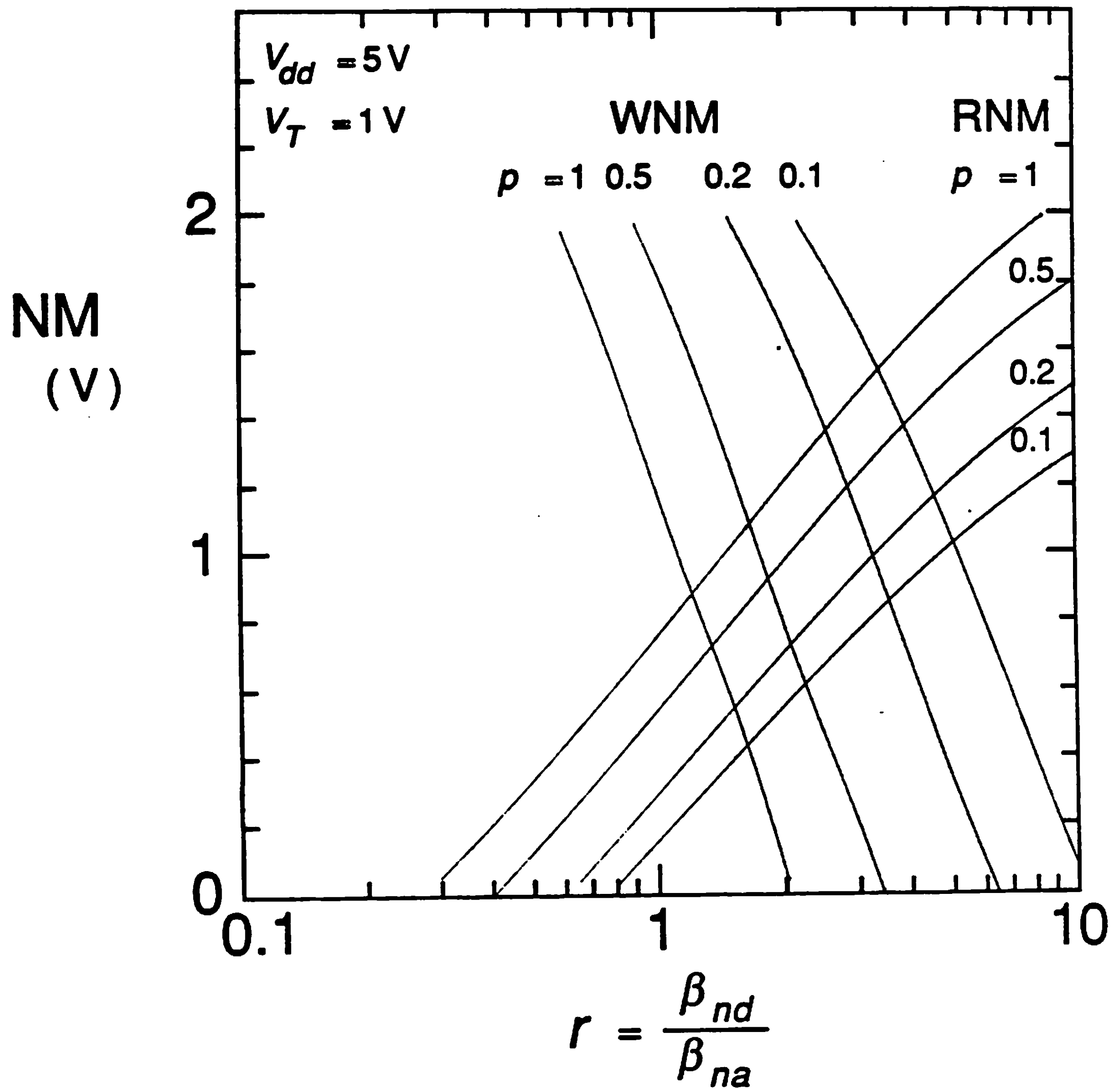


Figure 3.13 Read and write noise margins of differentially accessed cell shown as a function of beta ratio, r ,for different beta ratios, p .

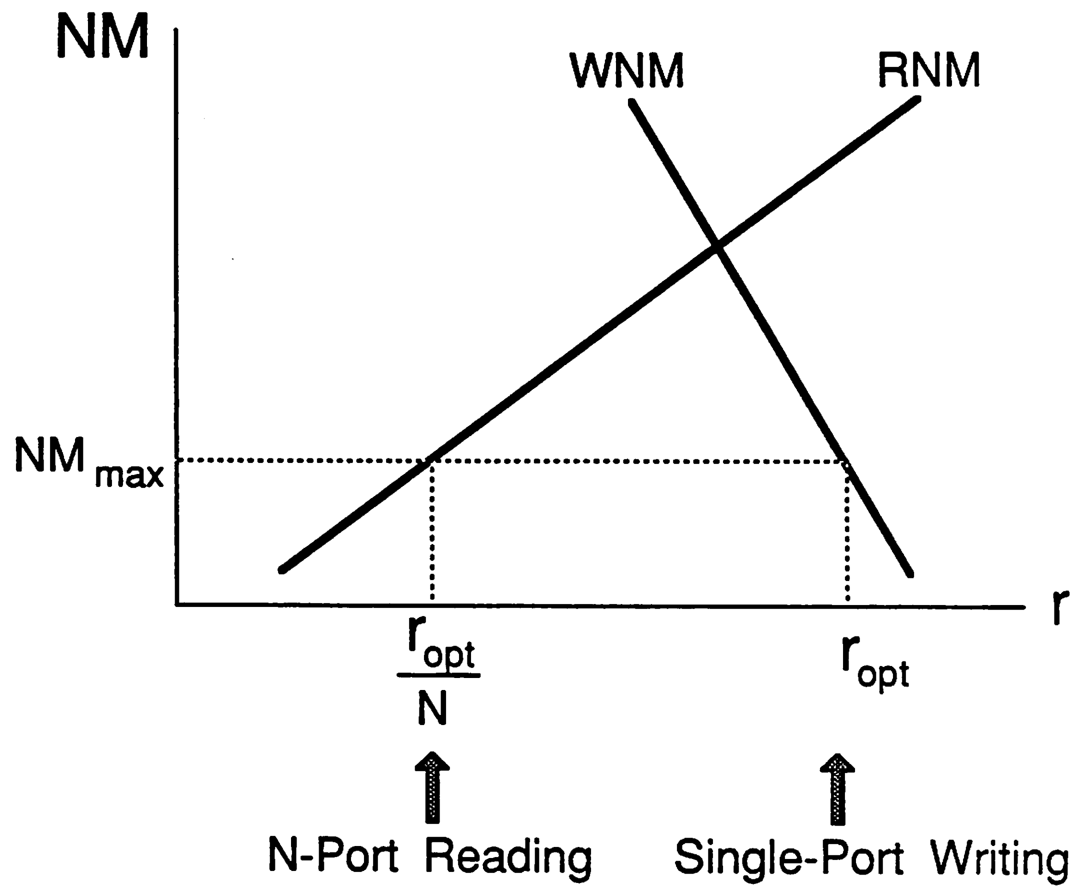


Figure 3.14 Definition of optimal beta ratio, r_{opt}, and maximal noise margin, NM_{max}, of N-port memory.

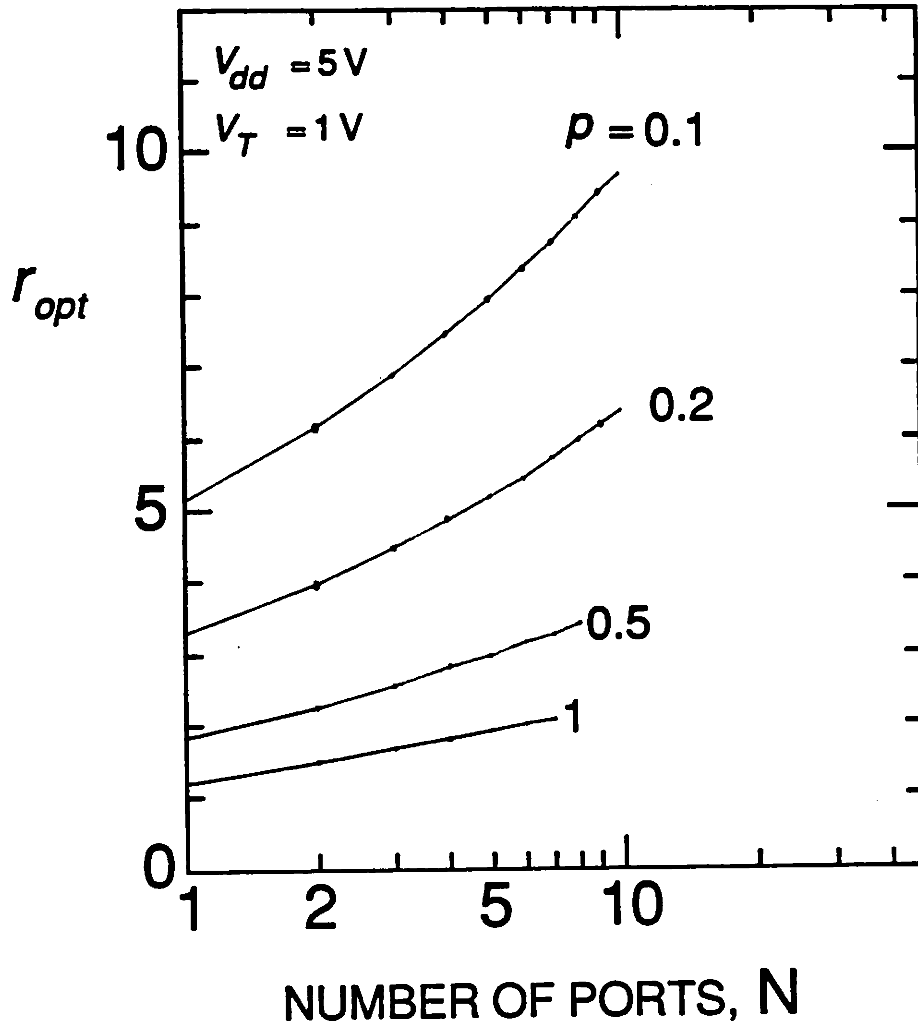


Figure 3.15 Optimal beta ratio, r_{opt}, for differentially accessed cell shown as a function of number of ports, N, for different beta ratios, p .

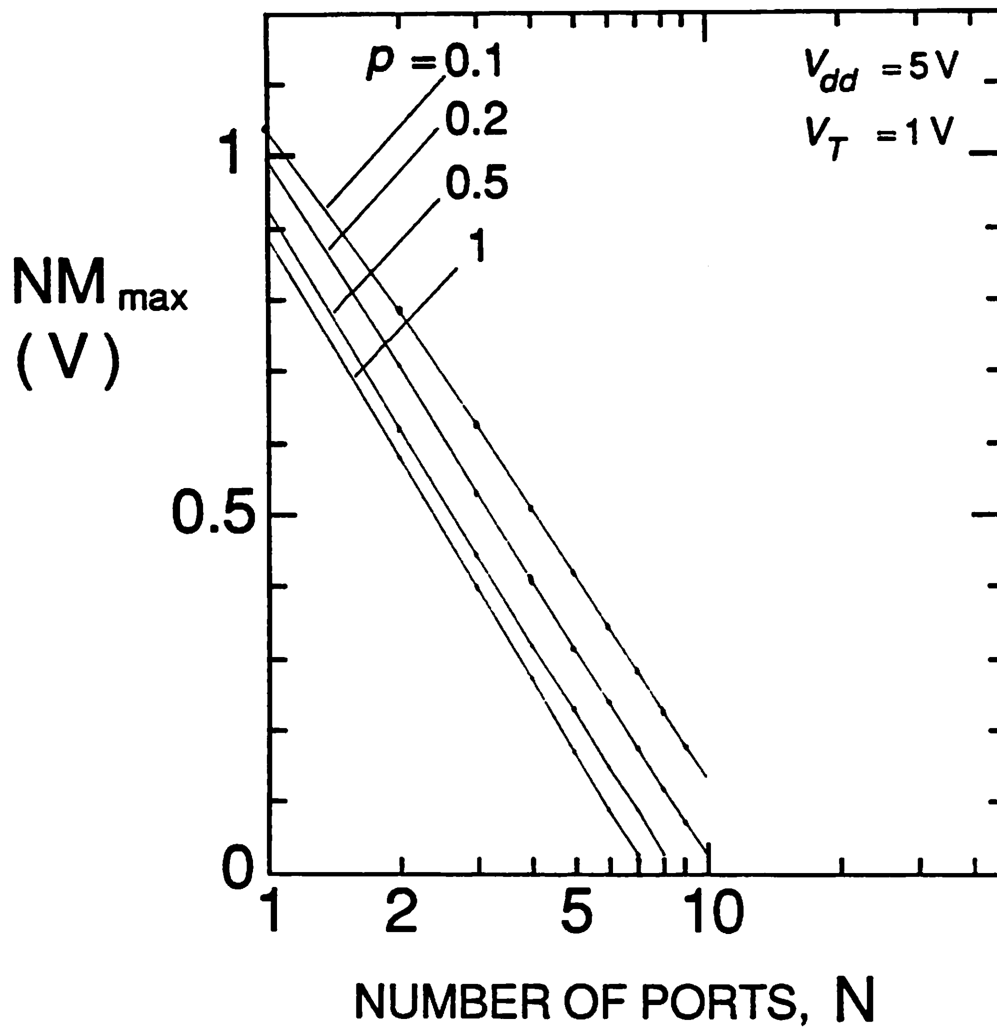


Figure 3.16 Maximal noise margin of differentially accessed cell shown as a function of number of ports, N, for different beta ratios, p .

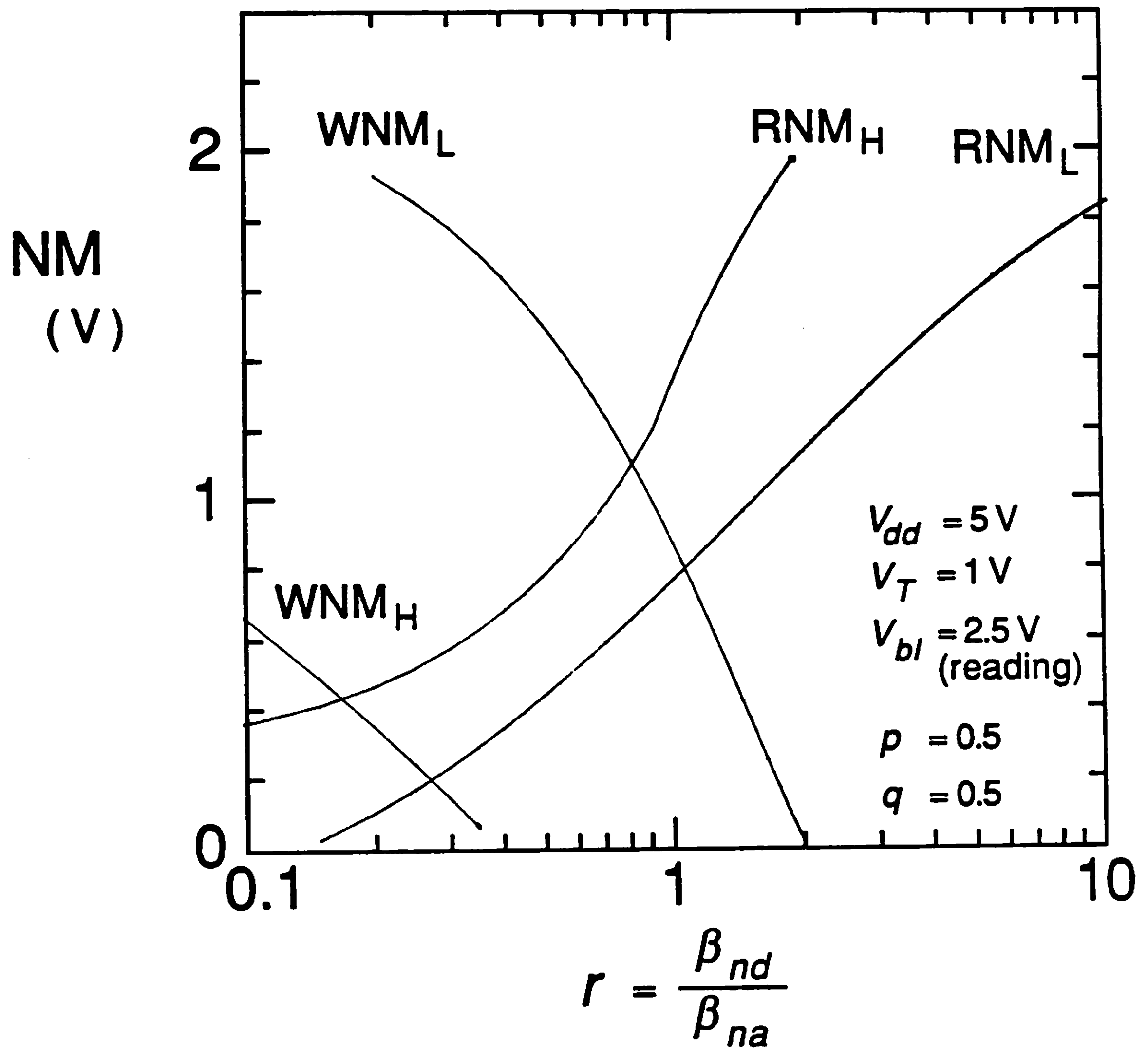


Figure 3.17 Calculated noise margins of single-ended access cell shown as a function of beta ratio, r .

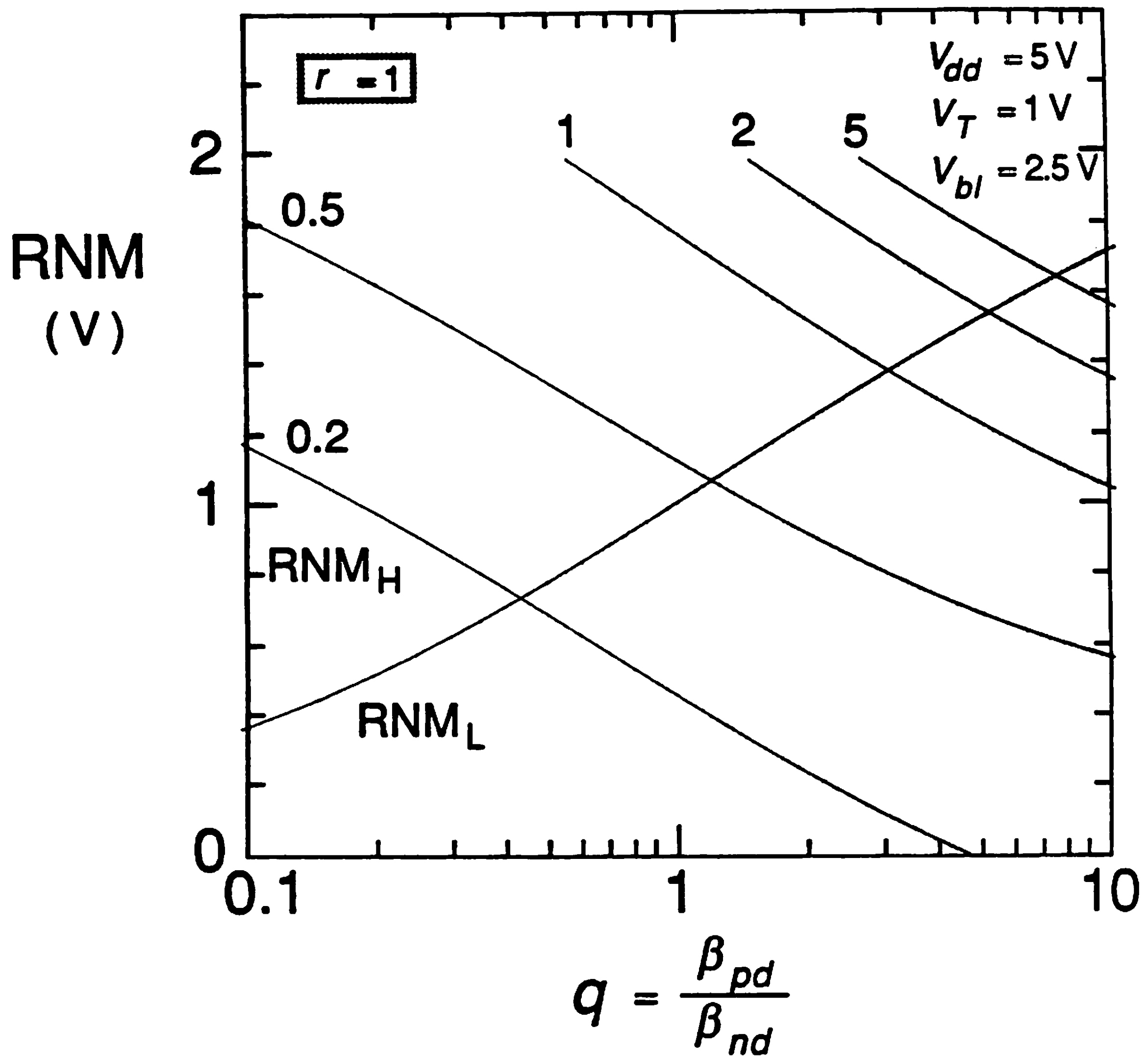


Figure 3.18 Calculated read noise margins of single-ended access cell shown as a function of beta ratio, q .

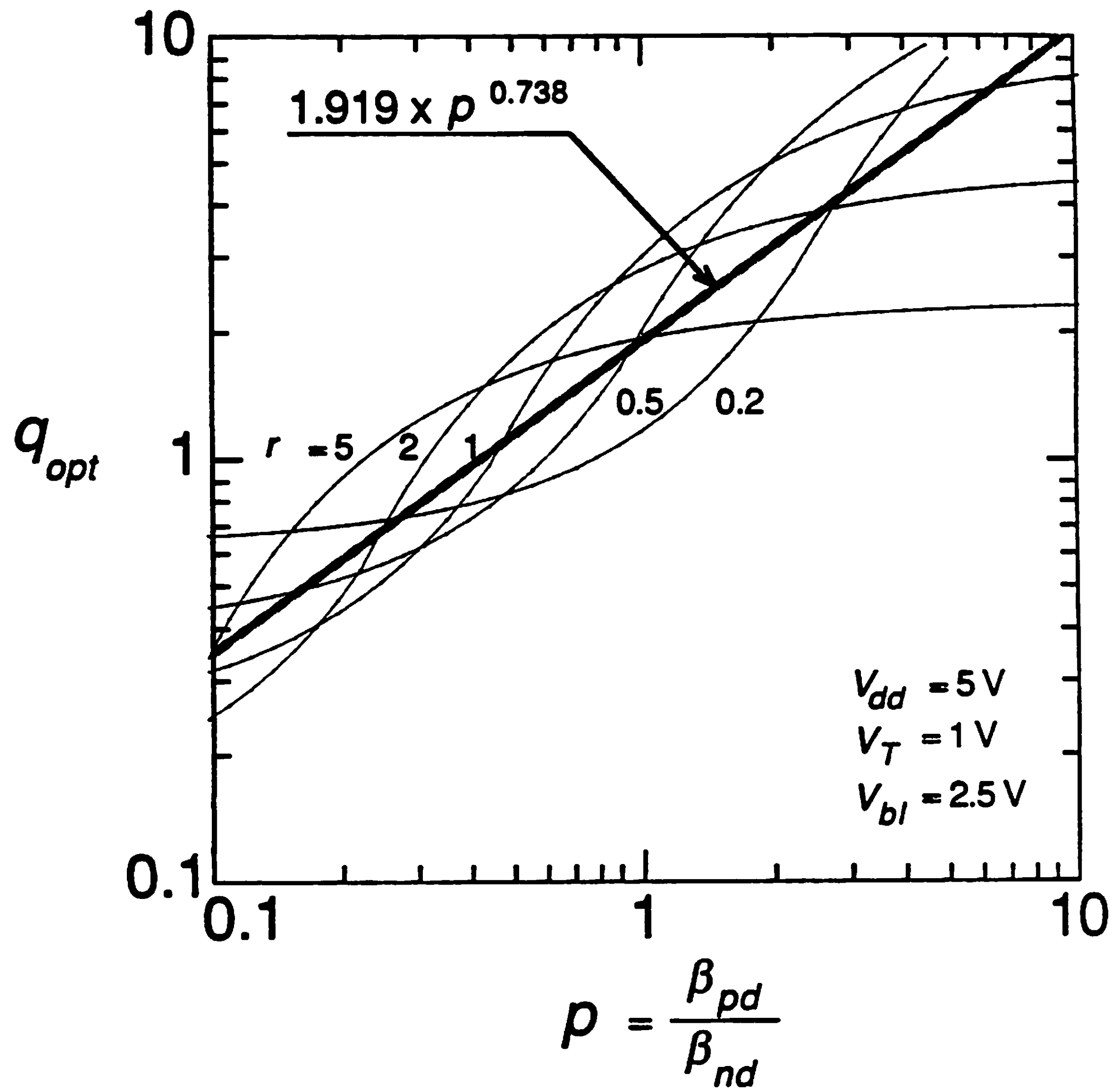


Figure 3.19 Calculated optimal beta ratio, q_{opt}, versus beta ratio, p, for different beta ratios, r.

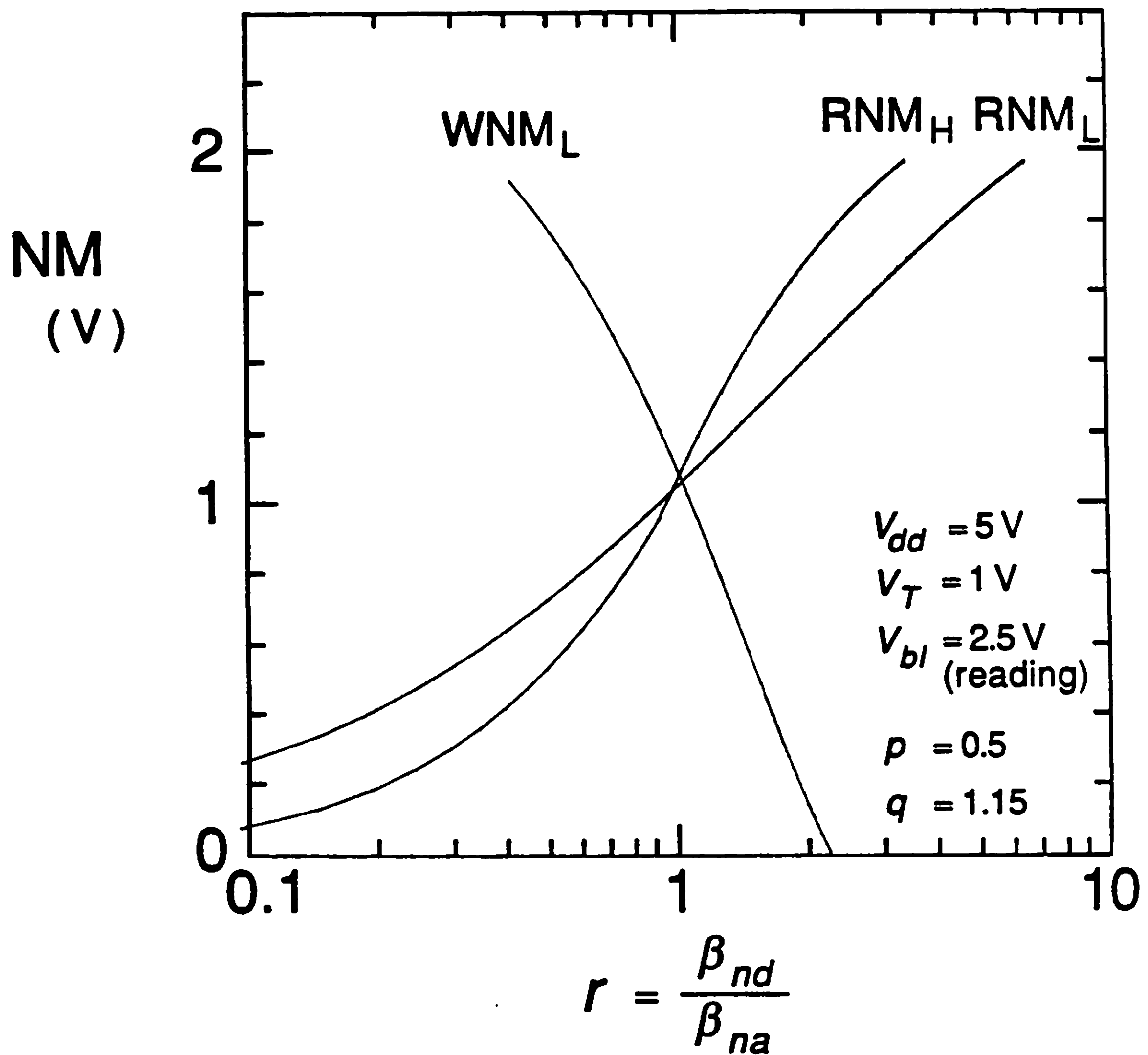


Figure 3.20 Calculated noise margins of pseudo-static single-ended access cell shown as a function of beta ratio, r, for optimal beta ratio, q_{opt} .

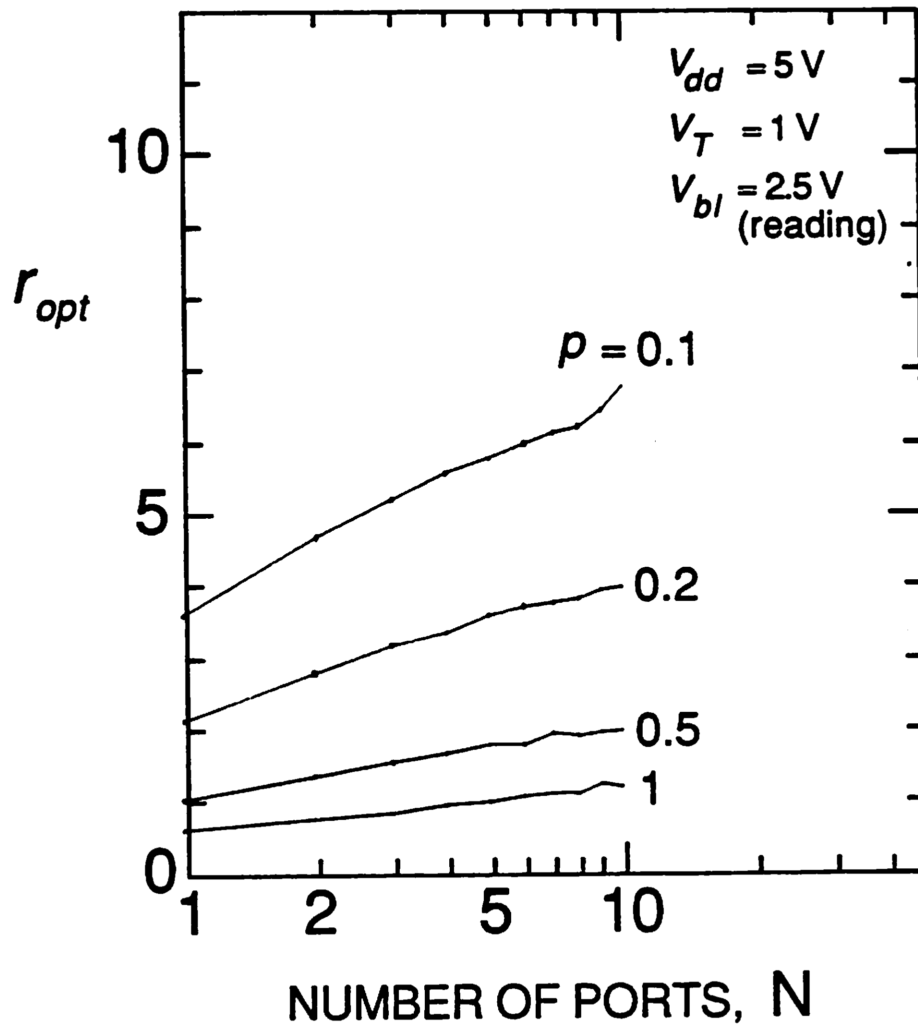


Figure 3.21 Optimal beta ratio, r_{opt}, for pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p.

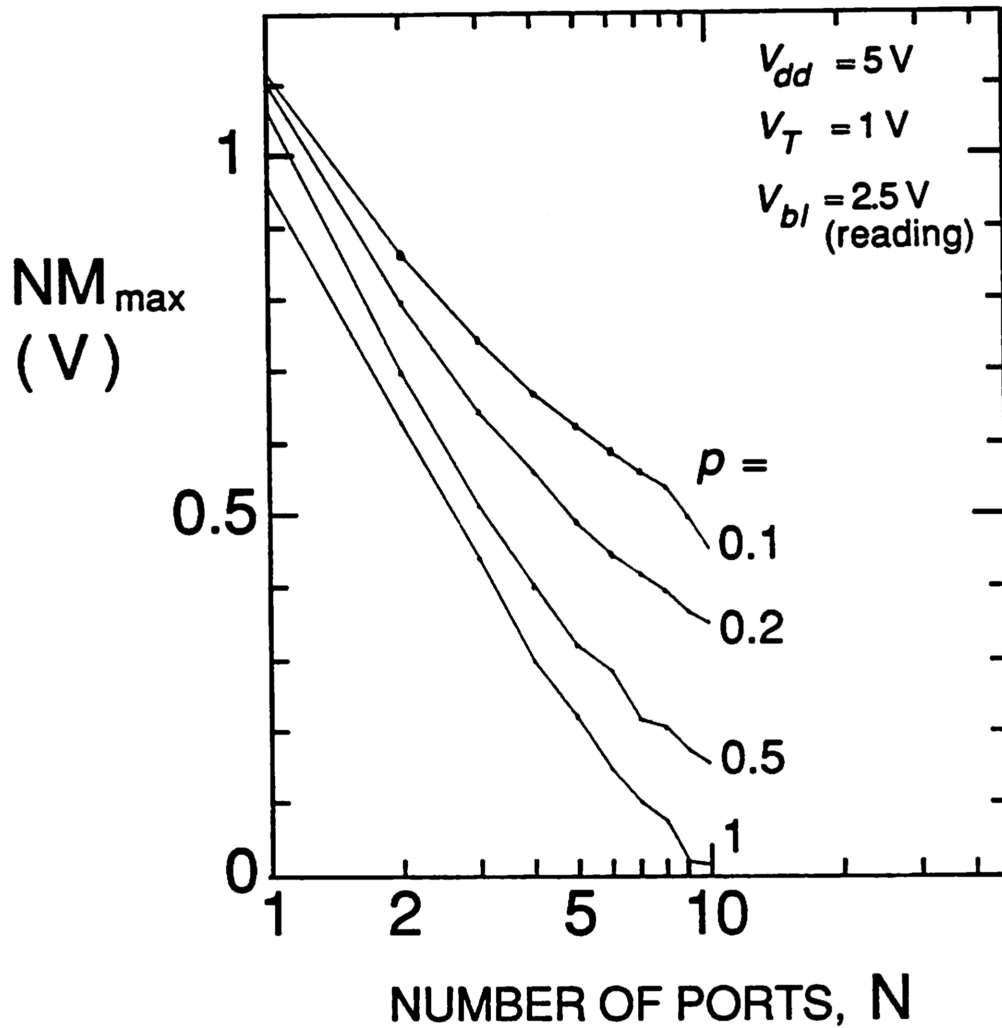


Figure 3.22 Maximal noise margin of pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p .

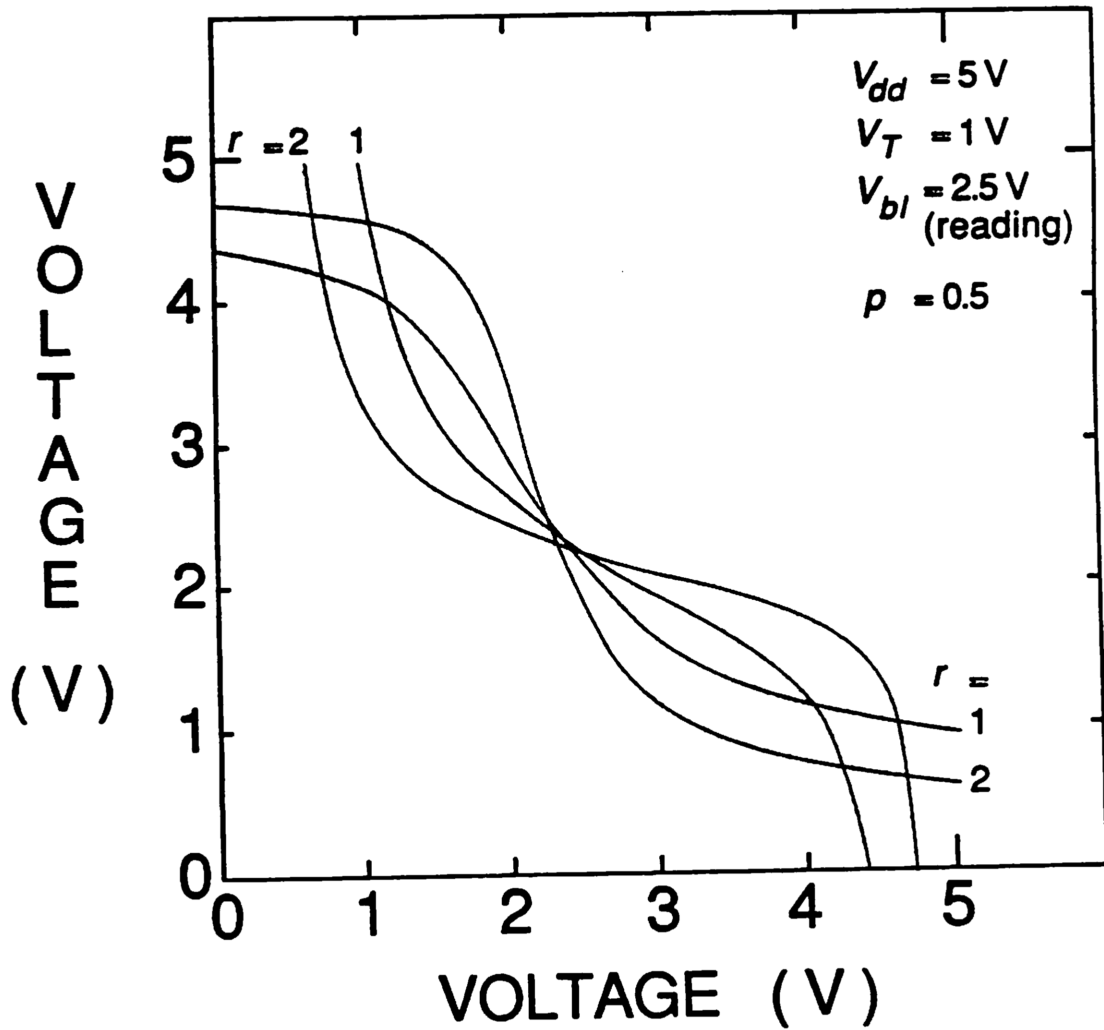


Figure 3.23 Transfer characteristics of modified single-ended access cell in read operation.

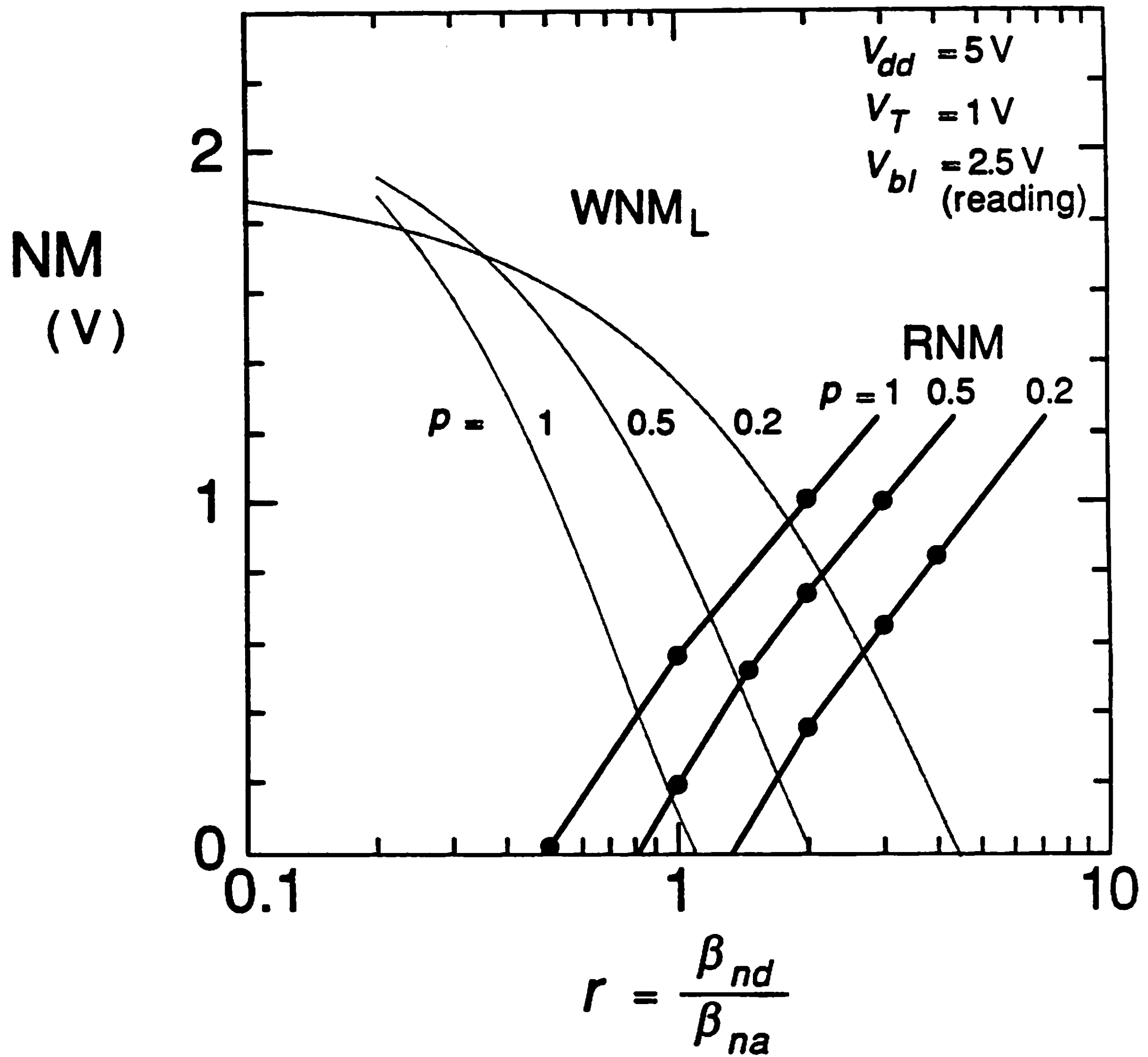


Figure 3.24 Calculated noise margins of modified single-ended access cell shown as a function of beta ratio, r, for different beta ratios, q_{opt}.

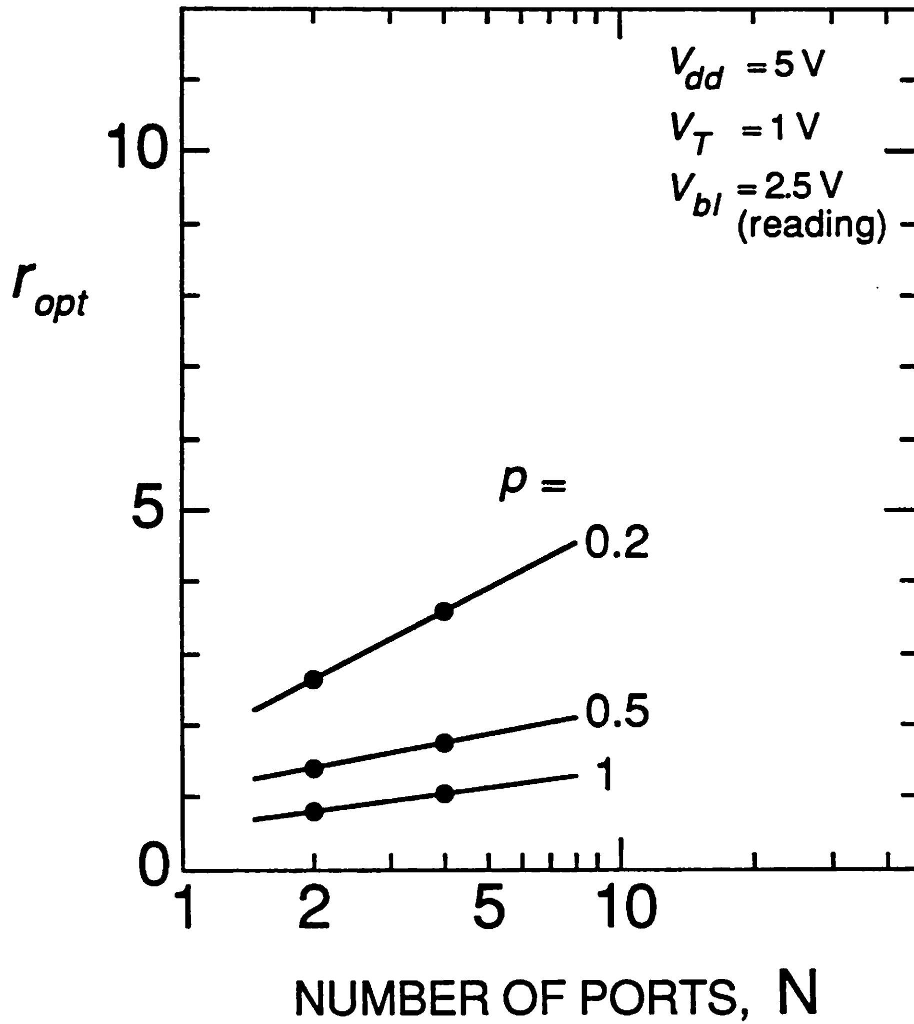


Figure 3.25 Optimal beta ratio, r_{opt}, for modified single-ended access cell shown as a function of number of ports, N, for different beta ratios, p.

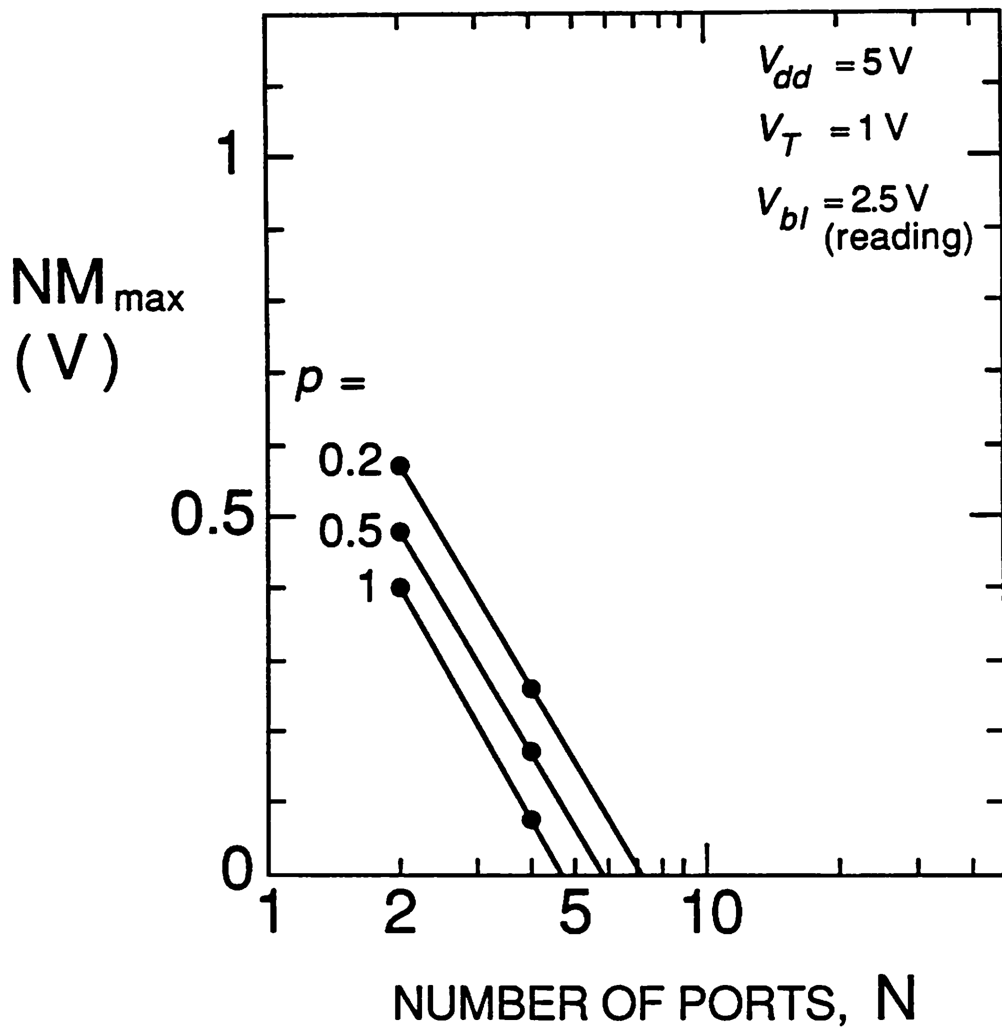


Figure 3.26 Maximal noise margin of modified single-ended access cell shown as a function of number of ports, N, for different beta ratios, p .

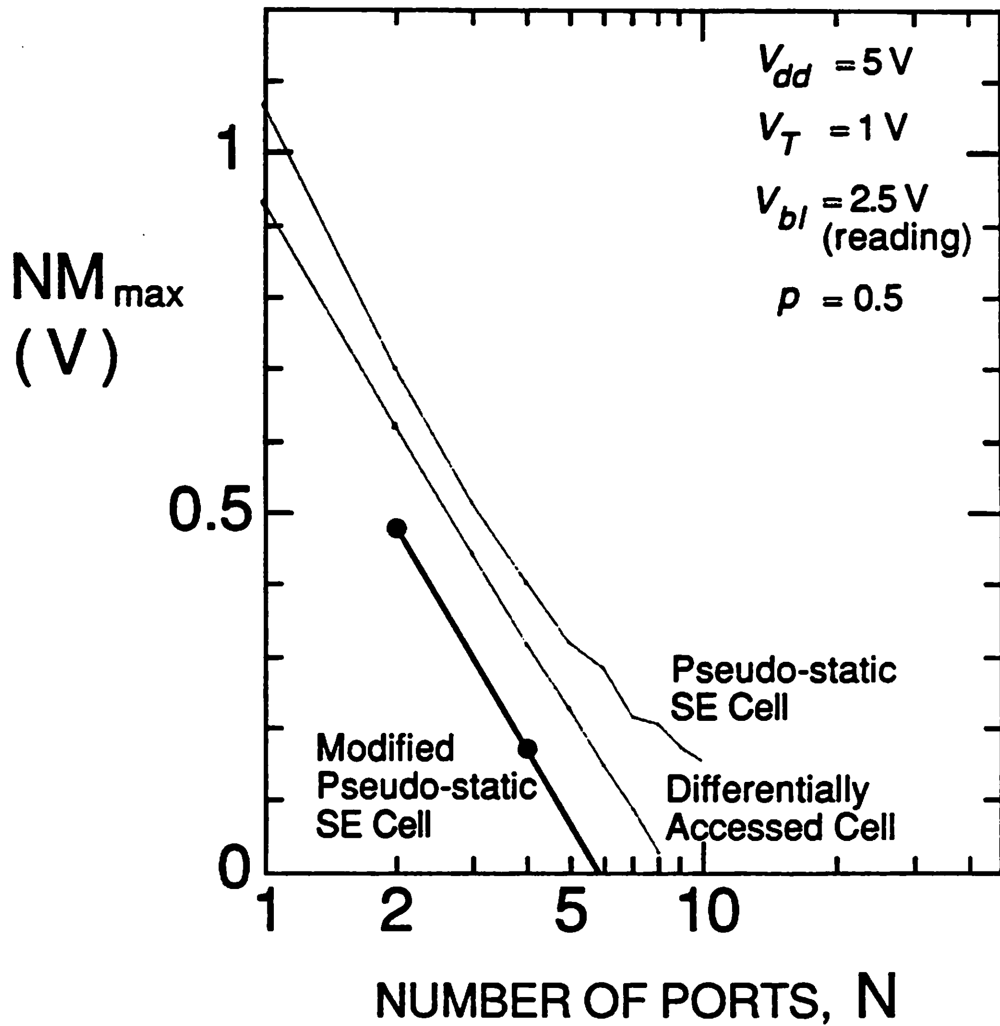
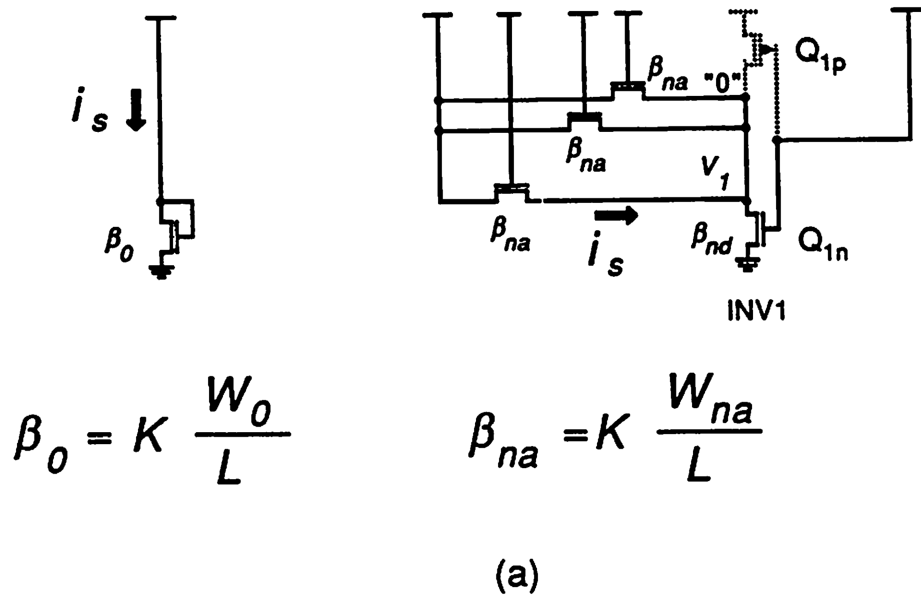
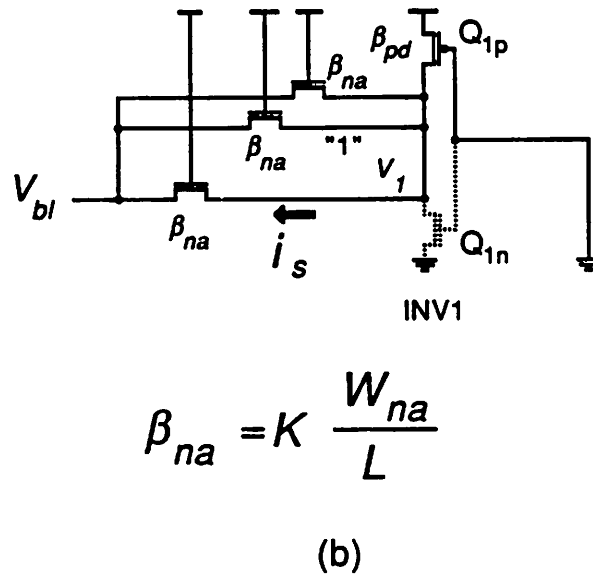


Figure 3.27 Maximal noise margin versus number of ports, N, for different cell structures.



$$\beta_0 = K \frac{W_0}{L} \qquad \beta_{na} = K \frac{W_{na}}{L}$$

(a)



$$\beta_{na} = K \frac{W_{na}}{L}$$

(b)

Figure 3.28 Definition of scale up factor, $F_s = \beta_{na} / \beta_0$:(a) differentially accessed cell and (b) single-ended access cell ("1" read).

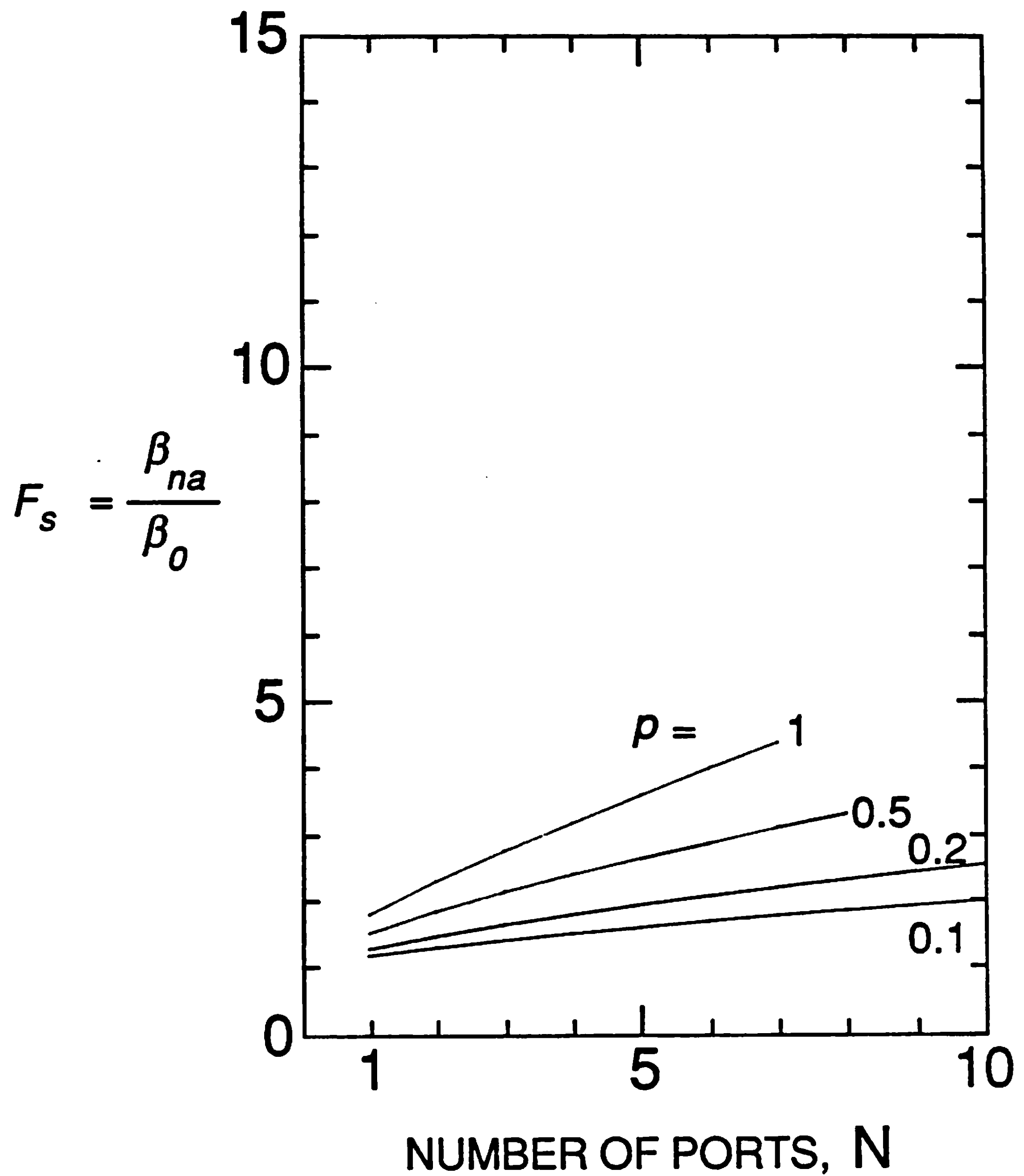


Figure 3.29 Scale up factor, F_s, of differentially accessed cell shown as a function of number of ports, N, for different beta ratios, p.

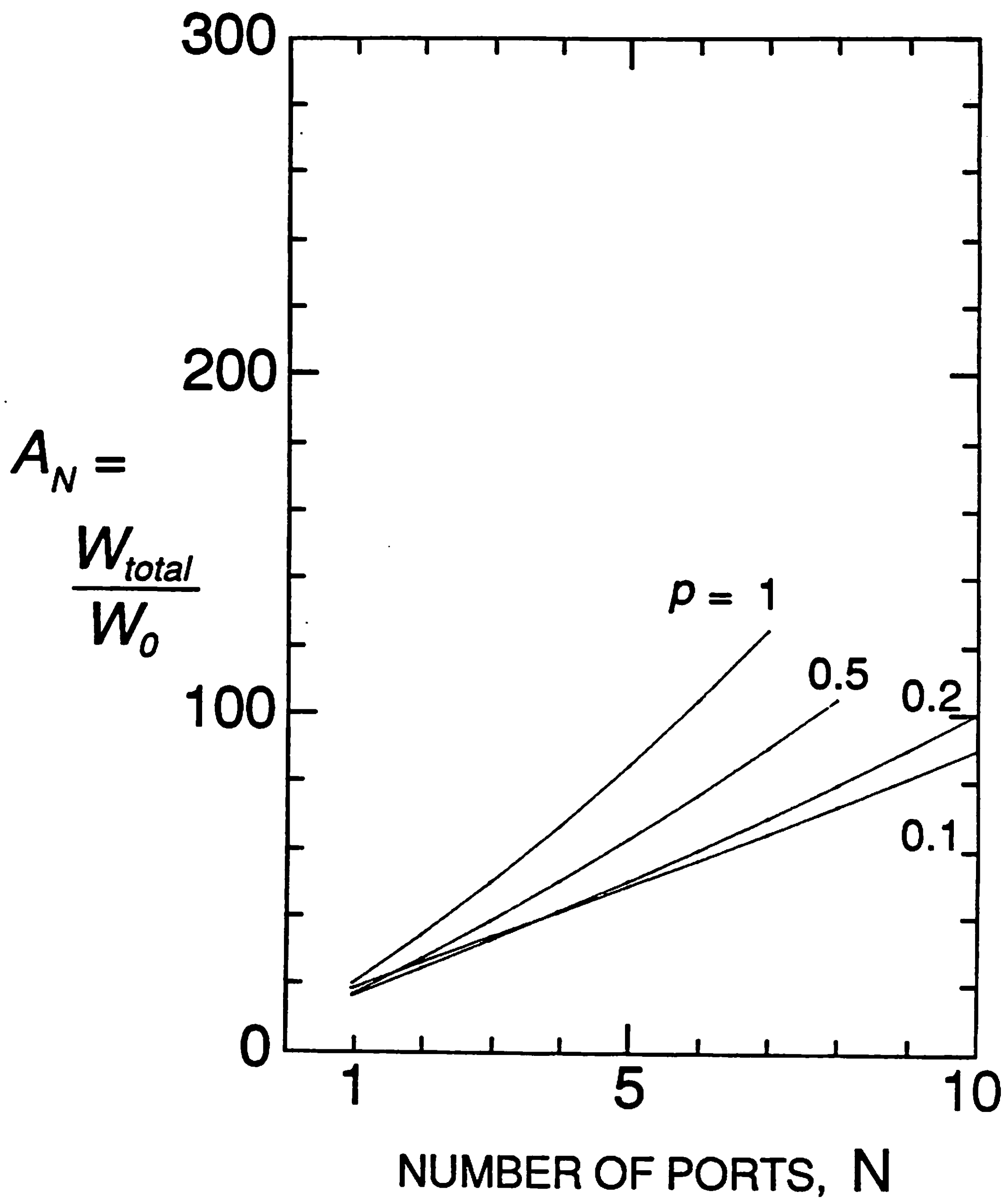


Figure 3.30 Normalized active area, A_N, of differentially accessed cell shown as a function of number of ports, N, for different beta ratios, p.

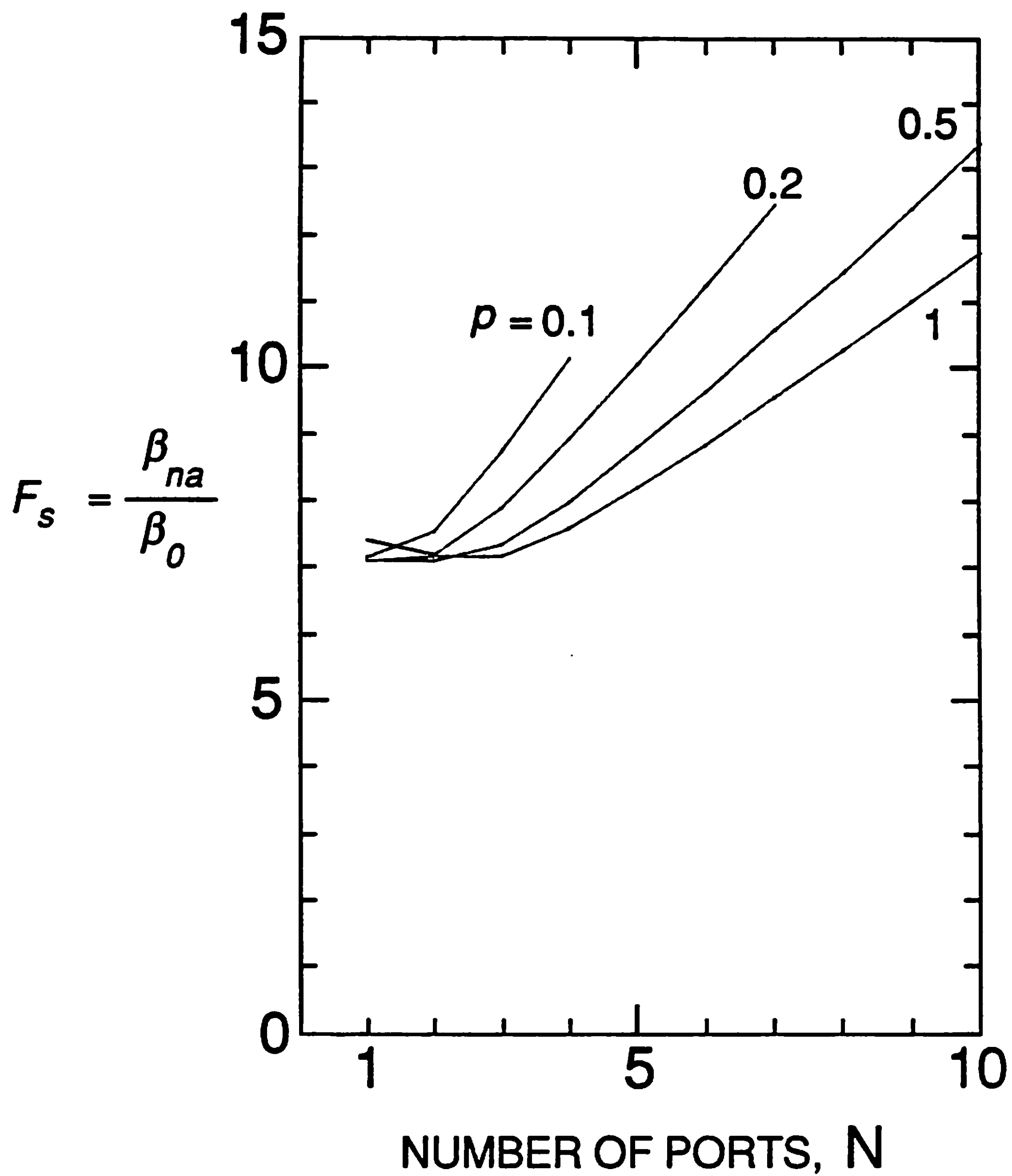


Figure 3.31 Scale up factor, F_s , of one-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p .

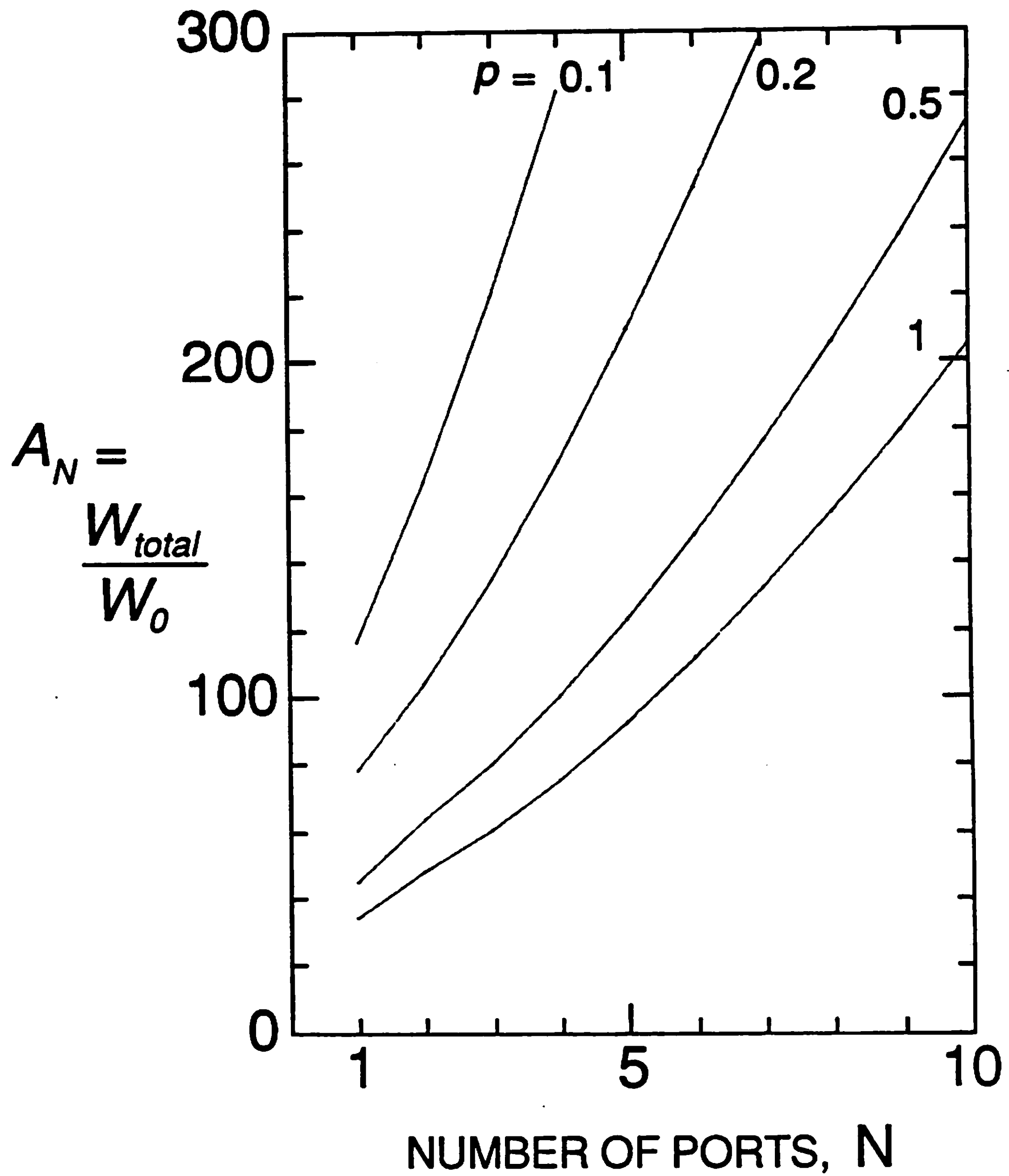


Figure 3.32 Normalized active area, A_N , of one-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p .

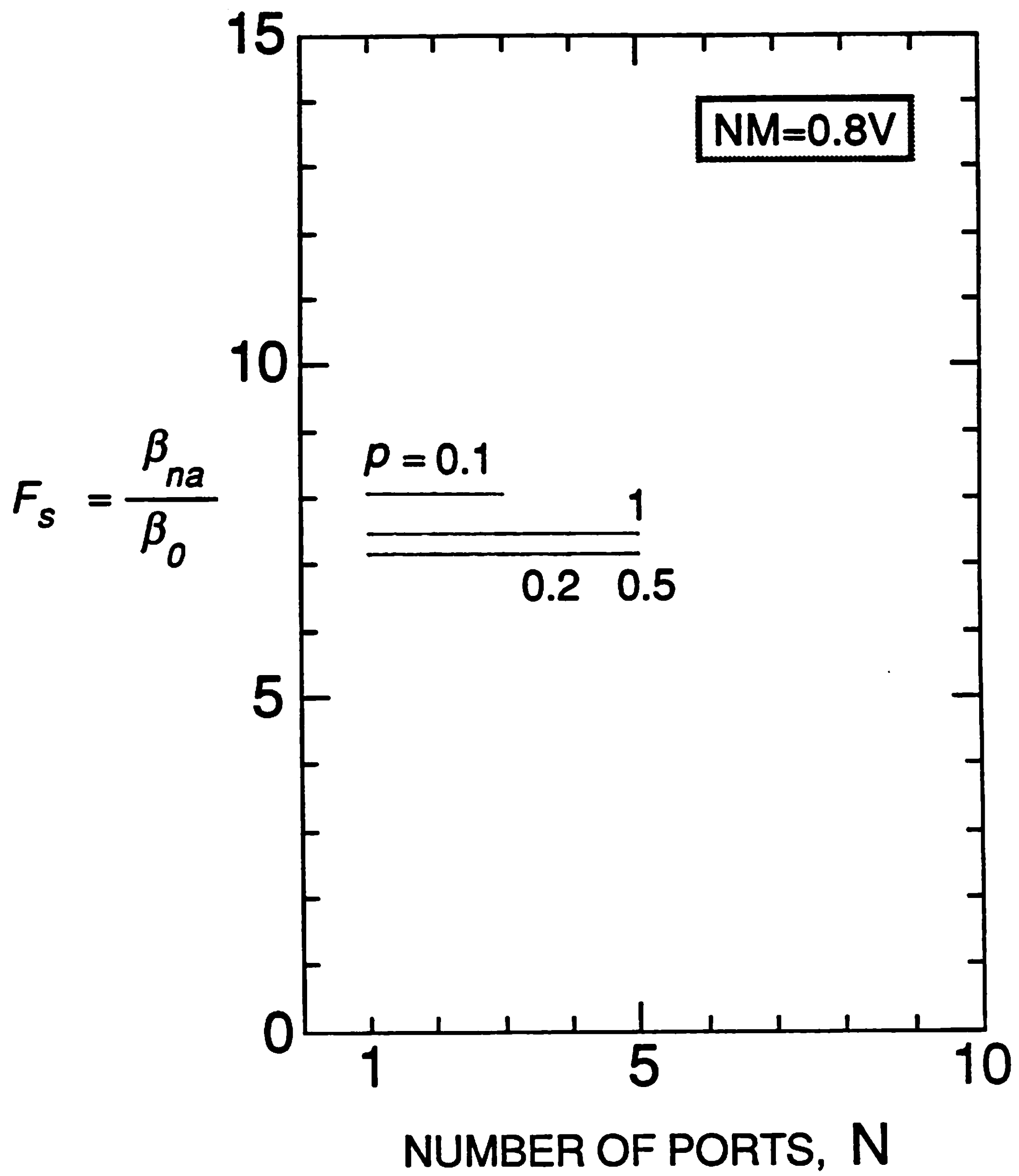


Figure 3.33 Scale up factor, F_s , of two-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p .

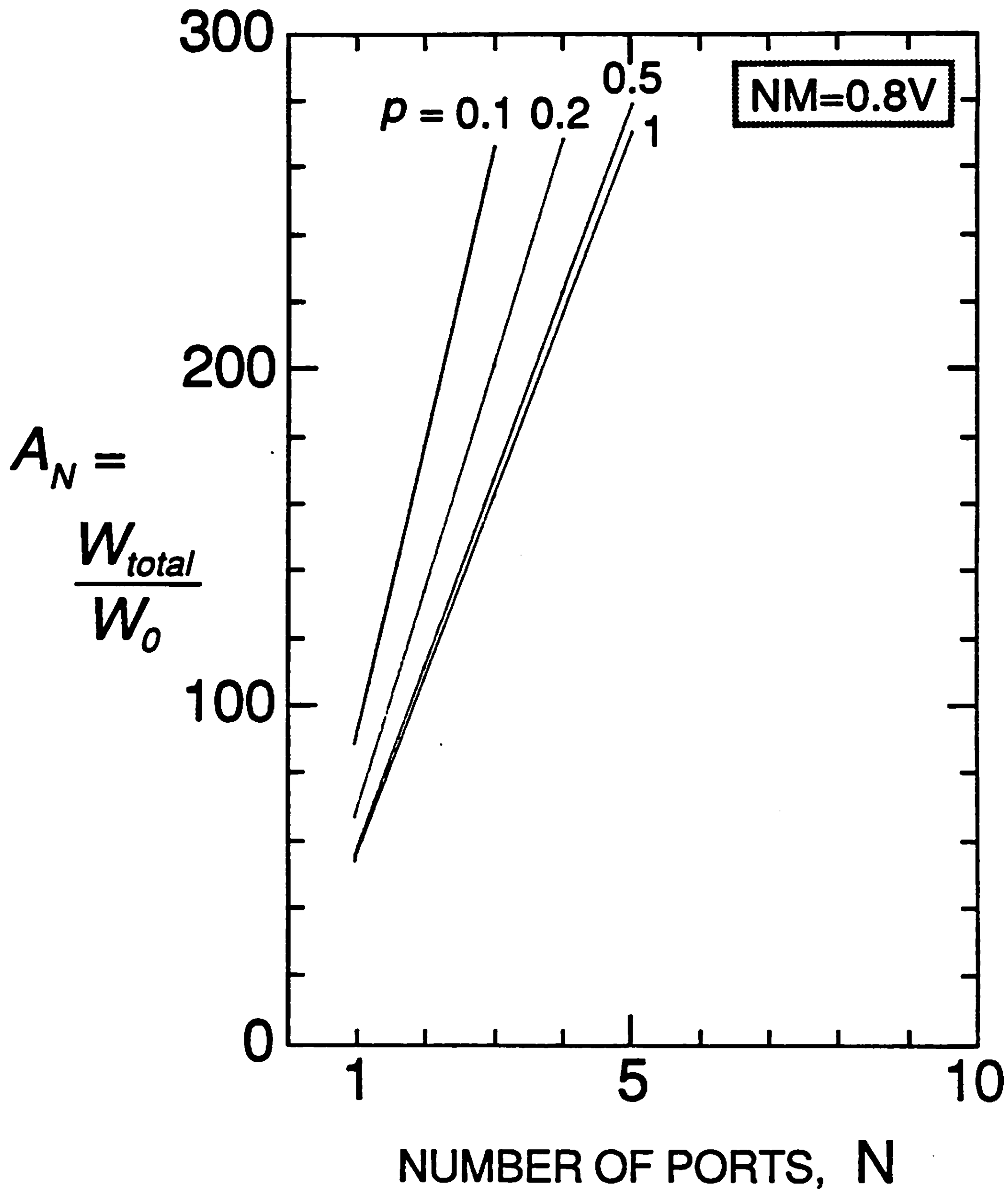
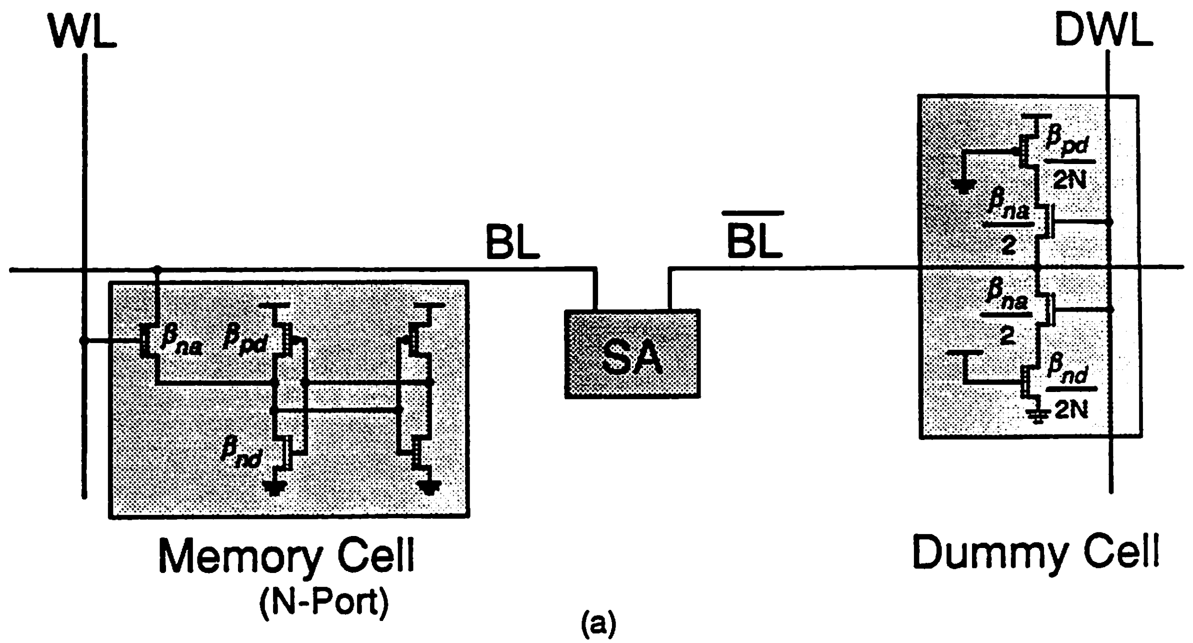


Figure 3.34 Normalized active area, A_N, of two-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p.



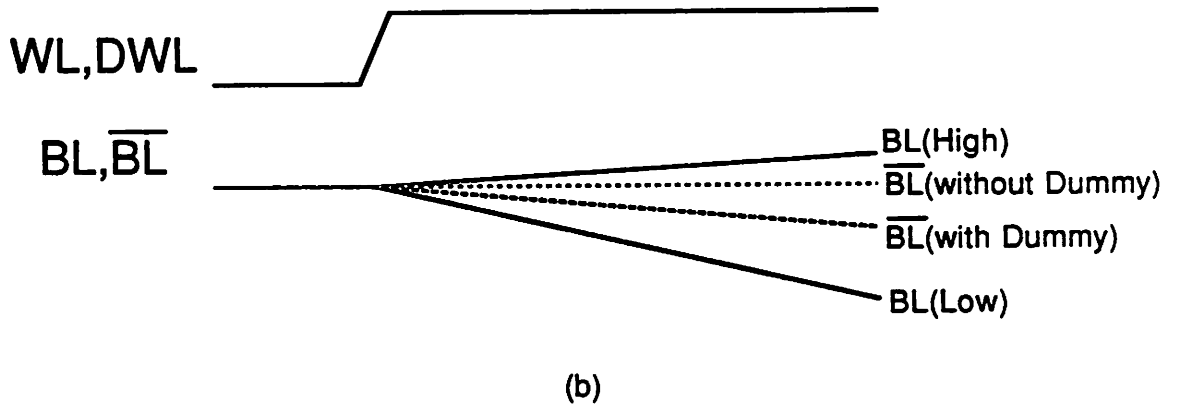


Figure 3.35 (a) Dummy cell circuit and (b) schematic of operation for signal/noise ratio improvement.

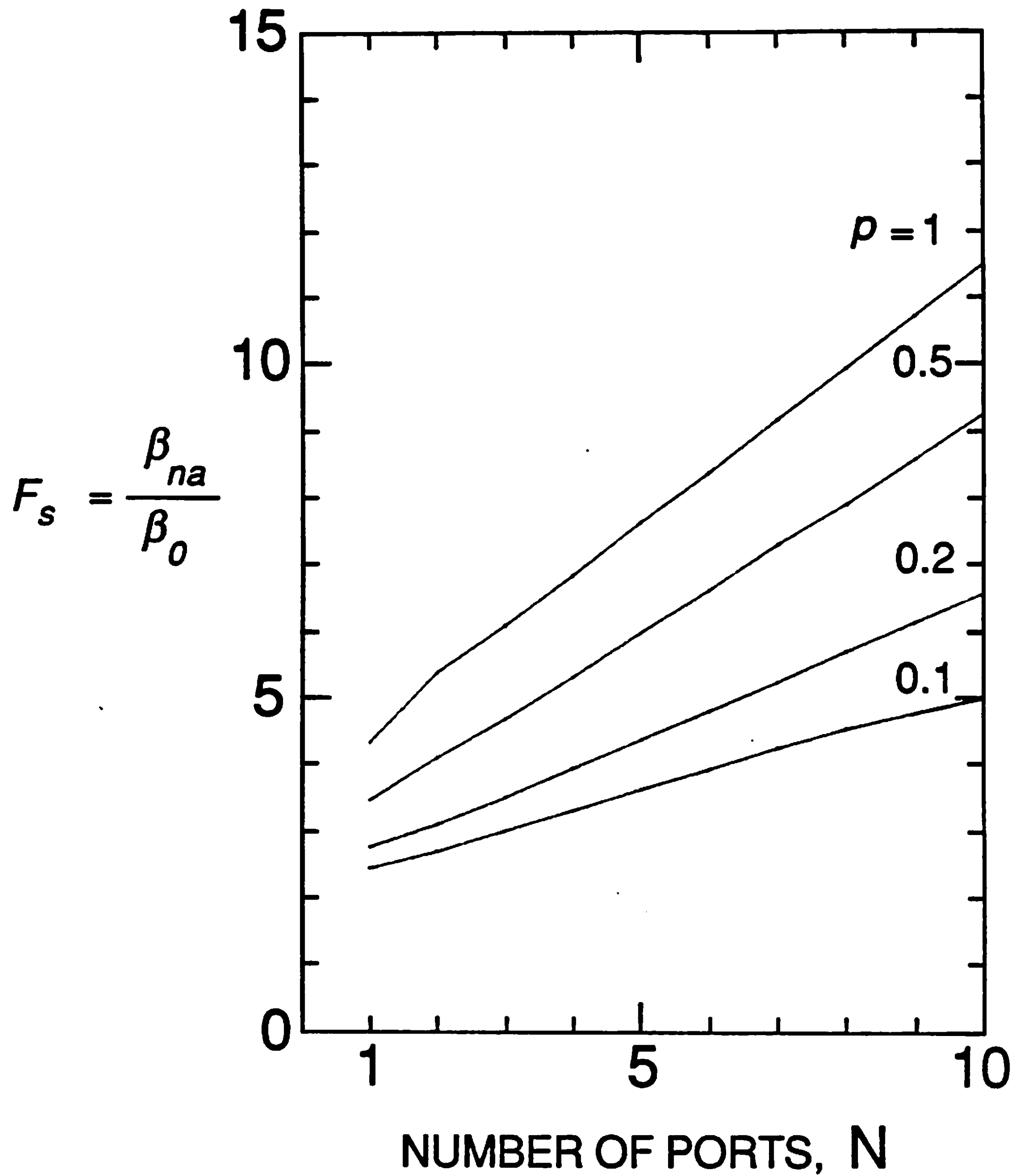


Figure 3.36 Scale up factor, F_s , of dummy-cell based one-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p .

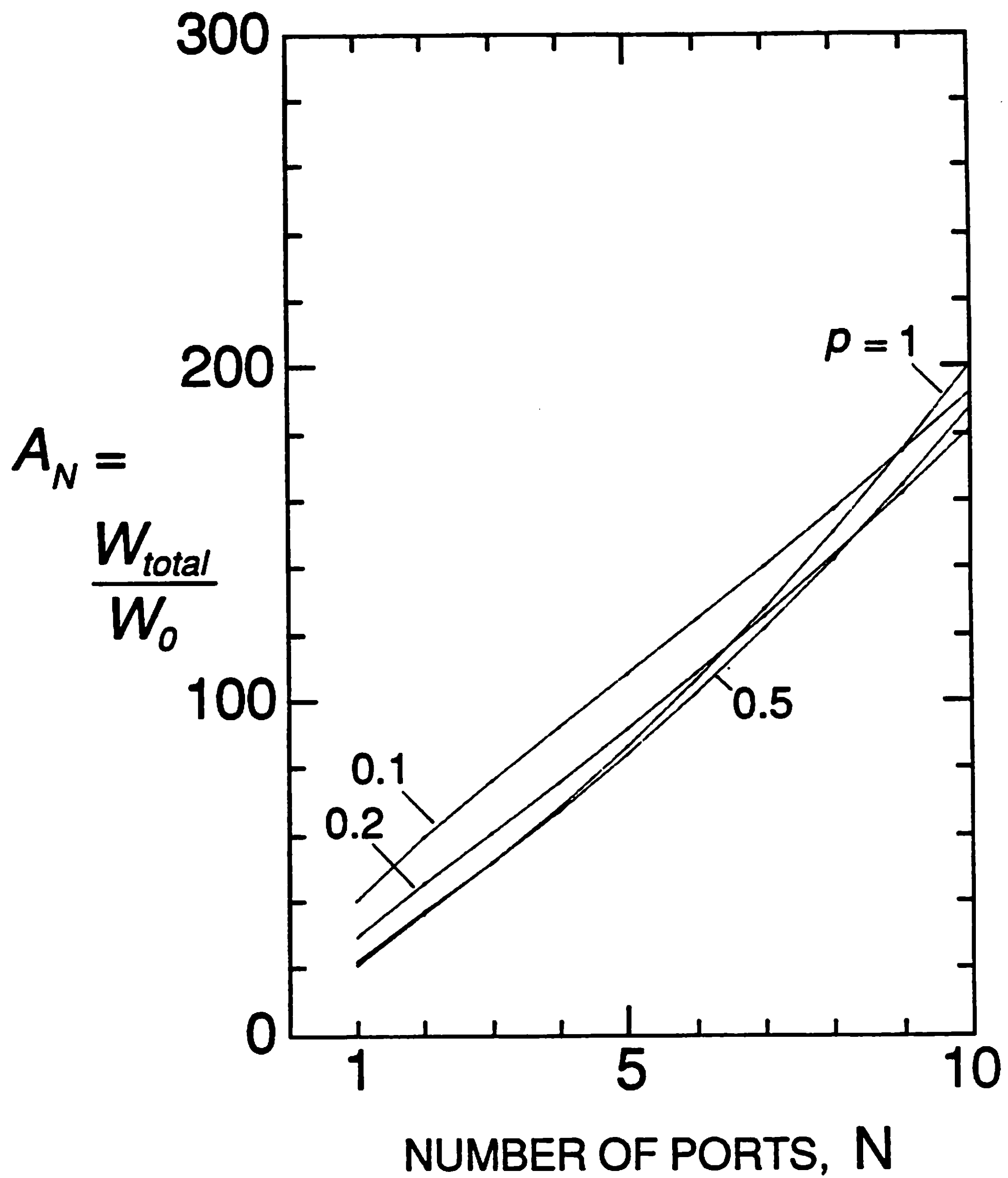


Figure 3.37 Normalized active area, A_N, of dummy-cell based one-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p.

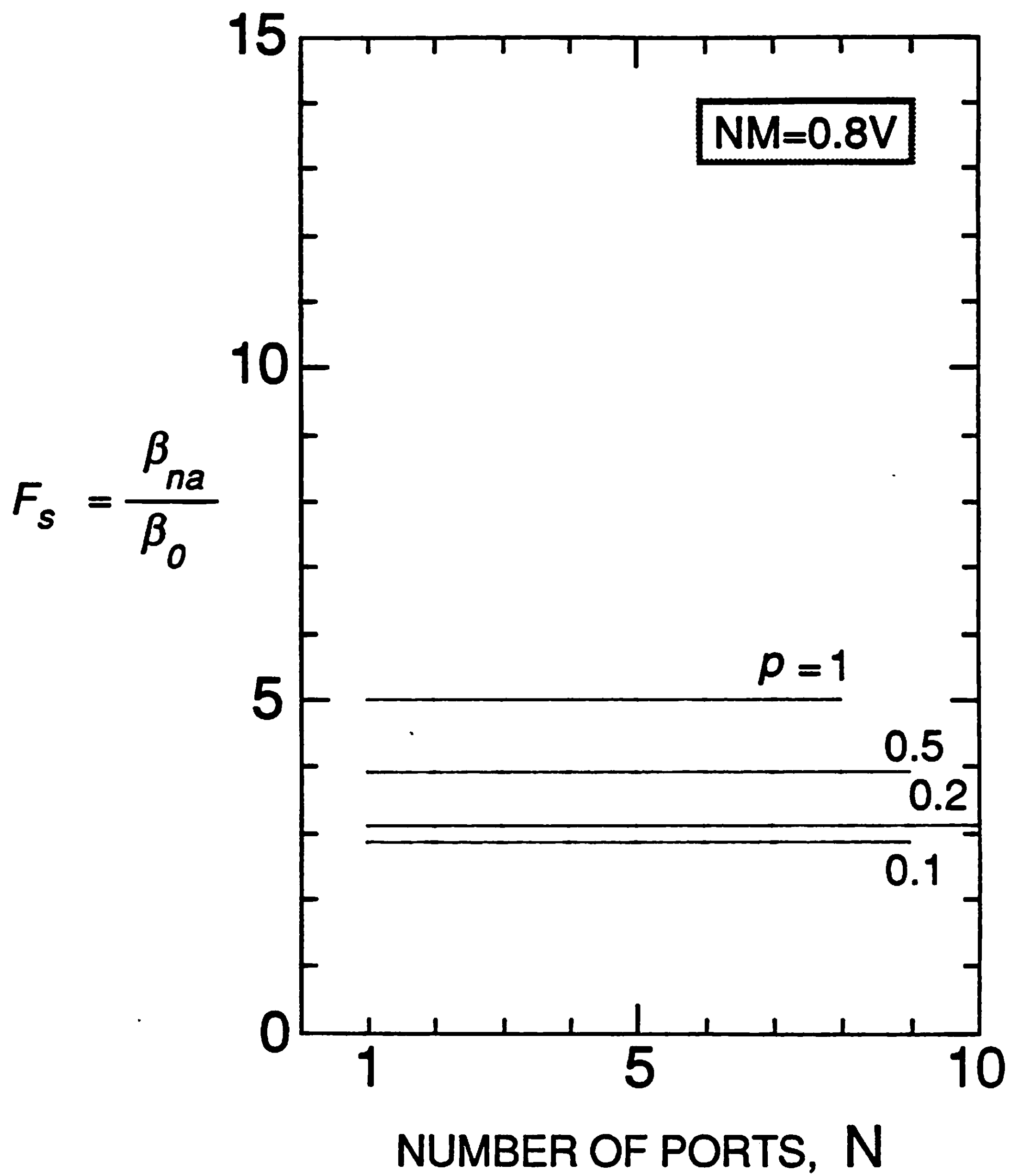


Figure 3.38 Scale up factor, F_s , of dummy-cell based two-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p .

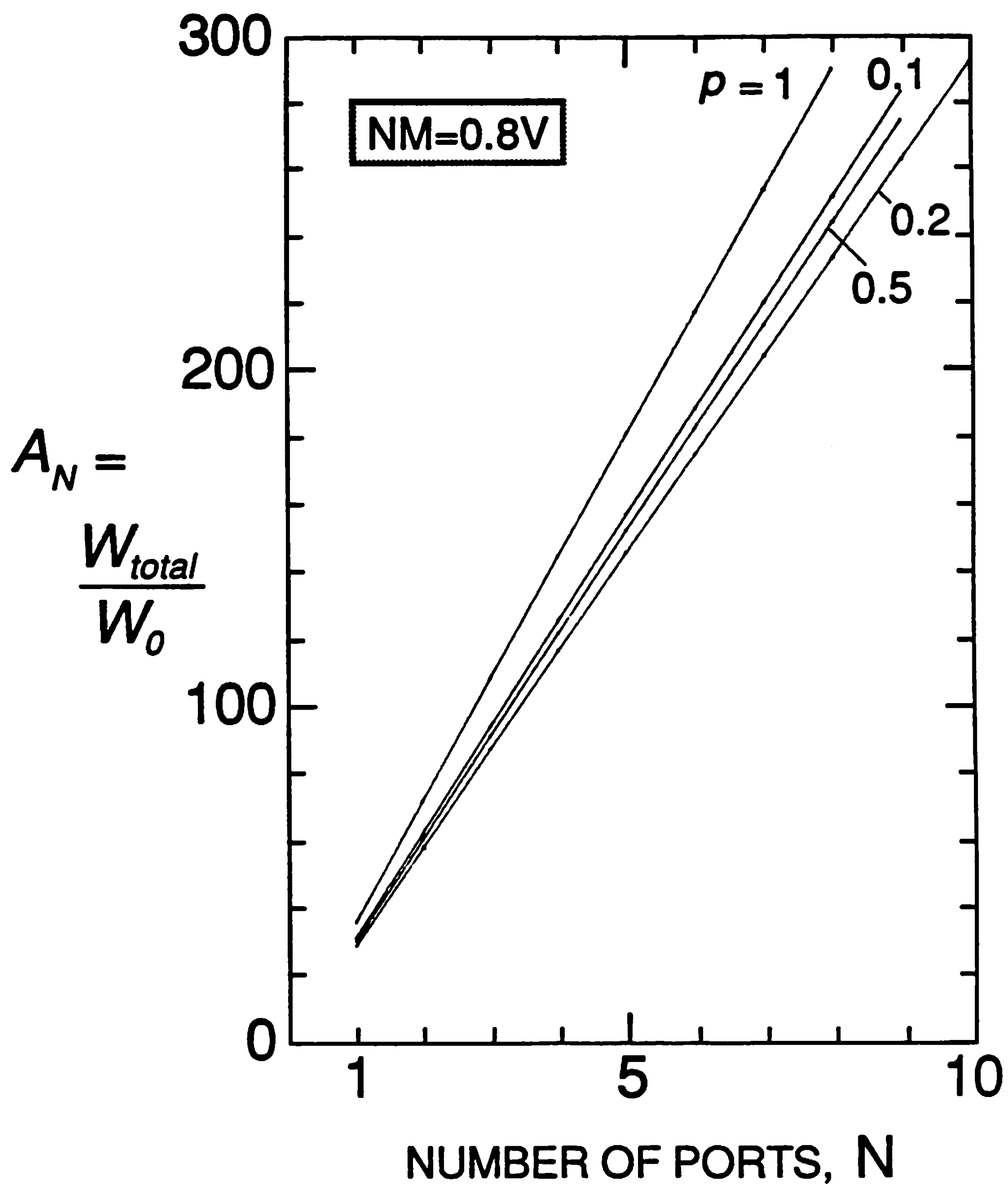


Figure 3.39 Normalized active area, A_N, of dummy-cell based two-side clocked pseudo-static single-ended access cell shown as a function of number of ports, N, for different beta ratios, p.

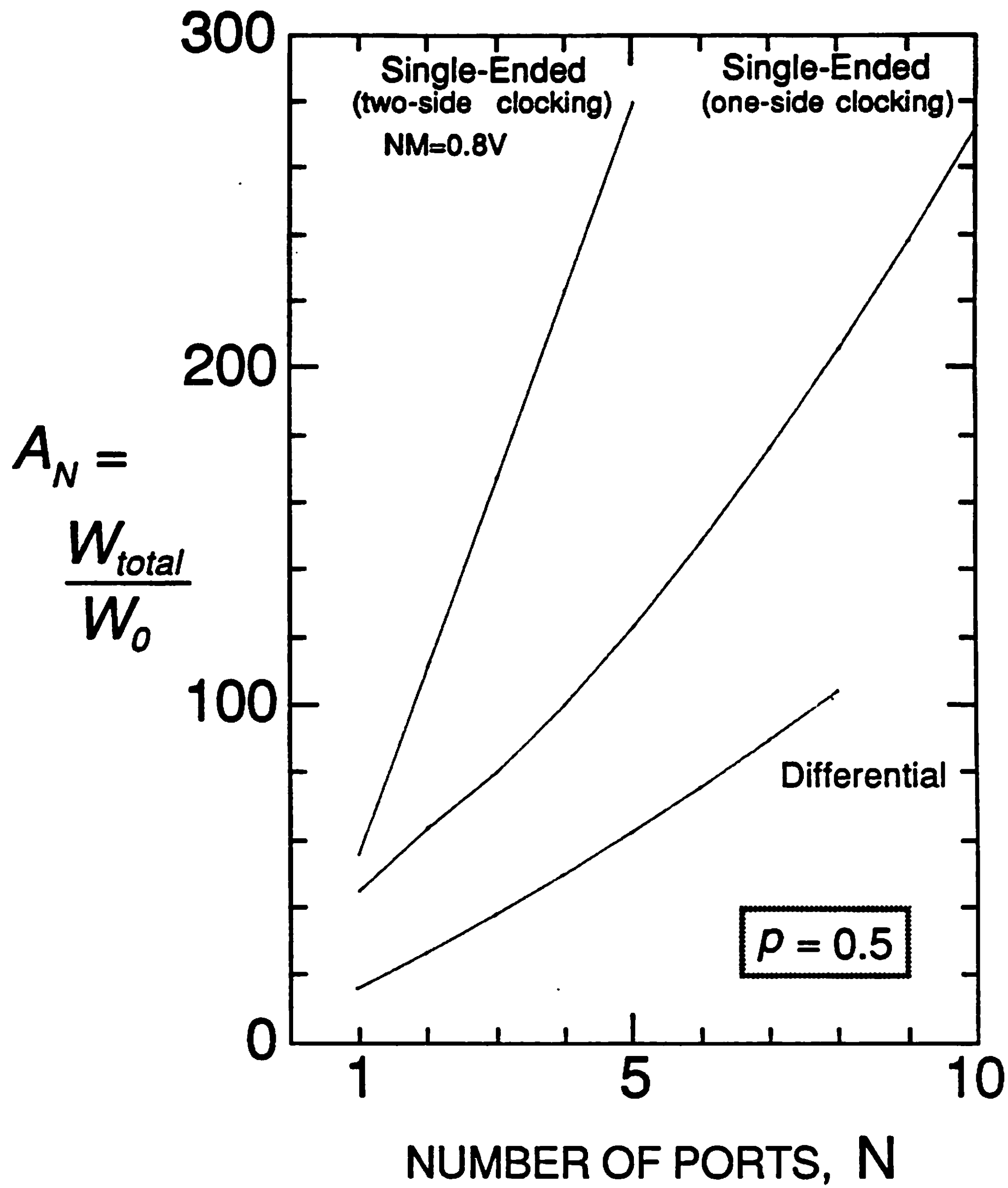
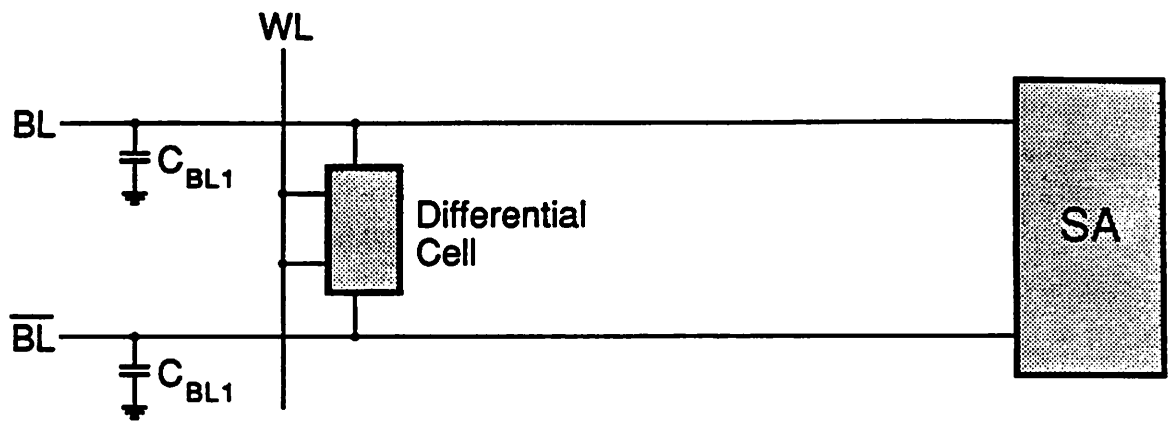
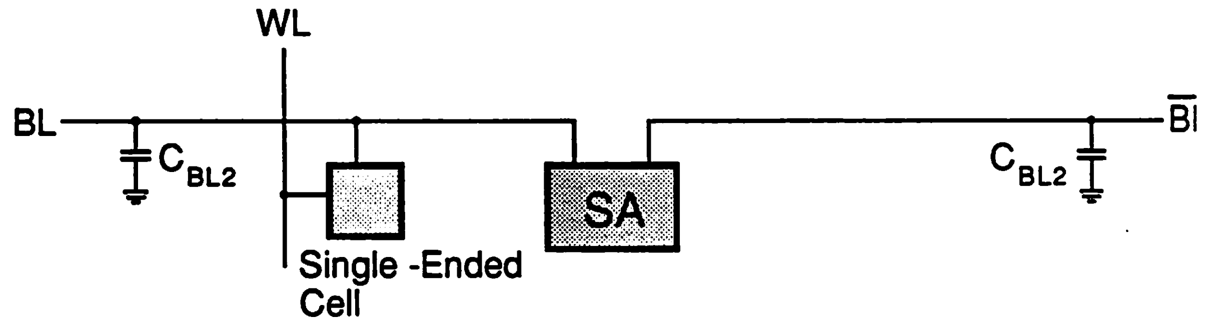


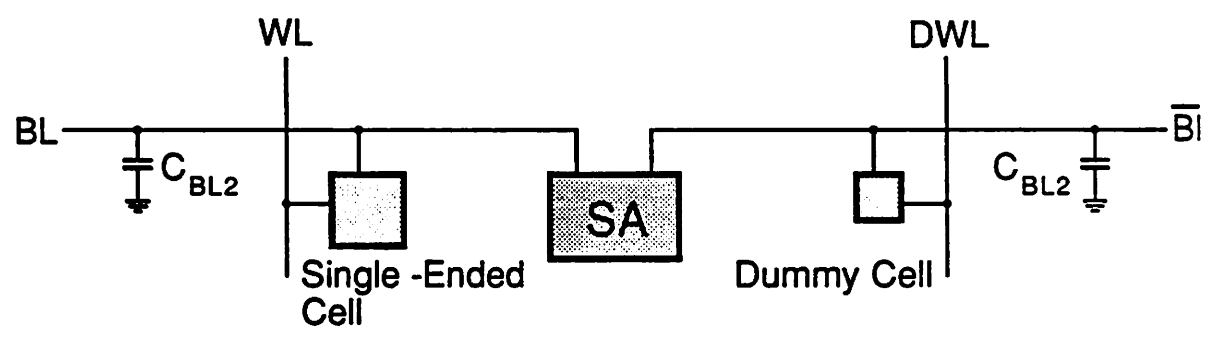
Figure 3.40 Normalized active area versus number of ports, N, for different cell structures.



(a)



(b)



(c)

Figure 3.41 Memory array configurations: (a) differential sensing, (b) split-bitline configuration using single-ended cell and (c) split-bitline configuration with dummy cell.

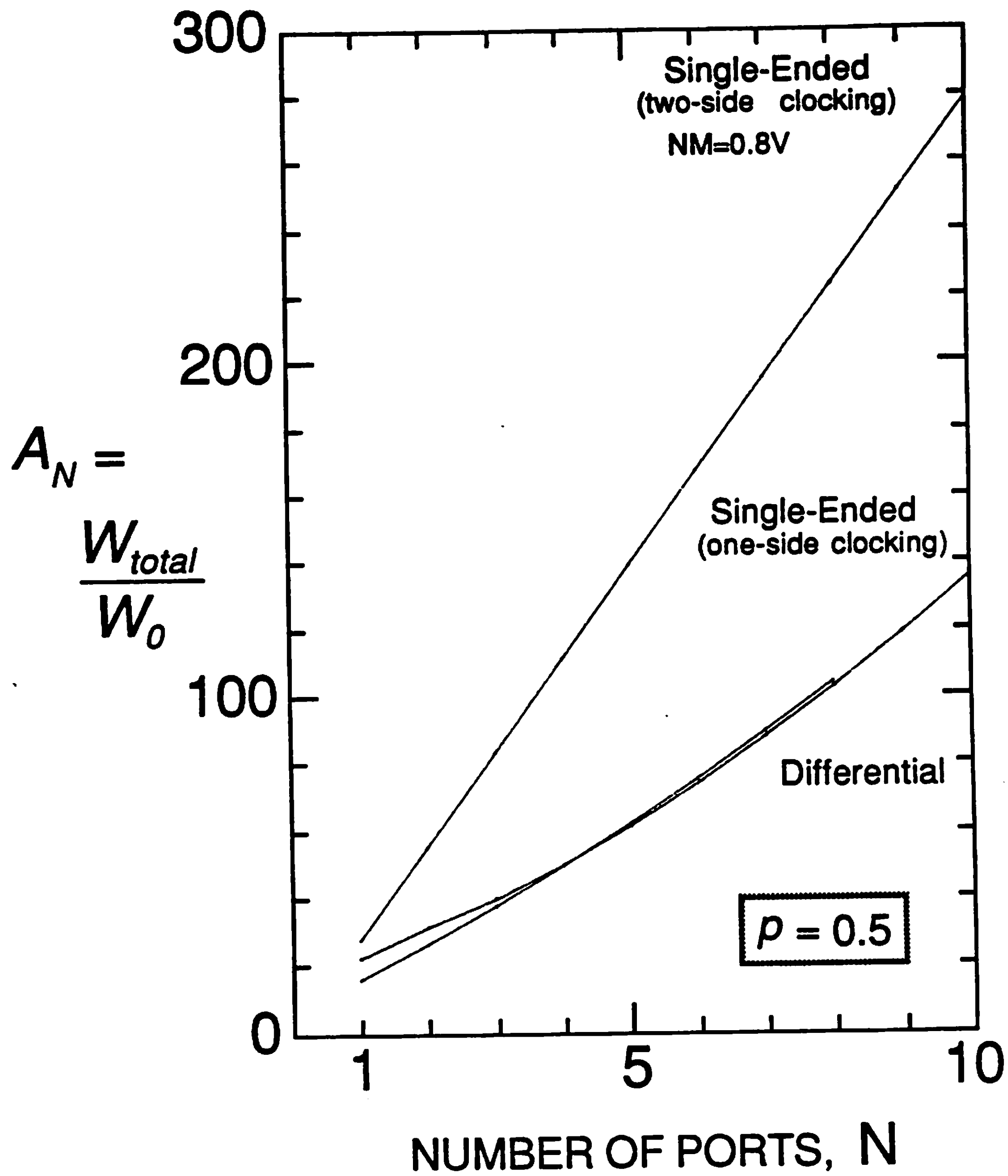


Figure 3.42 Normalized active area, A_N, versus number of ports, N, for different cell structures (reduced bitline capacitance for single-ended cell).

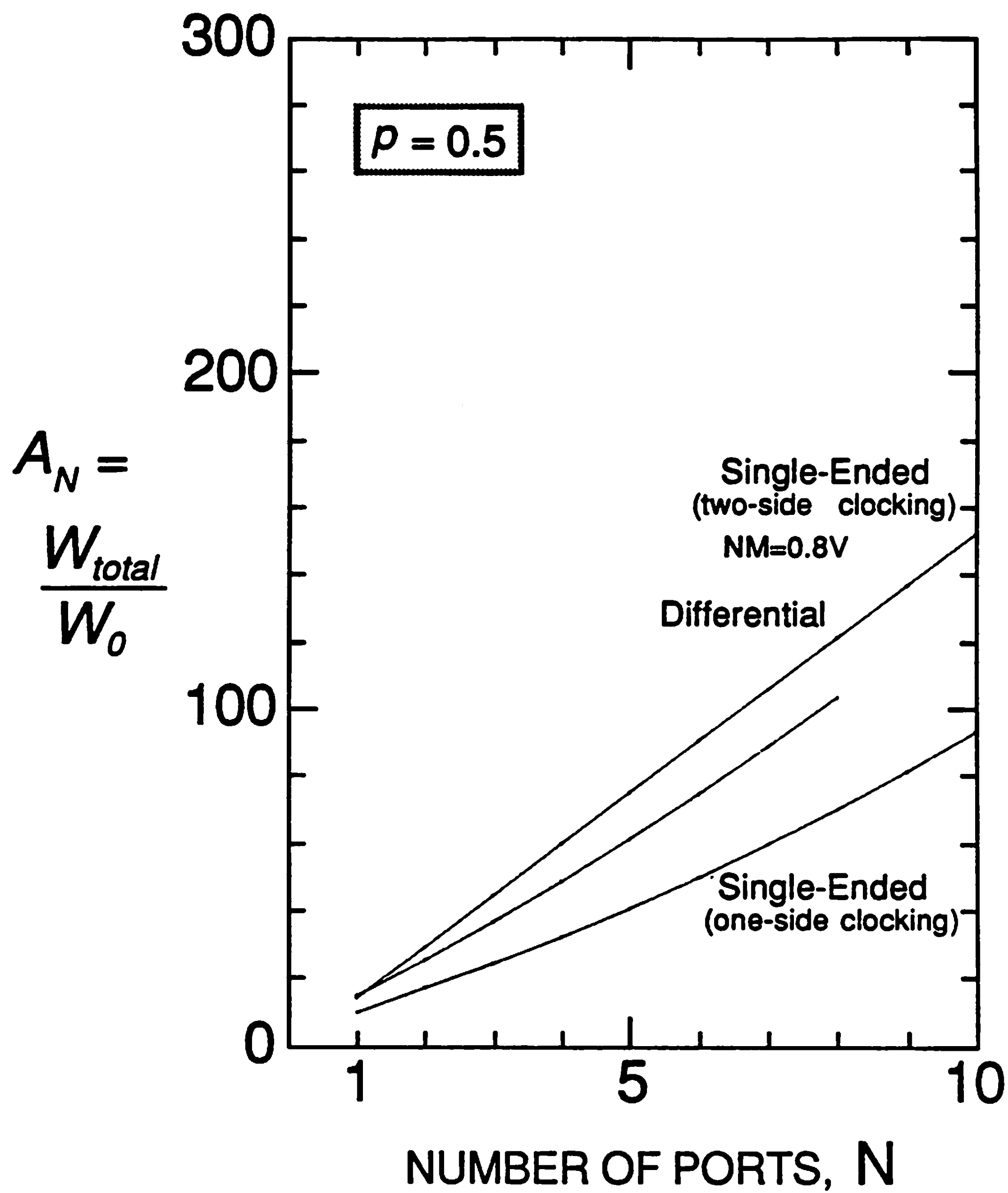
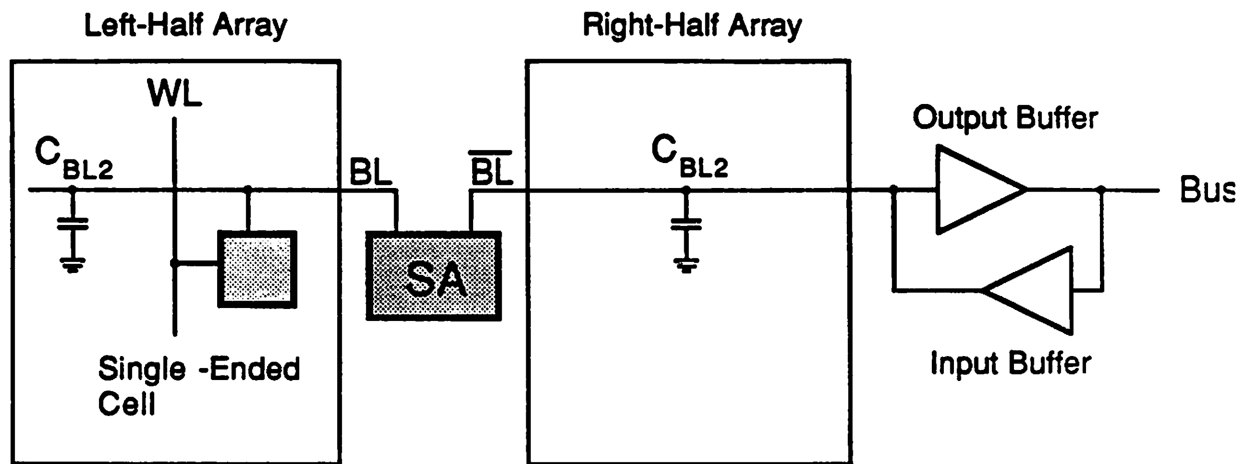
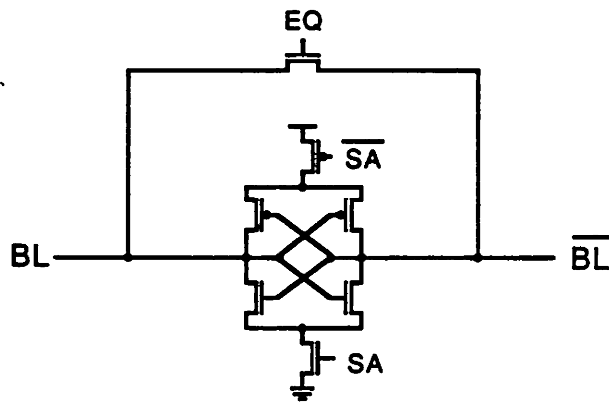


Figure 3.43 Normalized active area, A_N , versus number of ports, N, for different cell structures (improved signal/noise ratio due to reduced bitline capacitance and dummy cell for single-ended cell).

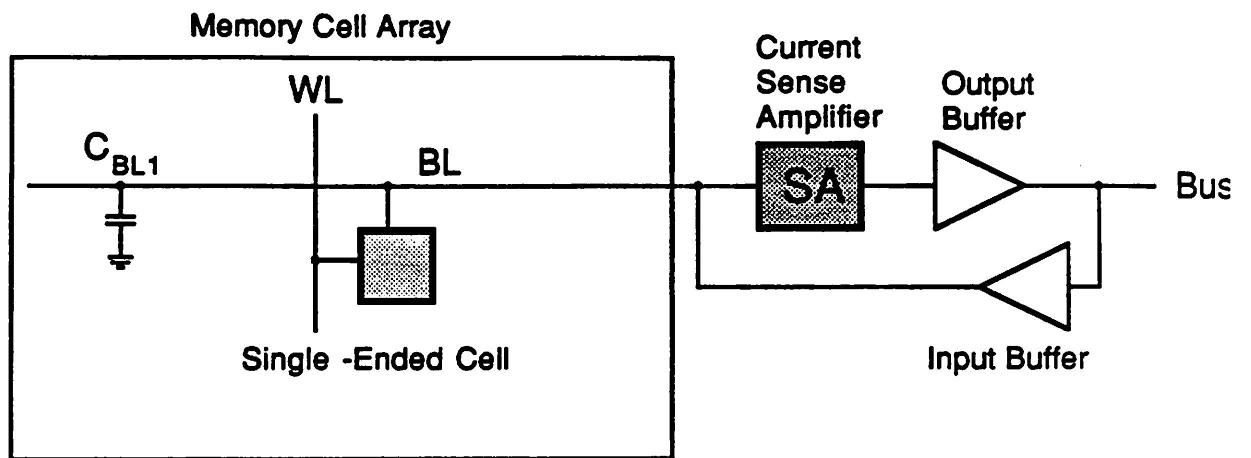


(a)

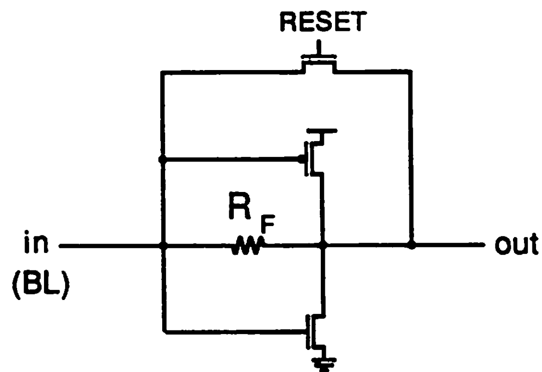


(b)

Figure 4.1 Voltage sensing scheme with:(a) split bitline configuration and (b) dynamic cross-coupled sense amplifier.



(a)



(b)

Figure 4.2 Current sensing scheme with:(a) one-sided bitline configuration and (b) transresistance sense amplifier.

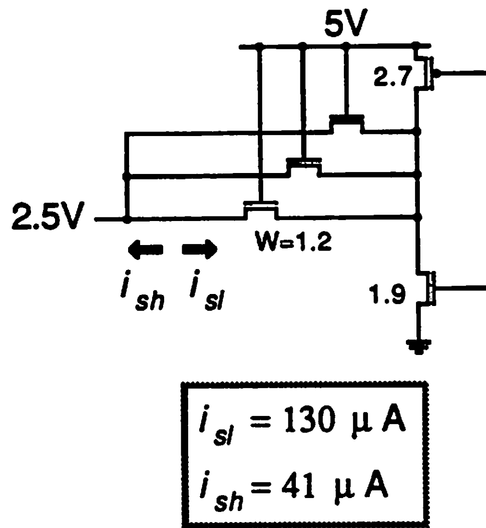


Figure 4.3 Signal current in minimum size single-ended triple-port cell.

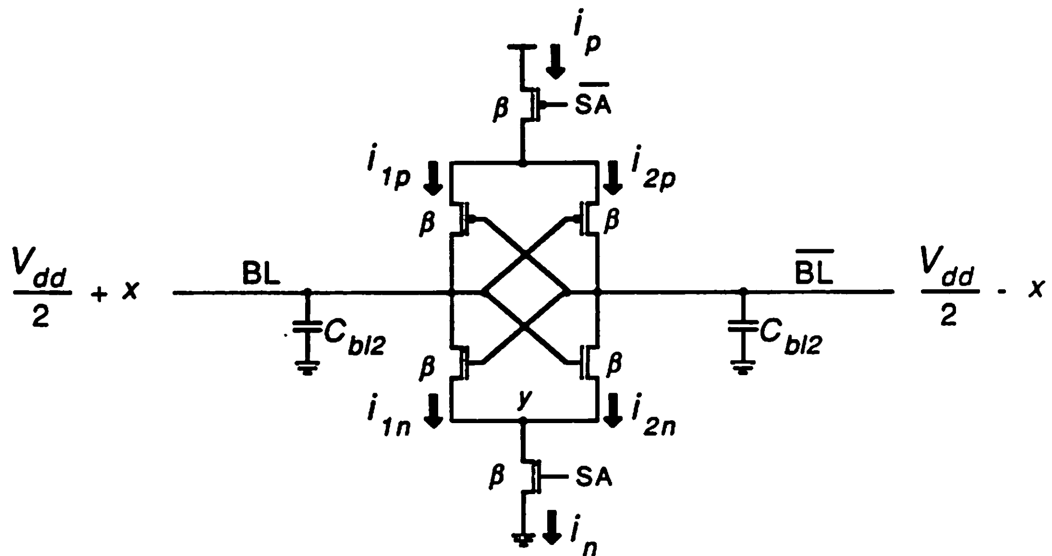


Figure 4.4 Dynamic cross-coupled sense amplifier circuit with notations for analysis.

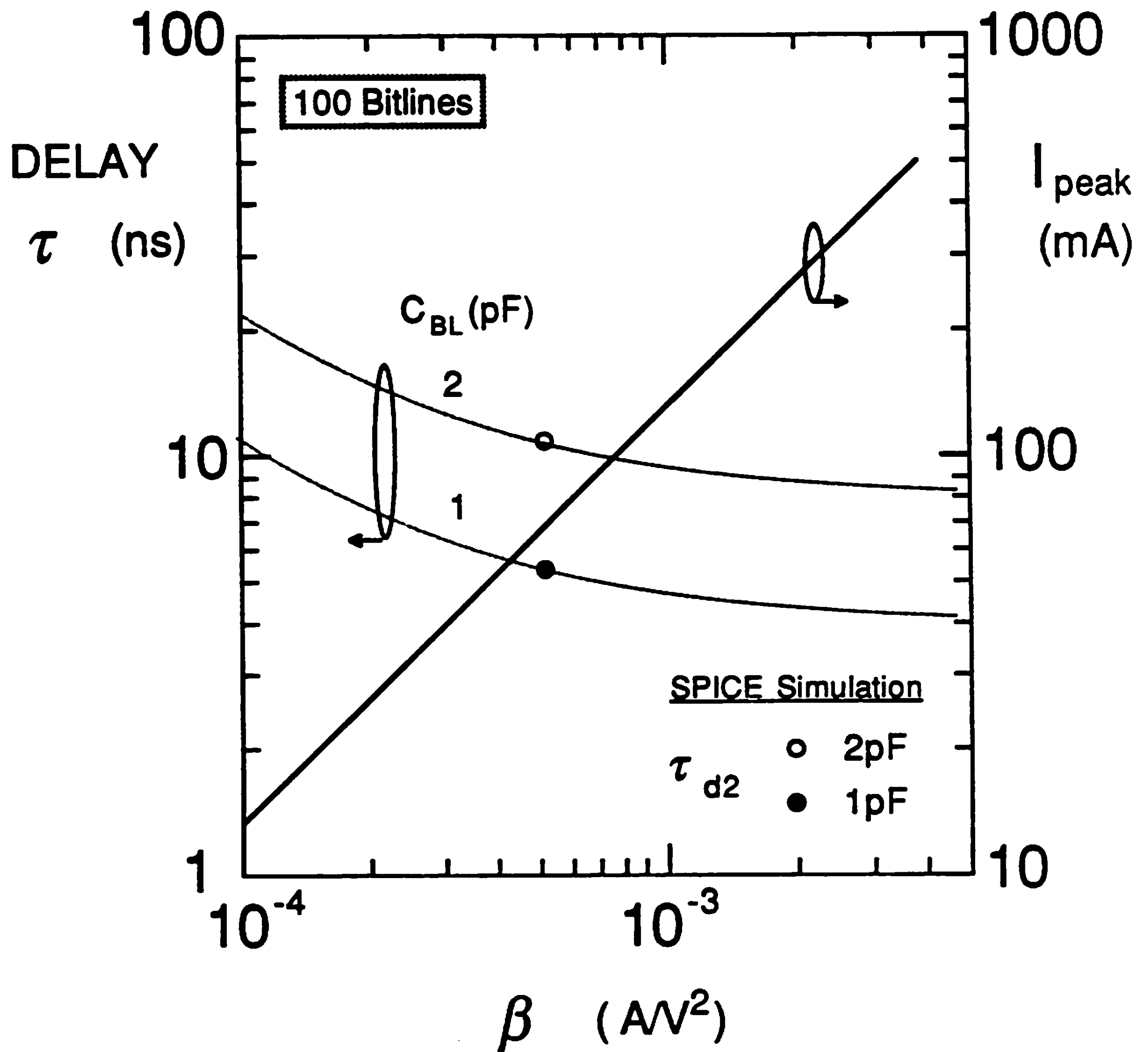


Figure 4.5 Calculated delay and peak current for dynamic cross-coupled sense amplifier shown as a function of transconductance of the transistor.

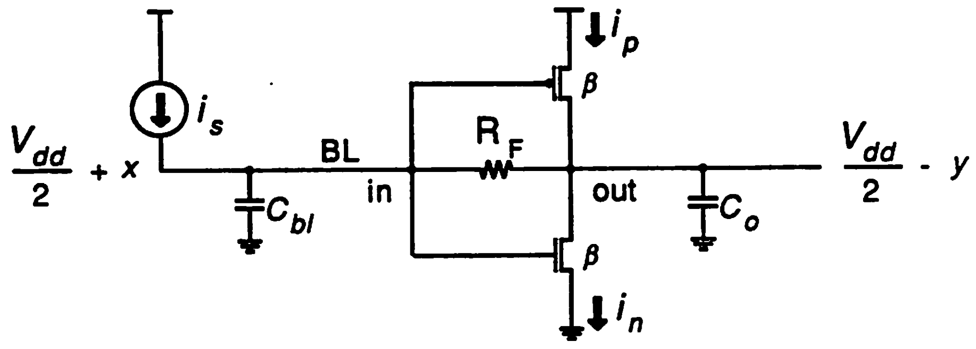


Figure 4.6 Transresistance sense amplifier circuit with notations for analysis.

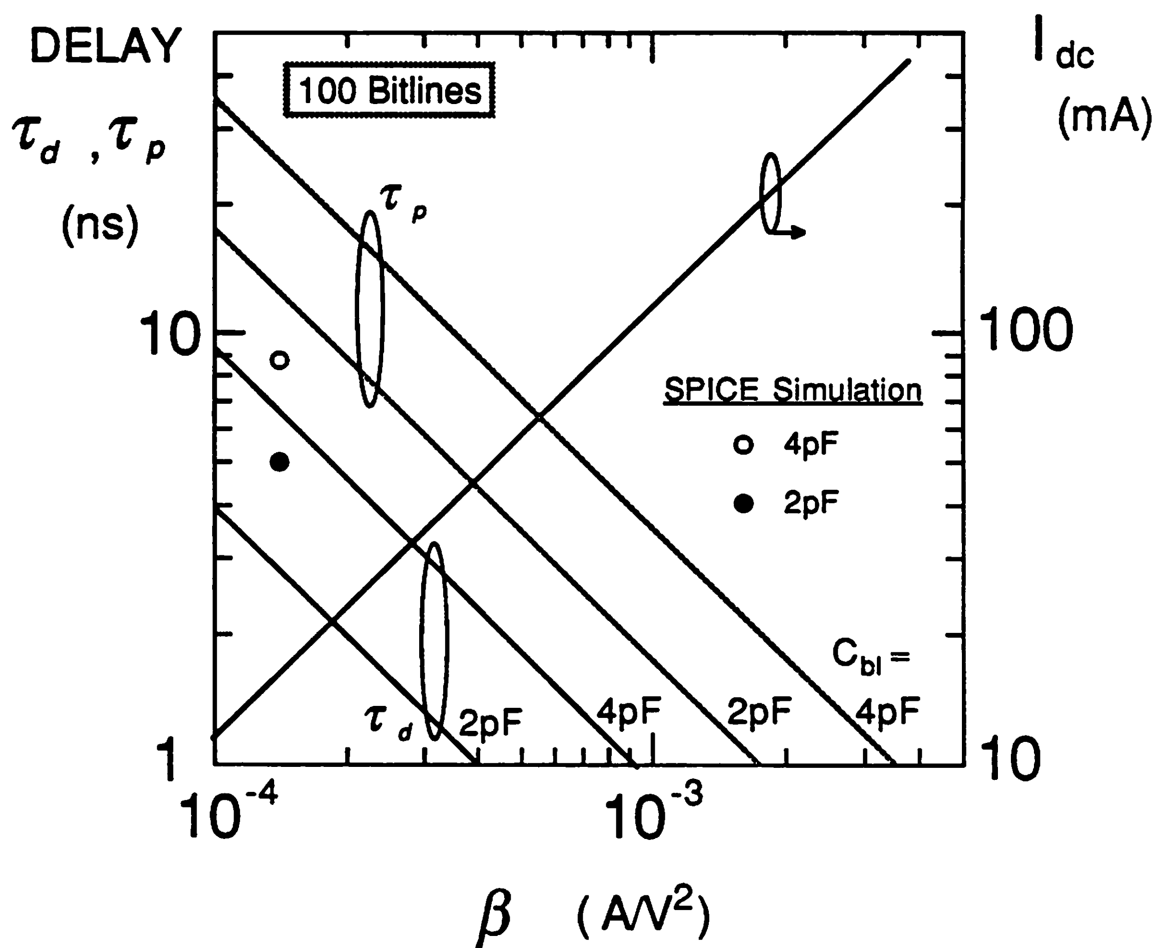


Figure 4.7 Calculated delay and peak current for transresistance sense amplifier shown as a function of transconductance of the transistor.

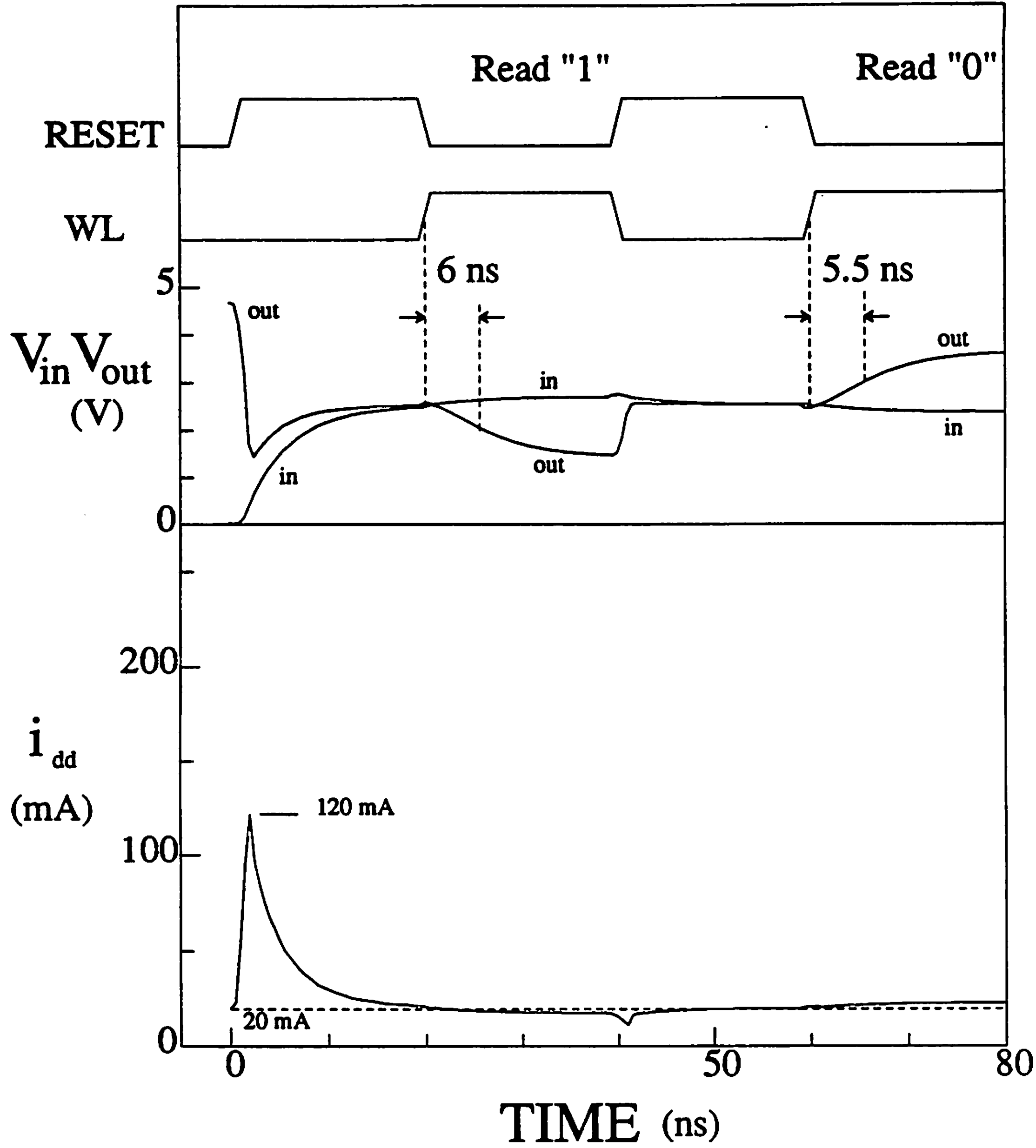


Figure 4.8 SPICE simulation results of the transresistance sense amplifier(Cbl=2pF).

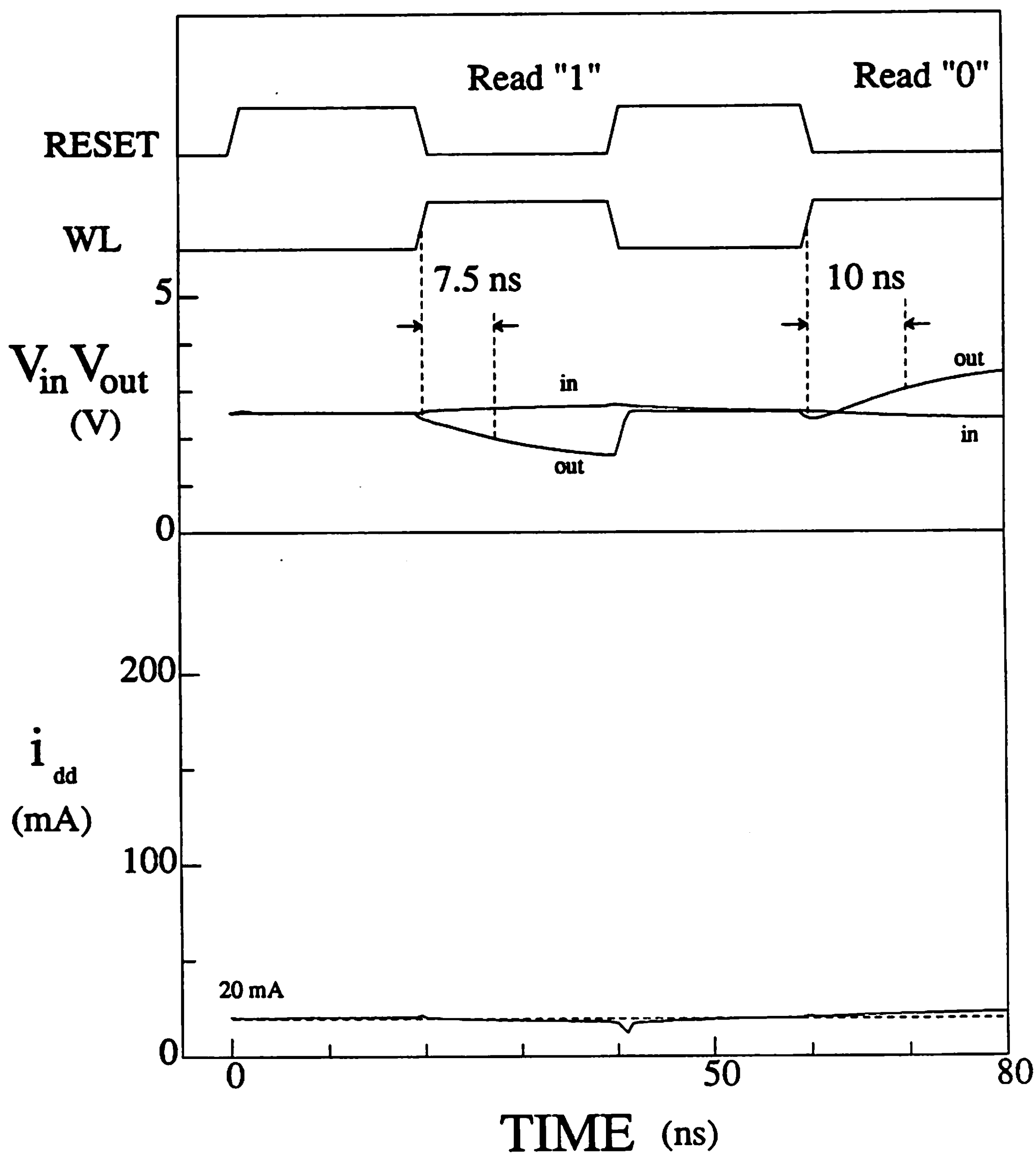
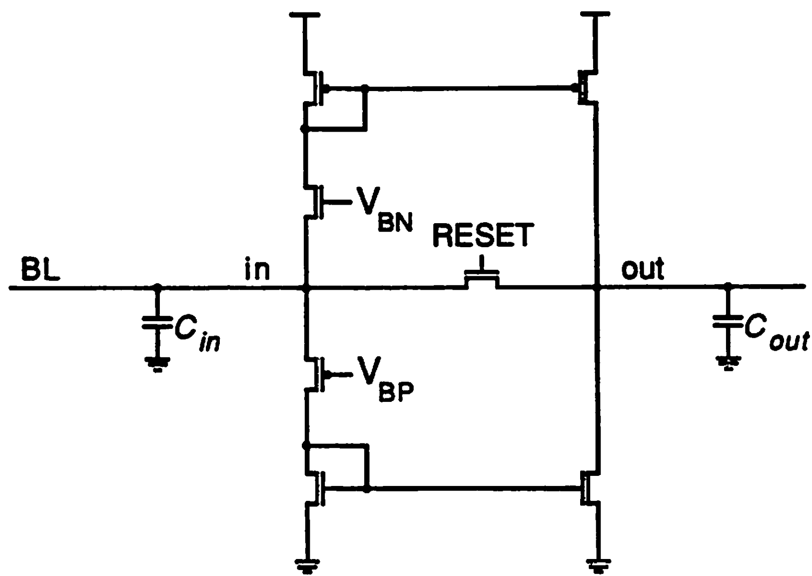
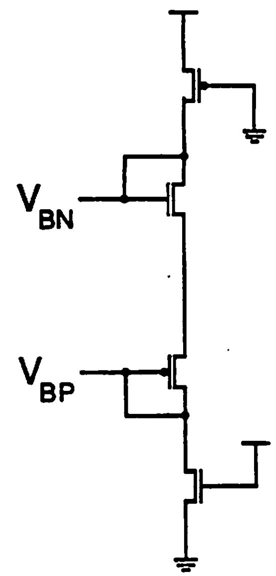
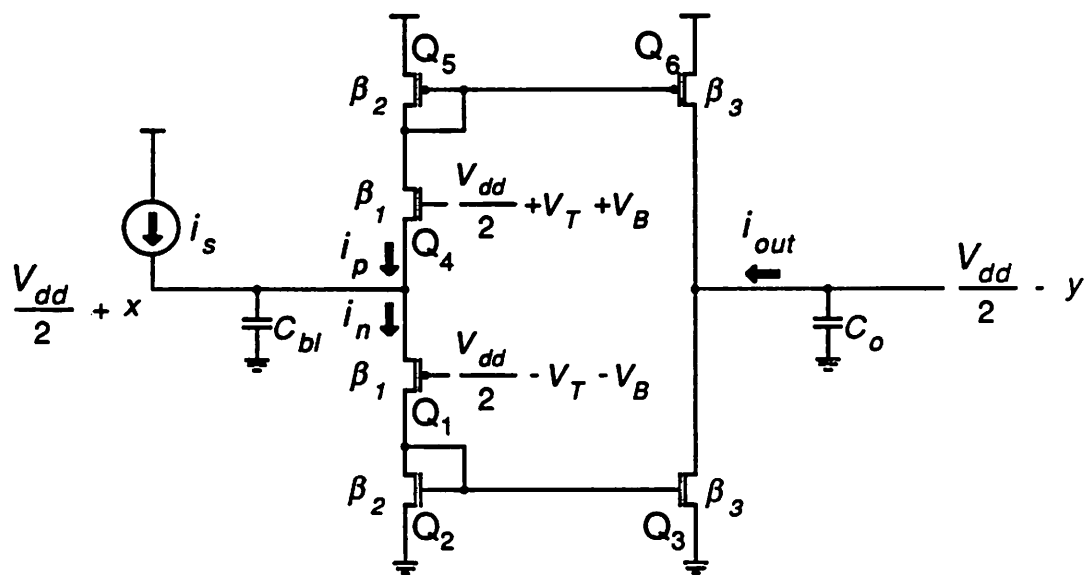


Figure 4.9 SPICE simulation results of the transresistance sense amplifier (Cbl=4pF).





(a)



(b)

Figure 4.10 Class-A push-pull current sense amplifier: (a) sense circuit and bias circuit , and (b) circuit diagram with notations for analysis.

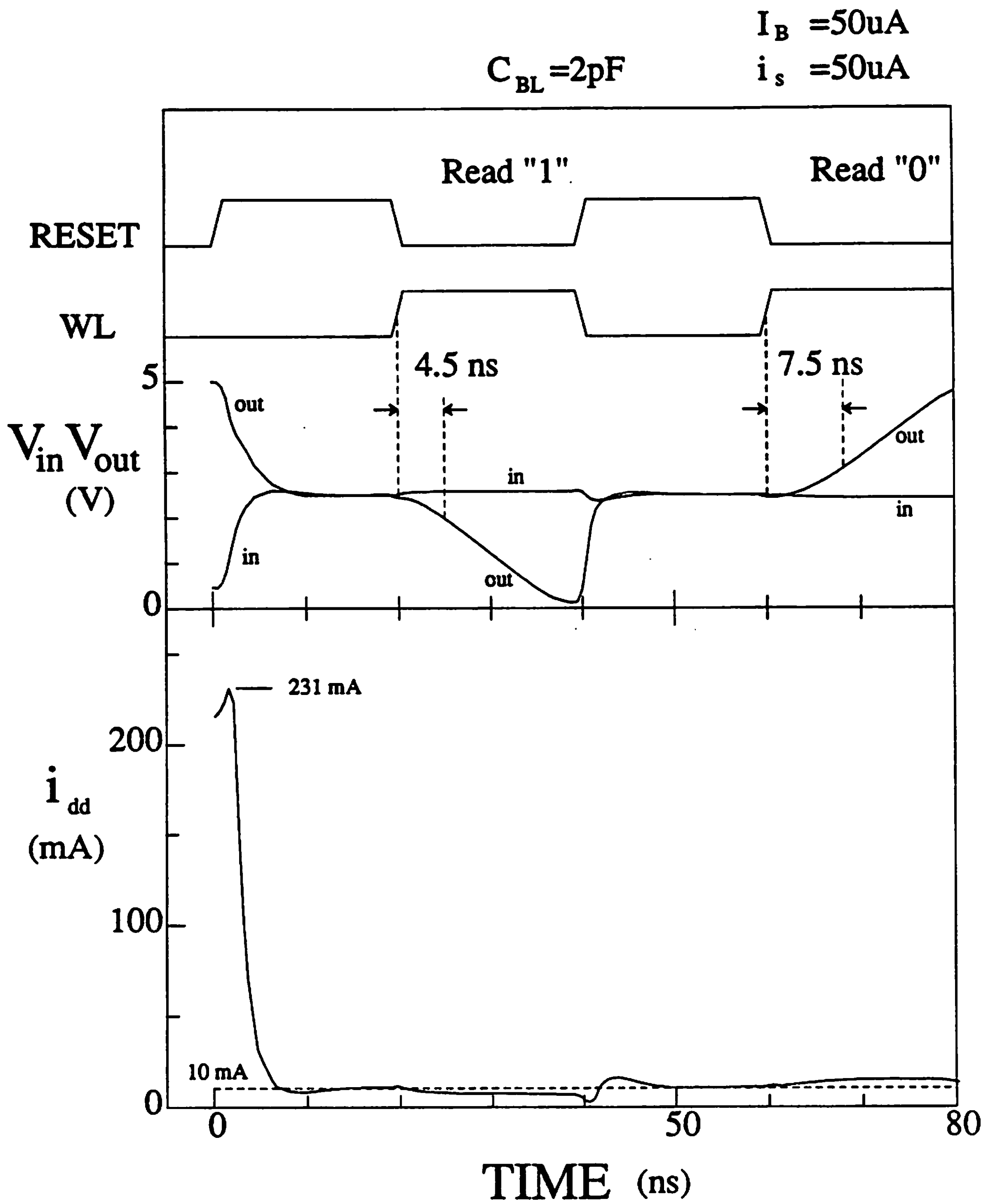


Figure 4.11 SPICE simulation results of the class-A push-pull current sense amplifier (Cbl=2pF).

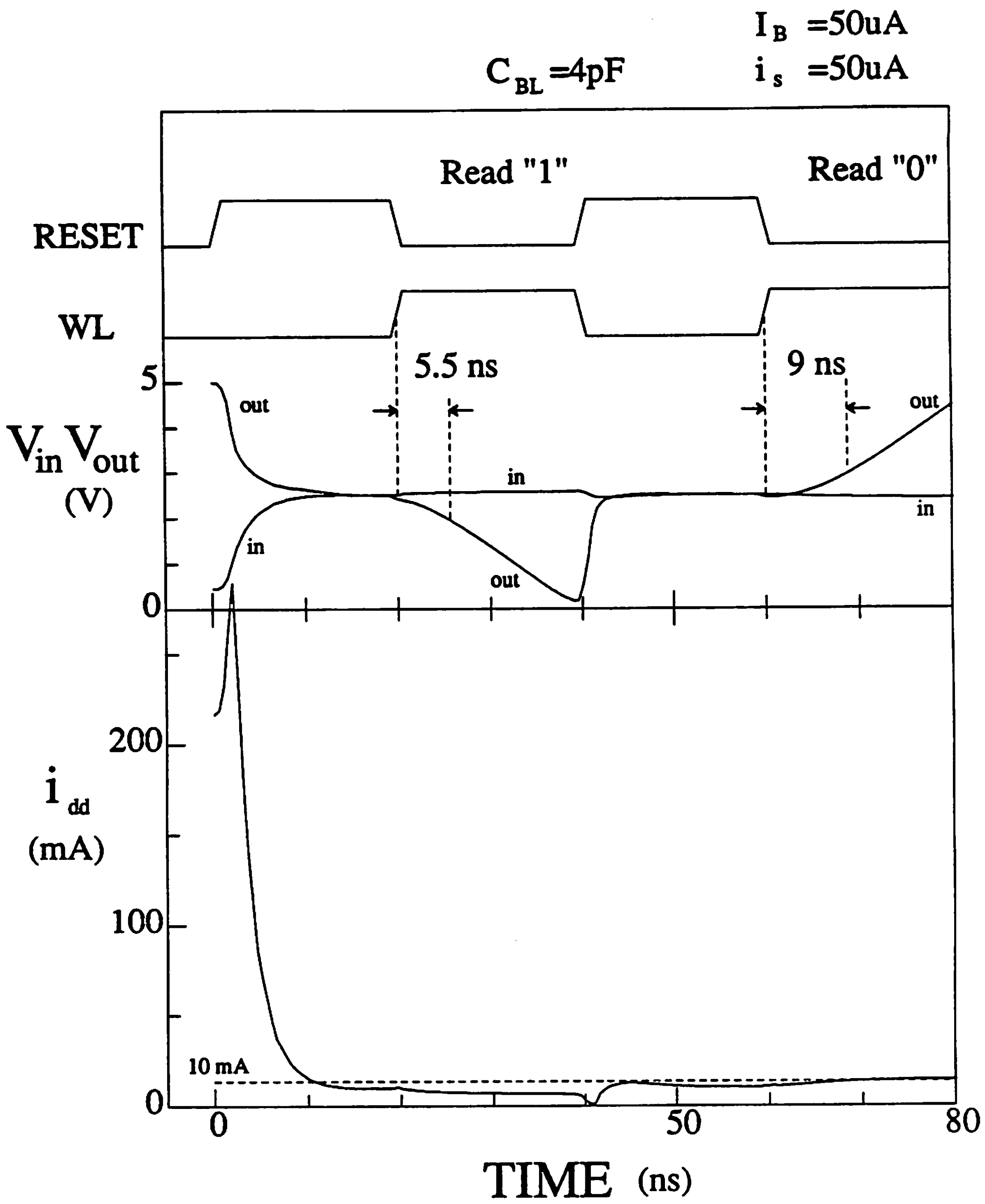


Figure 4.12 SPICE simulation results of the class-A push-pull current sense amplifier (Cbl=4pF).

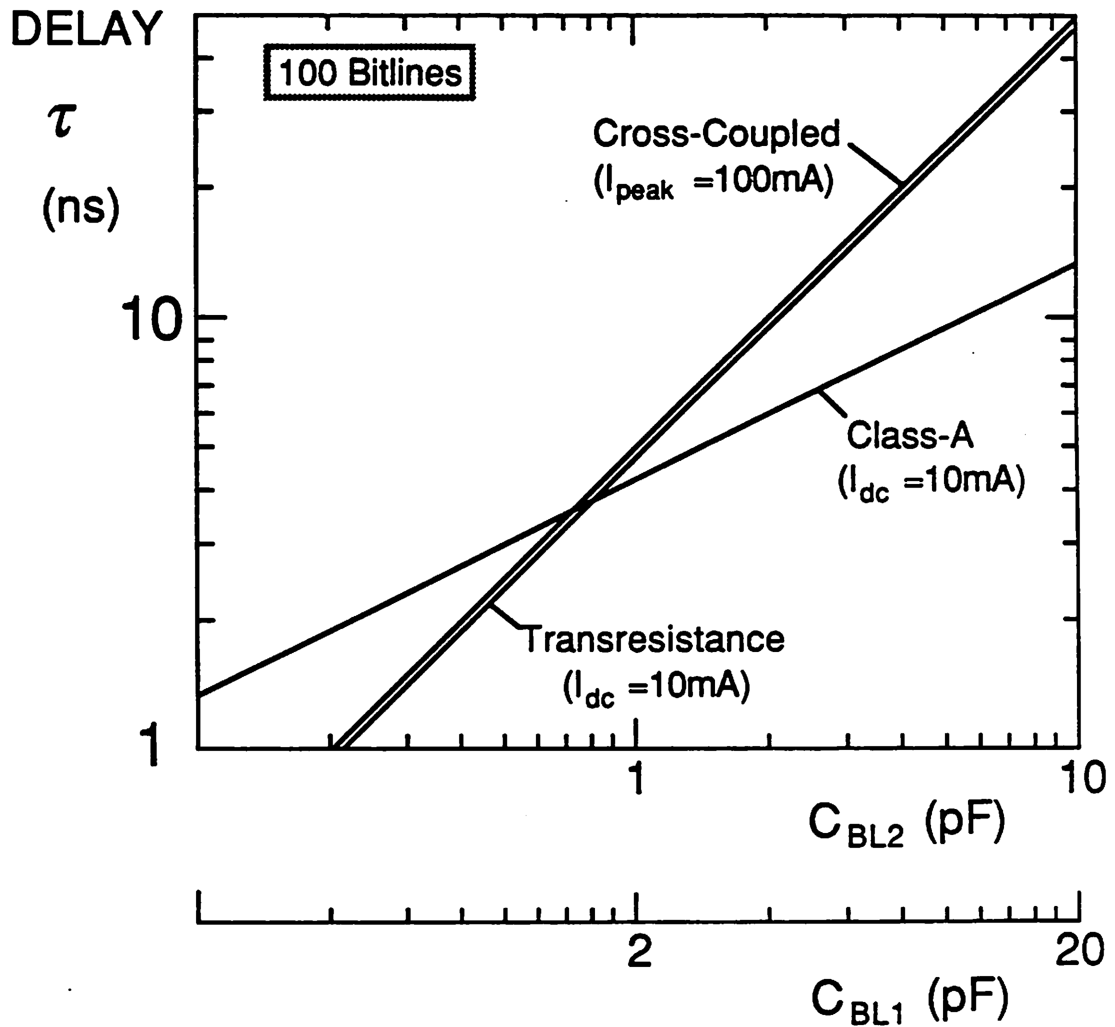


Figure 4.13 Calculated sensing delay as a function of bitline capacitance for different sense amplifiers.

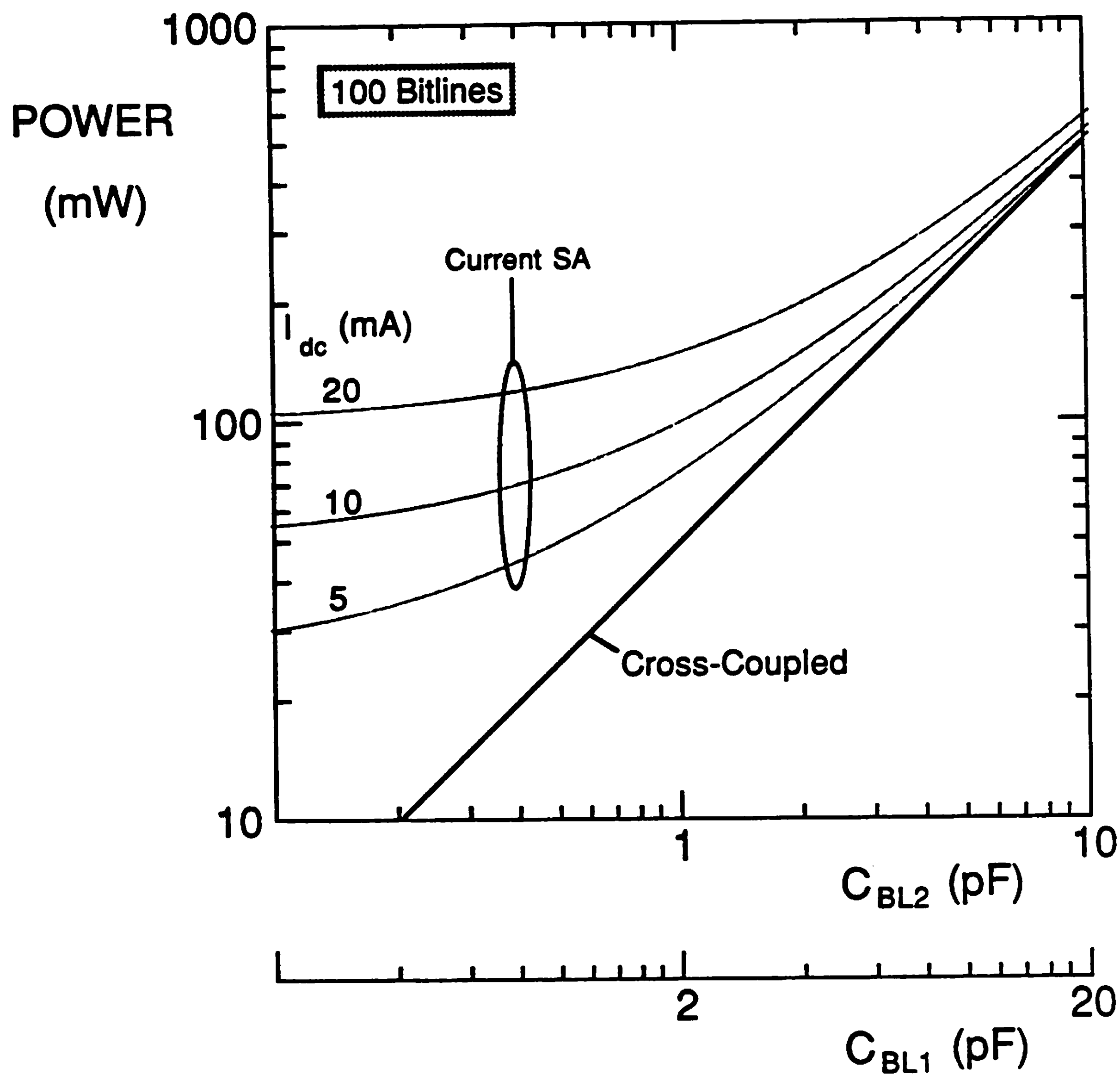


Figure 4.14 Calculated power dissipation as a function of bitline capacitance for different sensing schemes.

	Advantage	Disadvantage
Dynamic Cross-Coupled Sense Amp.	Easy to get Vdd/2 (but not accurate) Zero DC power	Needs timing margin for triggering SAs High peak current for high-speed sensing
Trans-resistance Sense Amp.	Simple configuration Small area	Needs prechrge Ckt. to set BL level after WRITE High DC Power for high-speed sensing High peak current for precharging BL
Class-A P.P. Current Sense Amp.	High-speed for larger BL capacitance Sensing speed is relatively insensitive to bias(DC) current	DC power High peak current for precharging BL Consumes large area

Table 4.1 Design tradeoffs for different sense amplifiers.

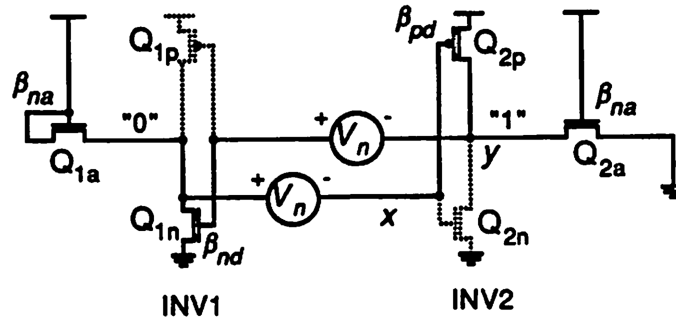


Figure A1.1 Circuit diagram of differentially accessed cell for analysis.

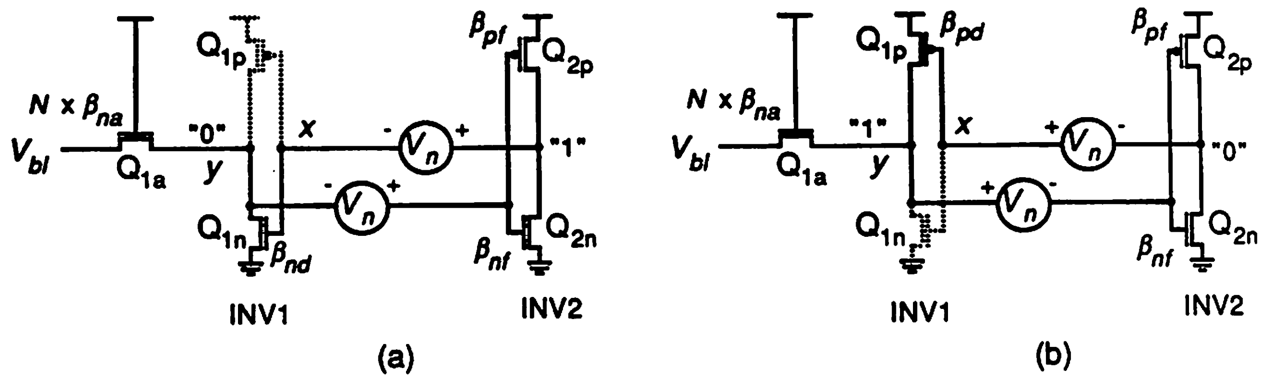


Figure A2.1 Circuit diagram of single-ended access cell for analysis: (a) "0" read noise margin, (b) "1" read noise margin.