Studies on Low Power Technologies for Battery-Operated Semiconductor Random Access Memories

山内，寛行

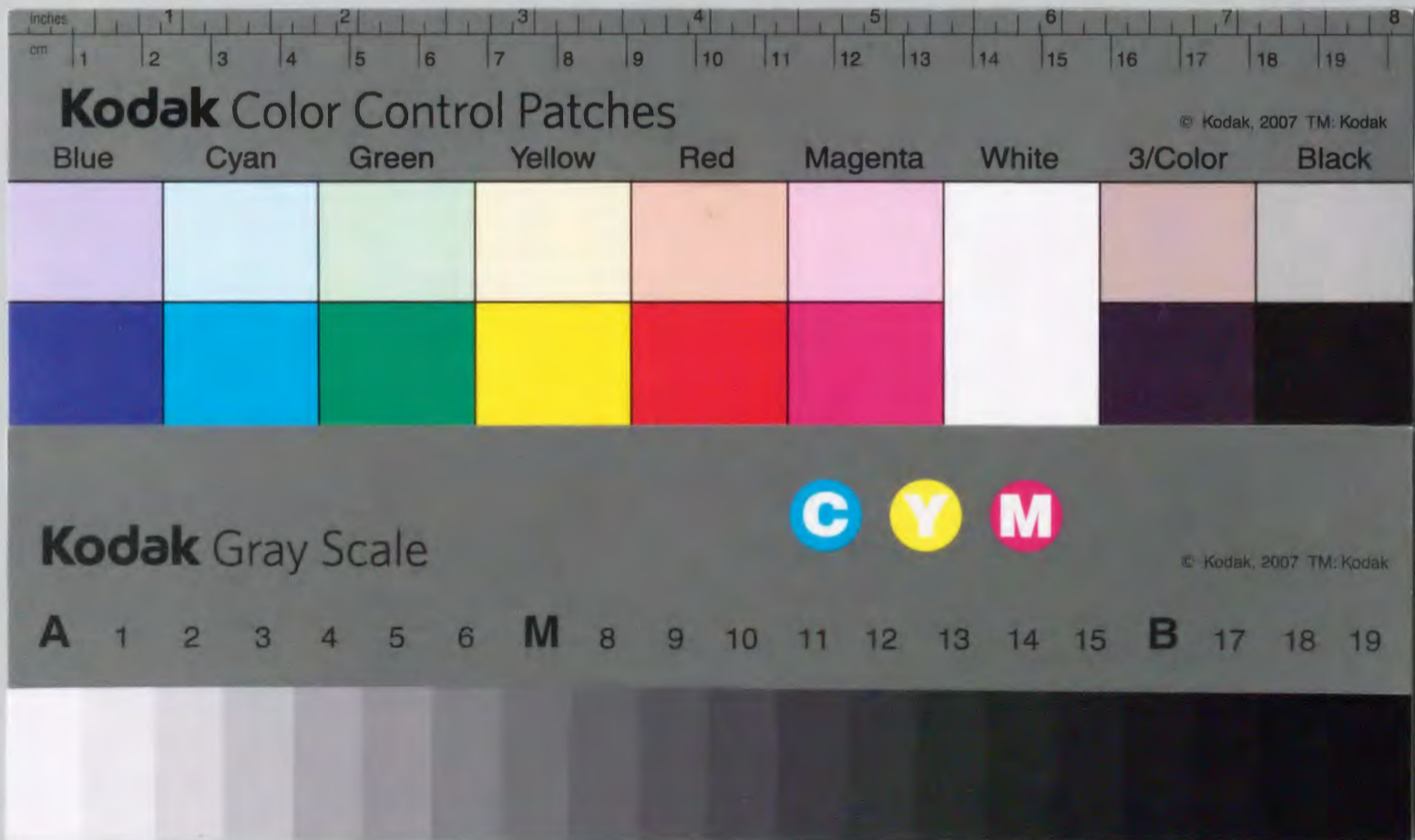# Studies on Low Power Technologies
for Battery-Operated Semiconductor
Random Access Memories

田中 利治

Studies on Low Power Technologies for Battery-Operated Semiconductor Random Access Memories

バッテリー駆動用半導体メモリの低消費電力化
に関する研究

May, 1997

Hiroyuki Yamauchi

山内　寛行

ABSTRACT

This thesis has reported the low power circuit technologies for battery-operated semiconductor random access memories and their systems, including 1) charge-recycling data transfer scheme and signal-swing suppressing time-multiplexed differential data transfer scheme, 2) data retention power saving technique for DRAM's down to as low as SRAM's by extending data refresh interval, resulting from using relaxed junction biased data retention scheme and plate-floating leakage-monitoring timer, and 3) operating voltage scaling techniques enabling to accommodate the operating voltages of DRAM's and SRAM's to 1.8V and 0.5V, which correspond to the supplied voltages from two Ni-Cd cells connected in series and a single solar cell, respectively, instead of using a standard supply voltage of Vcc=5.0V or 3.3V.

The main results of this thesis are as follows :

(1) To realize an over 3GB/s data transfer rate through the bus whose capacitance is as large as 14pF, the charge-recycling data transfer scheme has been developed, which enables to save the power consumption by the quadratic factor of suppressing ratio m of data bus swing. Assuming m is 8, conventionally consumed 1W-power have resulted in saving down to merely 16mW. Such dramatic power reduction has been verified by the simulated and measured data.

Furthermore, the time-multiplexed charge-recycling data transfer scheme has also been developed. According to the findings given through simulated and measured data, the proposed technique can reduce the bus power consumption to 1/11 and 1/3 of that when the bus activities are 100% and 25%, respectively, while reducing the number of signal wires by half, compared to the parallel architecture.

(2) To replace SRAMs with DRAMs in battery operated devices, the following circuit techniques have been developed : 1) the relaxed junction biasing scheme which enables to extend the data retention time by a factor of 3, resulting from relaxing the junction bias between the storage node and substrate and in turn, from reducing the junction leakage, and 2) plate floating leakage monitoring timer which can extend the refresh interval by a factor of 30, resulting from setting the optimum refresh interval based on the DRAM's temperature dependence. As a result, these have contributed to diminish the current consumption down to sub 0.4μA/MB, which is as low as SRAM. Furthermore, the following circuit techniques have been developed to suppress the DC current to less than 0.1μA/MB : 1) gate received level detector, which provides higher gain for the leakage current from or to the potential monitored node such as substrate, and 2) dynamically controlled reference generator, which cuts off the static current resulting from on and off switching of the power supply.   By utilizing such techniques, the world's smallest data

retention current of 0.5μA/MB has been accomplished by using experimental 16Mbit DRAM.

(3) To achieve the fast access time of less than 40ns even reducing the supply voltage to 1.8V, corresponding to the voltage of two Ni-Cd cells connected in series, five circuit techniques have been developed, as follows : 1) a parallel column access redundancy scheme featuring a current sensing address comparator, 2) a quasi-static cross-coupled data bus amplifier, 3) a gate isolated sense amplifier with low threshold voltage, 4) a layout that minimizes the length of the signal path by taking advantage of the lead on chip assembly technique, and 5) suppressing the asymmetrical characteristics in the sense amplifier when V_T and gate length are scaling. By utilizing such techniques, the world's fastest battery operated 16Mbit DRAM with the RAS access time of 20ns at 3.3V and also 36ns even at 1.8V have been developed, while keeping the standby current of only 5μA.

(4) To accommodate the operating voltage to the single battery power supply voltage, which should be scaled down to 0.5V ultimately, corresponding to solar cell, the V_T scaling have been chosen to compensate for the degradation in SRAM operation speed of 100MHz, while avoiding the exponentially increased subthreshold leakage as V_T is scaling. The key circuit technique to realize that is the offset data storage scheme, which enables to minimize the charge amount supplied from the embedded charge pump circuits. This provides the effective gate to source voltage (V_{GS}-V_T) up to 0.8V necessary to achieve 100MHz operation even at 0.5V single power supply. The possibility of realizing the 0.5V/100MHz SRAM operation, while suppressing the operating power of sub-5mW, has been verified by simulation.

According to the results of (1) through (4), it is expected that the low power circuit technologies proposed in this thesis enable to surmount the facing obstacles when try to meet the following requirements for: 1) saving power consumption down to sub-1W when data transfer of more than 3GB/s between memory and processor/graphics controller even for the bus capacitance of 14pF, 2) reducing DRAM data retention current to less than 0.5μA/MB as low as SRAM, and 3) accommodating the operating voltage to 0.5V of solar cell, while keeping 100MHz operation and sub-μA standby current. And as a result, these contribute to extend battery life-time and to accommodate operating voltage to single battery power supply voltage in battery-operated semiconductor random access memory systems. These result in increasing portability due to reducing battery size and weight and in getting a free from troublesome of quite often recharging necessary to recover battery supply voltage in portable battery operated devices.

ACKNOWLEDGMENTS

A TABLE OF CONTENTS

A List of Figures and Tables

[Chapter-3]

[Chapter-4]

A List of Technical Terms and Symbols

[Chapter-1]

DRAM: Dynamic Random Access Memory
SRAM: Static Random Access Memory
Ni-Cd: Nickel Cadmium, type of cell
Ni-MH: Nickel -Metal-Hydride, type of cell
Li-Ion: Lithium-Ion, type of cell
HDD: Hard Disk Drive
TFT: Thin Film Transistor
LC: Liquid-Crystal
PDA: Personal Digital Assistant
PC: Personal Computer
I/O: Input / Output terminal
UNIX: Software of operating system for workstation developed by AT&T
MS DOS: Software of operating system for PC developed by Microsoft Inc.
Windows95:   Software of operating system for PC developed by Microsoft Inc.
Windows NT: Software of operating system for PC developed by Microsoft Inc.
3D graphics: Three dimensional graphics
RAM: Random read/write Access Memory
ROM: Read-Only Memory
EPROM: Ultraviolet ray Erasable Programmable ROM
EEPROM: Electrically Erasable Programmable ROM
MOS: Metal-Oxide-Semiconductor
ISSCC: International Solid-State Circuit Conference
CICC: Custom Integrated Circuit Conference
MPEG: Moving Picture-coding Experts Group
PKG: Package
CPU: Central Processing Unit
DC-DC: DC to DC voltage transformer
L1/L2 cache: Primary/ secondary cache
Pentium/ Pentium Pro: Name of microprocessor produced by Intel Inc.
RISC: Reduced Instruction Set Computer, CISC: Complex Instruction Set Computer
HDTV: High Definition Television
Mb/Gb/MB/GB: Mega-bit/Giga-bit/Mega-byte/ Giga-byte, 1 byte = 8 bit
CG: Computer Graphics

Rambus DRAM: DRAM with high-speed interface, Trade mark of Rambus Inc.

SynchLink DRAM: DRAM with high-speed interface, now discussing in consortium

Vcc: Power supply voltage

Frq: Frequency

Tref: Refresh cycle time

Iss: Subthreshold leakage

LSI: Large Scale Integrated circuit

BW: Data transfer rate,   Bandwidth

P: Power consumption

F: Frequency of data transfer

N: Number of parallel data buses

C/ Cbus: Capacitance/ Bus capacitance

Vbus: Voltage of data bus-swing

MPU: Microprocessor unit

HZ: Unit of frequency

V_T: Threshold voltage

[Chapter-2]

CRB: Charge Recycling Bus

N: Number of bus width

Cbus: Capacitance of bus wiring

Vcc: Power supply voltage

Frq: Frequency

Super HD: Advanced level of High definition (HD)

CMOS: Complementary MOS, i.e., combination of PMOS and NMOS

MOSFET: Metal Oxide Semiconductor Field Effective Transistor

Psupp: Power consumption for suppressed swing bus

Pfull: Power consumption for full swing bus

m: Ratio of Vcc/Vbus

Vss: Source voltage of ground potential

Di/ XDi: Parallel complementary bus pair (i=integer)

Qi: Charge amount (i=integer)

Ti: A period in timing (i=integer)

Qtotal: Total charge amount

C_{Di}/ C_{XDi}:   Capacitance on complementary bus pair (i=integer)

INi/ XINi: Complementary of input of bus exchanger (i=integer)

SWi: Bus switches   (i=integer)

EQ: Signal of equalizing bus pair

V_{GS}:   Voltage difference of gate to source electrodes

I_{DC}:   DC idling current

ULSI: Ultra Large Scale Integrated circuits

ATi/ XATi: Complementary bus pair with imbalance capacitance

Kdi: Deviation amount of bus swing of #i-bus (i=integer)

ΔV: Noise amount injected into bus

CAD: Computer Aided Design

PCS: Personal Communication Service

[Chapter-3]

TMD: Time-Multiplexed Differential data transfer

DTD: Data Transition Detector

CRB: Charge-Recycling Bus

TM-CRB: TMD combined with CRB

SSL: Single Signal Line

VGA: Video Graphics Accelerator

HLP: Half Level Precharging

Vm: Suppressed voltage swing

m: Vcc/Vm

t_{ds}:   Delay time for SSL

t_{dn}:   Delay time for TMD

DCLK: Clock signal with doubled operating frequency

MCLK: Main clock signal

Ai/ Bi: Input signals

At/ Bt: Transferred bus signals

XOR: Exclusive-OR

A_{OT}/ B_{OT}:   Latched output data

RCLK: Clock signal for receiver

n: Number of elements in each stack

V_O: Over gate voltage, i.e., effective gate to source voltage

[Chapter-4]

RJB: Relaxed Junction Biasing

PFM: Plate Floating leakage Monitoring

V_{BB}:   Voltage of substrate

PDWD: V_{BB} pull-down word-line driver

GRD: Gate Received V_{BB} level Detector

PDA: Personal Digital Assistant

Self-refresh DRAM: Type of DRAM with built-in refresh controller including address counter and timer

V_N: Voltage of storage node

WL: Word Line

WDn: WL pull-down signal

V_A: Voltage of selected gate electrode

V_{REF}:   Voltage of reference level

V_{PLD}:   Reset signal of plate

I_{SB}:   Standby current

T_P: Pause period

I_{RF}:   Refresh current consumption

I_{BB}:   Substrate current

DCRG: Dynamically controlled reference generator

I_{RC}:   Data retention current $(I_{RC} = I_{RF} + I_{SB})$

T_B: Refresh period

T_P: Pause period

[Chapter-5]

SOJ package: Type of SOJ package

LOC: Lead On Chip assembly technique

PCAR: Parallel Column Access Redundancy

CSAC: Current Sensing Address Comparator

NPCA: N&PMOS Cross-coupled data bus Amplifier

GISA: Gate Isolated Sense Amplifier

S/A: Sense Amplifier

SCLm: Spare Column Line

NCLn: Normal Column Line

Td: Delay time

Tdc: Time to transmit S/A's data to read-bus-amplifier

V_{RN}:   Voltage of output node of dynamic NOR circuit

V_{RC}:   Voltage of gate of pull-down transistor

PPCA: P&PMOS Cross-coupled Amplifier

I_{SA}:   Current consumption through sense amplifier

DB/XDB: Complementary data bus line pair

DBI/XDBI: Complementary data bus line pair after column switch

DBII/XDBII: Complementary data bus line after first-stage sense amplifier

CX: Capacitance of V_{RN} node

SPY9:   Spare column line

LOCOS: Local Oxidation of Silicon

T_{RCD}:   Delay time of RAS to CAS enable timing

T_{RAD}:   Delay time of RAS to Address enable timing

T_{RAC}:   RAS access time

T_{AA}:   Address access time

I_{CC}:   Supply current, I_{CC1}: operating current, I_{CC2}: standby current

[Chapter-6]

BOGS: Boosted and Offset Grounded data Storage

NSD: Negative Source Drive

NWD: Negative Word-line Drive

CPU: Central Processing Unit

V_o: Effective gate to source voltage, $(V_{GS}-V_T)$

t_{WL-BL}:   Bit-line Access Delay

OSD: Offset-Source Driving

ΔV_{WR}:   Bit-line write signal swing

N_H/N_L: Storage node pairs of SRAM

C_{SN}:   Capacitance of common source node

Q_{SL}:   Charge amount for driving C_{SN}

Q_{PR}:   Charge amount for precharging C_{PR}

Q_{SD}:   Charge amount for driving C_{SN}

C_{BL}:   Capacitance of bit-line

C_{VPL}:   Capacitance of virtual source line

$\eta_A, \eta_B, \eta_C, \eta_D, \eta_E$:   supply efficiency of charge pump circuit A,B,C,D,E

I_{LK}:   Cell leakage current

Cj: Junction capacitance

C_{BL}:   Bit-line capacitance

Vj: Junction bias

V_{VPL}:   Potential of virtual source line

SL: Source-line

CHAPTER-1　Introduction

1-1　Backgrounds

1-1-1　Needs for Low Power

The need for low power has caused a major paradigm shift [1]. Previously, the major concerns in electronics devices were processing performance (throughput) and area, while low power was generally important only if some cooling limit were being exceeded. (Of course, there are exceptions to this rule. For example, there has been a long history of low power as a niche market for such applications as wrist watches, pocket calculators and heart pacers.) However, low power has emerged as a major theme today in the electronics industry.　This is because power consumption has become an increasingly important cost factor in terms of battery, chip packaging and chip cooling, when trying to increase computing capabilities which make possible powerful personal computers, sophisticated three-dimensional computer graphics, and multi-media capabilities such as real-time speech recognition and real-time video compression/ decompression.　Since the density and size of the chips and systems would continue to increase, it is clear that the difficulty in providing adequate cooling and battery operated life-time ultimately might either add significant cost to the system or provide a limit on the amount of functionality that can be provided.　Figure 1-1 shows the power consumption comparisons between battery-powered and nonbattery-powered systems. In battery-powered systems, there are notebook personal computers (PCs), personal digital assistant (PDA), handheld PC, memory card, wrist watch and heart pacer. On the other hand, category of nonbattery-powered system includes power hungry workstations and desktop PCs.

The current push toward lower power is likely based on the following three generic points [1]:　1) Battery operated portable systems such as PDA and handheld PC. Here, key issue is power saving to extend battery life within limited battery size as long as possible. This is because a major factor in the weight and size of these portable devices is the amount of batteries which is directly impacted by the power consumption in the electronic devices, including semiconductor random access memory systems. The first priority in designing these devices is for minimum power with the required level of performance, resulting in maximum battery life. Typical power consumption of handheld PC is 0.5W or less, as shown in Fig. 1-1.　2) High performance portable computers such as notebook PCs. Here, the target for low power is to reduce the power consumption of the electronics portion of the system down to negligibly small level compared to that of other parts of the system. This is because the whole system power is
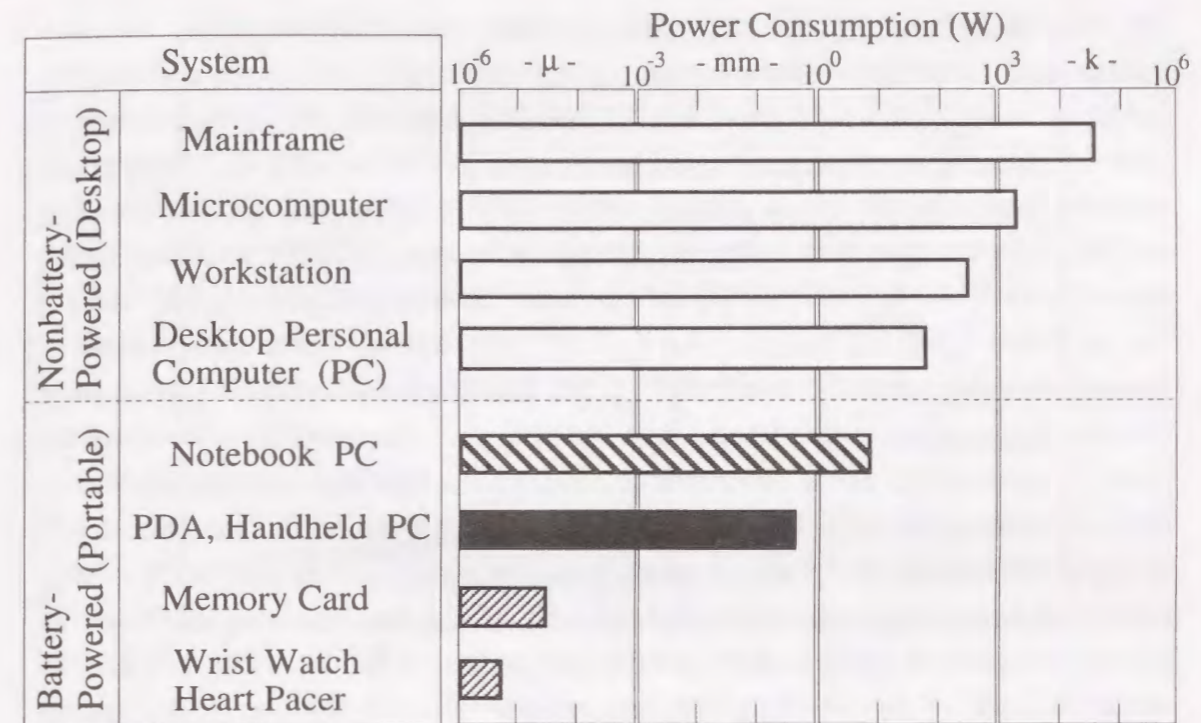
Fig.1-1　Power consumption comparison between battery-powered and nonbattery-powered systems.
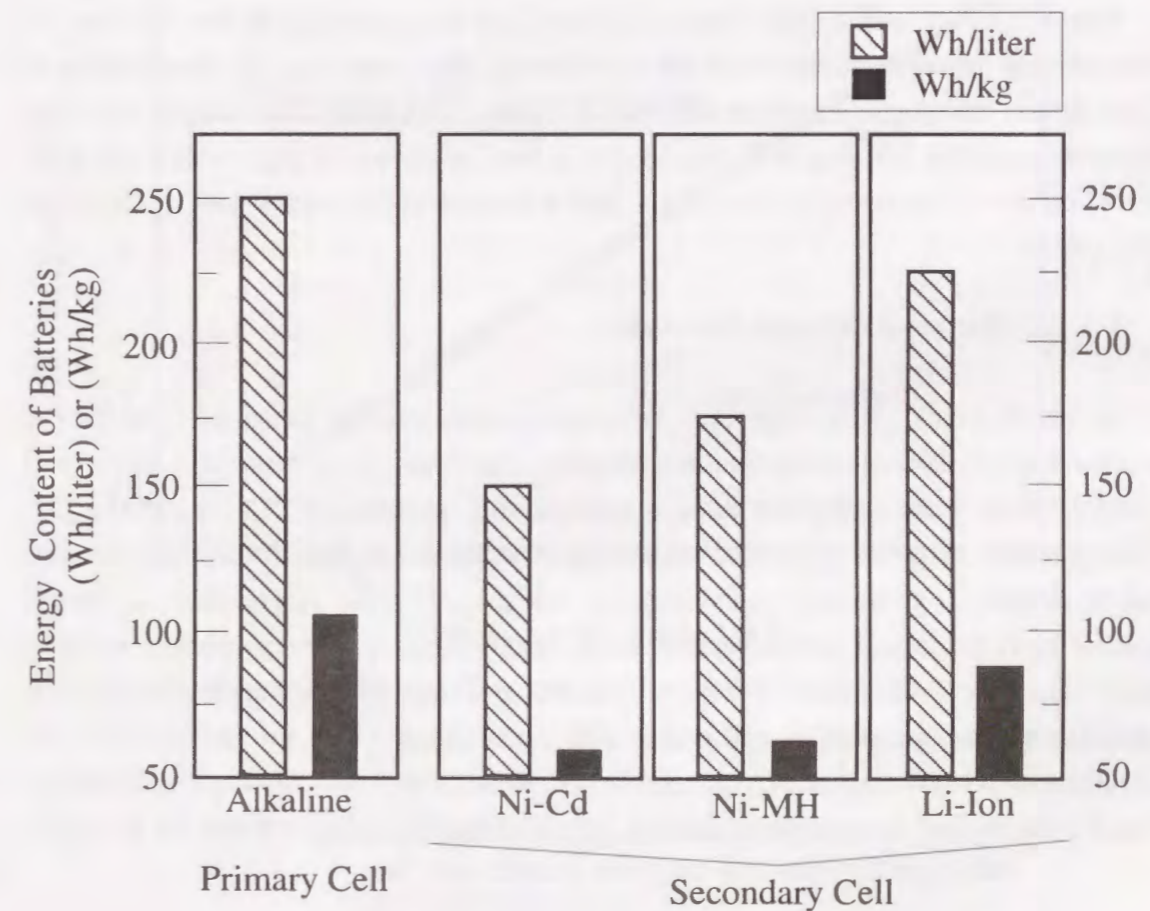
Fig.1-2　Energy content of typical primary cell and secondary cells.

not only determined by the electronics including semiconductor chips, but also determined by other parts of the system, such as the display and the hard disk drive (HDD). Thus, once a satisfactory level of power is achieved, this portable computer system is designed for further increased performance as close to desk top computers as possible. Typical power consumption of notebook PC is 10W or less [2]. Since energy content of typical cells with 0.4kg is expected to be only 25-30Wh according to the nominal data of energy content of typical secondary batteries of Ni-Cd, Ni-MH, and Li-Ion, as shown in Fig.1-2, it is clear that the battery-operated life time of these systems is limited below only about 2-3 hours [3]. 3) Nonbattery-powered systems such as main frames or large servers, workstations, desk top PCs, etc. (shown in Fig. 1-1), where the goal is to keep power below some limit imposed by the following requirements : a) to eliminate cooling fans from desk top to reduce system cost and office sound noise, and to improve reliability, b) to reduce a package cost for cooling a chip. For example, chip cooling by using expensive ceramic package with cooling fins results in an increase in the cost compared to plastic package without cooling fins, as shown in Fig. 1-3(a), c) to reduce the cost of power supply for heat removal by air conditioner. The system design point is for maximum performance for a given power level, but the power level are much higher compared to that of the above mentioned battery operated devices of 1) and 2) as shown in Fig.1-1.

Figures 1-3(a) and 1-3(b) show that why power consumption has become an increasingly important cost factor in terms of battery, chip packaging and chip cooling. Low power contributes to reduce the cost of battery, packaging, and cooling, resulting from reducing the number of battery in the system as shown in Fig. 1-3(b) and from avoiding to use an expensive ceramic package with cooling fins and/or fans as shown in Fig. 1-3(a).

1-1-2 Battery-Operated Systems

In recent years, it is clear that the most visible driving factor for low power semiconductor devices, in particular, memories has been the remarkable success and growth of the battery-operated devices market, such as portable PCs and PDAs [2]. Strong market growth for portable systems is expected to continue through this decade, while driven by technology enhancements which will make it possible to capture increasing function and performance in small, highly portable, battery-operated systems with long battery life time. Further increased capabilities such as speech recognition and handwriting recognition will enable new applications which go far beyond those available with today's systems. In near future, these battery operated devices should be needed for further increased performance as close to desk top computers as possible.

(a)

Assuming: Battery Cost and Weight ∝ Number of Battery

(b)
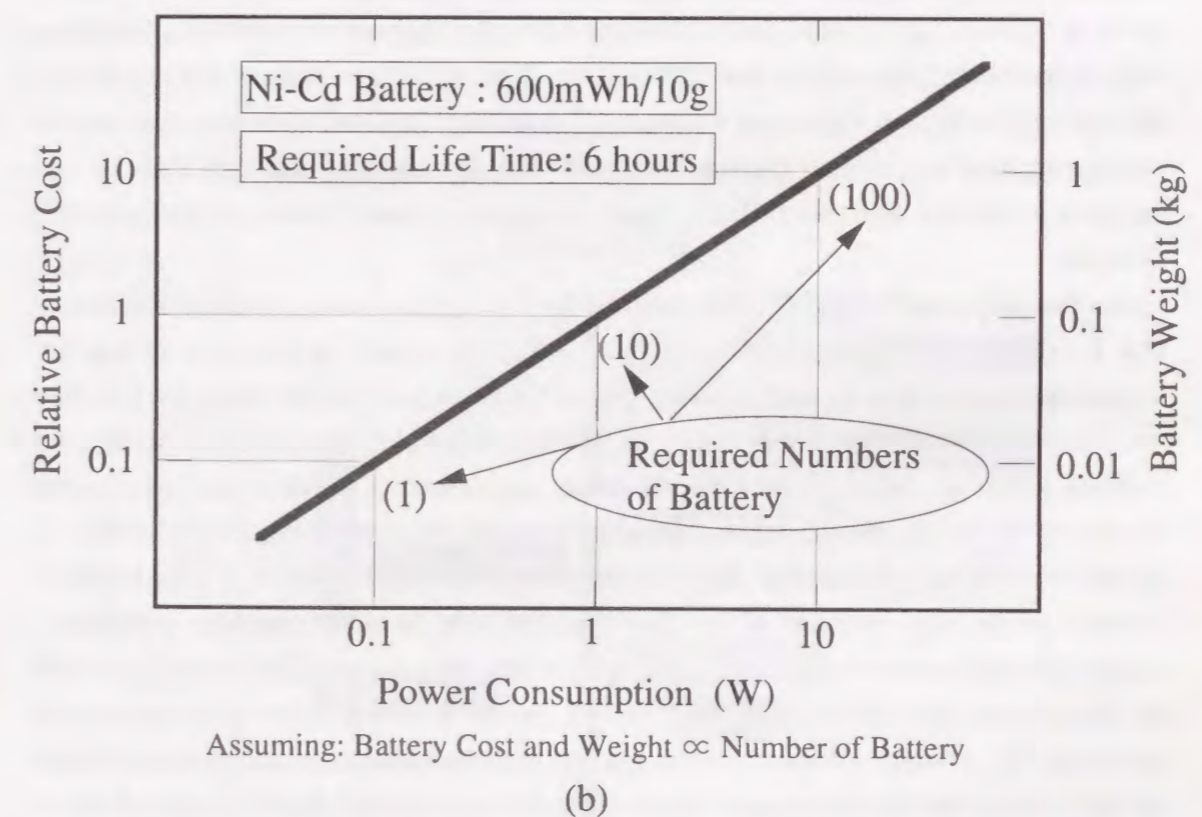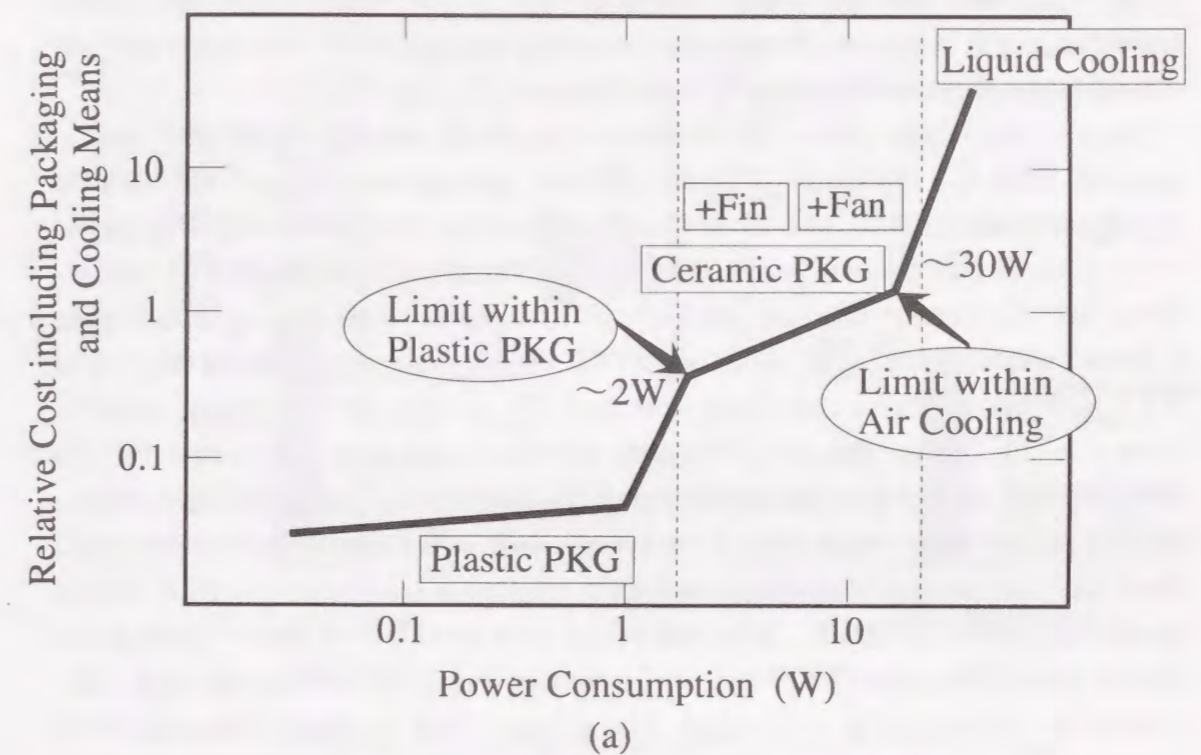
Fig.1-3 Relative cost versus power consumption :
(a) cost of packaging means and cooling means,
(b) cost of battery

Thus, it is clear that the further advanced low power technologies and battery technologies will be needed, because that portability and battery life have been, and will continue to be, key considerations for portable users.

Here, to clarify the reason for the above, the simple example of current battery-operated PCs is explained. Current portable computers with a 100-132MHz microprocessors, a HDD, and a color active matrix thin film transistor (TFT) liquid-crystal (LC) display typically have a 25Wh Ni-Cd (Nickel-Cadmium) or Ni-MH (Nickel-Metal-Hydride) battery and dissipate about 10W of power when running at full speed with hard disk spinning (at idle) and the TFT LC display on, thereby yielding about only 2-2.5 h of battery life in continuous operation [2]. A representative power budget is shown in Fig. 1-4. The whole system power is not only determined by the semiconductor electronics including microprocessors, memories, and other logic circuits, but also determined by other parts of the system, such as the display and the hard disk. Here, "full-on" refers to operation at maximum speed with LC display and HDD on and power management disabled. Note that in this case about 5W of power (50% of the total) is consumed in the CMOS logic and memories including main memory and video memory as shown in Fig.1-4(a) [2]. It is expected that the power consumption of memory portion continues to increase, resulting from increasing memory requirements in terms of capacity and access speed, when considering a support of multimedia functions such as hardware video accelerator, MPEG, etc. This is because storage and processing of moving picture and video and software of more sophisticated operating system and various applications, should require even more memory capacity and data transfer rate between memories and processors, resulting in more power hungry in portable PC systems.

Another important aspect of implementing low power in battery operated devices is that it is a power management technique, which are used to conserve power by monitoring system activity and reducing power levels in parts of the system when they are inactive, either by slowing or stopping the clock to the block in question or powering it down. For example, in the suspend mode, the system is inactive and takes some number of cycles to become active, but when resumed the activity commences where it left off when it was suspended. An example would be a system which is maintained in retention mode, with the state of the system stored in it. In many notebook computers, suspend mode is entered when the LC display panel lid is closed. The system turns off the processors, disk drives, and standard I/O, greatly reducing power consumption as shown in Fig. 1-4(b). When the lid is opened, full operation is resumed where it was left off. Power savings in this case can be 90% or more, enough to allow many days or weeks of battery life. The battery life time in the suspend mode tends to depend on the data retention power of dynamic random access memory (DRAM). This is because that
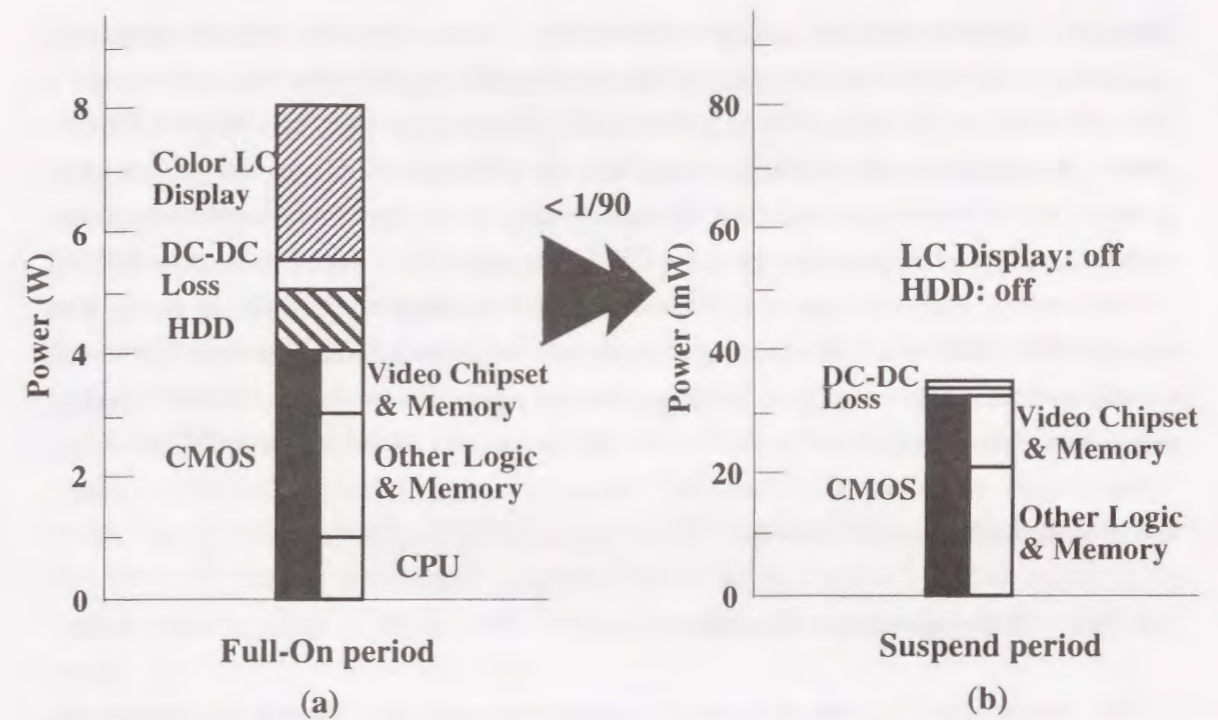
Fig.1-4 Typical power budget for a notebook computer in (a) full-on period and (b) suspend period
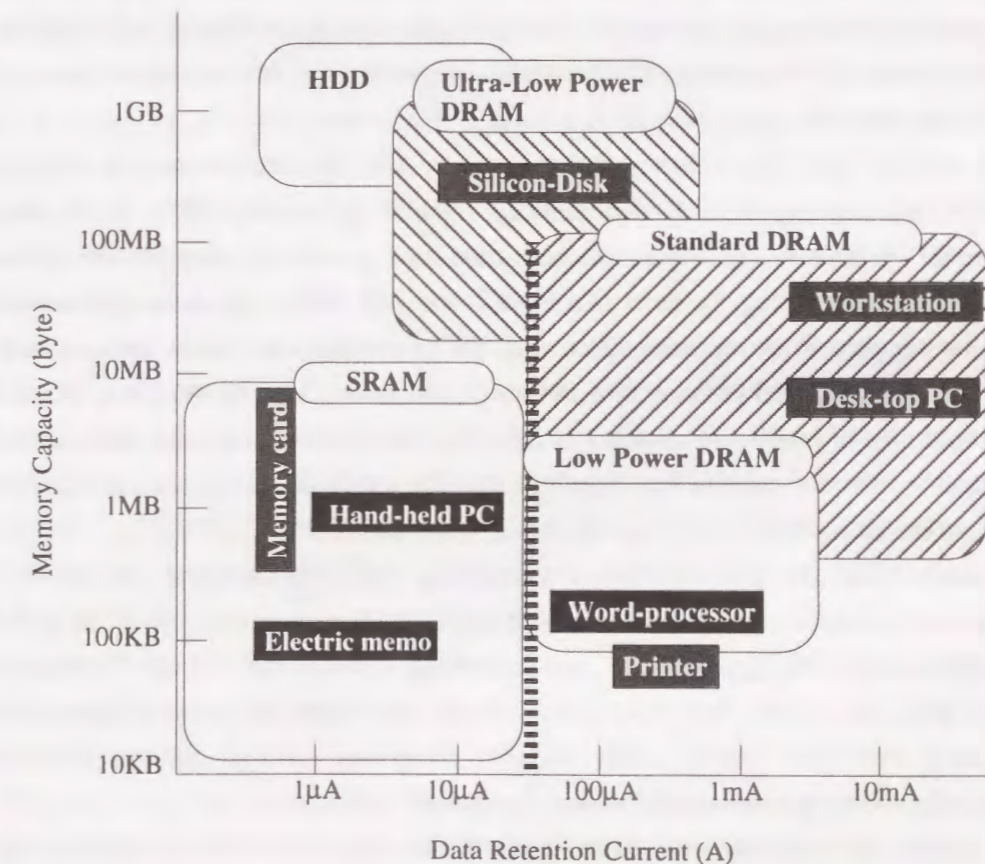
Fig.1-5 Comparison of requirements for data retention current between applications

DRAM is volatile, which do need to be refreshed. Figure 1-5 shows the comparison of requirements for data retention current between system applications for static random access memories (SRAMs), DRAMs and HDD. Since the refresh operation of DRAM should be performed by reading data of the all cells in order, resulting in an extra dynamic power consumption during the battery back-up period in the portable computer systems. If the data retention power of DRAMs could be reduced as low as SRAM (shown in Fig. 1-5) so as to extend the battery life time up to 10 years or more, it is expected that HDD would be no longer necessary in the notebook PCs, resulting in no more power and area hungry of HDD mechanical parts. This is because DRAM would play a role of high density and pseudo non volatile memory as substitute for HDD.

1-2 Semiconductor Memories

1-2-1 Retrospect and Prospect

The 1980s saw the dawn of the computer revolution. During this time, the workstation and personal computer (PC) pervaded our lives, i.e. office and home. The first awkward workstations and PCs grew into intelligent microcomputer systems on tiny silicon chips, rivaling the room size mainframe computers of the 1960s. All of these technological advances depended on the ability to store and retrieve massive amounts of data quickly and inexpensively. They all thus depended on the development of the semiconductor memories[4].

In the late 1980s, the following requirements for the semiconductor memories promised to increase significantly the memory value in the average PCs, workstations and consumer systems in the future: 1) large amount of system memory for the software such as computer operating systems like UNIX and MS DOS, which world standards began to be adopted in these areas, 2) frame buffer memory for video applications in both the engineering workstations and the consumer television. In addition, since the level of capability of intelligent systems is related to the memory capacity, the increased semiconductor memory content has been put into the range of an increasing number of systems applications.

Nowadays, in the mid 1990s, the capacity and performance demands for semiconductor memories are still driven by the system main memory for the larger and more sophisticated PC applications and operating environments like Windows95, Windows NT, etc., while they have been already truly ubiquitous, being present in almost every electronic system, such portable computers, mobile telecommunication systems, large memory smart cards, consumer games, and so on.

In the future, the semiconductor memories will be continuously driven by the various

PCs, all of which feature in more sophisticated operating environments, more human friendly user interfaces, longer battery operation, and more comfortable portability. Almost of those can support muti-media processing and communications, including 3D graphics processing for games and virtual reality, and interactive communications such as internet accessing and video conference, even for the portable PCs.

However that seemingly depends on the progress with respect to increasing density and saving power consumption in the semiconductor memories. This is because the following reasons: 1) Computer interfaces should be more human friendly with speech recognition and vision processing and character recognition, and communications interfaces, all of which require large amount of memory. 2) Higher density and lower power consumption memory will be needed and contributes to make the computer human friendly with more portability and longer battery life. 3) Smart card with multi-megabytes of further power saving memory will be the floppy disk replacement in battery operated palmtop and handheld computer systems- removal of mechanical driving units.

In particular, the above requirements will be more crucial for the portable PCs.

Thus, the ideal memory would be high density, non-volatile, low cost, with high speed random access and with low power dissipation. Those memory technologies which did not offer these advantages to some extent, were one by one successfully challenged by the MOS memories[4]. Unfortunately a single memory having all these characteristics are held by one or another of the MOS memories. For instance low power data retention as low as non-volatile memory is required even for dynamic random access memory, which features high density, low cost, and high speed read-write random access, but "volatile" - refresh operation is needed for data retention. The MOS memories fall into the following two broad categories[4]:

(1) Random read-write access memories (RAM);

Dynamic random access memory (DRAM), static random access memory (SRAM), allow the user both to read information from the memory and to write new information into the memory while it is still in the system,

(2) Read-Only memories (ROM);

ROMs, EPROMs, EEPROMs, are used primarily to store data: however, the EEPROMs can also be written into a limited number of times while in the system and it takes quite long time (longer by 2~3 orders of magnitude) to do that comparing to DRAM or SRAM. ROMs are non-volatile, that is, they retain their memory if the power is turned off whereas RAMs do not.

This thesis focuses on the low power and low voltage technologies for DRAMs and SRAMs. This is because DRAMs and SRAMs have become the dominant MOS memory

devices by taking advantage of higher speed read-write random access capability while keeping low cost comparing to other semiconductor memories, and these clearly will continue to do that in the future since from the usage point of view in the electronics devices with higher volume market, such as PCs.

1-2-2 Basics of DRAM and SRAM

A DRAM is a MOS memory which stores a bit of information as a charge on a capacitor. Since this charge decays away in a finite length of time (milliseconds), a periodic refresh is needed to restore the charge so that the DRAM retains its "memory". There are many advantages with DRAMs. The basic memory cell, which consists of a single transistor and a capacitor as shown in Fig.1-6(a), is small and a very dense array can be made using these cells. The major cost of a semiconductor memory is usually in the cost of the silicon wafer, thus the more chips on a wafer, the lower the cost of a single chip. DRAMs therefore have a lower cost per bit than memories with less compact arrays. DRAMs are also fast for a system to access, giving them a high performance rating.

The disadvantage of a DRAM is that it is volatile. The memory cells do need to be refreshed. They normally need additional circuitry to refresh the memory cells. However, to overcome this, low power DRAMs have the refresh control circuitry on chip. They have the important advantages of very low power consumption and entirely autonomous refresh. Consequently, when the system is idle, the memory controller does not have to periodically initiate refresh cycles - it can go to sleep with the rest of the system.

Historically, DRAMs have tended to be used in the main computer system memory and display frame memory for computer systems, such as PCs and workstations[4][5] as shown in Fig.1-7.

A SRAM cell consists of basic bistable flip-flop circuit which needs only a dc power applied to retain its memory as shown in Fig.1-6(b). It contains four transistors plus either two transistors as pull-up devices. The data, which is defined as a logic "1" or "0" is stored in either of pair of storage nodes (A and B) in flip-flop circuit.

No periodic refresh is required. This eliminates the need for external refresh circuitry as used in DRAMs. This lack of need for external support circuitry and consequent ease of use is a major advantage. Another main advantages of CMOS SRAMs are as follows: 1) their very low power standby characteristics which are used in battery back-up of large memory systems, 2) high speed read/write access capability. The only disadvantage remaining compared to DRAMs is that of size and as a result, of cost.

Historically, SRAMs have tended to be used in the on-chip primary (L1) cache and
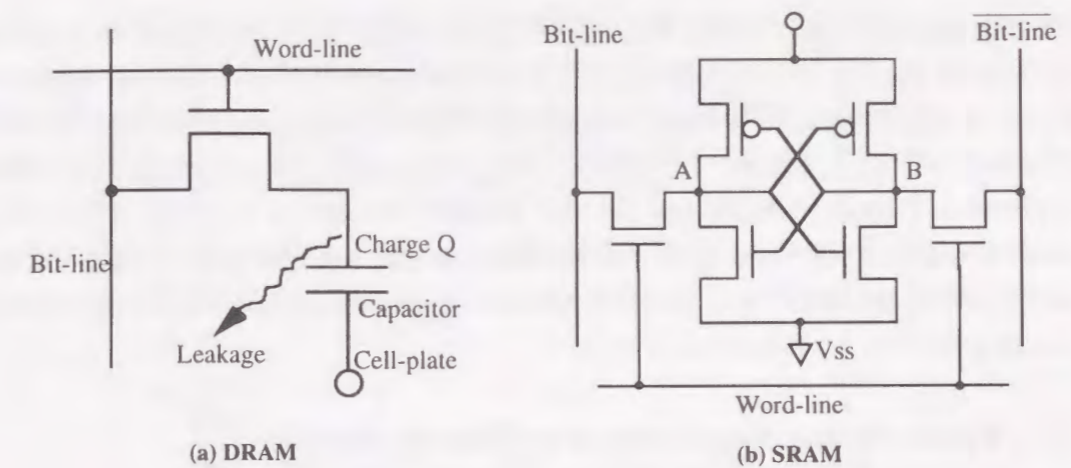
(a) DRAM   (b) SRAM

Fig. 1-6 Memory cell schematics of (a)DRAM (b) SRAM.
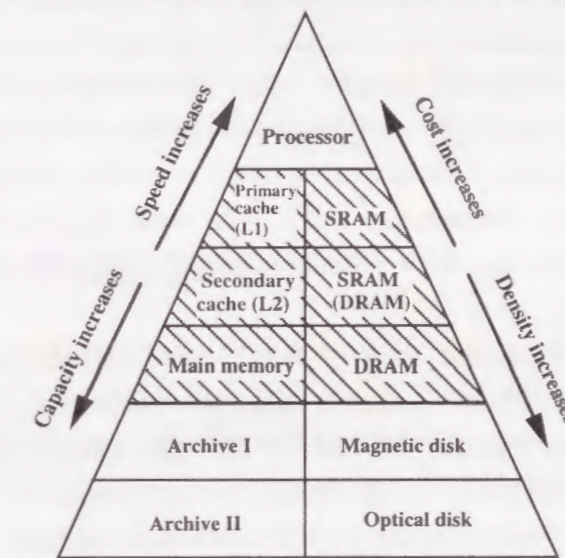
Fig. 1-7 Typical computer system data storage hierarchy (source [4]).
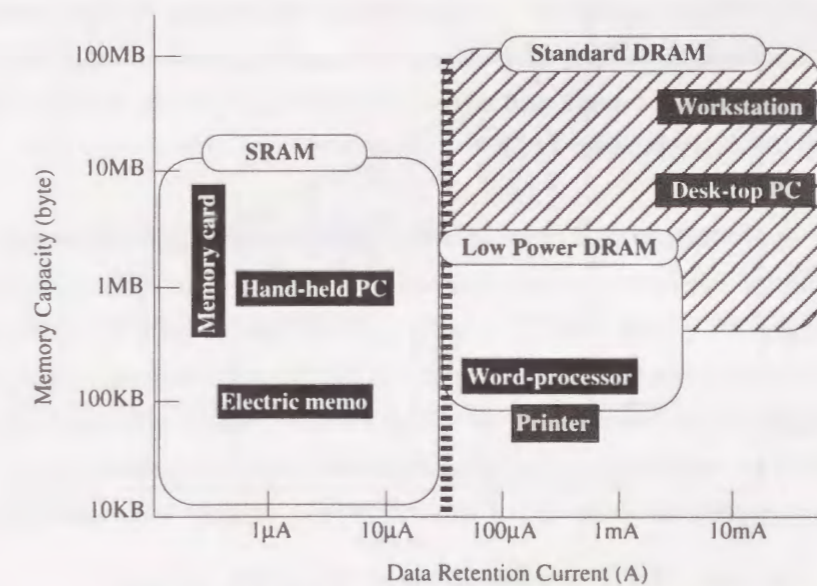
Fig. 1-8   Market comparison between DRAM and SRAM.

off-chip secondary (L2) cache, both of which need high speed random access capability but smaller capacity as shown in Fig. 1-7, and to be used for battery back up applications by taking advantage of very low power standby characteristics due to no need for refresh operation [4][5]. Figure 1-8 shows the market comparison of battery-operated applications between DRAM and SRAM. SRAMs are used in systems, which require small memory capacity and small data retention current, such as memory card and hand-held PCs. On the other hand, DRAMs are used in systems, which require large memory capacity, such as workstation and PCs.

1-3 Power Saving Requirements in Memory Systems

1-3-1 In Realizing Ultra-high Data Transfer Rate

Multimedia and main memory systems require exponentially increasing bandwidth to keep up with rising processor frequencies and demanding user applications as shown in Figs. 1-9 and 1-10. User applications are moving to include sophisticated features such as image compression/decompression (MPEG-2), real-time speech recognition and real time 3D graphics processing, high resolution images, resulting in systems compete for memory bandwidth [5].

To meet these requirements, most of the research and development efforts in the memory system have been oriented towards increasing the clock frequency synchronizing the data transfer [6][7][8][9] and the number of parallel data buses [10][11][12]. However, this has resulted in power hungry just like microprocessor shown in Fig. 1-11. Power consumption of individual memory components has been reaching the power limits of what can be dealt with by the following conditions:

1) economic packaging technologies, resulting in poor cooling capability causing the reduced memory devices reliability - exponentially decreasing of data retention time caused by the increased junction leakage with the increased junction temperature,

2) battery with adequate weight and size for portable equipment, and long battery life (operating life) to satisfy the user [13].

Since from the production cost point of view, such challenge for meeting the increased demand for higher memory system bandwidth should be overcome within system resource constrains (e.g. low cost packaging such as plastic package, no cooling fins and fan, batteries with adequate weight and size for potable devices), the low power data transfer technologies enabling an ultra-high memory bandwidth has been becoming more prerequisite for realizing sophisticated multimedia and main memory systems.

Thus, to try to overcome this, one of this study will be dealt with such challenge and
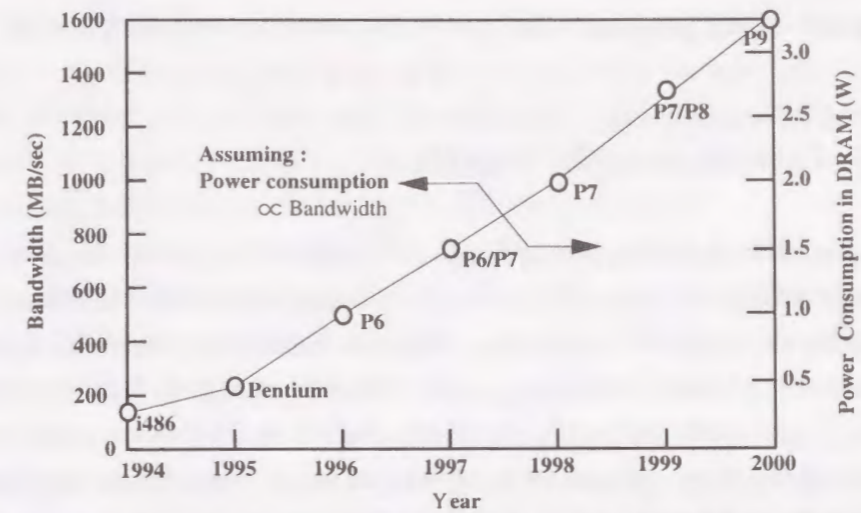
Fig. 1-9 Trends of processor bandwidth and DRAM power consumption
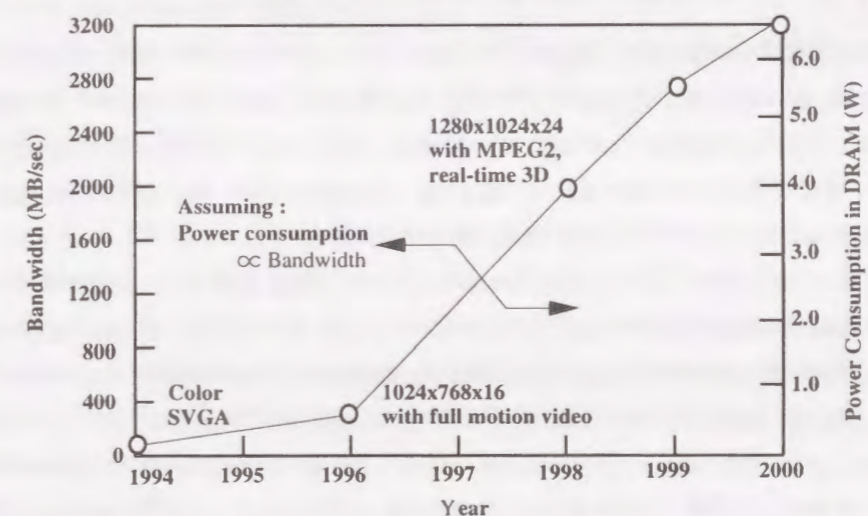
Fig. 1-10 Trends of graphics/multimedia subsystem bandwidth-requirements and DRAM power consumption
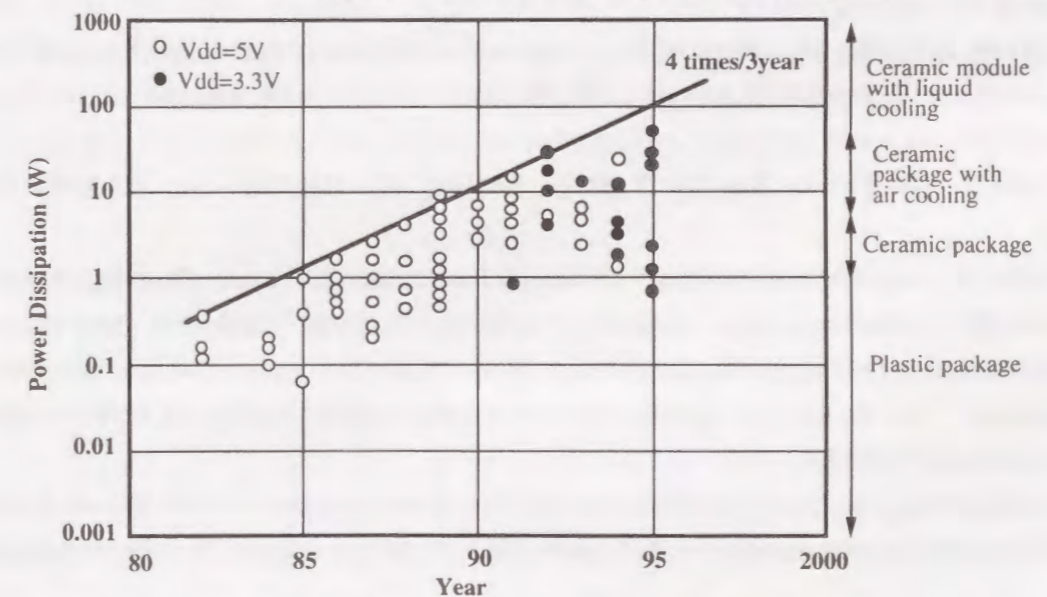
Fig. 1-11 Trend for power dissipation of MPU and DSP (source ISSCC, CICC)

will be focused on the proposed low power data transfer technologies for memory systems.

1-3-2   In Data Retention for DRAMs

DRAMs have been the most prerequisite semiconductor memories for the computer systems, because from the view point of cost advantage due to the high density, read-write random high-speed access capability.   However, the only remaining disadvantage of DRAMs is that the memory cells are volatile, which do need to be refreshed. Since the refresh operation should be performed by reading data of the all cells on a selected word-line and restoring them for each of all word lines in order, that unfortunately causes an extra power consumption during the battery back-up period in the portable computer systems [14].

On the other hand, although SRAMs no longer requires the refresh operation, which can realize low power data retention during the battery back-up period as shown in Fig. 1-12, the significant disadvantage of density, and as a result, of size in SRAMs seemingly never allowable for the consumer devices such as portable computers, because from the view points of cost and portability.

Thus, the demands for developing the low power data retention DRAM have been increasing to meet such requirements for semiconductor memories, - i.e., large capacity, but low cost, while keeping the capability of the read-write random high speed access.

Such low power DRAM has potential ability to cut into the magnetic disk market, which requires low cost and non-volatility, which means low power data retention [4]. Furthermore, since the DRAM no longer requires mechanical driving units for memory access unlike the magnetic disk, such DRAMs could provide more low power and high speed memory systems.

Thus, to realize this, some of this study will be dealt with such challenge and will be focused on the proposed low power DRAM data retention technologies.

1-3-3   In Power Supply Voltage Scaling for DRAMs and SRAMs

The 30 year world of standard 5V design has shifted into a mixed voltage landscape with 3.3V, lower voltages becoming mainstream design standards, and down to accommodated voltage to single battery cell as shown in Fig. 1-13.   This is seemingly caused by the following demands for the power supply voltage to LSIs, including DRAM and SRAMs:

1) Avoiding the power crisis which will need more expensive cooling equipment's (fins and fans) and non plastic package on the system boards (usually ceramic package is used, but it tends to require bigger footprint and more cost). With the progressive improvements in lithography and process techniques, more memory and logic circuits can be squeezed onto single chips, and operating speed with higher clock frequency can be increased. As a result, such chips, when operated at full speed, will consume as much as 1W for 16Mbit DRAM or 35W for high-end micro processors.

2) Extending the battery life, while keeping the adequate battery size and weight for portable and hand held systems.
Even though LSIs in portable and handheld computers have not yet required power consumption of 35W or more unlike that in high-end PCs, more power consumption will be needed in near future, to realize more human friendly interface and processing by using voice and visual recognition and 3D graphics processing. Thus, since from the battery life point of view, power saving for LSIs in portable and handheld computers has been becoming increasingly important.

Regarding the above requirements of 1) and 2), since the dominant factor of power dissipation for LSIs including DRAM and SRAM is proportional to the square of the supply voltage, operating circuits at the lowest voltage while keeping the required operating speed is the key to minimizing the power consumption in the LSIs.

3) Reduction of the number of battery connected in series so as to reduce the size and weight of the battery, while keeping the same current supply capability. For instance, to supply the voltage of 1.8V ~ 2.4V, only two Ni-Cd batteries connected in series is required. It is clear that a sub-1V operation will be needed ultimately, assuming use of the solar-cell of 0.5V. Single battery usage is important from the view points of size and weight of battery systems, besides the lower power consumption. Meeting such demand is thus the key to realize the portable computer systems with lower power consumption, as a result, smaller battery size and weight, and more portability.

4) Keeping the device reliability, while scaling the transistors with shorter gate-channel length, thinner gate-oxide layers, and finer geometry line length. This device scaling is needed to cram more functionality onto silicon chip, which provides lower cost and less power consumption LSIs boards or subsystems, including those of DRAMs and SRAMs.

However, at such lower supply voltages, the individual circuits elements such as driver, receiver, data-path circuits, and memory circuits, run significantly slower. Therefore, to meet such demands, this drawback must be compensated for by all means, for instance through appropriate circuit and architectural designs.

Thus, to realize this, some of this study will be dealt with such challenges in DRAMs and SRAMs and will be focused on the proposed low voltage operated high-speed
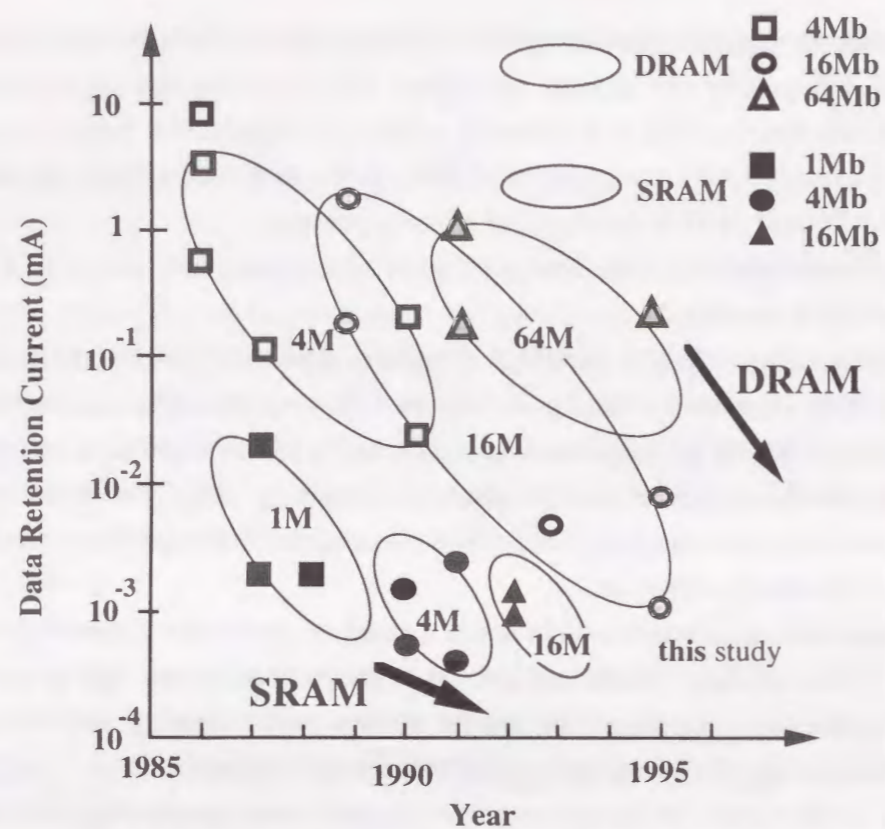
Fig. 1-12 Data retention current trends for DRAMs and SRAMs
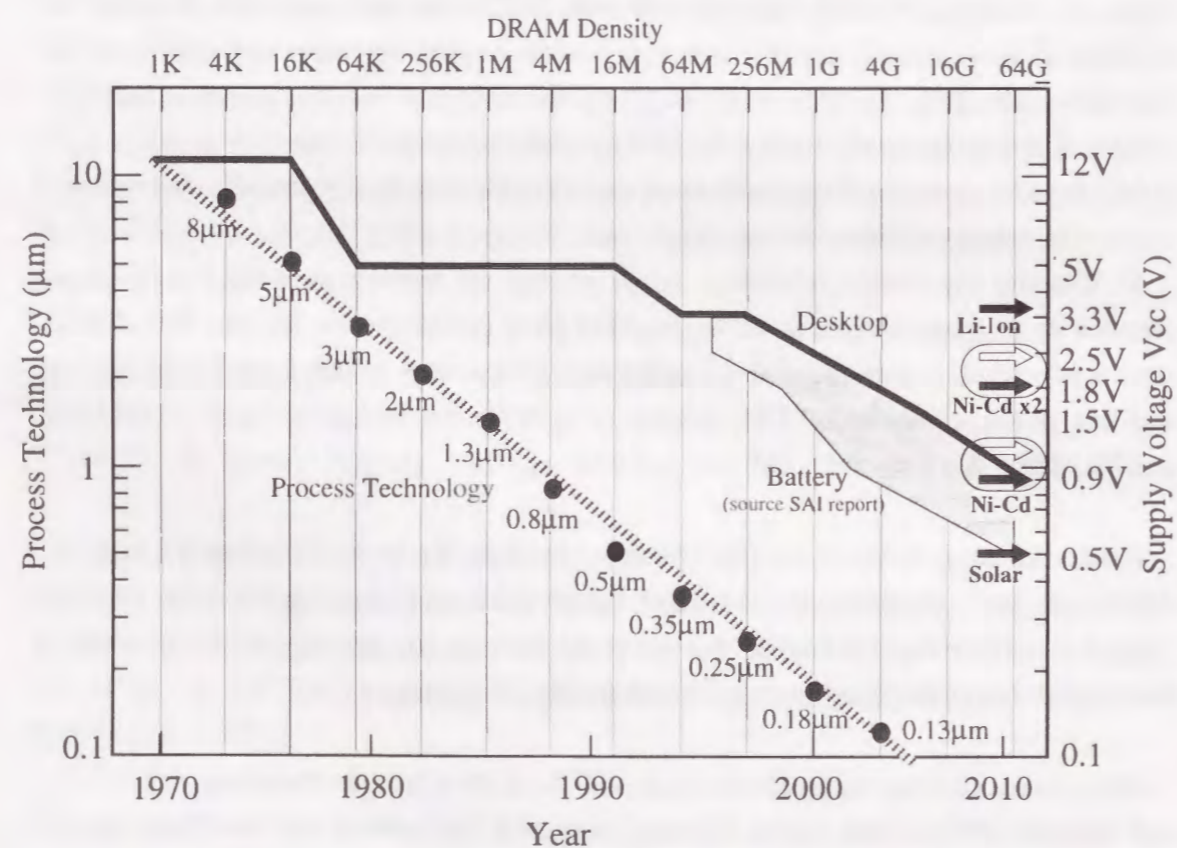
Fig. 1-13 Roadmap of supply voltage versus process technology

DRAMs and SRAMs circuit technologies.

1-4 Technology Trend for Low Power Memory System

To clarify the target position and originality of this thesis, low power technology trends for DRAMs and SRAMs chips and for high bandwidth data transfer schemes between memory and logic circuits, are discussed with respect to the following key issues :

1) Power consumption trend and prospect, while meeting the demands for increasing data transfer rate, are dealt with in the following section.

2) Data retention power consumption trend for DRAMs is discussed in section 1-4-2.

3) Operating voltage scaling trends for DRAMs and SRAMs, are discussed in section 1-4-3.

1-4-1 Power Consumption Trend
versus Data Transfer Rate Trend

Since the data transfer rate (bandwidth) is proportional to both of the clocking frequency of data transfer (F) and the number (N) of bits transferred in parallel, it is clear that increasing the data transfer rate tends to increase straightforwardly the power consumption in the memory systems, because of which is also proportional to both of the operating frequency of data transfer (F) and the number (N) of parallel data buses with significant amount of capacitance (C).

The power consumption (P) is approximately given by

$$P \approx F \times (N \times C) \times Vbus \times Vcc \qquad (1.1)$$

While, the data transfer rate (BW) is expressed as

$$BW = F \times N \qquad (1.2)$$

, where Vbus is the voltage of data bus-swing and Vcc is the power supply voltage.

Equations (1.1) and (1.2) show that the trend of the power consumption of that can be easily got a prospect by taking a view of the trend of required data transfer rate in the memory systems.   Thus, before discussing that, the trends of bandwidth demands for the PC main memory and graphics/multimedia subsystems are discussed.

Two factors are driving up the bandwidth requirements of PCs, one of which is the microprocessor speeds such as Pentium, Pentium Pro (P6), P7 and user applications. Microprocessor units (MPUs) have been moving to higher internal frequencies, wider internal data paths and sophisticated instruction processing techniques resulting in dramatically increased performance.

P6- and P7-class processors include very sophisticated instruction execution units that can process multiple memory transactions in parallel. Using many of the techniques demonstrated in advanced RISC processors, the P6 processor can issue multiple requests to the memory subsystem. At the same time these processors use 64-bit wide internal data paths that are clocked up to 200-300MHz. Thus, the bandwidth to the memory system is crucial point - substantial bottle-neck. As an example, a 64-bit bus operating at 100MHz can demand up to 800MB/s in bandwidth from its memory system as shown in Fig.1-9. Intel has announced that P7 in 1998 will require the 1.6GB/s in bandwidth from its PCs memory system, because a 64-bit bus should be operating at 200MHz.

The second trend is that user applications have moved beyond integer computationally based applications to data streaming multimedia based applications. Applications are taking advantage of high resolution full color displays, 3-D graphics and video based programs. As an example, 1280x1024x24 bits per pixel displays with MPEG-2 data decompression and real time 3-D graphics can demand beyond 2GB/s in bandwidth even from PCs memory system as shown in Fig.1-10.

According to trends of the bandwidth demands in PCs memory system, the data transfer rate of 2GB/s and more will be required in the 1998 timeframe. This means that the power consumption resulting from only the data transfer operation, can exceed 1.5W assuming the amount of operating load capacitance per bit (interconnecting the memory cells and memory controllers on a ULSI system chip), is 10pF and at the operating voltage of 3.3V.

Thus when have a view of the design requirements in realizing the future ULSI's for super high-definition (HD) full motion moving pictures and 3-D computer graphics applications in consumer electronics and handheld personal equipment's, such as palmtop PCs and portable multimedia access supporting full-motion digital video, it is evident that the low power technologies for high bandwidth data transfer of over 2GB/s in memory systems, are needed.

Since the power dissipation is direct proportion to the bus-swing voltage as expressed by equation (1.1), suppressing the bus-swing voltage is one of the attractive choice. However, even if suppressing the bus-swing voltage of Vbus from 3.6V down to 0.9V in order to reduce the charging and discharging amount, there still remains a bus-power dissipation of a far above 300mW at Vcc of 3.6V for the bus width of 512bit with bus-
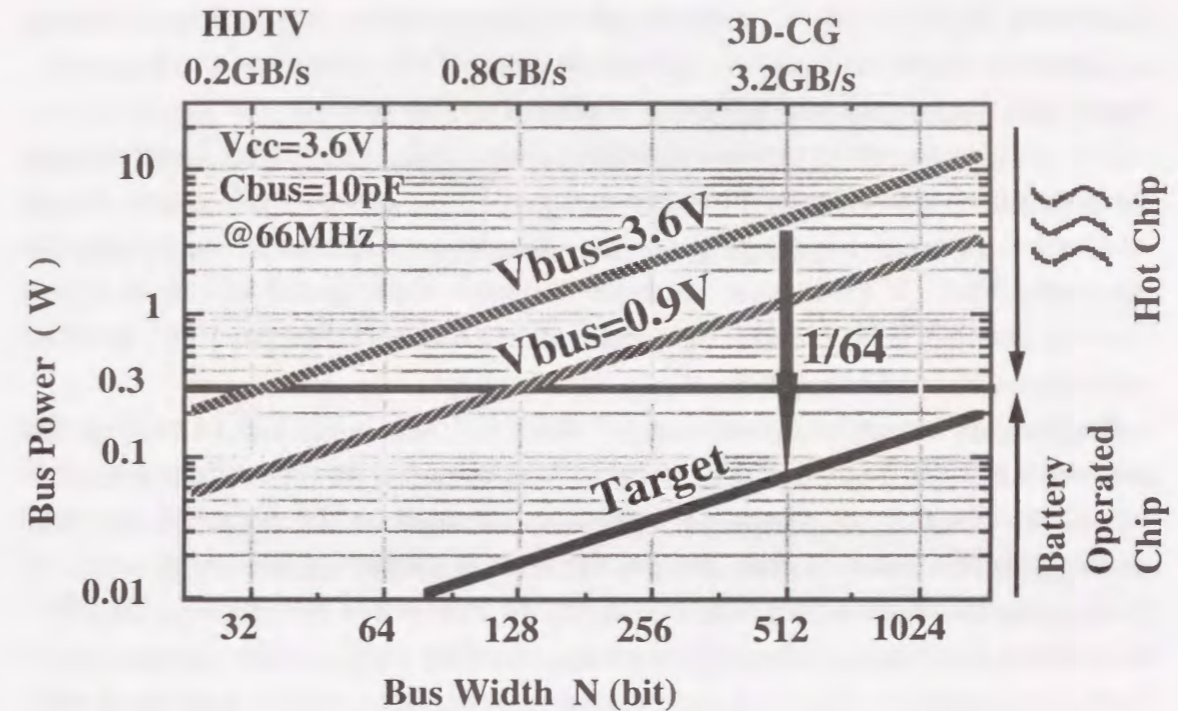
Fig. 1-14   Bus power consumption versus bandwidth
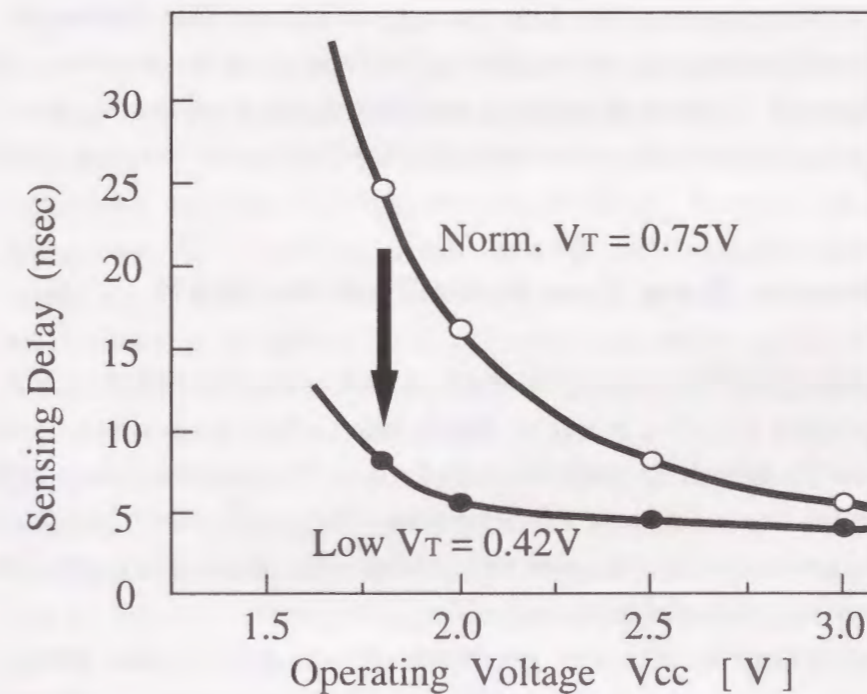
Fig. 1-15 Sensing delay versus operation Vcc
for low V_T and normal V_T.

capacitance of 10pF per bit operating at 66MHz, which is intolerable for battery operation as shown in Fig.1-14. This is because of the restrictions on battery-life, battery size and weight for requirements of adequate portability.

According to the above power estimation, we need the further effort for developing the power saving techniques. From the equation (1.1), we can easily expect that an another low power circuit design technology, besides the technique of suppressing the bus-swing Vbus, is prerequisite to realize the more sophisticated and more human friendly portable PCs, which can deal with true color, real-time, 3D graphics applications, and real time speech and voice recognition.

Breaking the conventional barrier of { F×N×C×Vbus×Vcc } in terms of power consumption is seemingly necessary to reduce continuously a power consumption, while consistently meeting an increasing bandwidth demands. The target of proposed technique in this thesis is thus to open the door to further power saving levels by breaking the above mentioned barrier. This is the main reason why some of this study focuses on the proposed low power charge recycling data transfer scheme, which enables to exploit the power saving benefits resulting from both of bus-swing Vbus suppressing and charge recycling, which enables to reduce the effective bus capacitance (N×C), simultaneously.

This saves the power by a factor of p×q, not p unlike conventional one, where the suppressing ratio of bus-swing voltage Vbus to supply voltage Vcc represents p (Vcc/Vbus = p), and the reducing ratio of the effective bus capacitance resulting from the charge-recycling denotes q. Assuming both of p and q are 8, the power saving factors increases up to 64 (p×q), which means conventional 1W-power results in saving down to merely 16mW.

1-4-2   Data Retention Power Consumption Trend for DRAM

The forthcoming target level for reducing the data retention current of DRAM is as low as the SRAM as shown in Figs.1-8 and 1-12, which inherently consumes less power resulting from no need for refresh operation to retain the data. This target could not only replace the SRAM in the low-end portable PCs market like PDA and handheld computers, but also might make inroads into the magnetic disk market. Thus from such target point of view, the power saving trend of this is discussed.

In the data retention period, DRAM is not accessed from the outside - memory controller, and the data are retained by the periodically refresh operation - destructive read-out characteristics of a DRAM cell necessitate successive operations of amplification and restoration for once read-out cells.

Before discussing the power saving trend of the DRAM data retention, the source of the data retention power consumption is briefly touched.

The refresh operation is performed by a latch-type CMOS sense amplifier on each bit-line connected the memory cells through the word-line controlling access transistor. Consequently, a bit-line is charged and discharged with a large voltage swing of DVbl - usually a half of Vcc, and with charging current of (Cbl×DVbl), where Cbl is the bit-line capacitance. Hence, the current consumption is approximately given by

$$ Icc \approx \{ (m \times Cbl \times DVbl) + (Cph \times Vcc) \} \times (n/Tref) + Idc \qquad (1.3) $$

where Cph is the total capacitance of CMOS logic and driving circuits in periphery, all of which is required for refresh operation. Tref is a refresh cycle time, and Idc is the static (DC) current, which has no dependence on Tref.
m represents the number of bit-lines activated at the same time- in one refresh operation, while n denotes the required number of refresh cycle to spread over all the memory cells.

Extending the refresh cycle time Tref and reducing the DC current Idc are vital to the power saving for the DRAM data retention, because the current consumption is dominated by the following factors, besides the supply voltage Vcc and the bit-line swing DVbl : 1) frequency of the refresh operation- being inversely proportional to the refresh cycle time, 2) DC current in the various supply voltage generators such as negative voltage generator to bias the potential of substrate, boosted voltage generator to drive the word-line, and half-Vcc generator to precharge the bit-lines.

Regarding the extension of the refresh cycle time, various efforts have been intensively made [14][15]. However, it is difficult to set the optimum refresh interval because the data retention characteristics of DRAM memory cells strongly depend on temperature. To solve this issue, memory cell leakage current and temperature monitoring schemes were developed [14]. However, the data retention current achieved was greater than 6μA/MB at 25°C. This value is not enough when considering SRAM data retention current is less than 0.5μA/MB, as shown in Figs.1-8 and 1-12.

Thus, to break the barrier of 0.5μA/MB in terms of data retention current, the following efforts have to be made : 1) extending the retention time by a factor of more than three, and 2) setting the optimum refresh interval based on the DRAM's temperature dependence, resulting in extension of refresh interval by a factor of more than five. This is seemingly vital to open the door to SRAM levels with respect to data retention current saving and to be the SRAM replacement in battery operated palmtop and handheld computer systems.

1-4-3   Operating Voltage Scaling Trend for DRAM and SRAM

Operating voltage scaling demands are driven by the following issues : 1) Power saving, i.e. - the operating voltage scaling is the most attractive choice to do that because of the quadratic influence of that [16], 2) Device reliability, i.e. - further cramming the functionality on a single silicon requires more device size scaling, resulting in shorter gate length and thinner gate oxide, and as a result, increasing an electric field in the device, and 3) Single battery operation, i.e. - the supply voltage from a single battery tends to be basically fixed depending on the type of the cell, i.e., about 2.7V~3.6V for Li-Ion cell, 0.9V~1.2V for Ni-Cd cell, and 0.5V for solar cell.

Regarding the voltage scaling to meet the demands of 1) and 2), nowadays 5V power is giving way to 3.3V and below like 2.5V or less in DRAM and Logic systems as shown in Fig. 1-13.

However, a reduction of supply voltage must be paid for a significant loss in maximum speed Fmax, according to the following well-known scaling rules :

in the case of long channel transistor,
$$Fmax\_long \propto (Vcc - V_T)^2 / Vcc \approx Vcc \qquad (1.4)$$

in the case of sub-μm short channel transistor,
$$Fmax\_short \propto (Vcc - V_T) / Vcc \approx 1 - V_T/Vcc \qquad (1.5)$$

To compensate for such a degradation in speed, the following elaboration have been chosen in this paper : that is the scaling of transistor threshold voltage V_T, i.e., - reducing V_T to accommodate the Vcc reduction, which enables to suppress the degradation of Fmax_short, according to the expression (1.5).

When considering the DRAM circuits, the degradation in speed of the sense amplifier amplification must be the worst case - i.e., being more strongly dependent on the ratio of V_T/Vcc as shown in Fig. 1-15. This is because the gate over voltage (potential difference between gate and source electrodes) in the cross-coupled pair transistors is only a half of Vcc at the most, while that of CMOS logic is full Vcc. To overcome this issue, gate isolated sense amplifier, which enables to reduce the V_T even with longer channel transistor, has been developed. This contributes to realize a higher speed operation and to provide a higher circuit reliability, resulting from using lower V_T, while suppressing asymmetric transistor characteristics such as V_T by using the longer gate channel length. On the other hand, it is difficult for the conventional sense amplifier to lower the V_T, because lowering V_T tends to increase an asymmetric transistor characteristics [17], while using longer gate channel length tends to increase V_T of the transistor [18].

Regarding the SRAM, further Vcc reduction to sub-1V or less, such as 0.5V, has already been ubiquitous as the designer's target [19][20]. However, since V_T has not

$$S = \frac{kT}{q} \ln 10 \left(1 + \frac{C_D}{C_{OX}} \right)$$

k : Boltzmann constant
T : Absolute temperature
C_D: Depletion layer capacitance
C_{OX}: Gate oxide capacitance
$\frac{kT}{q}$ *: ~ 70mV @ 75°C*
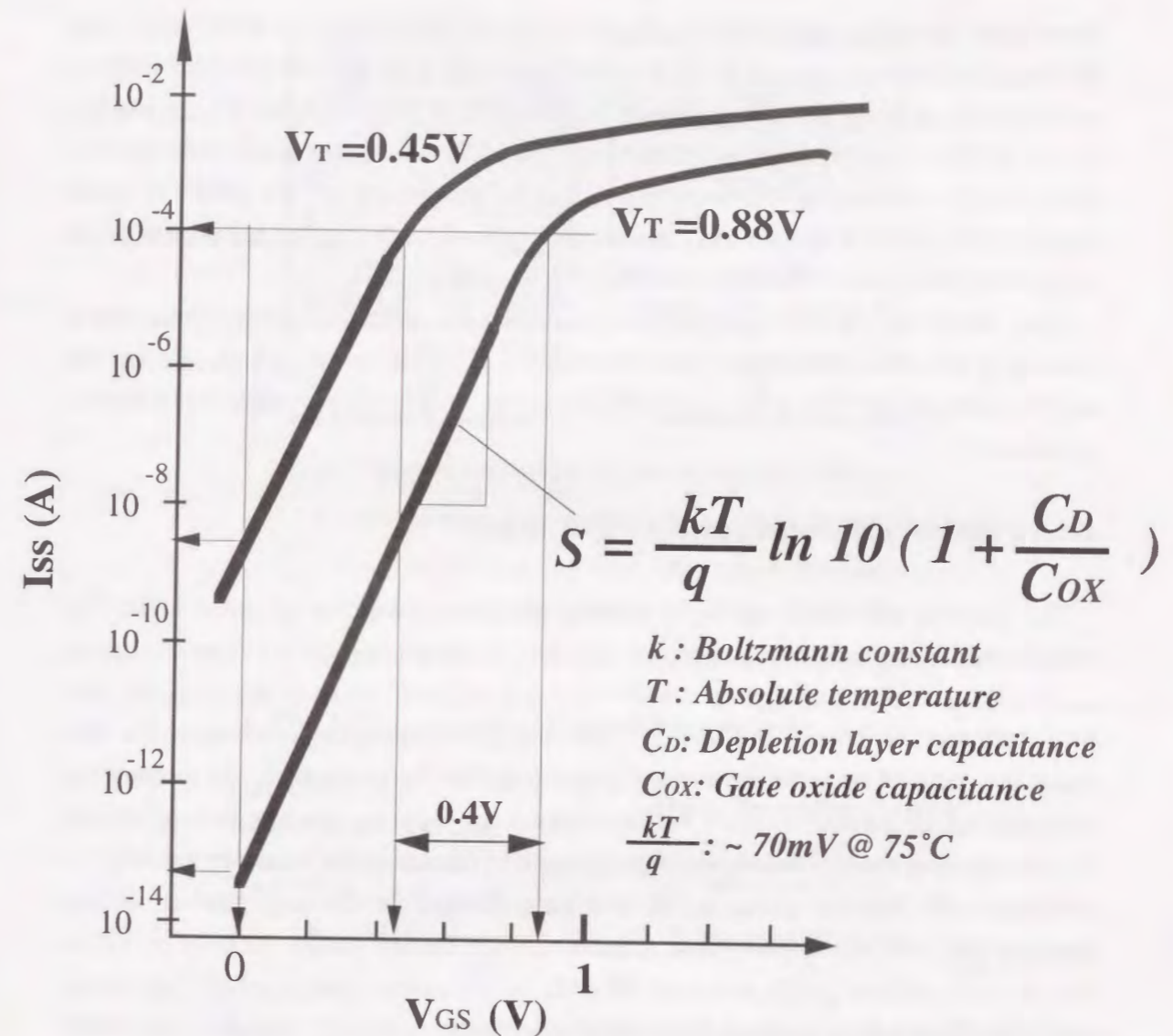
$$Iss = Io \, \frac{W}{Wo} \, 10^{-(V_T/S)}$$

W:Channel width of MOSFET
S: Slope, typ. 70mV ~ 100mV

Fig. 1-16 Relationship between leakage Iss and threshold voltage V_T.

been scaled down to zero or below yet, an intolerable degradation in access speed has not been overcome unfortunately. This is because circuit techniques which cope with the exponentially increase of leakage current (as shown in Fig. 1-16) when V_T approaches to zero or less, have not been developed yet. Since SRAM requires a power supply for data retention unlike non-volatile memory, it is difficult to cut off the power to avoid leakage (from a power source line through ground line). This is still big challenge for circuit engineer to avoid the leakage current, while retaining data.

Thus, one of target for circuit engineer seems that the subthreshold leakage current is reduced to less than 100nA/Mb, even if using V_T of 0V or below, which opening the door to ultimate level of solar-cell (0.5V) in terms of operating voltage for 100MHz operation.

1-5 Purpose and Significance of This Study

The purpose of this thesis is to provide the low power circuit technologies for semiconductor random access memories and their systems, including the power saving techniques in realizing a high bandwidth data transmission between the memories and their controllers as shown in Fig.1-17. The low power technologies discussed in this thesis are focused on what are needed to get over the facing obstacles in meeting the following requirements : 1) never ending demand for lowering power consumption per bit transmission and for increasing data throughput (data transfer rate) per watt (W) in semiconductor memory systems, 2) emerging demand in dynamic random access memory (DRAM) for reducing data retention current during battery back-up period as low as static random access memory (SRAM), which never requires refresh operation, and 3) ever increasing demand for accommodating the operating voltage to the scaled voltage supplied from a single battery - i.e. ~0.9V of Ni-Cd cell and ~0.5V of solar cell. The target positions of the low power technologies presented in this thesis are conceptually shown in Fig.1-18, with comparing to present production levels.

The significance of this study is to give the hint and/or the answer for surmounting the facing obstacles in realizing longer battery-life (less power hungry), more sophisticated, more human friendly, and well-miniaturized (single battery operated) mobile PCs.

To clarify the originality of this study, the still existing and the forthcoming obstacles are also briefly discussed.

Low power DRAMs and SRAMs have been a main technology driver of a low power circuit technology. In particular, successive circuit advancements in DRAMs have produced a power saving equivalent to 2~3 orders of magnitude over the last decade for a fixed memory capacity chip. This has contributed to avoid the power crisis

① — Charge-Recycling Data Bus Architecture discussed in Chapter-2

② — Time-Multiplexed Signal Swing Suppressing Strategy discussed in Chapter-3

③ — Extending Data Retention Time Scheme discussed in Chapter-4

④ — 1.8V/sub-40ns RAS Access DRAM Circuits discussed in Chapter-5

⑤ — Sub-1V/100MHz Operated SRAM Circuits discussed in Chapter-6
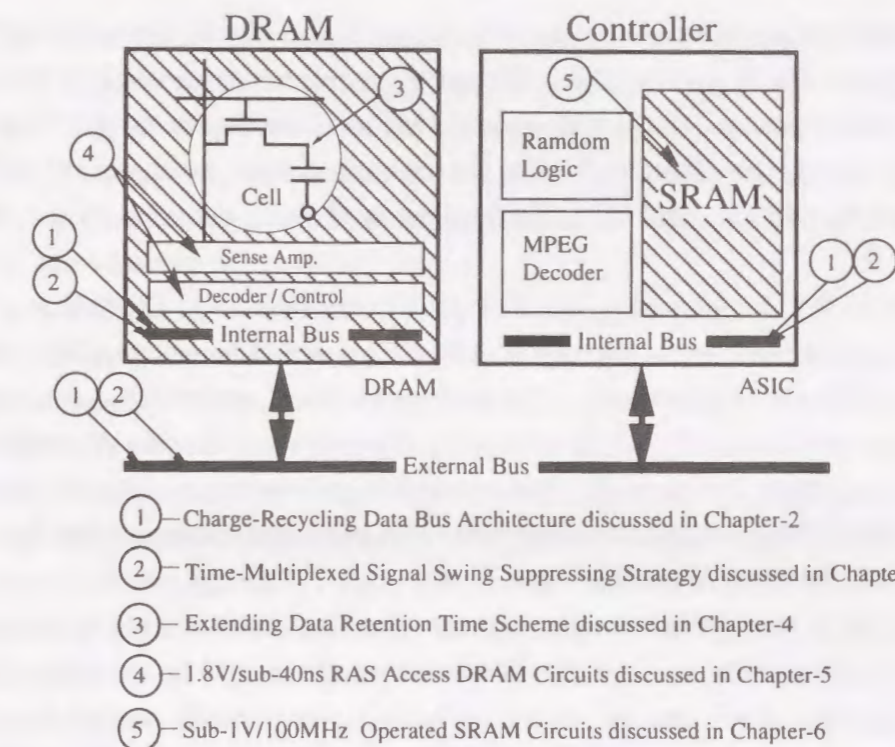
Fig. 1-17 Low power technologies in memory systems.
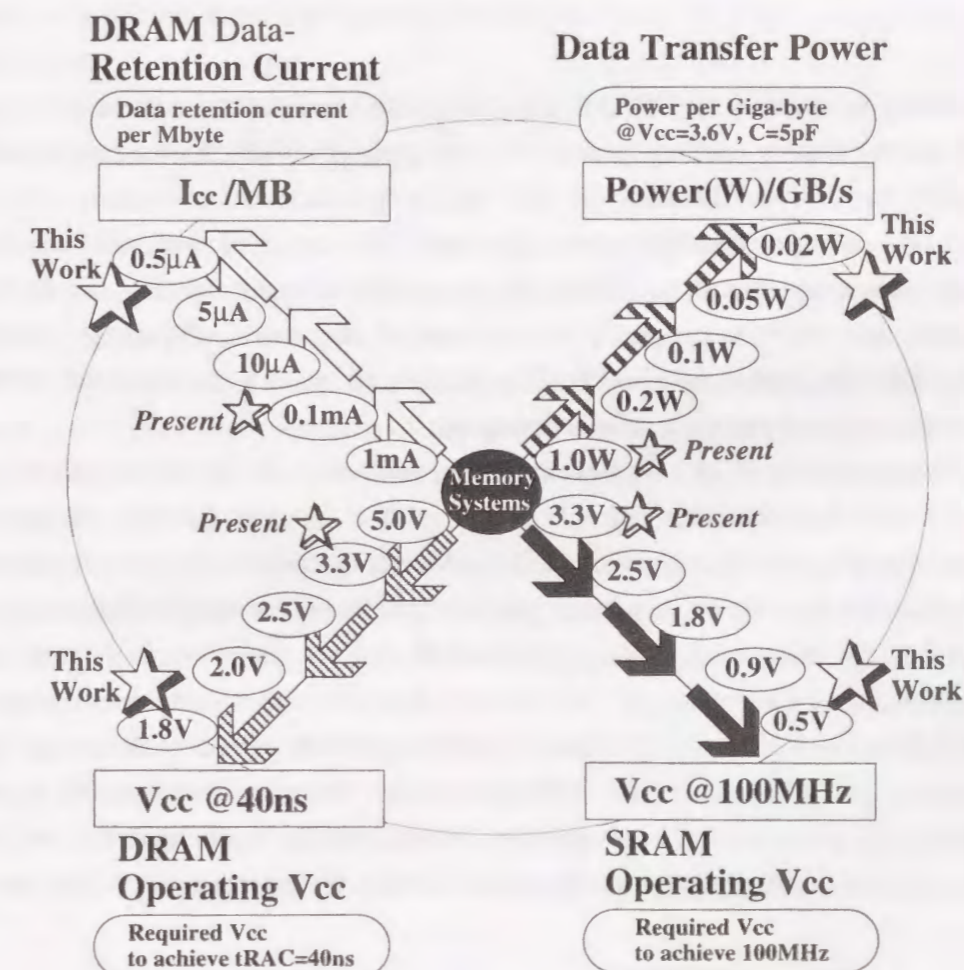
Fig. 1-18 Target position of this work.

mathematically expected by consistently increasing the capacity by four every three years, such as 1Mbit, 4Mbit, and 16Mbit. However, as mentioned earlier in sections 1-3-1 and 1-4-1, the demands for DRAMs and SRAMs has been beginning to orient towards increasing the data transfer rate from/to the memories, such as appearing in Rambus DRAM and SynchLink DRAM, while keeping to increase the capacity to 1Gbit and beyond.

This means that the increasing speed of power consumption in DRAMs and SRAMs should be accelerated such as micro-processors. This would be unacceptable prohibitive scenario of the power consumption since from the view point of the following : 1) DRAM data retention time, which is strongly dependent on the power consumption - junction temperature, 2) battery life time with an adequate battery weight and size, and 3) Cooling fins and fans, which is unacceptable requirements for the portable equipment's from the cost and space of point of view.

To bear up under power hungry requirements, the charge-recycling data transfer scheme, which enables to save the power consumption by the quadratic factor of suppressing ratio m of data bus swing. Assuming m is 8, the power saving factor dramatically increases up to 64 (8-squared), which means conventional 1W-power results in saving down to merely 16mW. This is the original and attractive point of this study.

One emerging demand for DRAMs is to reduce the data retention current up to sub-μA per MB during battery back-up period, the background of which is to replace SRAMs with such low power DRAMs in the mobile communication systems. As more sophisticated software capability from high-end PCs moves downstream overtime to handheld computers, it is expected that larger capacity of main memory, while keeping low power data retention capability will be needed, and the most leading candidate is probably DRAMs instead of SRAMs. The obstacle to attain such target for DRAMs is the necessity of power hungry refresh operation.

To overcome such issue, we developed the following circuit techniques to further reduce DRAM data retention current: 1) the relaxed junction biasing scheme which improves the retention characteristics of DRAM memory-cells resulting from relaxing the junction bias between the storage node and substrate to reduce the junction leakage, and 2) plate floating leakage monitoring timer which can set the optimum refresh interval based on the DRAM's temperature dependence. Such developed techniques contribute to diminish the current consumption down to sub 0.4μA/MB, which is as low as SRAM. Furthermore, the following circuit techniques is also the original and attractive point in this study : 1) gate received level detector, which provide higher gain for the leakage current from or to the potential monitored node such as substrate, and 2) dynamically

controlled reference generator, which cut off the static current resulting from the on and off switching the power supply. Such developed techniques contribute to suppress the DC current to less than 0.1μA/MB, which is negligibly small even when compared to SRAM. By utilizing such techniques, the world's smallest data retention current of 0.5μA/MB has been accomplished by using experimental 16Mbit DRAM. These are also the original and attractive points in this study.

Another emerging demand is to accommodate the operating voltage to the single battery power supply voltage, which should be scaled down to 0.9V of Ni-Cd cell and beyond, like 0.5V of solar cell. This contributes to make the battery-operated devices more human friendly since from the portability and recharging-free for solar cell points of view. However, a reduction of supply voltage must be paid for a significant loss in maximum speed or throughput.

To compensate for this, the V_T scaling have been chosen, while developing the circuit technology enabling to avoid the exponentially increased subthreshold leakage as V_T is scaling, and to suppress the asymmetrical characteristics in the sense amplifier when V_T scaling. By utilizing such techniques, which are the original and attractive points in this study, the world's fastest battery operated 16Mbit DRAM has been developed and the possibility of realizing the 0.5V/100MHz SRAM operation has been verified by using simulated data.

1-6   Constitution of This Paper

The goal of this study is to provide the low power circuit technologies to get over the facing obstacles when once a sophisticated and complex processing for multimedia systems conventionally provided from workstations/high-end PCs, try to move downstream to portable, palm-top, eventually well-miniaturized wrist watch type PCs.

Chapter-1 describes a introduction of this paper in which covers background of this study and technology trend for low power memory, followed by discussing the purpose and significance of this study.

In the Background, the following items are presented : 1) Needs for low power in battery-powered and nonbattery-powered systems are discussed from system cost point of view, in section 1-1-1 and 1-1-2, 2) What is the semiconductor memories ? -what is the usage and demands in electronics systems, in particular, computer systems ?-, is discussed in section 1-2-1, 3) Comparisons of semiconductor memories and basic concepts of dynamic random access memory (DRAM) and static random access memory (SRAM) are briefly given in section 1-2-2, 4) Background for increasing demands for power saving in realizing ultra-high data transfer and in providing the battery back-up

data retention, are discussed in sections 1-3-1 and 1-3-2, respectively, and 5) Background for increasing demands for operating voltage scaling is discussed in section 1-3-3.

To clarify the target position and originality of this thesis, low power technology trends for DRAMs, SRAMs chips and for high bandwidth data transfer schemes between memory and logic circuits, are discussed in terms of three key issues: 1) Power consumption trend and prospect for meeting required data transfer rate trend, is dealt with in section 1-4-1, 2) Data retention power consumption trend for DRAMs is discussed in section 1-4-2, and 3) Operating voltage scaling trends for DRAMs and SRAMs, are discussed in section 1-4-3. Purpose and significance of this thesis is discussed in section 1-5.

Chapter-2 will provide the charge recycling data transfer technology, which features virtual stacking of the individual bus-capacitance into a series configuration between supply voltage and ground. It enables to reduce not only each bus-swing but also a total equivalent bus-capacitance of the ultra-multibit buses running in parallel and as a result, to reduce the power consumption by the factor of the product of the bus-swing suppressing ratio and the number of stacking of the bus. After a brief introduction, the concept of charge recycling data transfer technology is explained in section 2-2. The principle of charge-recycle operation is discussed in section 2-3. In the following section, the circuit configuration of the bus architecture is described. The circuit operation and performance are demonstrated in section 2-5. Bus capacitance imbalances issues and noise issues are dealt with in sections 2-6 and 2-7, followed by conclusions of this chapter.

Chapter-3 will describe the signal-swing suppressing technology using time-multiplexed differential data-transfer scheme, which covers the evolution of the charge-recycle data transfer scheme proposed in Chapter-2.

After the discussion of the background and target in this low power strategies, the concept of the time-multiplexed data transfer (TMD) scheme is described in section 3-3. In section 3-4, the TMD scheme combined with charge-recycling bus architecture, enabling to further reduce the bus power consumption, are demonstrated. Before conclusions of this chapter, the power and area saving comparisons are given.

DRAM data retention power saving technology is discussed in Chapter-4.
In this chapter, a circuit technology to realize a self-refresh 16Mb DRAM with a sub-0.5μA per megabyte data retention current, allowing a 20-megabyte RAM disk to retain data for 2.5 years with a single button-shaped 190mAh lithium battery, is presented. Part

of the circuit design technology incorporates a relaxed junction biasing scheme, which enables to shift the storage node voltage to a lower potential to relax the junction bias, and in turn, to suppress the junction leakage. Details of this is discussed in section 4-2. Leakage monitoring circuit that helps compensate for a speed difference in charge decline between the few short-retention cells and normal retention cells, is described in section 4-3. In section 4-4, the gate-received substrate level detector scheme, which makes possible the avoidance of DC-idling current, is presented. The combined result of these improvements is summarized and the contribution of the proposed schemes to the measured performance is clarified in section 4-5, followed by the conclusions of this chapter in section 4-6.

Circuit technology for high speed battery-operated DRAMs is discussed in Chapter-5. This chapter covers the four circuit techniques, which contributes to realize the 16Mb DRAM with the RAS access time of 20ns at 3.3V and also 36ns at 1.8V. Their values of access speed were the fastest among ever reported 16Mb DRAMs. A parallel column access redundancy scheme coupled with a current sensing address comparator is presented in section 5-2, followed by a quasi-static signal sensing circuits in section 5-3. The gate-isolated sense amplifier with low threshold voltage is described in section 5-4. In section 5-5, suppressing of asymmetrical characteristics in scaled DRAM sense amplifier is dealt with. The chip architecture, device features, and the chip performance are demonstrated in section 5-6, followed by the conclusions of this chapter in section 5-7.

Chapter-6 will describe the circuit technology for high-speed battery-operated SRAMs. This chapter shows the possibility in realizing the sub-1V of (0.5V~0.8V) / 100MHz operated SRAMs based on the simulated data.
A sub-1V operated high-speed SRAM cell strategy including the boosted offset grounded data storage scheme and charge-recycle source over-driving scheme, is discussed in section 6-2. Power consumption comparisons and discussions are given in section 6-3. This is followed by the conclusion of this chapter in section 6-4.

Conclusions including technical prospect are discussed in Chapter-7.

References

[1] Lewis M. Terman, "Scanning the Issue, Special Issue on Low Power Electronics", Proceedings of IEEE, Vol.83, No.4, pp.495-496, April, 1995.

[2] Erik P. Harris, et al, "Technology Direction for Portable Computers", Proceedings of IEEE, Vol.83, No.4, pp.636-658, April, 1995.

[3] Robert A. Powers, "Batteries for Low Power Electronics", Proceedings of IEEE, Vol.83, No.4, pp.687-693, April, 1995.

[4] Betty Prince, Semiconductor Memories- second edition, A Handbook of Design, Manufacture and Application, by John Wiley & Sons Ltd, England, 1991.

[5] Betty Prince, High Performance Memories, New Architecture DRAMs and SRAMs-ecolution and function, by John Wiley & Sons Ltd, England, 1996.

[6] Y. Takai, et al., "250Mbytes/s synchronous DRAM using 3-stage-pipelined architecture, IEEE Journal of Solid State Circuits, Vol.29, No.4, April, 1994.

[7] T. Saeki, et al., "A 2.5ns Clock Access 250MHz 256Mb SDRAM with a Synchronous Mirror Delay", ISSCC Digest of Technical Papers, pp.374-375, Feb.,1996.

[8] N. Kushiyama, et al, "500Mbyte/sec Data Rate 512Kbits x 9 DRAM Using a Novel I/O Interface", Symposium on VLSI Circuits, Digest of Technical Papers, pp.66-67, June, 1992.

[9] S. Przybylski, New DRAM Technologies, 2nd Edition - A Comprehensive Analysis of the New Architectures, by Micro Design Resources, 1996.

[10] S. Miyano, et al, "A 1.6GB/s Data-Transfer-Rate 8Mb Embedded DRAM", ISSCC Digest of Technical Papers, pp.300-301, Feb., 1995.

[11] Y. Aimoto, et al, "A 7.68GIPS 3.84GB/s 1W Parallel Image-Processing RAM Integrating a 16Mb DRAM and 128 Processors", ISSCC Digest of Technical Papers, pp.372-373, Feb., 1996.

[12] Y. Nitta, et al, "A 1.6GB/s Data-Rate 1Gb Synchronous DRAM with Hierarchical Square-Shaped Memory Block and Distributed Bank Architecture", ISSCC Digest of Technical Papers, pp.376-377, Feb., 1996.

[13] L. Geppert, "Solid sate," IEEE Spectrum, pp. 35-39, Jan. 1995.

[14] K. Sato, et al., "A 4-Mb Pseudo SRAM Operating at 2.6±1V with 3-μA Data Retention Current', ISSCC Digest of Technical Papers, pp.268-269, Feb., 1991.

[15] Y. Kagenishi, et al, "Low power self refresh mode DRAM with temperature detecting circuit", Symposium on VLSI Circuits Digest of Technical Papers, pp.43-44, June, 1993.

[16] A. Chandrakasan, Low Power Digital CMOS Design, by Kluwer Academic Publishers, 1995.

[17] H. Yamauchi, et al, "A Circuit Design to Suppress Asymmetrical Charcteristics in High-Density DRAM Sense Amplifiers", IEEE Journal of Solid-State Circuits, Vol. 25, No.1, pp.36-41, February, 1990.

[18] H. Yamauchi, et al, "A Circuit Technology for High-Speed Battery-Operated 16Mb CMOS DRAM's", IEEE Journal of Solid-State Circuits, Vol.28, No.11, November 1993.

[19] H. Mizuno, et al, "Driving Source-Line (DSL) Cell Architecture for Sub-1V High-Speed Low-Power Applications", Technical Digest of Symposium on VLSI circuits, pp. 25-26, June, 1995.

[20] K. Itoh, et al, "A Deep Sub-1V, Single Power-Supply SRAM Cell with Multi-Vt, Boosted Storage Node and Dynamic Load" Technical Digest of Symposium on VLSI circuits, pp.132-133, June, 1996.

CHAPTER-2   Charge Recycling
Data Transfer

Abstract

An asymptotically zero power charge recycling bus (CRB) architecture, featuring virtual stacking of the individual bus-capacitance into a series configuration between supply voltage and ground, has been proposed. This CRB architecture makes it possible to reduce not only each bus-swing but also a total equivalent bus-capacitance of the ultra multi-bit buses running in parallel.   The voltage swing of each bus is given by the recycled charge-supplying from the upper adjacent bus capacitance, instead of the power line. The dramatically power reduction was verified by the simulated and measured data. According to these data, the ultra-high data rate of 25.6Gb/s can be achieved while maintaining the power dissipation to be less than 100mW, which corresponds to less than 10% that of the previously reported 0.9V suppressed bus-swing scheme, at Vcc =3.6V for the bus width of 512bit with the bus-capacitance of 14pF per bit operating at 50MHz.

2-1   Introduction

The ultra-high data rate of 25Gb/s and beyond is one of the most important design requirements in realizing the future ULSI's for super high-definition (HD) moving pictures and three-dimensional(3-D) computer graphics applications in consumer electronics and handheld personal equipment's, such as palmtop PC's and portable multimedia access supporting full-motion digital video. The most effective means to achieve such a data rate is to employ a large number of buses, corresponding to the parallelism of N-bit, interconnecting the embedded memory, the graphics controller, *etc*, on a ULSI system chip, instead of increasing operating frequency as shown in Fig.2-1. For example, even at an operating frequency of 50MHz, parallel buses of more than 512bit are required to achieve the ultra-high data rate of 25Gb/s and beyond as shown in Fig.2-2.   However, a drastic increase of the power dissipation which is direct proportion to the number of bus width N is inevitable due to the increased total bus-capacitance (N • Cbus), where Cbus is the bus-capacitance of each bus. Even if suppressing the bus-swing voltage of Vbus from 3.6V down to 0.9V in order to reduce the charging and discharging amount[1], there still remains a bus-power dissipation of a far above 300mW at Vcc of 3.6V for the bus width of 512bit with bus-capacitance of 14pF per bit operating at 50MHz, which is intolerable for battery operation as shown in
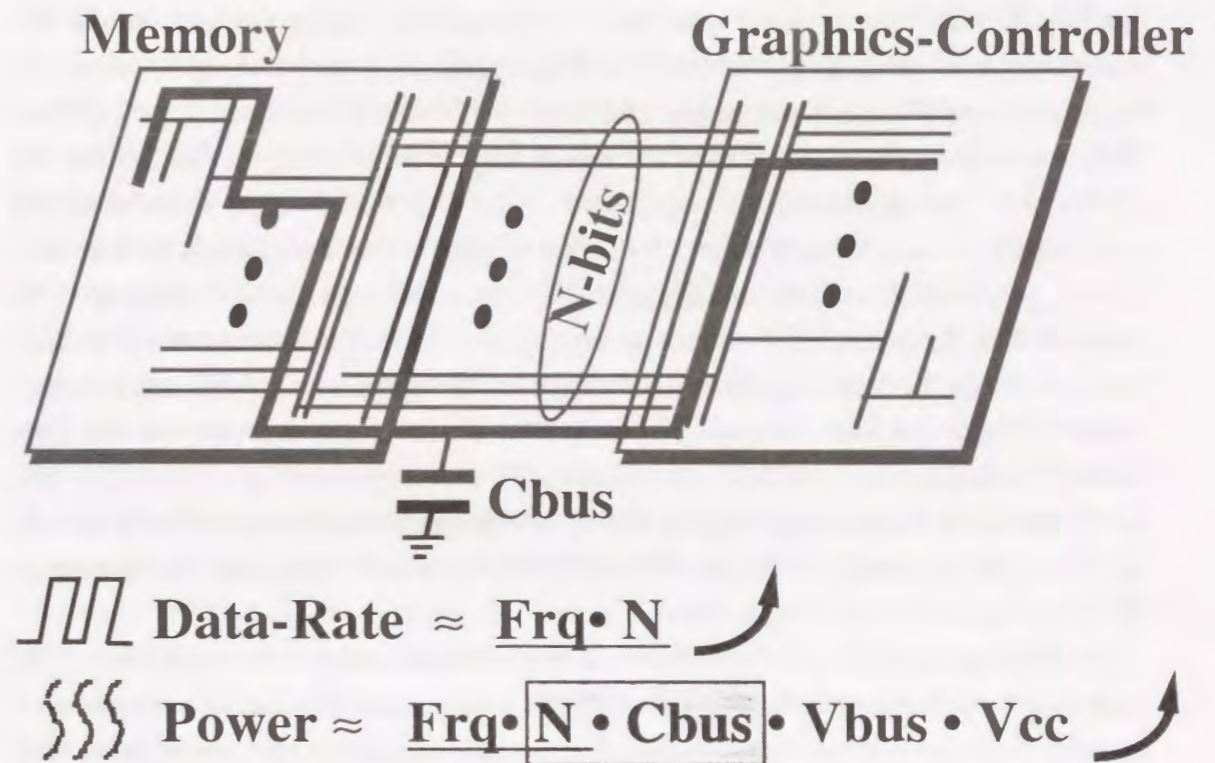
Fig. 2-1. Background for this low-power bus architecture.
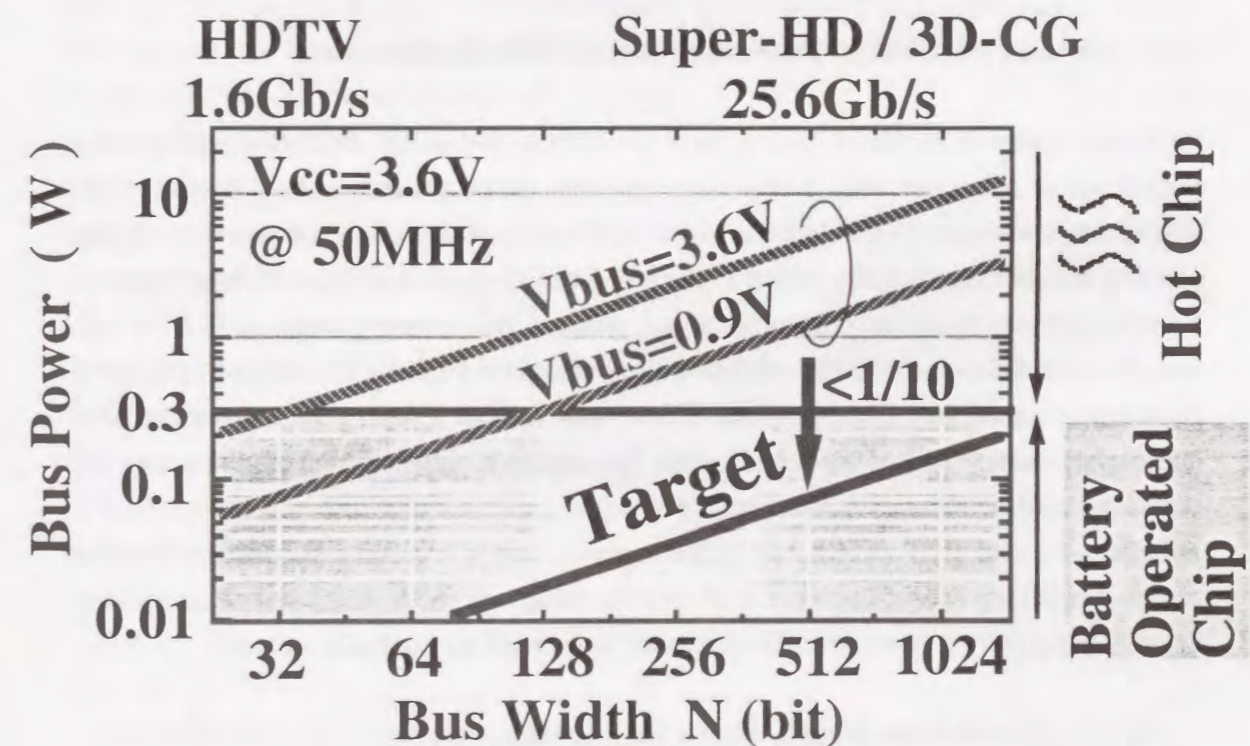
Fig. 2-2.   Target on bus-power consumption for this work.

Fig.2-2. This is because of the restrictions on battery-life, battery size and weight for requirements of adequate portability[2]. This paper proposes a complete Charge-Recycling Bus(CRB) architecture that can reduce the bus-power dissipation to less than 10% of the conventional suppressed bus-swing scheme while realizing the data rate of 25Gb/s[3]. This asymptotically zero power CRB architecture features virtual stacking of the individual bus-capacitance into a series configuration between supply voltage and ground. It enables to reduce not only each bus-swing but also a total equivalent bus-capacitance of the ultra multi-bit buses running in parallel. The voltage swing of each bus is given by the recycled charge-supplying from the upper adjacent bus capacitance, instead of the power line. According to the simulated and measured data of the test chip fabricated by using 0.5μm CMOS technology, the ultra-high data rate of 25.6Gb/s can be accomplished while maintaining the power dissipation to be less than 100mW at Vcc =3.6V for the bus width of 512bit with the bus-capacitance of 14pF per bit operating at 50MHz.

In the following section, the concept of CRB architecture is explained. In Section 2-3, the principle of charge-recycle operation in CRB is discussed. The circuit configuration of CRB is described in Section 2-4. The circuit operation and performance are demonstrated in Section 2-5. Bus capacitance imbalances issues and power-on state and noise issues are discussed in Sections 2-6 and 2-7, respectively. Conclusions are given in Section 2-8.

2-2 Concept of Charge Recycling Bus (CRB) Architecture

As mentioned in Section 2-1, to achieve the ultra-high data rate of 25Gb/s and beyond, parallel buses of more than 512bit are required, even at an operating frequency of 50MHz as shown in Fig.2-2. Thus, in order to reduce the charging and discharging amount, suppressing the bus-swing voltage of Vbus down to less than 1V is necessary. However, even if using conventional suppressed bus-swing (Vbus=0.9V) scheme coupled with a down converter, the bus power dissipation cannot be reduced adequately for battery operation. This is because there still remain a contribution from the total bus-capacitance (N • Cbus), where N and Cbus are the number of total bus pairs and the capacitance of the individual bus, respectively.

Before the charge recycling bus (CRB) architecture is conceptually described, what causes wasteful-power dissipation in the conventional scheme is discussed by comparing the conventional two types of full-swing and suppressed-swing bus schemes.

2-2-1 Conventional Data Transfer Scheme

Schematics of conventional "full-swing" and "suppressed swing" bus scheme are shown in Figs.2-3(a) and 2-3(b), respectively. In these figures, the operating waveforms of the three-parallel complementary bus pairs in the three cycles are illustrated. Each operating waveform in one cycle is divided into two periods. One half of the cycle is for the equalizing and the precharging at the midpoint of the swing voltage of each complementary data bus pair, and another is for the developing of the complementary bus-swing voltage and establishing of the transferred data.

In this paper, unless otherwise noted, the conventional "full-swing" bus scheme shown in Fig.2-3 (a) features the following two bus controls: 1) half-level precharging in the former half of the cycle and 2) developing from the 1/2Vcc to the "High"(Vcc) or "Low"(Vss) level in the latter half of the cycle. This bus control can save the bus power by a half of that compared with a output bus power of static CMOS logic gate undergoing of the "High" to "Low" CMOS level transition.

In Fig.2-3 (b), the two symbols of resistance between Vcc and Vbus and between Vbus and Vss represent the turn-on-resistance of MOSFETs composing the output-driver and the reference voltage generator, respectively in the voltage down-converter. This voltage down-converter including reference voltage generator basically functions as a resistance potential divider suppressing the bus swing voltage from Vcc down to Vbus. However, this scheme induce a large amount of the power-loss within the resistance. For example, that value corresponds to 75% of the total power consumption, when the Vbus is 0.9V for Vcc=3.6V. This is because of the Joule-Heat power loss between the Vcc and the Vbus. In addition, DC idling current to Vss (IDC) also induce a wasteful DC power-loss as shown in Fig.2-3(b).

As mentioned before, the suppressed bus swing voltage of Vbus is a key factor to reduce the bus-power consumption. For example, the power reduction ratio of the suppressed bus scheme to the full-swing bus scheme (Psupp / Pfull) can be reduced with the suppressed bus swing ratio m (Vcc / Vbus) in Fig.2-4. However, the power loss ratio of the Joule-Heat power to the total bus power (Ploss / Ptotal) increases with the m (Vcc / Vbus). For the m=16, corresponding to the Vbus=0.2V for the Vcc=3.2V, the (Ploss / Ptotal) reaches over 90%. This result implies that a new suppressed bus scheme is required for realizing a dramatic bus power reduction to an asymptotically zero compared with the conventional one, without a resistance voltage divider inducing a Joule-Heat power loss.

2-2-2 Charge Recycling Bus (CRB) Architecture

The proposed approach to low-power bus design features the CRB architecture, making it possible to meet both of the following two key bus challenges, simultaneously:
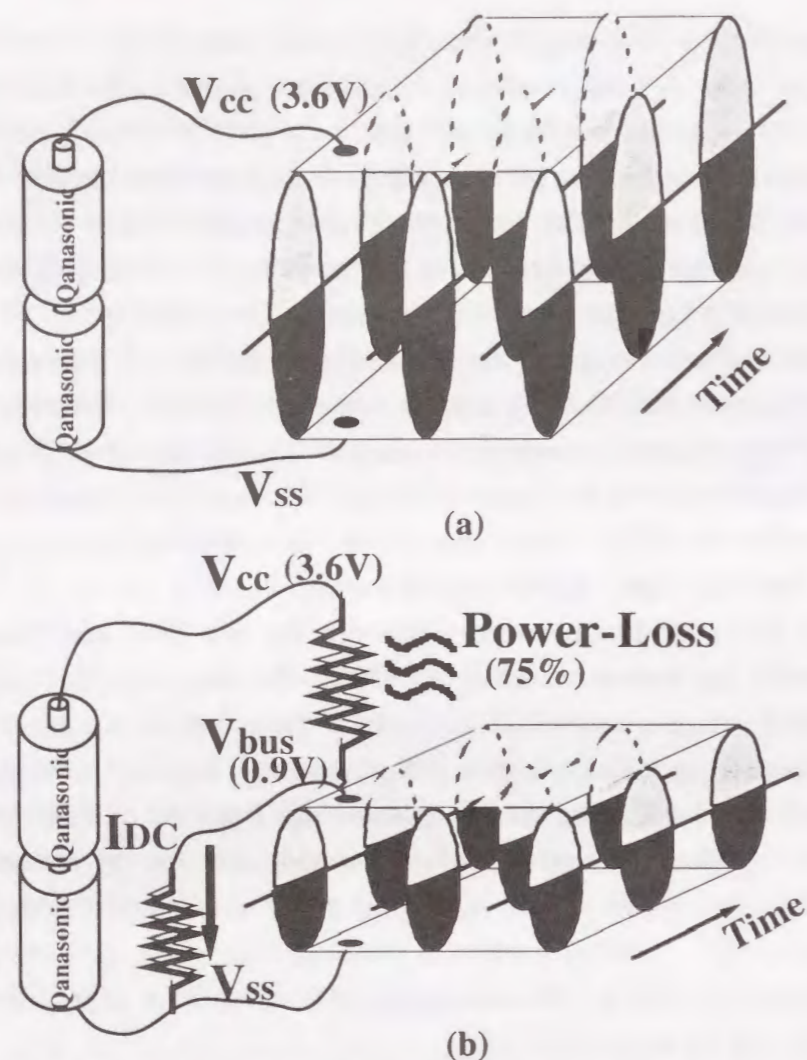
Fig. 2-3. Comparison of conventional bus schemes :
(a) Full-swing bus scheme,
(b) Suppressed-swing bus scheme with on-chip voltage down-converter.
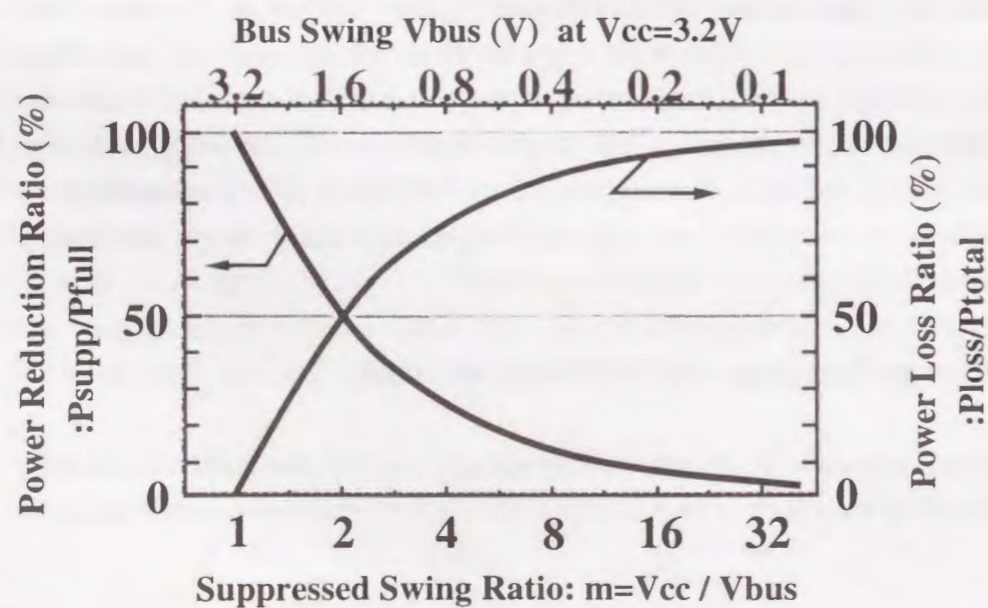
Fig.2-4. Ratios of power reduction and power loss as a function of suppressed-swing ratio (m).

1) the first key point is the suppressed bus swing scheme using the capacitance-voltage-divider, which no longer requires wasteful DC idling current and Joule-Heat power loss, unlike conventional resistance voltage divider, such as voltage down converter. 2) another key point is the charge-recycling[3][4] among the multi-bit buses running in parallel in every cycle, that is, the used charge for establishing of high level in each bus is never dumped to ground every time.

The key concept of the CRB architecture is "virtual stacking" of the individual bus-capacitance Cbus into a series configuration between Vcc and Vss, in order to reduce not only the bus-swing but also the total equivalent bus-capacitance. "Virtual stacking" means the following points: 1) in fact, the individual bus-capacitance is mainly composed of the bus wiring capacitance to substrate and the junction capacitance of the MOSFETs that is connected to the bus. Thus, each bus capacitance Cbus cannot be directly connected in series configuration as shown in Fig.2-5. 2) However, a simple bus control, featuring the stepwise precharging of each bus pair and inter-bus charge sharing among the stepwise precharged adjacent bus capacitance's in the former and the latter half of the clock cycle, respectively, makes it possible to realize the "virtual stacking" of each bus capacitance. The conceptual description is illustrated in Fig.2-8 and the detail discussion about that is made in Section 2-3.

The point in favor of this bus architecture is that the resistance voltage divider is no longer necessary to realize the suppressed small bus swing. This is because that this virtual stacking of the individual bus capacitance Cbus can play a role of virtual capacitance voltage divider, just like a three capacitors connected in series between Vcc and Vss as shown in Fig.2-5. Basically, each stepwise-divided voltage V1 and V2 is determined only by the ratio of the distributed bus capacitance values of Cbus. This is because that an establishing of the high or low level of each bus pair is given by the charge sharing with the adjacent capacitance, which is precharged to one-step higher or lower than own potential, respectively.

2-2-3 Dissipated Charge Amount Comparison

To compare the charge dissipation among the conventional full-swing and suppressed bus scheme and the proposed charge-recycling bus architecture, the illustrations of operating waveforms for three parallel complementary bus pairs (Di,XDi i=0,1,2) in three cycles of T1,T2, and T3 are shown in Fig.2-6, emphasizing on charge supplying from the power line and charge dumping to ground. In these illustrations, let water-faucet represents the power supplying line of Vcc and let garbage-can represents ground line of Vss. Funnel with valve represents the voltage down converter making it

Fig. 2-5. Concept of charge-recycling bus architecture.

(a) Full-swing bus scheme.

(b) Suppressed-swing bus scheme with on-chip down-converter.
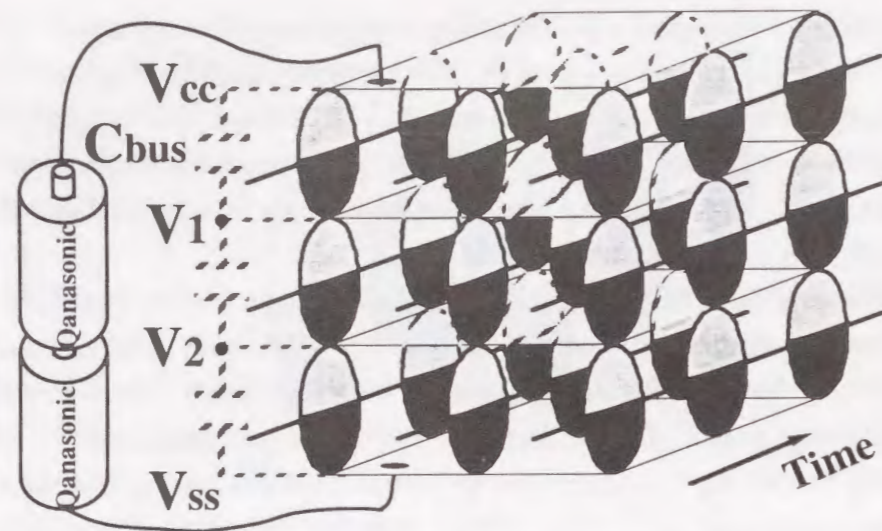
(c) Charge-recycling bus architecture.
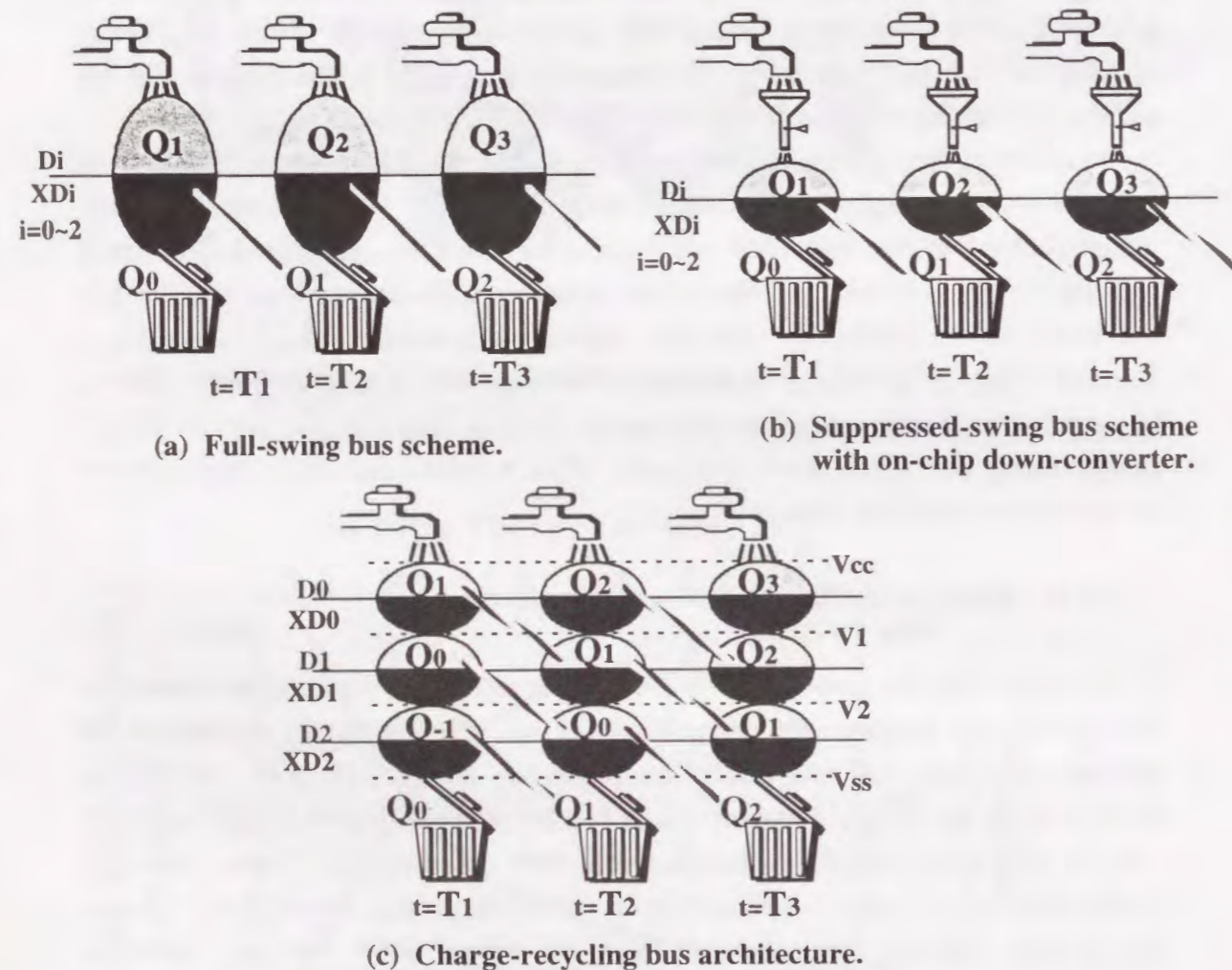
Fig. 2-6. Comparison of charge-dissipation among three types of bus schemes :
(a) Full-swing bus scheme, (b) Suppressed-swing bus scheme with on-chip
down-converter,
(c) Charge-recycling bus scheme.

possible to reduce the charging and discharging amount by suppressing the bus swing voltage as shown in Fig.2-6(b).

The important point in the conventional bus scheme shown in Figs.2-6 (a) and 2-6(b) is that every bus undergoing of transition from the precharged level of midpoint to high level or low level (that is to say, charging and discharging) in every cycle, the new charge supplying of Q_1, Q_2, and Q_3 are required at T_1, T_2, and T_3, respectively and the used charge after only one playing a role of high-level establishing in the former cycle is threw away to the ground line just like a garbage.   This implies that every time charge supplying from the power source line is required to transmit every one bit in every cycle, just like Q_1, Q_2, and Q_3 shown in Fig.2-6.   Thus, for example, in order to transmit the 25Gb/s, an intolerable large amount of charging and discharging is needed in the conventional bus architecture shown in Figs.2-6(a) and 2-6(b).

On the other hand, in the case of the charge recycling bus (CRB) architecture shown in Fig.2-6(c), from the top data bus pair (D_0, XD_0) through the bottom data bus pair (D_2, XD_2) are virtually connected in series between Vcc and Vss. Thus, for example, the intermediate bus pair (D_1, XD_1) and the bottom bus pair (D_2, XD_2) never connect to the power line directly.   In stead of power line, each bus pair except for the top bus pair recycle the used-charge stored on the adjacent upper bus capacitance pair, in order to undergo each high level transition.   For example,   the charge Q_1 can be recycled again and again by rolling down to the lower adjacent bus pair in every cycle, T_2 and T_3, until this Q_1 reaches the ground line as shown in Fig.2-6(c). As a result, regarding the parallel transmission of the three bits in the cycle of T_2, only one charge supplying of Q_2 is needed for the top bus pair and the other two bits can be transmitted by using the recycled charge Q_1 and Q_0 for the intermediate bus pair (D_1, XD_1) and the bottom bus pair (D_2, XD_2), respectively, where Q_0 represents the charge supplied from the power line at the one cycle before the cycle of T_1.   This implies that additional charge supplying is no longer necessary for stacking bus pairs except for the top bus pair, and an asymptotically zero power data transmission can be realized with the increasing number m of stacking bus pairs. This is because the bus swing voltage of the top bus pair can be reduced with the increasing number of the stacking bus pairs and the amount of only one charge supplying can be suppressed to the negligible small value (that is asymptotically zero) compared with that of the conventional full swing bus scheme.

Table 2-I. shows the power comparison based on the dissipated charge amount among the following three types: 1) conventional full-swing of Vcc scheme.(Conv.full) 2) conventional suppressed swing of Vcc/m scheme. (Conv.suppr.)   3) the CRB architecture with stacking bus number of m. (CRBm)

Regarding the total amount of the dissipated charge for the CRB architecture,   except for the charge supplying for the C1 of the first top data bus pair,   a new charge

supplying is no longer necessary for the other bus pairs as mentioned before repeatedly. In addition, the bus swing voltage is suppressed to the $1/m$ of Vcc. Thus, the total dissipated charge (Qtotal) can be expressed as shown in Table 2-I. On the other hand, for the conventional two bus schemes, all parallel bus pairs numbering from the first through the m-th require the new charge supplying simultaneously. In addition, the bus swing voltage is full-swing of Vcc for the full swing bus scheme, and the $1/m$ of Vcc for the suppressed swing bus scheme. Therefore, the Qtotal for each conventional one can be shown by each expression in Table 2-I. According to these expressions, the power reduction to $1/m^2$ of that compared with the conventional full swing scheme can be realized by using the CRB architecture. Even comparing with the conventional suppressed swing (Vcc/m) scheme, the CRB scheme can reduce the bus power dissipation to $1/m$ of the conventional one.

2-3 Principle of CRB Operation

In this section, why the CRB control makes it possible to realize the suppressed small bus swing and charge recycling among the parallel bus pairs, simultaneously, is discussed. In addition, how "virtual stacking" of the bus capacitance to substrate in series configuration between Vcc and Vss can be provided, is described.

2-3-1 Charge-Recycling Operation Among the Stacking Bus Capacitance

The charge recycling operation among the stacking of numerous bus capacitance's in series configuration between Vcc and Vss is discussed by using the illustration of behavior of the stored charge on each bus shown in Fig.2-7.

Figure 2-7 shows the behavior of the stored charge on each bus in the period of the time t_0, t_1 and t_2. The latter half of the cycle (t_0 and t_2) are for establishing of the complementary high and low levels of each bus pair. The former half of the cycle (t_1) is for equalizing and precharging of each bus pair. The C_{Di} and C_{XDi} of $i=0,1,2$ shown in Fig.2-7 represent the bus capacitance-name of the top bus pair (D0, XD0), the intermediate bus pair (D1, XD1), and the bottom bus pair (D2, XD2), respectively.

The height of each bar-graph represents the charged potential level of each bus. For example, the potential values of C_{D0} and C_{XD0} are Vcc and V1 at t_0, respectively and when the time becomes t_1, the potential values of those become the same precharged level V_{P1} of the top bus pair. When the time becomes t_2, the potential values of C_{D0} and C_{XD0} become V1 and Vcc, respectively, where new charge Q3 added on the capacitance C_{XD0} is supplied from Vcc line, and the used charge Q2 added on the bus

Table 2-I. Power comparison among three types as follows:
1. conventional full-Vcc-swing, 2. conventional suppressed Vcc/m-swing , 3. m-stacked CRB architecture.

	Conv. full	Conv. suppr.	CRBm
Total-Dissipated Qtotal	$m \cdot C_i \cdot V_{cc}$	$m \cdot C_i \cdot V_{cc}/m$	$C_1 \cdot V_{cc}/m$
	all for (1 ~ m-th)	all for (1 ~ m-th)	only for top (1-st)
	full-Vcc-swing	1/m-Vcc-swing	1/m-Vcc-swing
Norm. Power	1	$1/m$	$1/m^2$
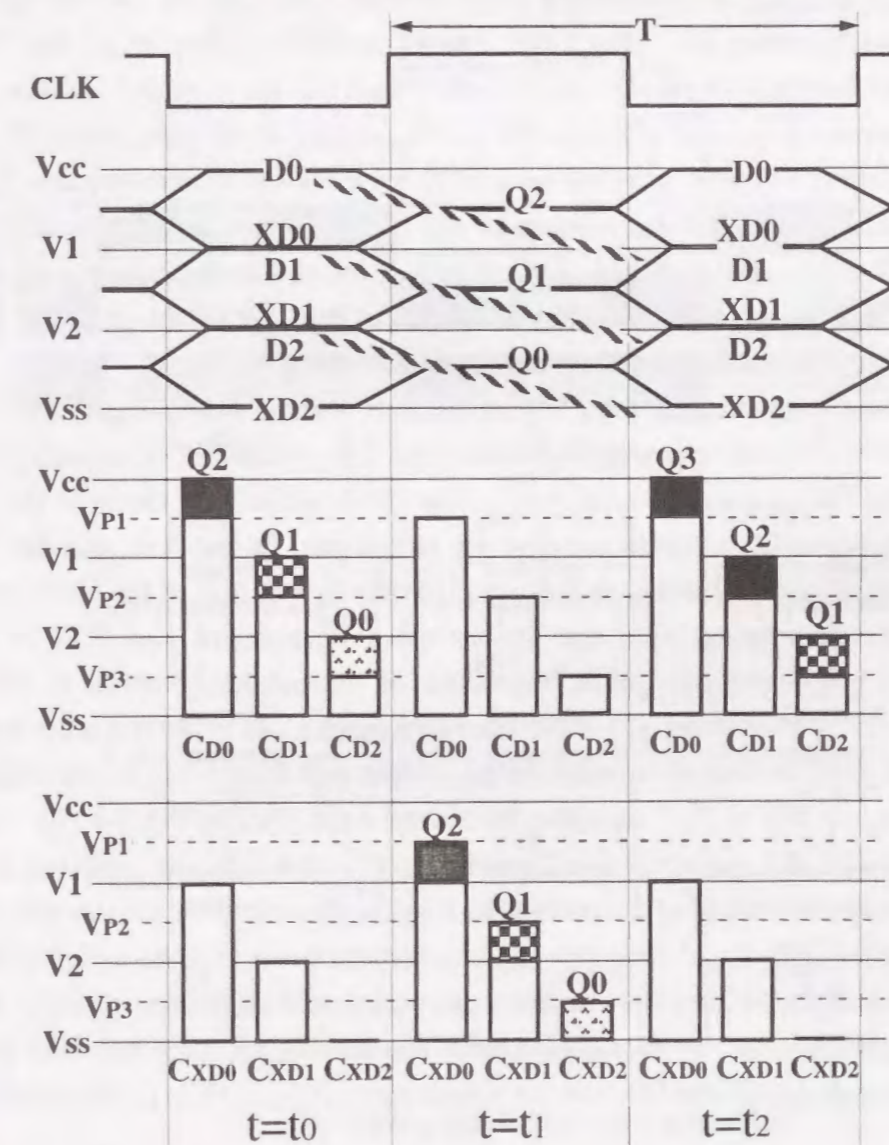
Fig. 2-7. Concept of charge-recycling operation among the stacking bus capacitance.

capacitance C_{D1} is transferred from the upper adjacent bus capacitance C_{XD0}. In the same way, the recycled charge Q_1 added on the bus capacitance C_{D2} is given from the upper adjacent bus capacitance C_{XD1}. The used charge Q_0 stored on the bus capacitance C_{XD2} at t_1 is dumped to the ground line at t_2.

2-3-2　Concept of Bus Control for CRB Architecture

The concept of bus control for CRB architecture are shown in Fig. 2-8. Schematic in Fig. 2-8(a) shows how the bus capacitance to substrate of the three bus pairs running in parallel (C_{Di} and C_{XDi} of i=0~2) connect in series each other between Vcc and Vss, where the input of the bus exchanger (INi of i=0~2) controls either of each bus pair connects to either of the upper or lower adjacent bus pair. As mentioned before, in the former half of the clock cycle, each bus pair is equalized by the signal of EQ (shown in Fig. 2-8(a)) and each stepwise equalized potential level (5/6, 3/6, or 1/6 of Vcc) is retained as shown in Fig. 2-7. Fig. 2-8(b) shows the timing diagram of the CRB operation. In the latter half of the clock cycle, one of each bus pair connects to one of the adjacent bus pair by controlling of bus exchanger, depending on the input signal (INi of i=0~2), for example, when INi of i=0~2 become (0,1,1), one of the top bus pair (C_{XD0}) connects to Vcc, and another (C_{D0}) connects to one of the intermediate bus pair (C_{D1}) as shown in Fig. 2-8(b). The switches (SWi of i=0~2) for connecting between adjacent bus pairs and between the top/bottom bus and Vcc/Vss line, are turned on in the latter half of the clock cycle, and are turned off during the equalizing period.

The block diagram of showing the CRB architecture, and the timing diagram and the truth table of the bus control signal are shown in Figs. 2-9.(a) and 2-9(b), respectively.

Regarding the i-th complementary bus pair in the CRB architecture, the input INi and the equalization signal EQ establish not only the output pair, Di and XDi, bus also the bus-level signals at nodes H and L as follows: 1) in the former half of the clock cycle, the EQ signal synchronized with the system clock equalize Di and XDi of the complementary bus-output pair, while the nodes of the bus-level signals H and L become open, that is Hi-Z state; 2) in the latter half of the clock cycle, the input signal INi switches higher (or lower) level of the bus-output pair Di and XDi to the node of the bus-level signal H (or L) according to the the truth table in Fig. 2-9 (b). The series connection of the numerous bus capacitance (C_{Di} and C_{XDi} of i=0~(m-1)) is realized by joining the node H of the (i+1)-th bus pair to the node L of the i-th bus pair. For example, when INi=0 and INi+1=1, the bus-output Di connects the nodes of L of the #i block, and the bus-output Di+1 connects the nodes of H of the #i+1 block. As a result, the Di and Di+1 are connected each other and become the same potential level, that is, the stored charge Qi and Qi+1 on the capacitance's (C_{Di} and C_{Di+1}) of the bus Di

(a) Schematic of CRB

(b) Timing diagram of CRB operation
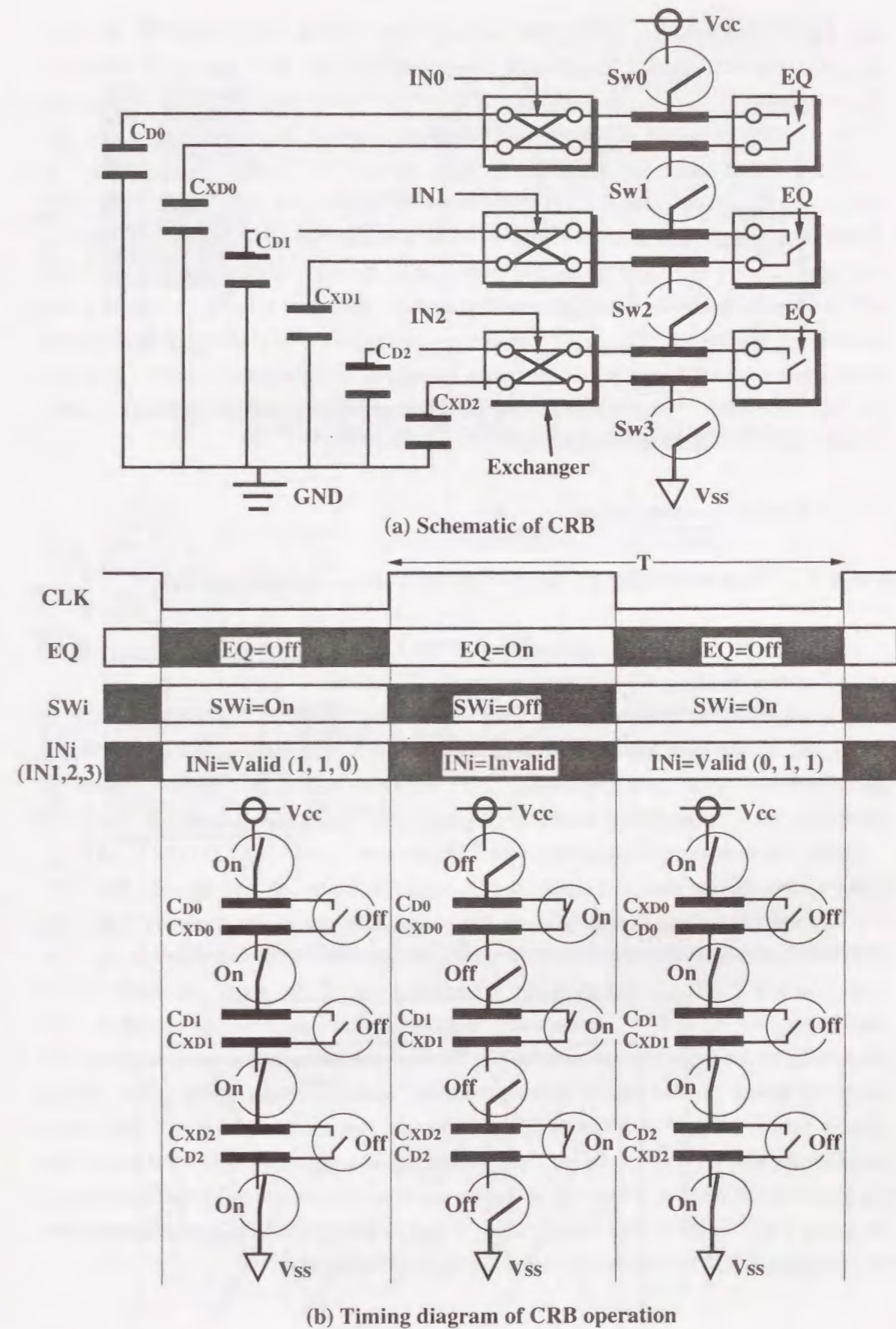
Fig. 2-8. Concept of charge-recycling bus (CRB) architecture :
(a) Schematic of CRB. (b) Timing diagram of CRB operation.

and Di+1, respectively, are shared between the two. If the values of the both capacitance's are equal, the Di and Di+1 undergo the low and high transition symmetrically.

As a result, by using the above mentioned bus control and configuration, the bus capacitance to substrate of the three bus pairs running in parallel (C_{Di} and C_{XDi} of i=0~2) can be virtually connected in series between Vcc and Vss as shown in Fig.2-8 (b). This bus architecture performs charge-recycling between the upper and lower adjacent bus pairs in every cycle and the equivalent bus capacitance of the parallel operating buses can be reduced to $1/m^2$ of that compared with full swing bus scheme, where m is the number of the buses running in parallel. In addition, this charge-recycling bus architecture play a role of virtual capacitance voltage divider making it possible to realize the suppressed bus swing without using the conventional equivalent resistance voltage divider that may induce the intolerable DC idling current[5].

2-4   Circuit Configuration of CRB

2-4-1   Transistor Level Circuit Configuration of CRB Driver

The transistor level circuit configuration of the CRB driver and their timing diagram and schematic operating waveforms are shown in Fig.2-10.

This driver circuit consists of the bus switch and equalizer. The bus switch is composed of the two MOSFET pairs, and each gate electrode is connected to the complementary input pair (INi/XINi) , respectively and each common source of MOSFET pair is connected to the bus-level node (Hi or Li) as shown in Fig.2-10 (a).

Taking a look at the operating waveforms of the output pairs (Di/ XDi ,Di+1/ XDi+1), in the former half of the clock cycle, the EQ signal is turned on and the intra-bus pairs (Di/XDi and Di+1/XDi+1) are equalized and in the latter half of the clock cycle, the bus switches are controlled by the complementary input pair (INi/XINi and INi+1/XIN+1) as shown in Fig.2-10(b). For example, when the input of the upper and lower driver circuits become "0" and "1", respectively, the four circled transistors are turned on, and the bus Di in the upper circuit and the bus Di+1 in the lower circuit are connected and equalized based on the charge sharing between each bus capacitance. This charge sharing between the Di and XDi+1 of the upper and lower circuits makes it possible to establish the low level of the Di for the upper circuit and the high level of the XDi+1 for the lower circuit without a new charge supplying from the power line, simultaneously as shown in Fig.2-10(b). This charge transfer and recycling from the upper to the lower bus capacitance can be realized for the all cases in this same way.
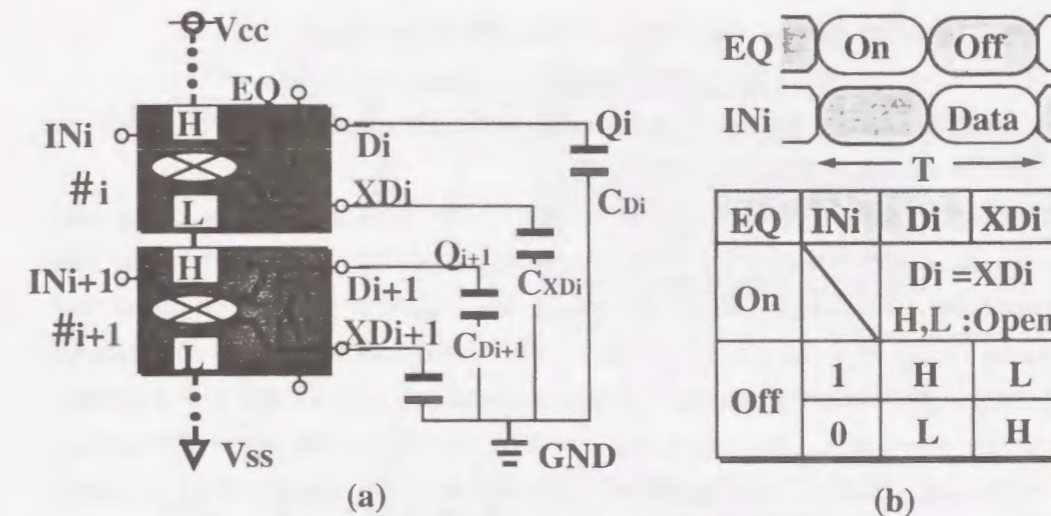
Fig. 2-9.   Bus control for CRB architecture:
(a) Block diagram of CRB, (b) Timing diagram and truth-table.
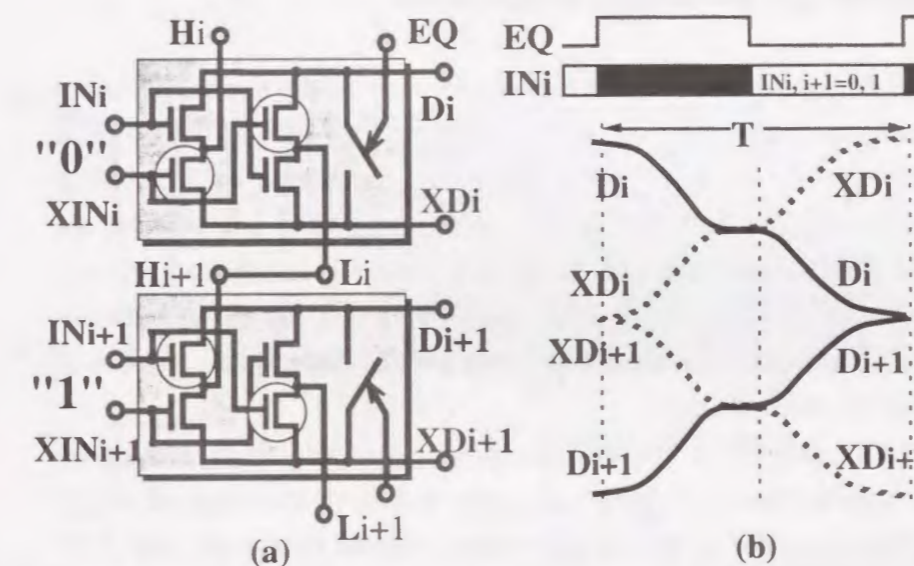
Fig. 2-10.   Transistor-level circuit configuration of CRB architecture:
(a) Circuit configuration of CRB, (b) Timing diagram and operating waveforms
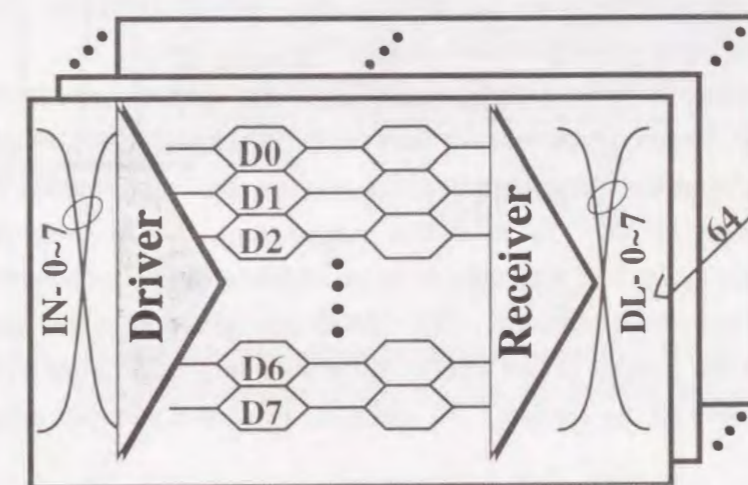
Fig. 2-11.   Bus configuration of CRB architecture
for ultra-multi-bit (512bits) buses.

2-4-2 Bus Configuration for Ultra-Multi-Bit Buses

The bus configuration of the CRB architecture with ultra-multi-bit buses, such as 512bit bus pairs is shown in Fig.2-11.

Parallel buses of 512bit are divided into 64 blocks. One block includes the driver and receiver circuit for the bus wirings for 8-bit, numbering from D0 through D7. The bus driver suppress the bus swing voltage to 1/8 of Vcc and the receiver amplifies the suppressed bus swing to a full swing of Vcc. The most important issue in realizing charge-recycling is prevention of wasteful charge dissipation, such as DC idling current to ground. To solve this, the input gate pairs of the upper and lower drivers are controlled by the signals of full-swing of Vcc (INi of i=0~7) so that a pair of the input gates never turn on simultaneously[6].

2-4-3 CMOS Driver and Receiver Configurations

The relation between the gate types of the driver and receiver in the upper and lower circuits is complementary as shown in Figs 2-12(a) and 2-13(a), that is, PMOS gate drivers and NMOS gate receiver are used for the upper bus pairs whose each operating bus potential is higher than 1/2Vcc, and NMOS gate drivers and PMOS gate receiver are used for the lower bus pairs whose each operating bus potential is lower than 1/2Vcc. The circuit configurations of NMOS and PMOS gate drivers and the complementary input and output signal waveforms for the NMOS and PMOS drivers are shown in Figs. 2-12 (b) and 2-12 (c), respectively.

In the case of PMOS driver, to prevent the input gate pair from turning-on simultaneously, the voltage of the input signal pair never undergo a transition of Vcc to Vss, simultaneously, that is, either of the complementary signals retains the Vcc level constantly. In equalizing period, both of the two signals set the Vcc level, and either of the complementary input undergoes a high to low transition in the rest of the cycle. For the NMOS driver, either of the input pair maintain the Vss level every time, and after equalizing period, either input signal goes high level of Vcc.

The circuit configurations and the complementary input and output signal waveforms for the NMOS and PMOS receivers are shown in Figs. 2-13(b) and 2-13(c), respectively. The bus receiver is required for stable amplifications and low power operations over the wide input-voltage range of Vcc. To meet this requirement, the receiver circuit is composed of a dynamically latched type current sense amplifier and a gate-receiver that functions as a voltage to current converter. The CMOS configuration in the upper and lower receiver circuits can provide a fast operation by obtaining of a larger V_{GS} more than Vcc/2 for both types of the circuits. In addition, since the parasitic capacitance
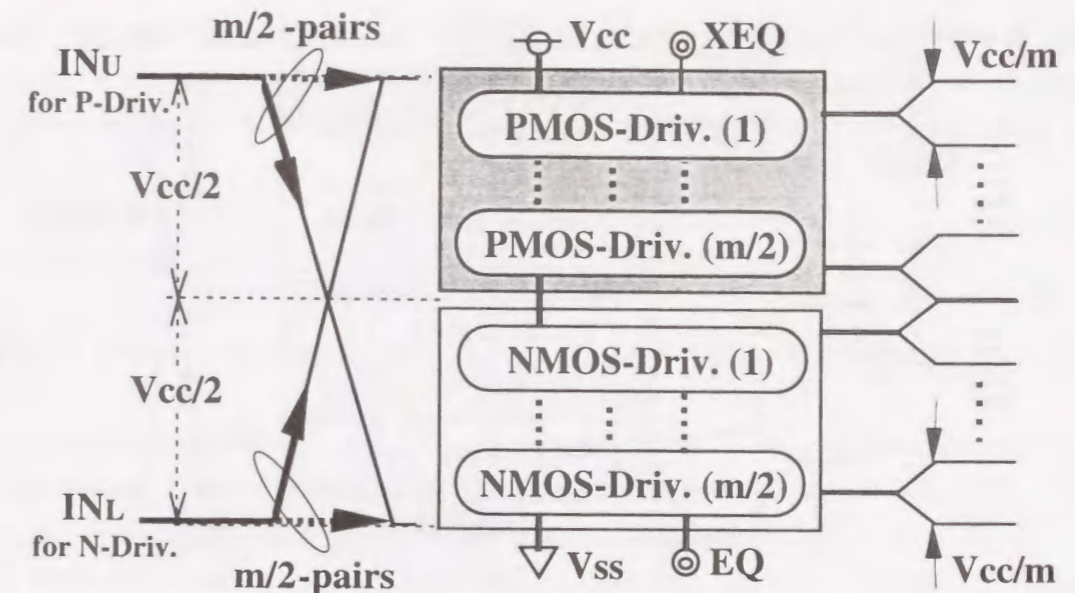
(a)
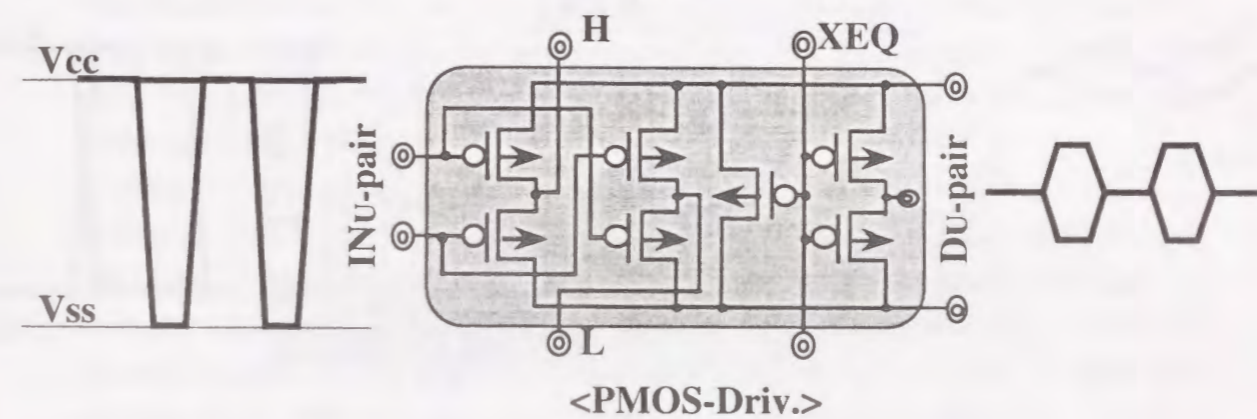
(b)
<PMOS-Driv.>
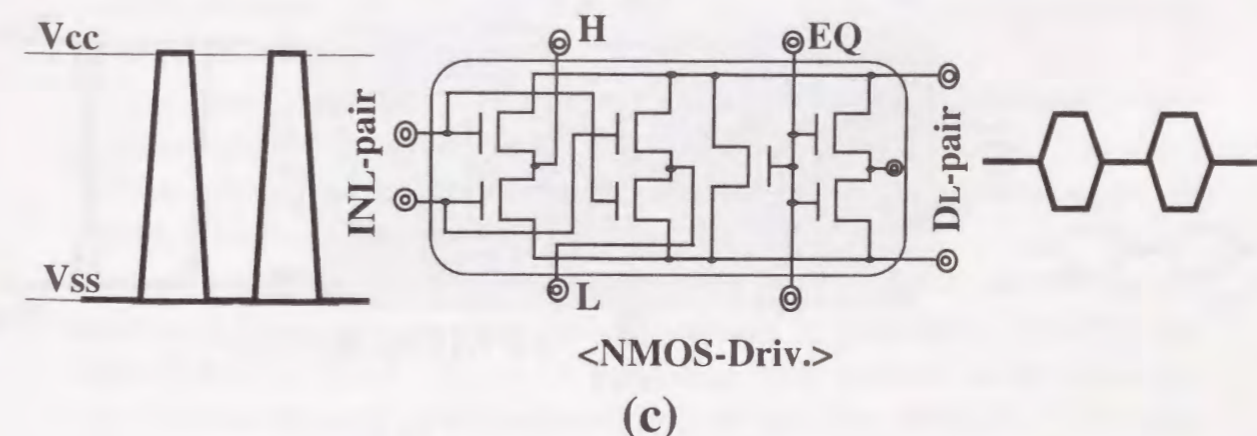
(c)
<NMOS-Driv.>

Fig. 2-12. Transistor-level circuit configuration of CRB architecture:
(a) CMOS driver configuration, (b) PMOS driver, (c) NMOS driver.
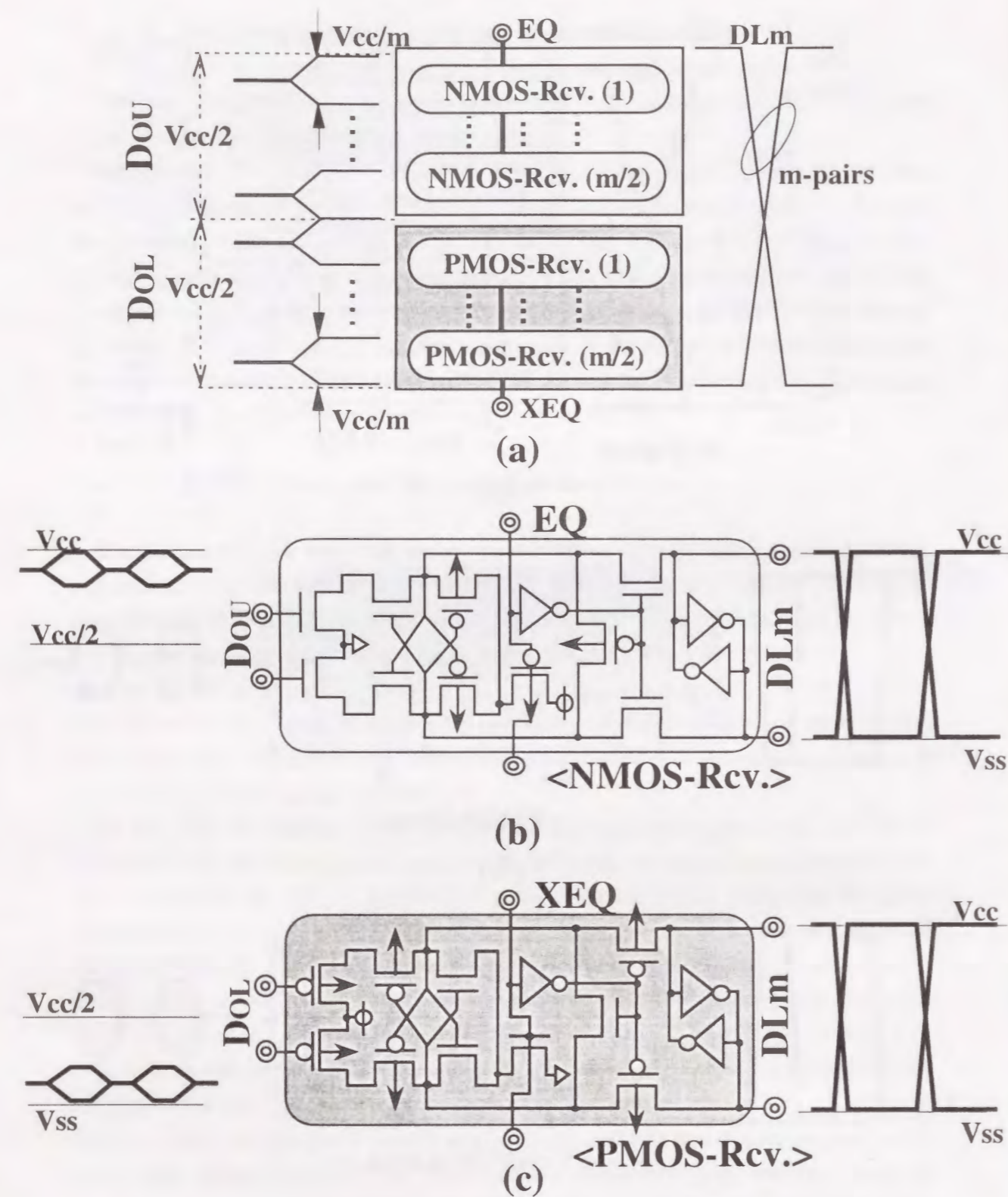
Fig. 2-13. Transistor-level circuit configuration of CRB architecture:
(a) CMOS receiver configuration, (b) NMOS receiver, (c) PMOS receiver.

value of the output node in the receiver circuit and the input gate capacitance value in the driver circuit are less than 5% of that compared to the bus capacitance of several pF, the total power consumption is hardly increased even if including the receiver's and driver's one.

2-5 Circuit Operation and Performance

2-5-1   Simulated Results and Discussions of CRB Operation

The simulated operating waveforms of the 8bit stacking buses, numbering from D_0 through D_7 are shown in Fig.2-14 (a). When the operating voltage Vcc is 3.3V, the swing voltage of each bus is automatically fixed to about 0.4V as shown in Fig.2-14 (a). Stable data transfer can be achieved at the operating frequency of 50MHz. The important concern in this architecture is how the precharged potential of each bus pair is determined between Vcc and Vss. Basically, each precharged potential is determined only by the distributed bus capacitance values.   When the capacitance values of all buses are equaled, a difference of each precharged potential is distributed equally, just like a staircase.

This is because that potential transition of high-to-low or low-to-high in each bus is undergone by only the charge sharing between the adjacent bus capacitance precharged at a different potential, except for each one of the top and bottom bus pairs connecting Vcc and Vss, respectively.   Needless to say, when the "power-on state", that is, when the potential of power supply line is changed from Vss to Vcc, several dummy clock cycles are required for achieving the above mentioned stable precharge operation.   This is because that it takes several clock cycles corresponding to the number of stacking bus pairs, for the stored charge on the top bus capacitance supplied from the power line, to be transferred from the top bus pair down to the bottom bus pair.

When introducing the CRB architecture to the practical chip is considered, several ways can overcome the above mentioned "power-on" problem, for example, the simple resistance voltage divider that is activated only when power-on state, can supply the suitable initial potential for each bus.

Another important concern in this architecture is how the charge sharing and transferring speed between both adjacent bus pairs is distributed. Basically, the required time for charge sharing and transferring from the upper to the lower bus capacitance is depended on the drive-ability of the bus driver MOSFET. The speed differences, among the top bus pair D_0 through the intermediate bus pair D_3 for PMOS type driver and among the intermediate bus pair D_4 through the bottom bus D_7 pair for NMOS driver are suppressed to less than 2ns that are negligibly small compared with
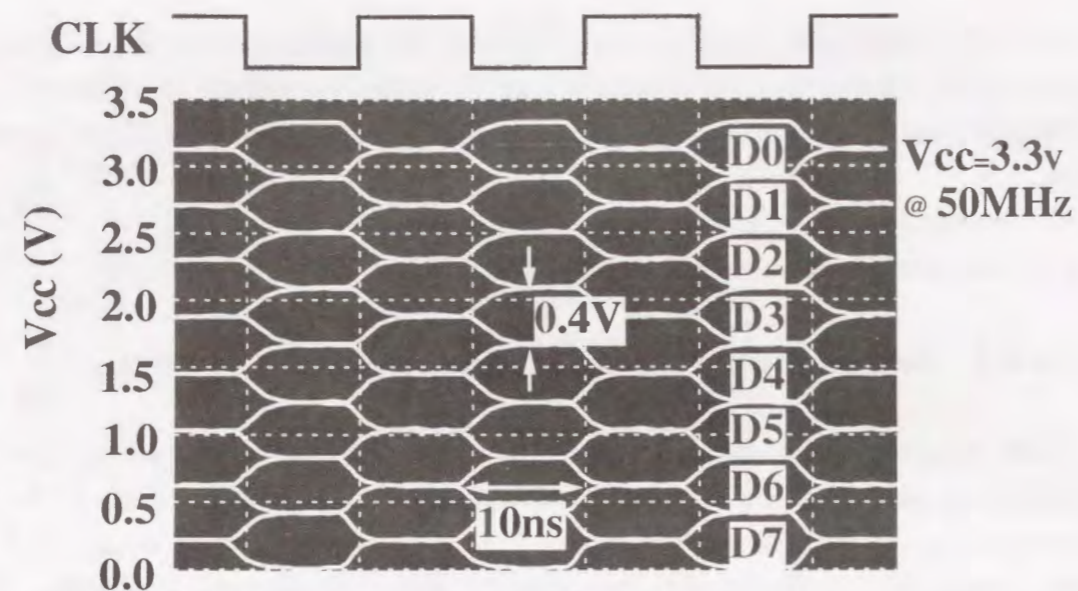
(a) 50MHz Operating waveforms of 8-stacked CRB at Vcc=3.3V.
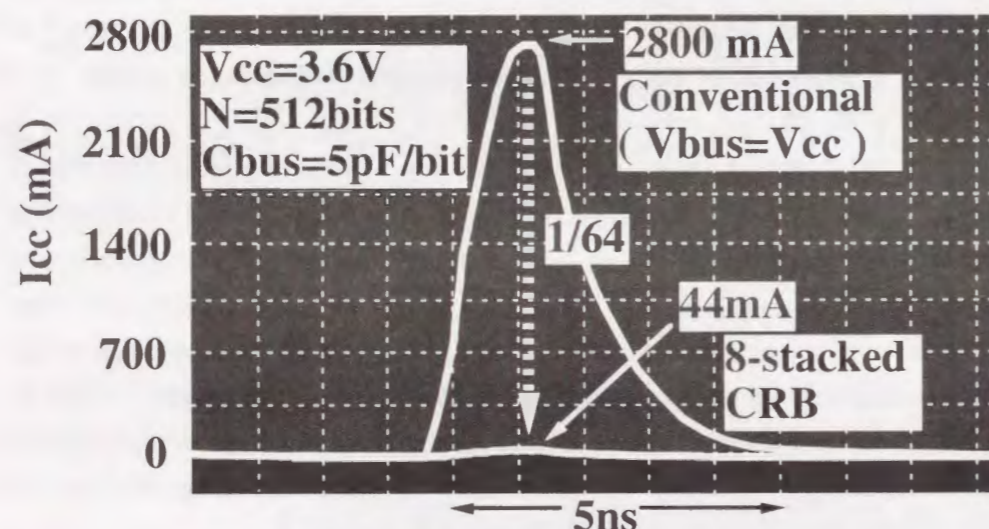
(b) Comparison of operating current waveforms

Fig. 2-14.   Simulated results:(a) 50MHz-operating waveforms, (b) Comparison of operating current waveforms.

Table.II   Device Characteristics

Process technology	0.5μm twin-well CMOS DRAM technology
	Triple poly Si / Single polycide / Double metal
Transistor	Tox=12nm
	Ln=0.55μm (Nch-MOS)
	Lp=0.60μm (Pch-MOS)
	Vtn=0.5 V (Nch-MOS)
	Vtp=0.6 V (Pch-MOS)

10ns of half of the clock cycle corresponding to 50MHz as shown in Fig.2-14(a). This speed difference is caused by the body effect   (threshold voltage V_T increasing with voltage of source to substrate potential) of each driver MOSFET. For example, regarding the PMOS driver, source potential of each driver MOSFET is as follows: 1) Vs=Vcc for D0 bus pair driver, 2) Vs=5/8 Vcc for D3 bus pair driver, thus, the threshold voltage difference between the two becomes about 100mV for Vcc=3.3V, and results in inducing of the speed difference of less than 2ns.

Next, operating current amount comparison between the conventional full-swing bus scheme and the CRB architecture with stacked bus structure every 8 bit are given in Fig.2- 14(b).   The condition parameter is assumed as follows: 1) Vcc=3.6V, 2) the total operating bus number N=512bit, 3)   bus capacitance per bit Cbus=5pF and 4) stacking bus number for CRB architecture m=8.

The peak current of about 2800 mA (2.8A) for the conventional full-swing bus scheme, can be reduced to 44mA by using the CRB architecture, that corresponds to 1/64 of the conventional one.   This is because that regarding the 8-stacked parallel bits in each block of 64 blocks, zero power data transmission can be realized based on the charge-recycling bus architecture, except for the only one bit of the top data bus pair. In other words, the CRB architecture can make the most of the "hot" new charge supplied from power line in each block, without waste charge dissipation, such as every time charge dumping to ground.

2-5-2   Measured Results and Discussion

The measured typical operating waveforms of the 4-stacked CRB (m=4)   shown in Fig.2-15(a).   In addition, the micro-photograph of the test-site of the CRB circuits is given Fig.2-15(b).   This test-site was fabricated by using the 0.5μm CMOS DRAM technology for a fabrication convenience.   The features of process technology and transistor parameter such as threshold voltage and channel length, *etc*, are shown in Table 2-II.

A stable precharging and developing of each bus pair can be observed at the operating frequency of 50MHz and Vcc=3.0V. Each precharged potential and each swing voltage can be achieved constantly, just same as simulated results.

2-5-3   Comparison of Bus Power Consumption

In Fig.2-16, the comparison of the bus power consumption among the three types bus scheme of the conventional full-swing bus, the suppressed swing bus, and the CRB, are shown as a function of number of the bus width.
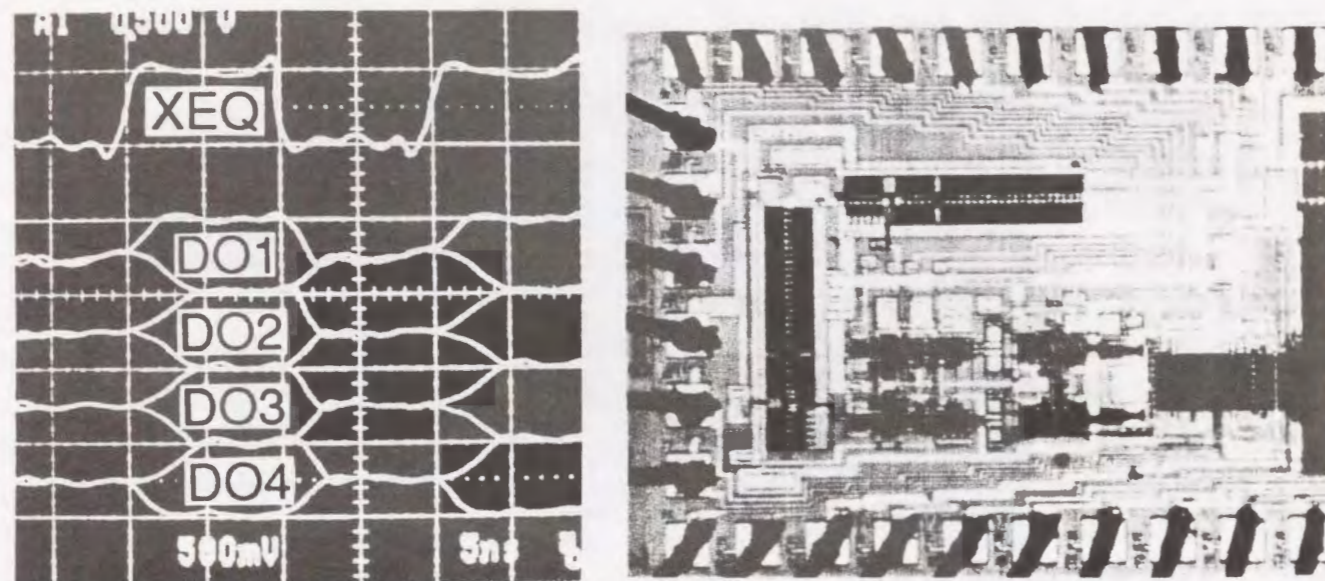
50MHz
Vcc=3.0V

(a)

(b)

**Fig.2-15 Experimental results: (a) Typical operating waveforms,
and (b) Microphotograph of CRB test-site.**

HDTV
1.6Gb/s

Super-HD/3D-CG
25.6Gb/s

Vbus=Vcc
Conv.

Vbus=Vcc/8

1/8

1/64

CRB-8

Vcc=3.6V
Cbus=14pF
@ 50MHz
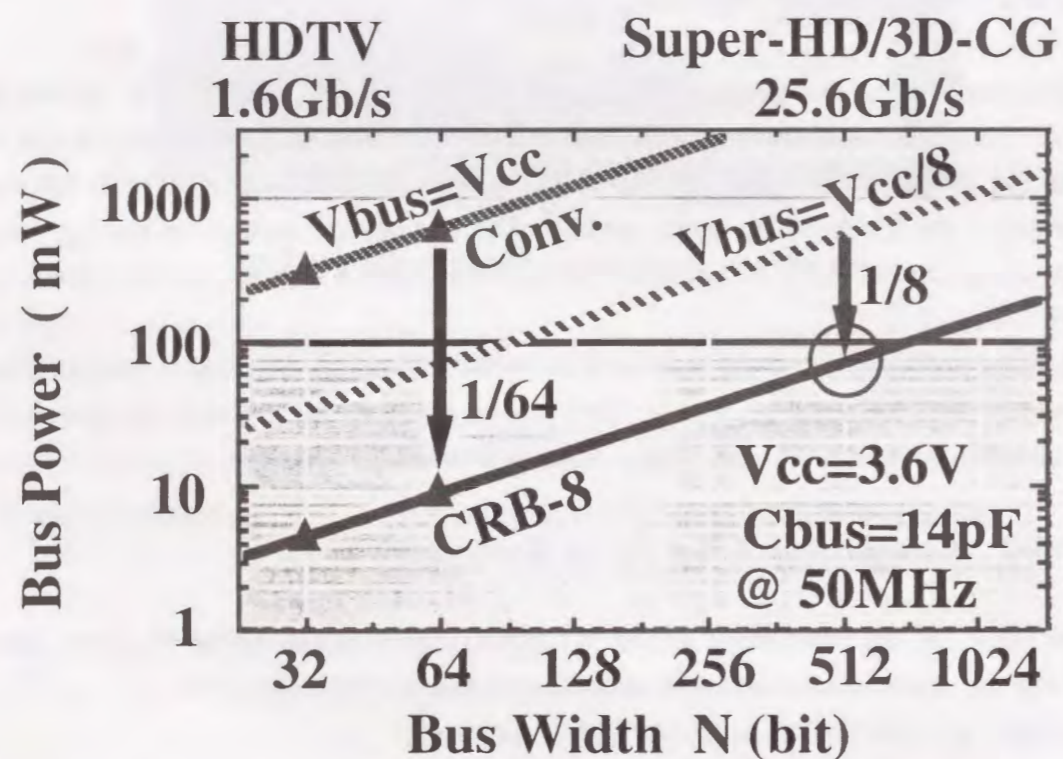
Fig. 2-16. Comparison of bus-power consumption between conventional and CRB.

The most upper line show for the conventional full-swing bus scheme with Vbus of full Vcc (3.6V). The intermediate line show the bus power consumption for the conventional suppressed bus scheme with Vbus=1/8Vcc (0.45V). The lower line represent the dissipated bus power of the 8-stacked CRB (m=8) architecture. The measured data are denoted by using triangular marks.

This CRB scheme can reduce the bus power to 1/64 and 1/8 of that compared with conventional full swing and suppressed bus scheme, respectively.

According to these data, the ultra high data rate of 25Gb/s can be achieved, while maintaining the power dissipation to be less than 100mW at Vcc of 3.6V, where the bus width is 512bit and the bus capacitance is 14pF/bit, and the operating frequency is 50MHz. The proposed CRB architecture is the most promising candidate for interconnecting the embedded memory and graphics controller, *etc*, on battery-operated ULSI's system chip for super HD applications.

2-6   Bus Capacitance Imbalances Issues

Before the discussion about the bus capacitance imbalances issues, we clarify the relation between the bus capacitance distribution and the bus swing distribution. Schematic in Fig.2-8(a) shows how the bus capacitance to substrate of the three bus pairs running in parallel (C_{Di} and C_{XDi} of i=0~2) connect in series each other between Vcc and Vss, where the input of the bus exchanger (INi of i=0~2) controls either of each bus pair connects to either of the upper or lower adjacent bus pair. As mentioned before, in the former half of the clock cycle, each bus pair is equalized by the signal of EQ (shown in Fig.2-8(a)) and each stepwise equalized potential level (5/6, 3/6, or 1/6 of Vcc) is retained. Fig.2-8(b) shows the timing diagram of the CRB operation. In the latter half of the clock cycle, one of each bus pair connects to one of the adjacent bus pair by controlling of bus exchanger, depending on the input signal (INi of i=0~2), for example, when INi of i=0~2 become (0,1,1), one of the top bus pair (C_{XD0}) connects to Vcc, and another (C_{D0}) connects to one of the intermediate bus pair (C_{D1}) as shown in Fig.2-8(b). The switches (SWi of i=0~2) for connecting between adjacent bus pairs and between the top/bottom bus and Vcc/Vss line, are turned on in the latter half of the clock cycle, and are turned off during the equalizing period.

As a result, by using the above mentioned bus control and configuration, the bus capacitance to substrate of the three bus pairs running in parallel (C_{Di} and C_{XDi} of i=0~2) can be virtually connected in series between Vcc and Vss, as shown in Fig.2-8(b). Thus, for easy understanding about the relation between the bus capacitance distribution and the bus swing distribution, the equivalent circuits and the expressions are shown in Figs. 2-17(b) and 2-18(b). This implies that each bus swing of Vm is determined only
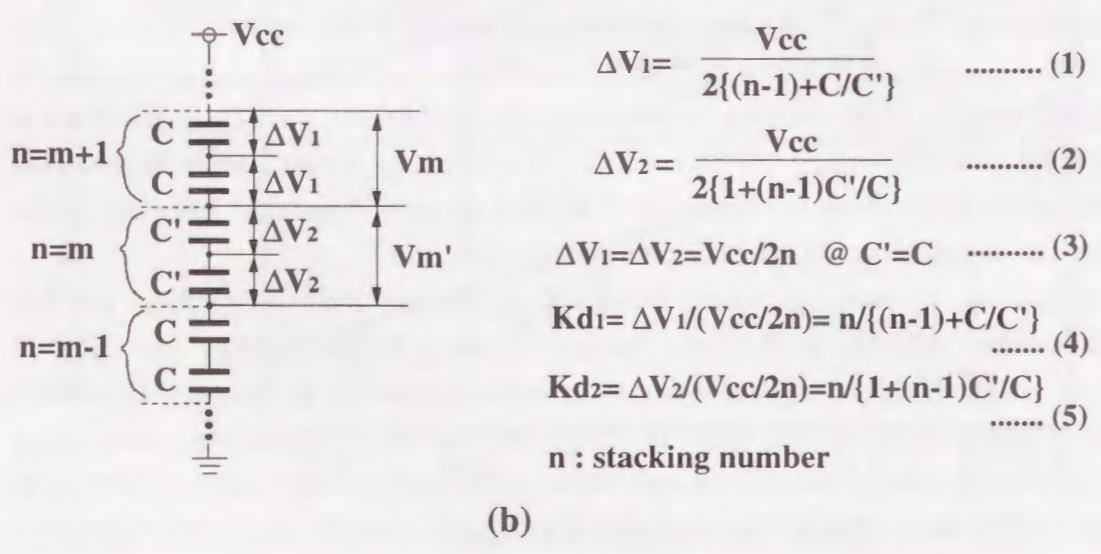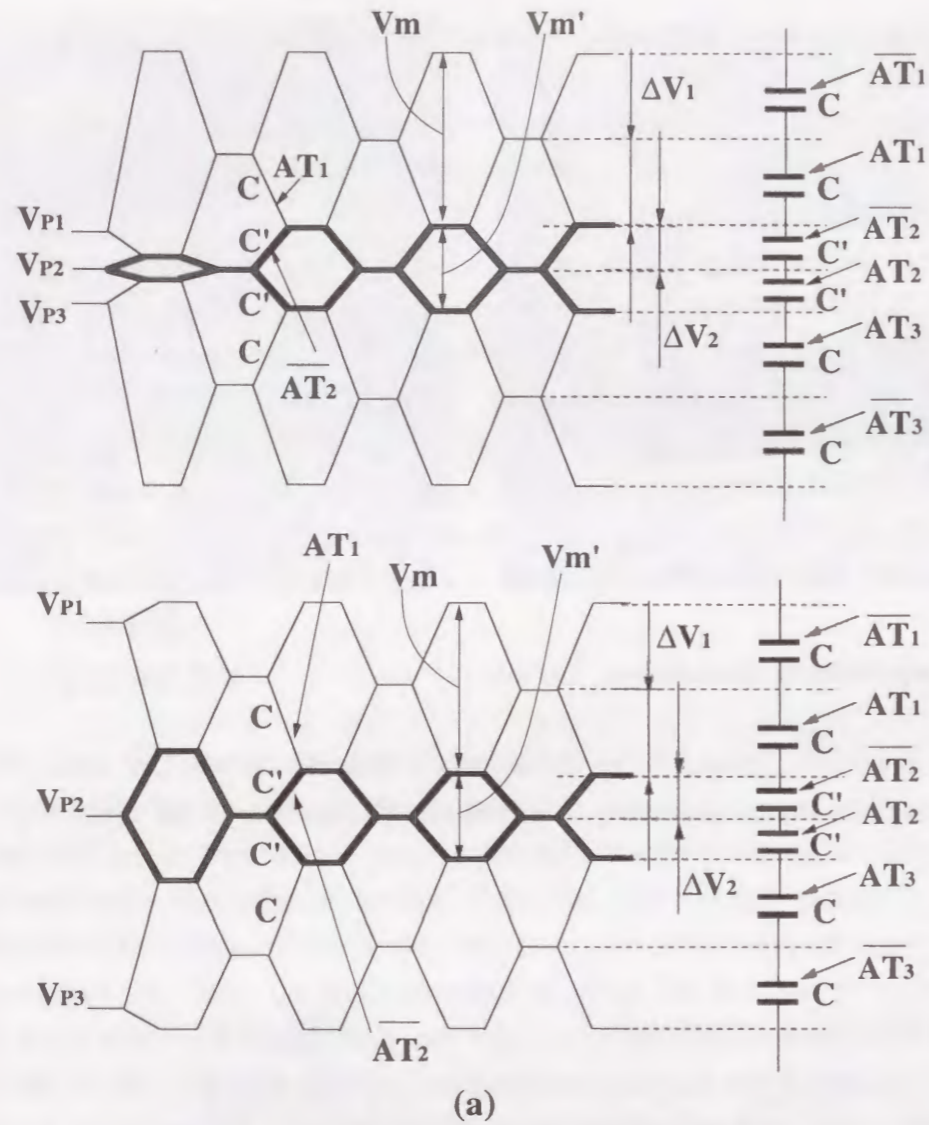
Fig. 2-17 (a) Vm distribution depending initial position of VPn
(b) Expressions of Vm distribution depending on C/C'
@ C-imbalance between inter-buses (e.g. b/w AT1 and AT2)

$$\Delta V_1 = \frac{Vcc}{2\{(n-1)+C/C'\}} \quad \cdots\cdots (1)$$

$$\Delta V_2 = \frac{Vcc}{2\{1+(n-1)C'/C\}} \quad \cdots\cdots (2)$$

$$\Delta V_1 = \Delta V_2 = Vcc/2n \quad @ \ C'=C \quad \cdots\cdots (3)$$

$$Kd_1 = \Delta V_1/(Vcc/2n) = n/\{(n-1)+C/C'\} \quad \cdots\cdots (4)$$

$$Kd_2 = \Delta V_2/(Vcc/2n) = n/\{1+(n-1)C'/C\} \quad \cdots\cdots (5)$$

n : stacking number
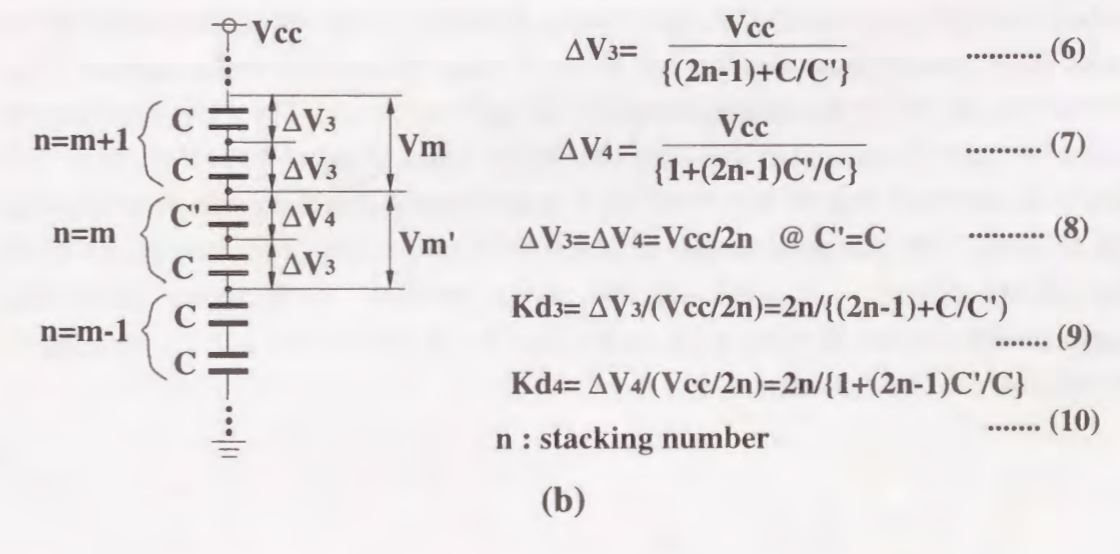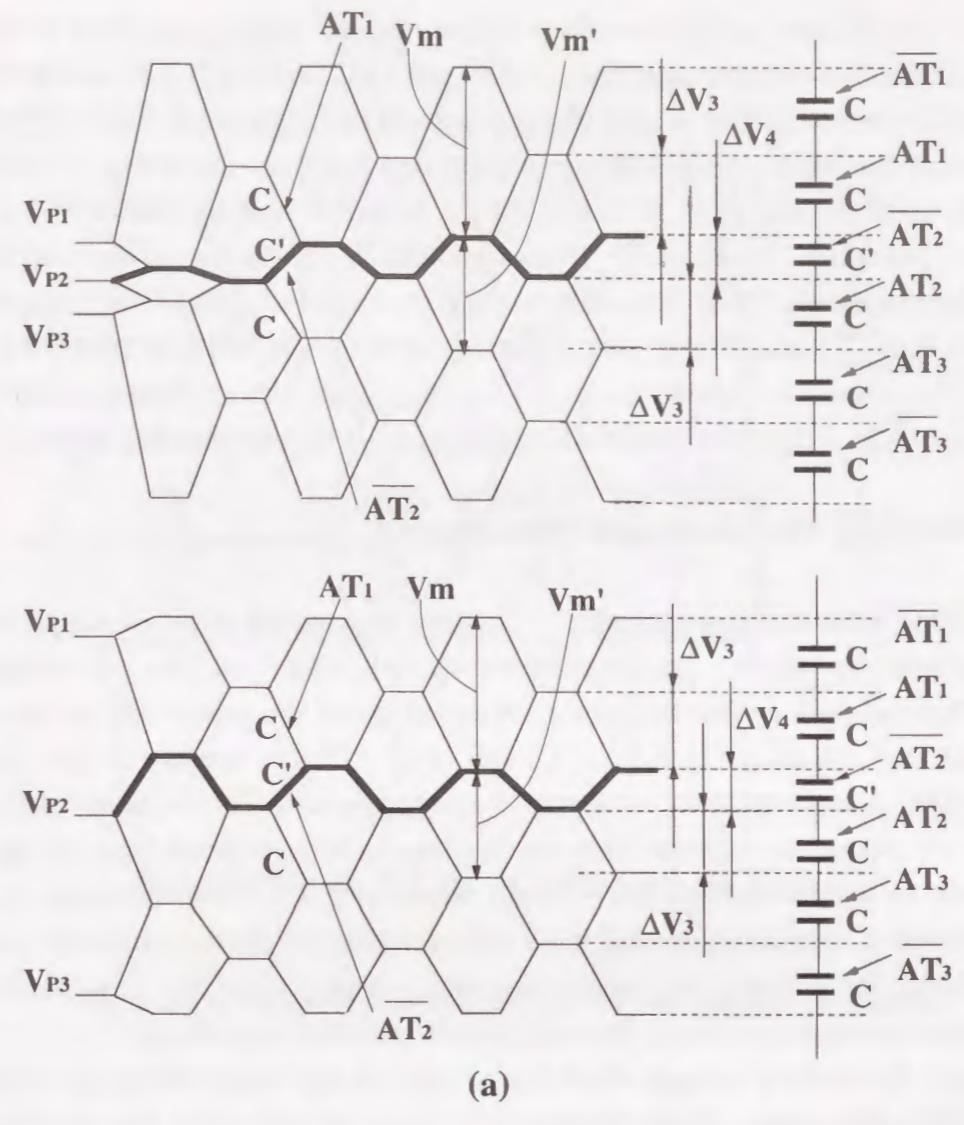
Fig.2-18 (a) Vm distribution depending initial position of VPn
(b) Expressions of Vm distribution depending on C/C'
@ C-imbalance between intra-buses (e.g. b/w AT2 pairs)

$$\Delta V_3 = \frac{Vcc}{\{(2n-1)+C/C'\}} \quad \cdots\cdots (6)$$

$$\Delta V_4 = \frac{Vcc}{\{1+(2n-1)C'/C\}} \quad \cdots\cdots (7)$$

$$\Delta V_3 = \Delta V_4 = Vcc/2n \quad @ \ C'=C \quad \cdots\cdots (8)$$

$$Kd_3 = \Delta V_3/(Vcc/2n) = 2n/\{(2n-1)+C/C'\} \quad \cdots\cdots (9)$$

$$Kd_4 = \Delta V_4/(Vcc/2n) = 2n/\{1+(2n-1)C'/C\} \quad \cdots\cdots (10)$$

n : stacking number

by the distributed bus capacitance values. Figures 2-17(a) and 2-17(b) show the swing (Vm) distribution when the capacitance of AT2 pair (AT2 and /AT2) is C', while that of others is C. The swing Vm' of AT2 pair is expressed by (2) shown in Fig.2-17(b). The deviation of Vm' (Kd2) from ideal case is expressed by (5) shown in Fig.2-17(b). For example, assuming that the C'/C is 0.9 and 1.1 when the stacking number n is 8, the Kd2 is 1.1 and 0.92, respectively. Thus, sensitivity of the bus swing deviation to the capacitance imbalance is less than 10%, causing the variation of 40mV for a bus swing of 400mV at Vcc=3.3V and n=8. Figures 2-18(a) and 2-18(b) show the Vm distribution when the capacitance of /AT2 is C', while that of others is C. The deviation of Vm' (Kd4) from ideal case is expressed by (10) shown in Fig.2-18(b).

2-7   Power-on State Issue and Noise Issue

Regarding when the "power-on state" (i.e. when the potential of power supply line is changed from Vss to Vcc, precharge levels of V_{P1}, V_{P2}, and V_{P3},. is unsettled), several dummy clock cycles stabilize the swing voltage of Vm independent of the initial precharge level, as shown in Figs. 17(a) and 18(a). This is because it takes several clock cycles, depending on the number of stacking bus pairs, for the stored charge on the top bus capacitance supplied from the Vcc line, to be transferred from the top bus pair down to the bottom bus pair. When introducing the CRB architecture to the practical chip is considered, several ways can overcome the above mentioned "power-on" problem, for example, the simple resistance voltage divider that is activated only when power-on state, can supply the suitable initial potential for each bus.

Figure 2-19 show the concept of self-biased and self-recoverable precharge operation in the CRB architecture. When the noise ΔV is injected into a bus line as shown in Fig.2-19, the precharge level and swing voltage of successively transferred data are reduced to $\Delta V/4$ and $(Vm-\Delta V/2)$, respectively. However, when taking the worst case of noise issue into consideration for the practical chip, the injected noise amount (e.g. 40mV) directly affects the sensing margin of the receiver circuits. Thus, it is necessary to suppress the noise amount to less than 100mV for a bus swing of 400mV at Vcc=3.3V and n=8, assuming that the rest (300mV) is a sufficient to detect the transferred data in the receiver. The twisted data-line arrangement[5] which was previously proposed in the DRAM circuits, can overcome the above problem. Furthermore, decoupling capacitance connected in series between Vcc and Vss as shown in Fig.2-19, contributes to suppress the noise problem.

Fig.2-19 Concept of self-biased and self-recoverable precharge.

2-8　Conclusion

A CRB architecture, featuring virtual stacking of the individual bus-capacitance into a series configuration between supply voltage and ground, has been proposed. This can reduce not only each bus-swing but also a total equivalent bus-capacitance of the ultra multi-bit buses running in parallel. The charging and discharging of each bus are based on the charge-sharing among the upper and lower adjacent bus capacitance equivalently connected in series, instead of the power line. According to the 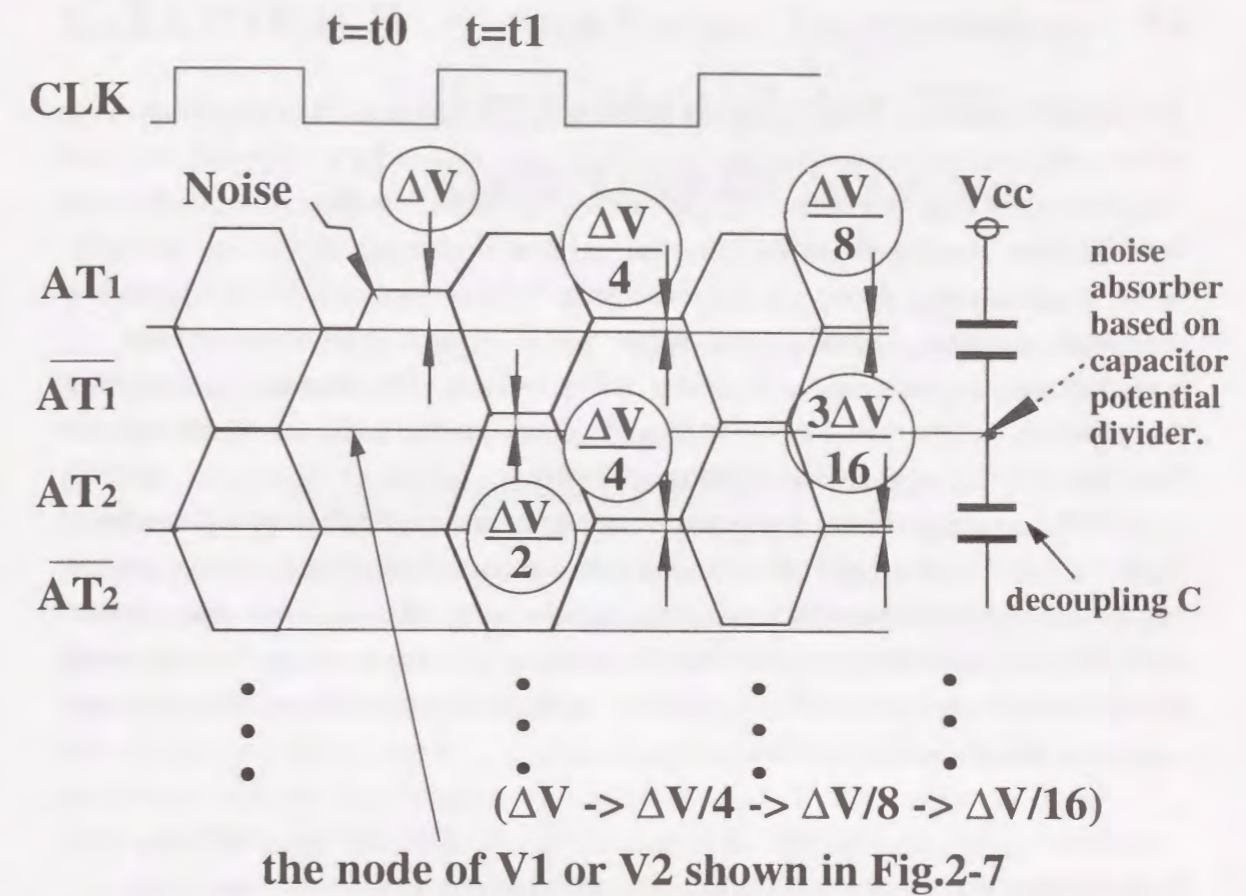simulated and measured data, the data rate of 25.6Gb/s can be realized while maintaining the power dissipation to be less than 100mW at Vcc =3.6V for the bus width of 512bit with the bus-capacitance of 14pF per bit operating at 50MHz.

After the several problems, such as the noise issue and the CAD issue in introducing to the practical ULSI's, have been overcome, this proposed architecture will become the most promising candidate for interconnecting the embedded memory and graphics controller, *etc*, on battery-operated ULSI's, making it possible to realize future personal communications services (PCS's) applications with universal portable multimedia access supporting full motion high-definition digital video.

References

[1] Y.Nakagome, et al, " Sub-1V swing bus architecture for future low-power ULSI's " , in Symposium. on VLSI Circuits Digest of Technical Papers , pp. 29 - 30 , Jun. 1992.

[2] T.Bell " Incredible shrinking computers ", IEEE Spectrum , pp. 37 - 43, May. 1991.

[3] H.Yamauchi, et al " A Low Power Complete Charge-Recycling Bus Architecture for Ultra-High Data Rate ULSI's " , in Symposium. on VLSI Circuits Digest of Technical Papers , pp. 21 - 22 , Jun. 1994.

[4] T.Kawahara, et al., " A Charge recycle refresh for Gb-scale DRAM's in file applications " ,in Symposium on VLSI Circuits Digest of Technical Papers , pp. 41-42, May 1993.

[5] D.Takashima, et al, "Low Power on-chip supply voltage conversion scheme for 1G/4G bit DRAMs" ,in Symposium on VLSI Circuits Digest of Technical Papers , pp. 114-115, June 1992.

[6] H.Yamauchi, et al, "A 20ns Battery-Operated 16Mb CMOS DRAM", ISSCC Digest of Technical Papers , pp. 44-45, Feb. 1993.

CHAPTER-3   Signal-Swing Suppressing Time-Multiplexed Differential Data-Transfer Scheme

Abstract

This paper presents a signal-swing suppression strategy which uses a time-multiplexed differential data-transfer (TMD) scheme combined with a data-transition detector (DTD) circuit, featuring shared complementary wires, which are originally allocated to adjacent signal bits, respectively. TMD can be exploited to reduce the signal voltage-swing and to realize a charge-recycling bus (CRB) architecture[1]. This enables a dramatic power reduction without the throughput-loss due to time-multiplexing, while maintaining the same number of signal wires compared to a single signal line (SSL) scheme. This is because the differential transfer scheme inherently has a more capability in terms of throughput and noise tolerance compared to SSL. To demonstrate the effectiveness of TMD with DTD and TMD with CRB(TM-CRB), power consumption comparisons were made between SSL, the parallel architecture[2], TMD with DTD, and TM-CRB. For all measurements, the same throughput conditions were used based on the simulated and measured data of the 0.5μm CMOS devices. This paper presents why TM-CRB can reduce the power dissipation on heavily loaded bus lines to less than 1/31 and 1/8, when the bus activity of 100% and 25%, respectively, while maintaining the same number of signal wires, compared to SSL.

3-1　Introduction

In recent years, the parallel architecture[2] has become the centerpiece of low-power techniques. This is because the parallel architecture can be exploited to scale the power-supply voltage, enabling quadratic power reduction while maintaining the same throughput. The parallel architecture was used in the data path circuits which include logic gates, such as adder, comparator, latch, *etc.* [2]

However, the method of power savings by using the parallel architecture has the overhead of increased layout area, and would not be suitable for area-constrained designs. For example, when exploiting the parallel architecture in designing the data-bus circuits for interconnecting the embedded memory and the graphics controller (as shown in Fig.3-1(a)) , there inevitably comes an increase in the number of signal wires to recoup the throughput-loss due to the scaled supply voltage and the lower operating
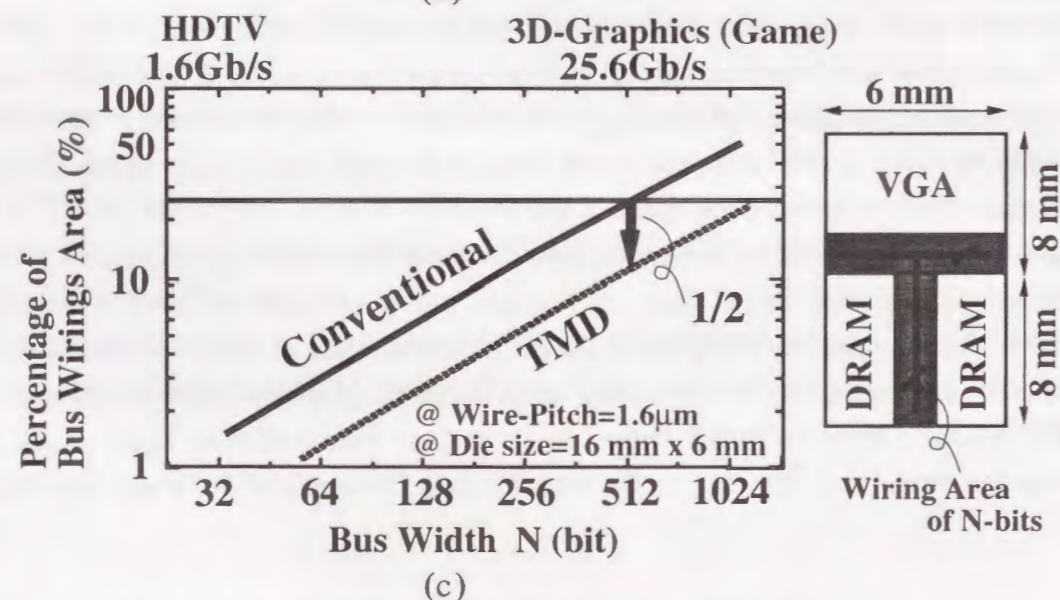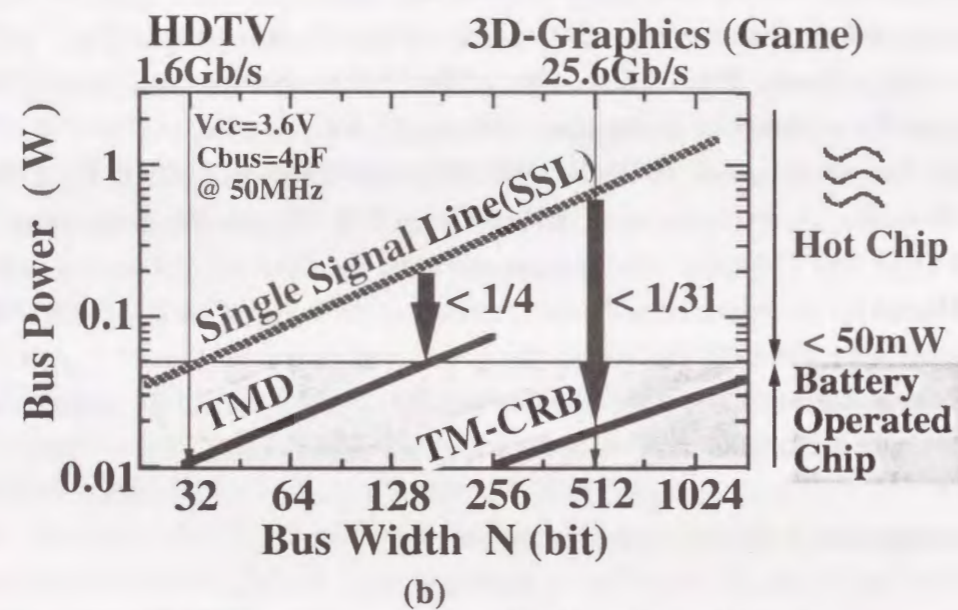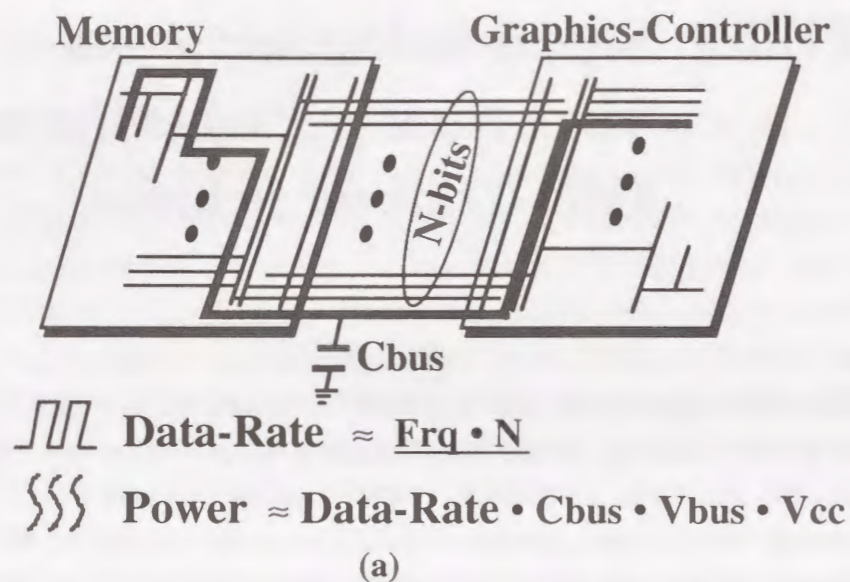
Memory   **Graphics-Controller**

Cbus

$\square$ **Data-Rate** $\approx$ **Frq • N**

$\lesssim$ **Power** $\approx$ **Data-Rate • Cbus • Vbus • Vcc**

(a)

(b)

(c)

Fig.3-1. (a) Background for low-power bus architecture.
(b)Target on bus-power consumption for this work.
(c) Percentage of wiring area versus the number of wirings.

frequency. This problem becomes more serious as the supply voltage approaches the sum of the threshold voltage. This is because the speed-degradation due to the reduced supply voltage, will be intolerably increased. (eventually, it will be exponentially increased.)

On the other hand, the differential data transfer scheme, which is almost used in the memory circuits, including DRAM[3] and SRAM, is clearly a very effective way to reduce the bus power consumption on heavily loaded data lines. This is because the differential data transfer scheme inherently has higher noise margin and can reduce the signal voltage-swing when compared against single signal line (SSL) scheme (shown in Figs. 3-2(a) and 3-3(a)). However , the method of power savings by using the differential data transfer scheme has the overhead of increased the number of signal wires by a factor of 2 just same as using the parallel architecture.

Hence, to solve this, we proposed a time-multiplexed differential data-transfer (that's called TMD) scheme in this paper. This features a shared complementary wires between adjacent signal bits, which are originally allocated to adjacent signal bits, respectively. This can realize a virtual differential data transmission by using the time-multiplexed adjacent wires, while maintaining the same number of signal wires and the same throughput[4].

In this paper, the background and target in our low power strategies are briefly discussed in the following section. After that, the concept of a time-multiplexed differential data-transfer (TMD) scheme is described in Section 3-3. In Section 3-4, the TMD scheme combined with Charge-Recycling Bus (CRB) architecture, making it possible to further reduce the bus power consumption, is discussed. The circuit operation and performance are demonstrated in Section 3-5. Conclusions are given in Section 3-6.

3-2   Background and Target

3-2-1   Bus Power Consumption

The ultra-high data rate of 25Gb/s and beyond is one of the most important design requirements in realizing the future ULSI's for real-time three-dimensional(3-D) computer graphics applications in consumer electronics including video game machine and handheld personal equipment's, such as palmtop PC's and portable multimedia access supporting full-motion digital video. The most effective means to achieve such a data rate is to employ a large number of buses, corresponding to the parallelism of N-bit, interconnecting the embedded memory, the graphics controller, *etc*, on a ULSI system chip, instead of increasing operating frequency as shown in Fig.3-1(a). For example,
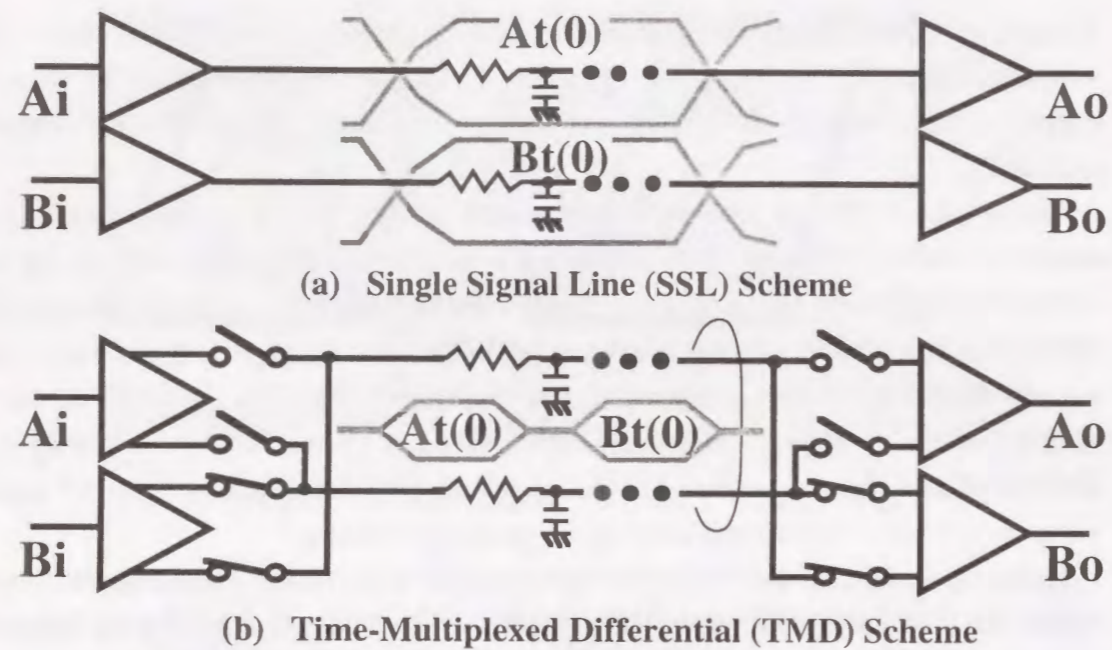
(a) Single Signal Line (SSL) Scheme

(b) Time-Multiplexed Differential (TMD) Scheme

Fig. 3-2 Concept comparison between (a) SSL and (b) TMD schemes.
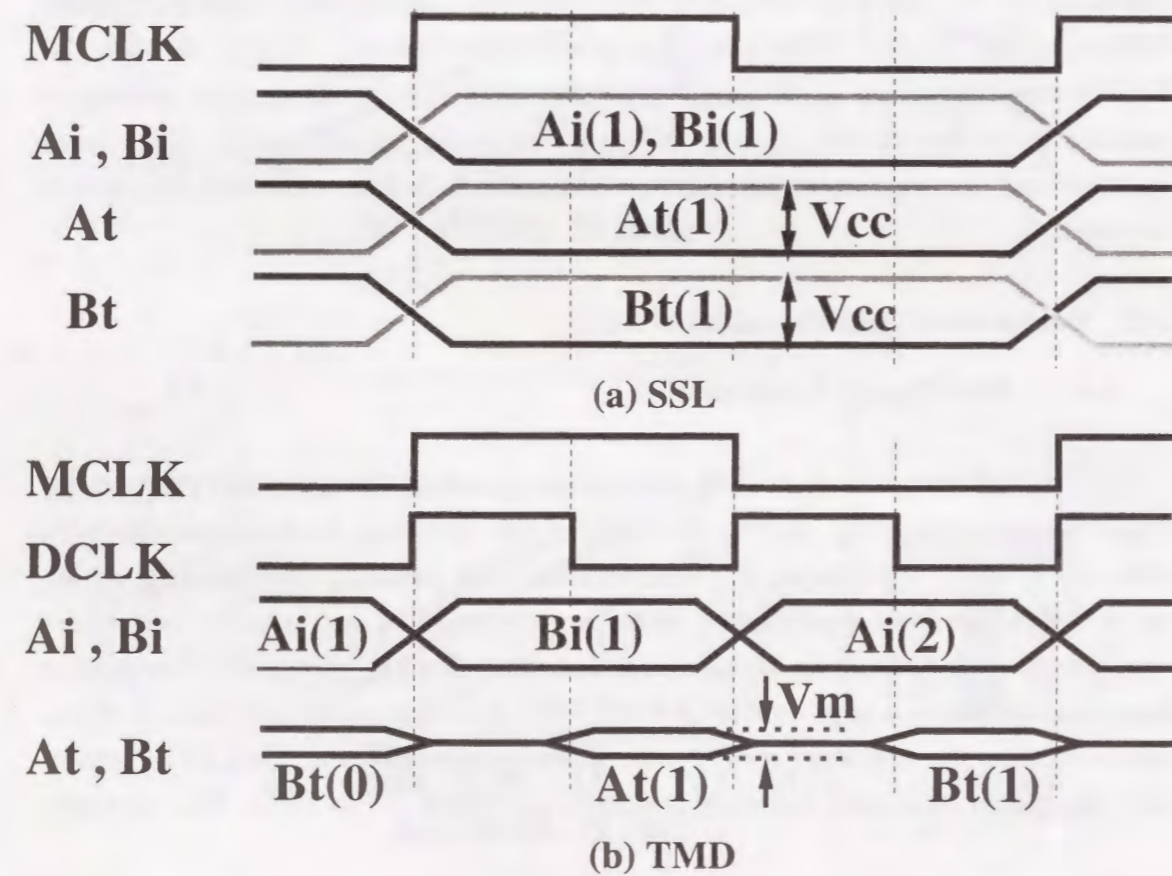

(a) SSL

(b) TMD

Fig. 3-3 Timing comparison between (a) SSL and (b) TMD schemes.

even at an operating frequency of 50MHz, parallel buses of more than 512bit are required to achieve the ultra-high data rate of 25Gb/s and beyond as shown in Fig.3-1(b). However, a drastic increase of the power dissipation which is direct proportion to the number of bus width N is inevitable due to the increased total bus-capacitance (N•Cbus), where Cbus is the bus-capacitance of each bus. Even if scaling the supply-voltage of Vcc down to 3.6V, there still remains a bus-power dissipation of a far above 50mW for the bus width of 512bit with bus-capacitance of 4pF per bit operating at 50MHz, which is intolerable for battery operation as shown in Fig.3-1(b). This is because of the restrictions on battery-life, battery size and weight for requirements of adequate portability.

As you can see from Fig.3-1(b), it is quite obvious that the new bus architecture is prerequisite to reduce the bus power to less than 50mW, instead of a single signal line (SSL) scheme(shown in Fig.3-2(a)). Thus, this paper proposes a time-multiplexed differential data-transfer (TMD) scheme that can realize the data-rate of around 2-Gbits/s and the TMD scheme combined with a Charge-Recycling Bus(CRB) architecture (that's called TM-CRB) that can further increase the data-rate to 25-Gbits/s, while meeting the requirements for bus-power dissipation of less than 50mW.

Before describing the proposed TMD and TM-CRB architecture, the comparisons of conventional power reduction techniques are briefly touched.

Table 3-I shows the comparisons of power reduction between the SSL scheme, parallel architecture, and non-time-multiplexed differential data-transfer scheme, assuming that the activity factor of data-lines (data-transition probability) is 100%.   As you can see from this table, the differential data transfer scheme is the most effective means to save the data-transmission power on heavily loaded data lines, rather than parallel architecture. This is because that the parallel architecture can save the power consumption to merely 36% of SSL even reducing the both supply voltage Vcc and bus swing voltage Vbus to 60% at the sacrifice of layout area due to the twice the number of signal wirings.

On the other hand, the differential data transfer scheme inherently can reduce the bus swing voltage Vbus up to 1/8 of the Vcc=3.3V, resulting in the 75% power reduction even at the twice the number of data transition due to the complementary data transitions at both of the falling and rising edges (Frq=2, shown in Table 3-I), compared to SSL. This is because that can provide a sufficient sensing noise margin of less than 100mV. Thus, it is quite obvious that the proposed TMD and TM-CRB architecture had better take advantage of the differential data transfer scheme for signal transmission through heavily loaded data lines, instead of parallel architecture.   This is the reason why the proposed schemes are based on the differential transfer scheme.

3-2-2   Layout Area Consumption

Table-3-I Comparisons of conventional power reduction techniques.

@Bus Activity = 100%

	SSL	Parallel	Differential
Frq.	1	0.5	2
N•Cbus	1	2	1
Vbus	1	0.6	0.125 (1/8)
Vcc	1	0.6	1
Power	1	0.36	0.25 (1/4)
Wire	1	2	2

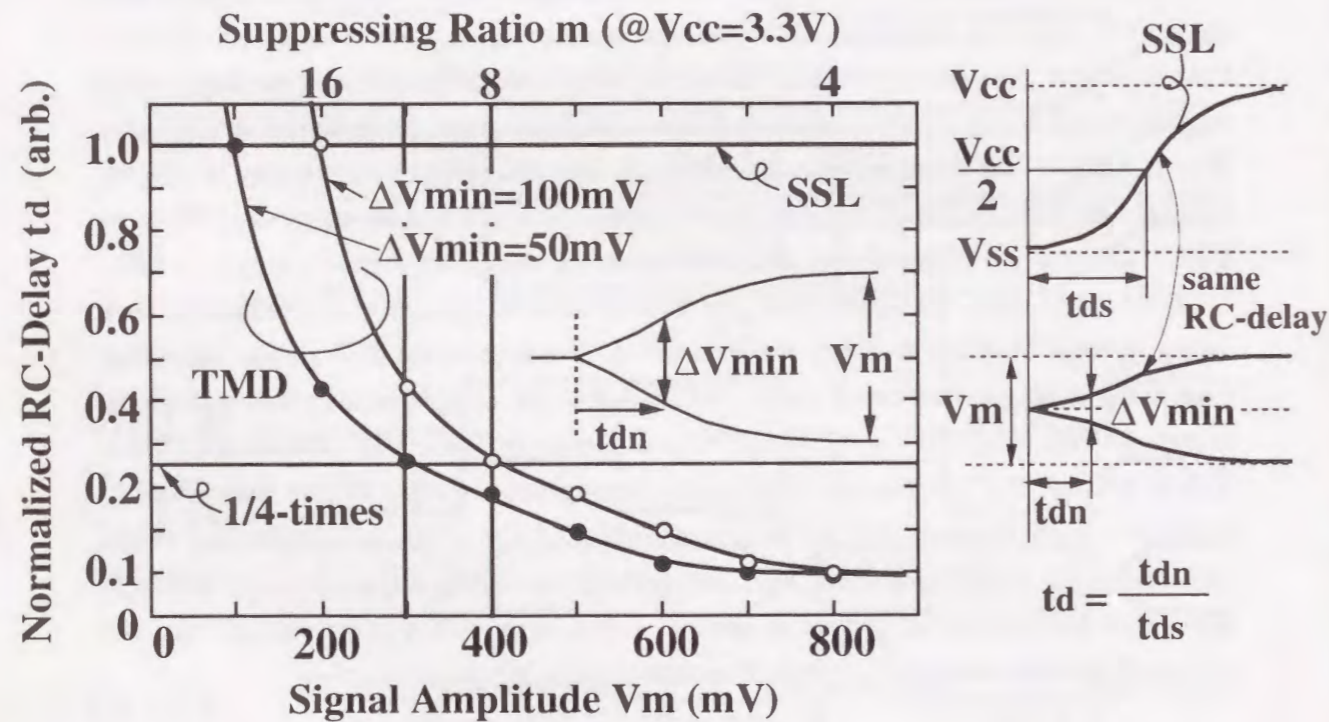❏ Power ≈ Frq • N• Cbus • Vbus • Vcc   @ Vcc=3.3V

Fig. 3-4 Normalized RC-delay vs. Transferred signal amplitude.

The most disadvantageous point in using the non-time-multiplexed differential transfer scheme or parallel architecture is the requirements of twice the number of signal wirings.   That directly causes serious penalty of wiring layout area as shown in Fig.3-1(c).   For example, when realizing the data-rate of 25-Gbits/s by using the data lines of 512bit , the bus wiring area would occupy up to 25% of 96 mm^2 even when the wiring pitch is scaled to 1.6μm and the data buses are multiplexed for read and write operation, as shown in Fig.3-1(c). This is really intolerable for the area-constrained chip, such as a integration chip of DRAM and VGA graphics controller (shown in Fig.3-1(a)).

This is the reason why the TMD and TM-CRB architectures that features the wiring-shared architecture have been proposed in this paper.

Furthermore, to meet the challenges of maintaining the same throughput without increasing the number of wires and of reducing bus power consumption, simultaneously, when compared against SSL, we have also proposed a half level precharging scheme (that's called HLP) and data-transition detector (that's called DTD), those of which are used in the TMD scheme.   DTD scheme is not employed into TM-CRB.   This is because DTD scheme disturbs the charge-recycle operation. (i.e. when DTD works, the charge will not be transferred to lower stage)

3-3   Concept of Time Multiplexed Differential Data Transfer (TMD) Scheme

3-3-1   TMD architecture

The key concept of TMD is "sharing of time and wiring" between the two adjacent wires, which are originally allocated to each signal (Ai,Bi) in a clock period (T) to realize a virtual differential data transmission as illustrated in Figs.3-2(b) and 3-3(b). That is, a differential data transmission can be realized without doubling the number of wires as shown in Fig.3-3(b) by multiplexing the two adjacent wires which are originally allocated to the signal of Ai and Bi between the former half and the latter half of a clock cycle, respectively.

TMD can meet the following requirements simultaneously : 1) suppressed voltage-swing of Vm, 2) no increase in the number of signal-wires, and 3) no throughput-loss.

The most important concern in the TMD circuit is the throughput-loss due to the time-multiplexing (shown in Fig.3-3(b)). To clarify this, a RC-delay comparisons between the SSL and TMD schemes are carried out, as shown in Fig.3-4. The delay-time t_{ds} for SSL is defined as the time required to undergo a signal transition of a half of a CMOS level, as shown in Fig.3-4. (e.g. from Vss to 1/2Vcc or from Vcc to 1/2Vcc).   On the

other hand, the t_{dn} for TMD is defined as the time required to achieve the differential swing of $\Delta Vmin=100mV$, or 50mV, independent of signal amplitude Vm, as shown in Fig.3-4. For all comparisons between the t_{ds} for SSL and the t_{dn} for TMD, the same RC-delay conditions (i.e, the same values of resistance's of transistor and interconnect, and the same value of interconnect capacitance) were used. Here, $\Delta Vmin=50mV$ or 100mV is the required voltage difference necessary to detect a signal value with a sufficient noise margin of the latched current sense amplifier [1] shown in Fig.3-7(b). The t_{dn} of TMD can be reduced to less than 1/4 of the SSL's t_{ds} when $Vm \geq 400mV$ and $\Delta Vmin=100mV$. This implies that a suppressing ratio of m=8 (m=Vcc/Vm) is possible at Vcc=3.3V even while undergoing 4 transitions (i.e. 2 equalizing and 2 developing) in the period T, which are required in the TMD scheme as shown in Fig.3-3(b). In other words, TMD can recoup the throughput-loss due to the time-multiplexing, by exploiting the DCLK clock of the doubled operating frequency (shown in Fig.3-3(b)), which can be achieved by synchronizing with the both of rising and falling edge of MCLK, while making it possible to reduce the bus power consumption to 1/4 of the SSL.

3-3-2 Half-level Precharging (HLP) Scheme

The increased power problem by a factor of 2 due to the doubled operating frequency for TMD, is briefly discussed here. The TMD scheme just solves this problem by using the half-level precharging (HLP) scheme. The HLP scheme features the pre-equalization of the differential bus in the former half of the clock cycle. That makes it possible to precharge to the half level of swing voltage Vm with almost zero power resulting from the charge-sharing (i.e. charge-recycling) between the differential bus pair as shown in Fig.3-5. Thus, HLP can reduce the charging amount by half per bit. As a result, in totality, the increased charging amount caused by the doubled operating frequency is just canceled resulting from using the HLP scheme, when compared against a static operation without the pre-equalizing.

3-3-3 TMD with data transition detector (DTD) scheme

Another important concern in the HLP scheme is the power loss due to the pre-equalization in the former half of the clock cycle when data successively transferred are the same.

As mentioned earlier, the HLP scheme can save the charging amount by half necessary to establish the "H" level on heavily data lines resulting from using the HLP scheme, when compared against a static operation without pre-equalizing. However,
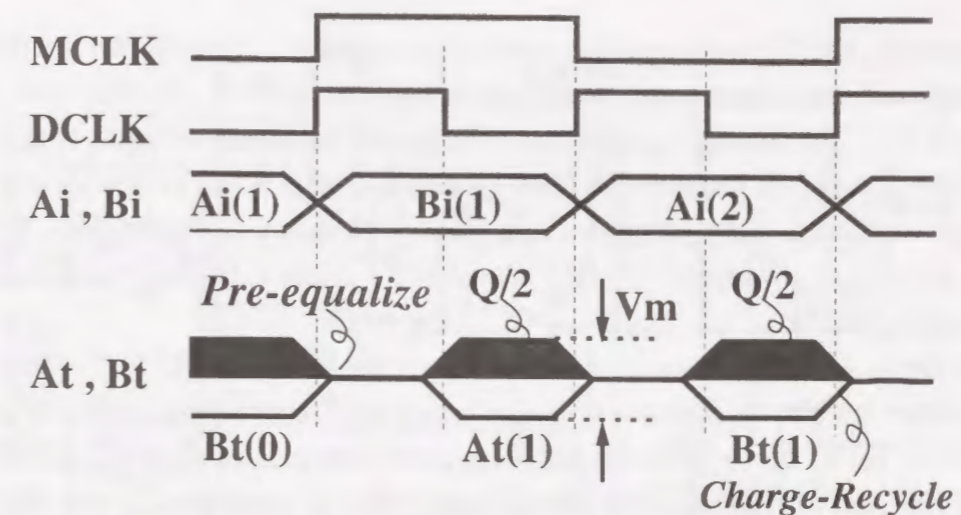

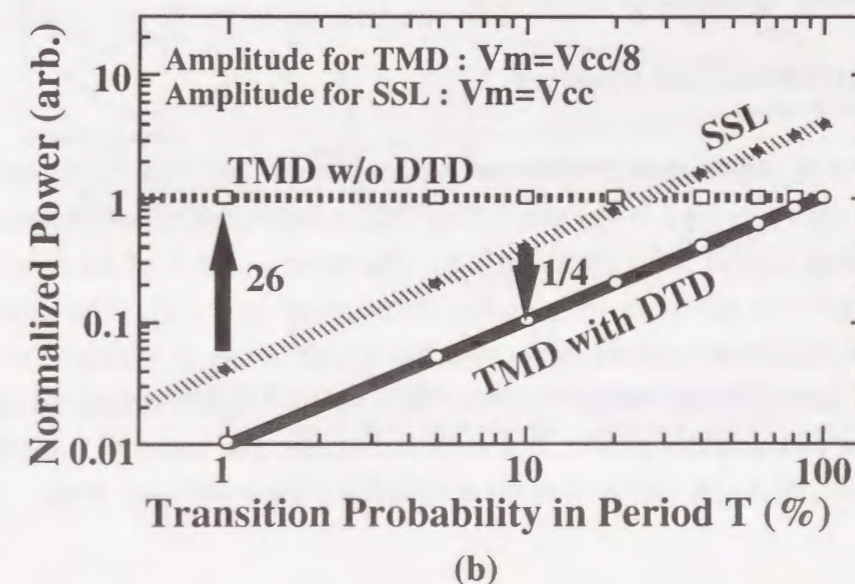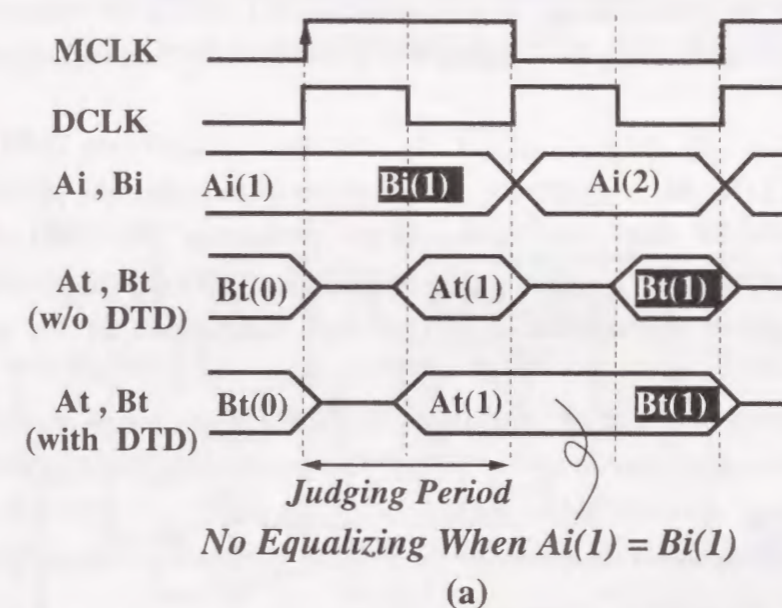Fig. 3-5. Concept of Half-Level Precharging (HLP) scheme

(a)

(b)

Fig.3-6. (a) Concept of TMD with data transition detector (DTD), (b) Power reduction capability of TMD with DTD.

unfortunately, that depends on the transition probability. That is, the precharging operation will be wasted when data successively transferred are the same (e.g., Ai(1)=Bi(1)). For example, even when suppressing the signal voltage-swing to 1/8, the power dissipation becomes 26-times larger than SSL with full swing when the transition probability in the period T is less than 1% as shown in Fig.3-6(b). To solve this, a data transition detector (DTD) (shown in Fig.3-7(a)) has been developed and employed in the TMD.

Figure 6(a) shows the concept of the TMD combined with DTD. The TMD with DTD scheme features the pipe-line operation between the input latching of Ai, Bi and the line driving of At, Bt, as shown in Fig.3-6(a). For example, taking a look at the input signal of B, the latching operation of input data Bi(1) is synchronized with the rising edge of main clock MCLK. On the other hand, line driving operation of transferred data Bt(1) is synchronized with the falling edge of MCLK. Thus, the half of the clock cycle can be allocated for the time necessary to judge whether Ai(1) equals the successive data Bi(1) or not. As a result, when Ai(1) equals Bi(1), successive pre-equalizing operation is just canceled.

Figure 3-6(b) shows the comparisons of bus power consumption of the TMD between the cases of using DTD and without that, as a function of the transition probability in successively transferred data transmission. (e.g., probability for TMD of when Ai(1)=Bi(1), Bi(1)=Ai(2),...) Comparing with SSL, TMD combined with DTD reduces the bus power consumption to 1/4 of that, independent of the transition probability.

The transition detector consists of a exclusive-OR (XOR) whose inputs are data Ai, Bi successively transferred as shown in Fig.3-7(a). The equalizing signal EQ cancels the successive equalizing operation when the output of the XOR is "L" (when Ai=Bi). Thus, TMD can save up to 1/4 the power of the SSL, independent of the combination of input data.

3-3-4 TMD Driver and Receiver

To realize the TMD scheme, time-multiplexing driver and receiver circuits were developed as shown in Figs.3-7(a) and 3-7(b). These circuits are synchronized with the rising and falling edge of main clock (MCLK). The driver consists of the pair of input-data latch circuits, the pair of driver switches, an equalizer, and DTD. The latch circuits hold the input-data (Ai,Bi) and the driver switches connect either of wiring pairs (At, Bt) with either of scaled supply voltages (Vu for "H", Vb for "L") depending on the input-data except for an equalizing period, as shown in Fig.3-8. The equalizer precharges the signal wires (At, Bt) to the half level of the differential voltage-swing of Vm.
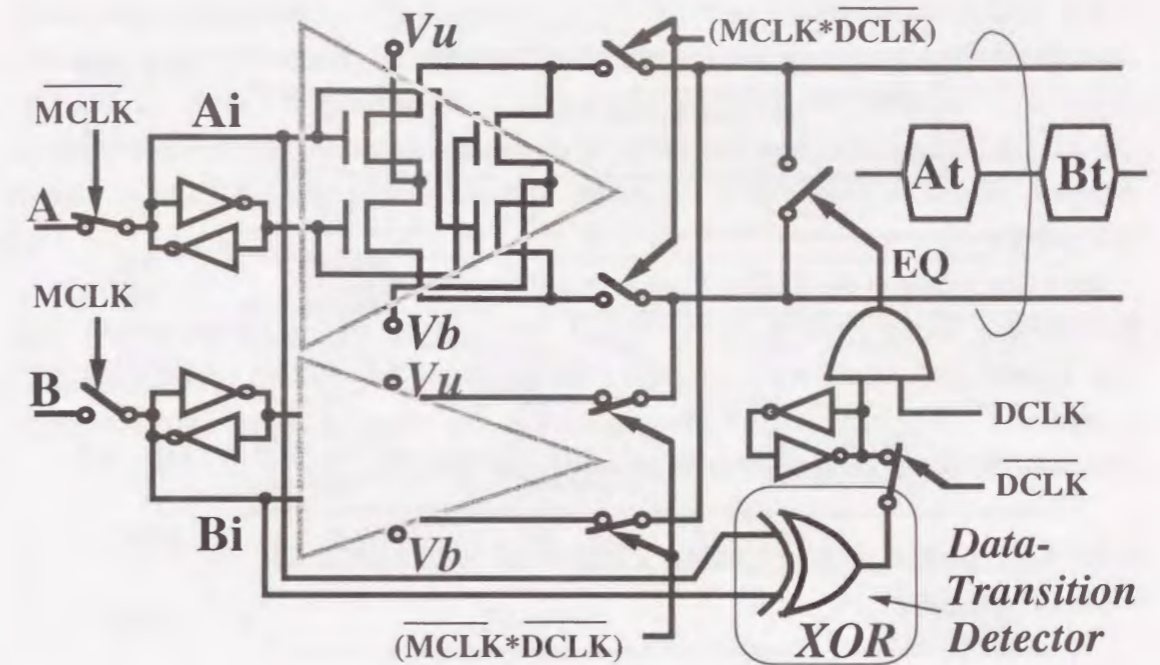
(a)
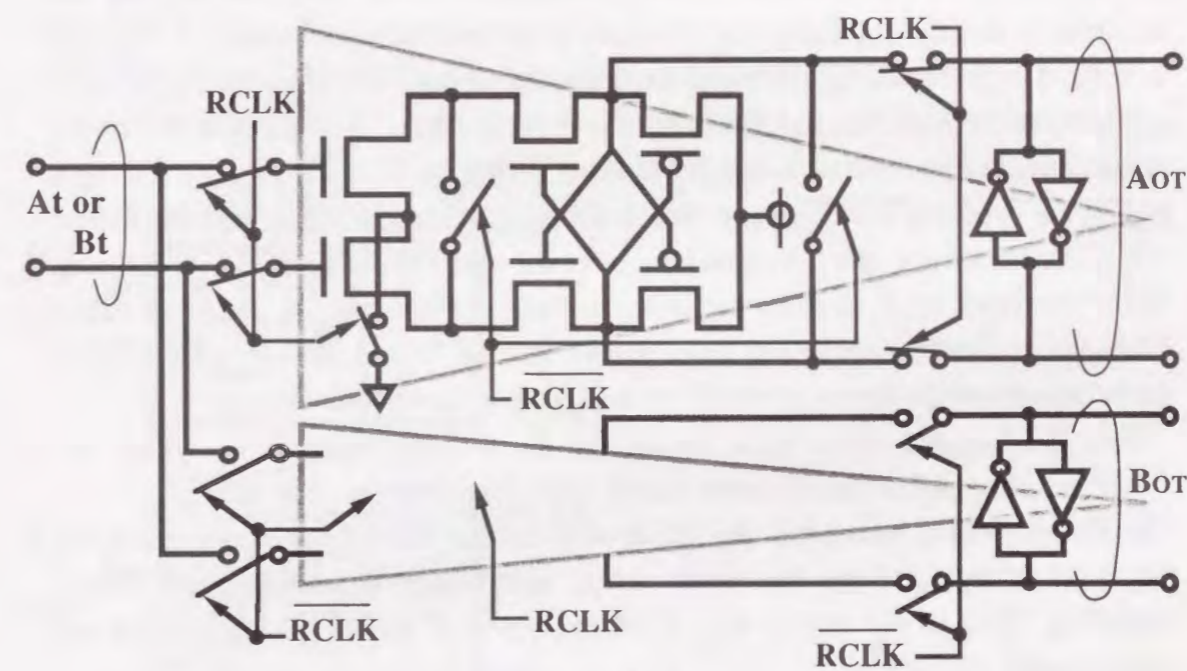
(b)

Fig.3-7. (a) TMD circuit configuration of driver and (b)TMD circuit configuration of receiver

The receiver circuit shown in Fig.3-7(b) is composed of a latched type current sense amplifier and a gate-receiver that functions as a converter of voltage difference to current difference[1] between the differential signal lines.　The sensing of transferred data (At, Bt) and the latching of output data (A_{OT}, B_{OT}) are synchronized with the RCLK clock, which is delayed by a factor of quarter of the clock cycle compared to MCLK, as shown in Fig.3-8.

Here, the timing of the TMD scheme is briefly described.

For example, taking a look at the signal of B, the input data B(1) is latched as Bi(1) and that is driven to signal wirings as Bt(1) at the rising and falling edge of MCLK, respectively.　The transferred Bt(1) is latched as the output of $B_{OT}(1)$ just one clock later from the timing of input latching as shown in Fig.3-8.

3-4　Low power strategy using TMD combined with CRB (TM-CRB)

3-4-1　Concept of TM-CRB

The most attractive point of the TMD in realizing low-power data transmission is the capability of directly employing our previously proposed CRB architecture[1]. This can be realized by just stacking numerous driver circuits (shown in Fig.3-7(a)) into a series configuration between Vcc and Vss as shown in Fig.3-10(a).　For example, Vbl of the upper driver is connected with Vu2 of the lower driver, as shown in Fig.3-9. Taking a look at the operating waveforms of the output pairs (shown in Fig.3-9) in the former half of the clock cycle, the differential bus pairs are equalized, respectively. In the latter half of the clock cycle, the bus switching is controlled according to the input value. When the inputs of the upper and lower circuits become "0" and "0", respectively, four circled transistors are turned on as shown in Fig.3-9.

And, the bus At1 in the upper circuit and the bus /At2 in the lower circuit are connected and equalized based on the charge-sharing between the two.

This charge-sharing establishes the "L"-level of the At1 for the upper circuit and the "H"-level of the /At2 for the lower circuit, simultaneously, without new charge supplying. This is just reason why CRB can save the power dissipation for data transmission.

This implies that the time-multiplexed CRB (TM-CRB) (shown in Fig.3-10(a)) can be realized without increasing the number of wires per bit and throughput-loss, unlike the conventional CRB architecture[1]. Since each suppressed voltage-swing Vm is achieved by only the charge sharing between adjacent signal-bits(e.g. when "L" of A0 and "H" of A1 transitions) instead of the power line, TM-CRB can further reduce the
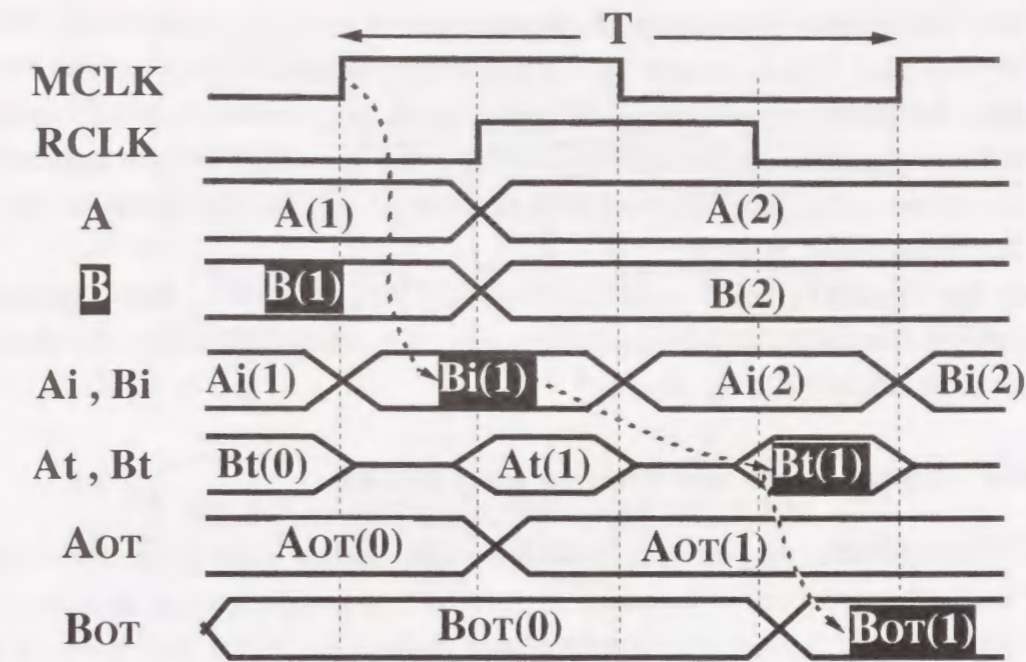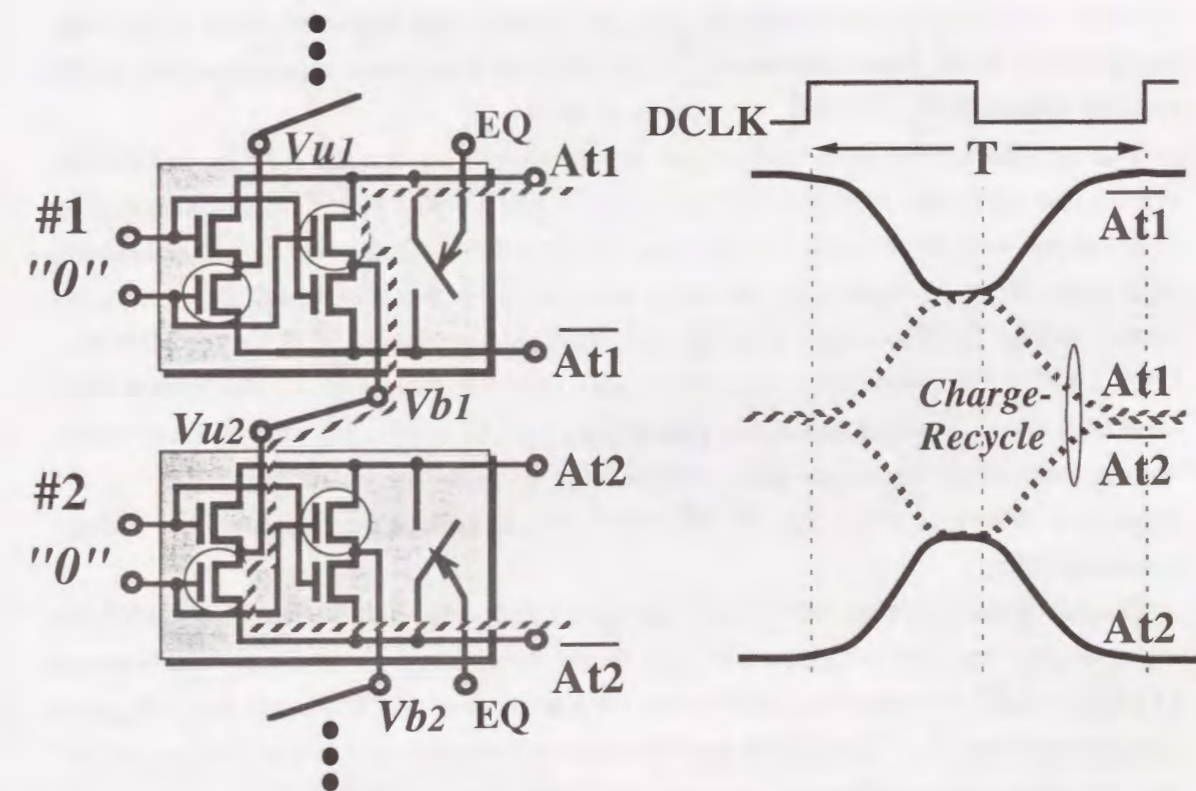

Fig. 3-8　Timing diagram of TMD scheme operation

Fig. 3-9　Concept of Charge-Recycling Bus (CRB) architecture

power to 1/n (n equals the number of elements in each stack) compared to the TMD without CRB case. This is because TM-CRB no longer requires charge supplying from the power line necessary to undergo a "H" transition in each element, since the recycled charge from the adjacent upper element is supplied to the lower element so as to undergo a "H" transition in that, except for the top element (for A0 and B0) that constitutes 1/n of all of the elements in each stack.

Note that the DTD scheme can not be employed into TM-CRB. This is because DTD scheme disturbs the charge-recycle operation. (i.e. when DTD works, the charge will not be transferred to lower stage.)

3-4-2 CMOS Driver and Receiver for TM-CRB

The relation between the gate types of the driver and receiver in the upper (operated in $\geq 1/2$Vcc) and lower circuits (operated in $\leq 1/2$Vcc) is complementary as shown in Fig.3-10(b) and Fig.3-10(c), that is, PMOS gate drivers and NMOS gate receivers are used for the upper bus pairs, whose each operating bus potential is higher than 1/2Vcc, and NMOS gate drivers and PMOS gate receivers are used for the lower bus pairs whose each operating bus potential is lower than 1/2Vcc. Note that the driver and receiver shown in Figs.3-7(a) and 3-7(b) are named "NMOS gate driver" and "NMOS gate receiver", respectively, associated with the "N"-channel type transistor pairs, composing bus-switches in the driver (shown in Fig.3-7(a)) and composing input-gate pairs in the receiver (shown in Fig.3-7(b)).

The complementary input and output signal waveforms for the NMOS and PMOS drivers and receivers are shown in Figs. 3-10(b) and 3-10(c). The CMOS configuration in the upper and lower driver circuits can provide a low-voltage operation by obtaining of a larger V_{GS} of input-gate pair more than Vcc/2 for both types of the circuits, as shown in Fig. 3-10(b). Here, note that low-threshold voltage (e.g. 0.3V for NMOS, - 0.3V for PMOS) is introduced into the drivers except for the drivers of the bottom stage (NMOS-Driv.(1)) and the top stage (PMOS-Driv.(1)). Since the potential of the common source terminal of input-gate pair is shifted up to over Vm, resulting in the V_{GS} of negative (-Vm) value when V_G=0V, the low-threshold voltage never induces the leakage current problem.

The bus receiver is required for constant speed and stable amplification over the wide input-voltage range of Vcc. To meet this requirement, the receiver circuit is composed of a latched type current sense amplifier and a gate-receiver that functions as a voltage to current converter[1]. The CMOS configuration in the upper and lower receiver circuits can provide a fast operation by obtaining of a larger V_{GS} of MOSFET more than Vcc/2 for both types of the circuits, as shown in Fig.3-10(c).
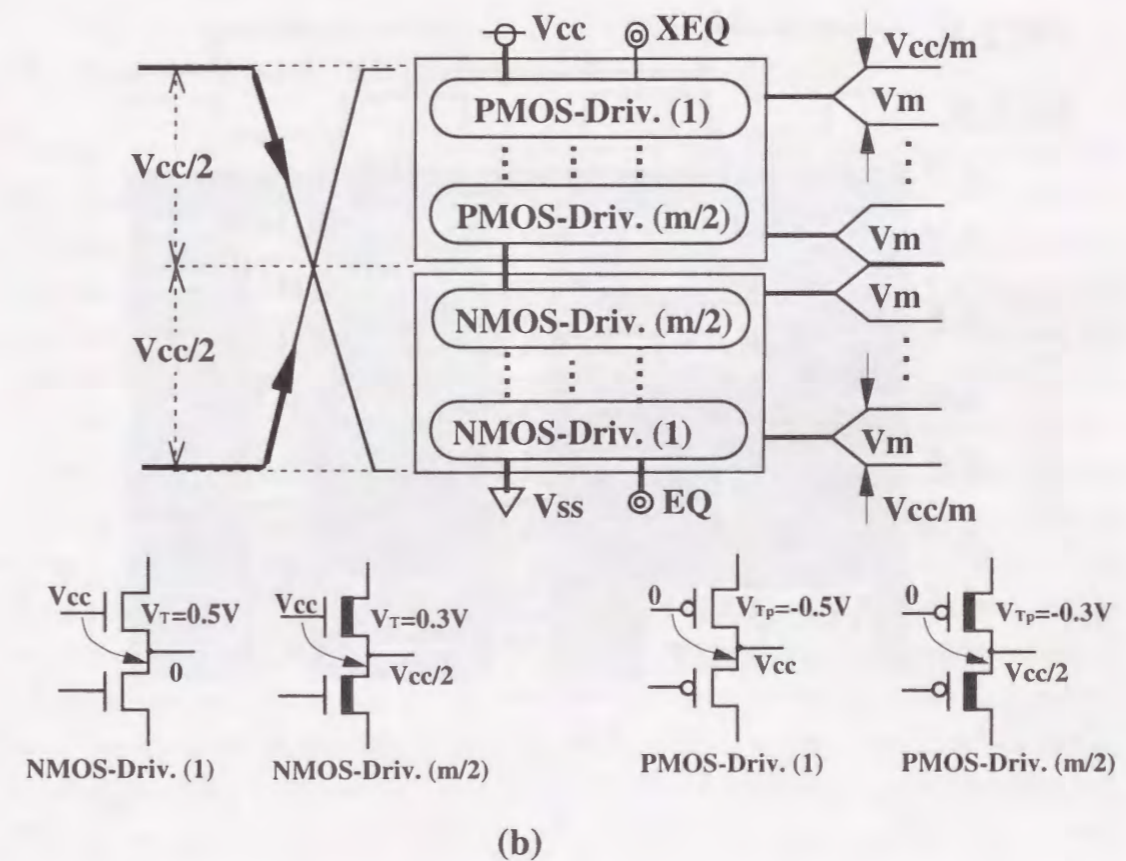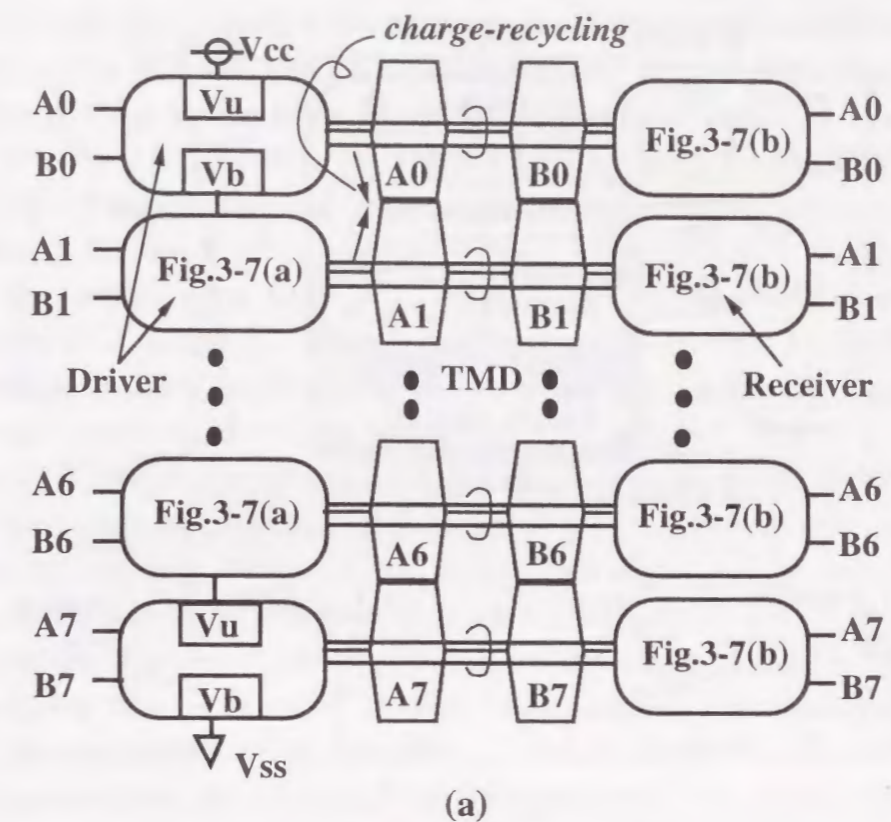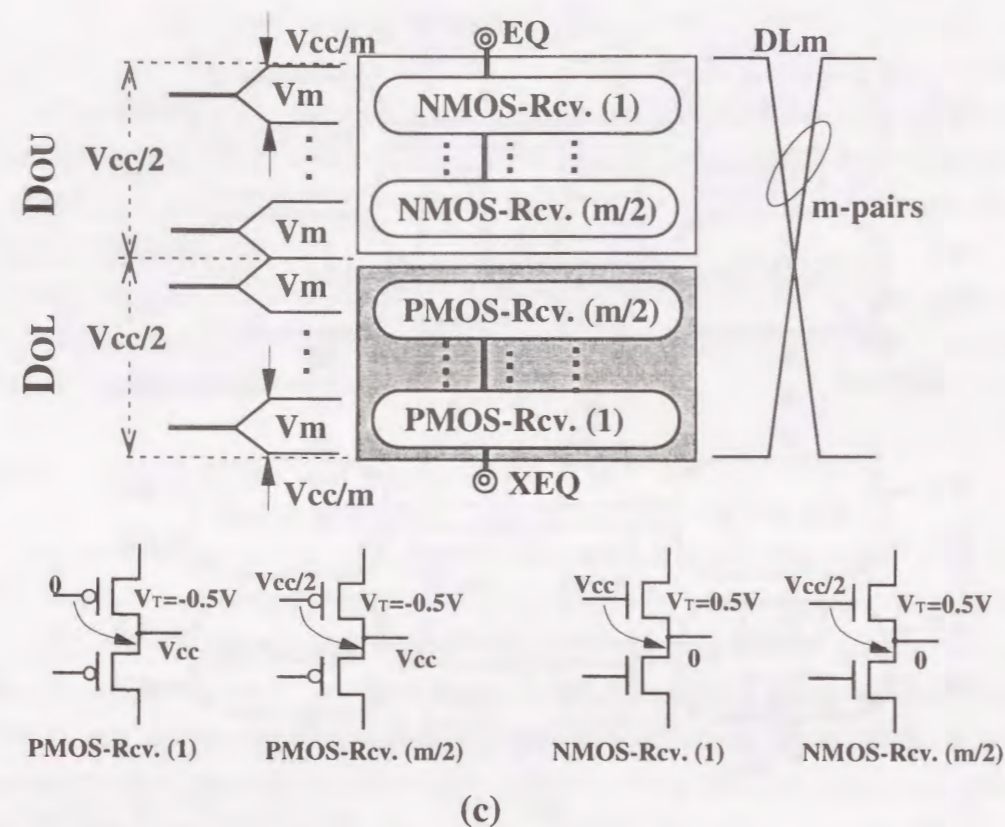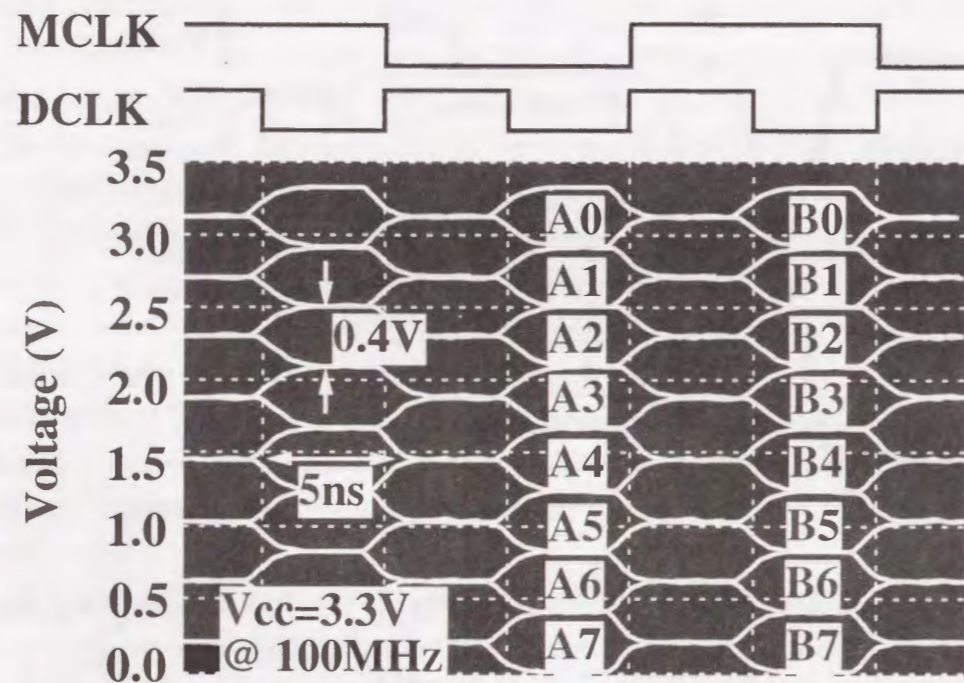

(a)

(b)

Fig.3-10. (a) Time-multiplexed CRB (TM-CRB) configuration and (b) CMOS driver configuration.

Fig.3-10. (c) CMOS receiver configuration and
(d) Simulated operating waveforms of TM-CRB.

The CMOS configuration in the upper and lower driver circuits enables to lower the voltage supply limitation to 1.2V, assuming that the required effective gate to source voltage $V_O=V_{GS}-V_T$ necessary to provide a stable operation is 0.3V. Note that $V_{GS}=Vcc-Vcc/2$ and $V_T=0.3V$ for NMOS-Driv. (m/2) shown in Fig.3-10(b). On the other hand, for the receiver circuits, the voltage supply limitation is reduced to 1.2V by employing the low-V_T (0.3V for NMOS, -0.3V for PMOS) into the input-gate transistor, assuming that $V_O=0.3V$. Since the number of transistor with the low-V_T is limited to 1024, resulting in the total channel width of less than 10 mm even if 512bit, the leakage amount is less than $0.1\mu A$, which is negligible small when compared against the total current dissipation of 6mA (shown in Fig.3-1(b)).

Figure 3-10(d) shows the simulated operating waveforms of the 16-bits, numbering from A0 through A7 and from B0 through B7.
When the operating voltage of Vcc is 3.3V, the swing voltage of each bus was automatically fixed to about 0.4V as shown in Fig.3-10(d).

Stable data-transfer of 16-bits with the bus-capacitance of 4pF per bit, which is synchronized with the DCLK of 100MHz, can be achieved, while maintaining the power dissipation to be less than 0.4mA (including 0.2mA in the receiver circuits). The speed distribution between the A0 through the B7 is negligibly small, when compared against 5ns of the half of the clock cycle.

3-5 Power and Area Comparisons

Table 3-II shows the comparisons of power and layout-area consumption, between SSL, parallel architecture, TMD, and TM-CRB at the Vcc of 3.3V. According to the simulated data based on the test devices designed by using 0.5μm CMOS technology, TMD can save up to 1/1.4 the power and 1/2 the Si-area (wiring number) of the parallel architecture at Vcc=3.3V assuming that the bus activity is 100%. This is due to the dramatic reduction in the signal voltage swing on heavily loaded data lines resulting from using the time-multiplexed shared differential transmission lines between adjacent bits.

TM-CRB can further reduce the power consumption to 1/11 of the parallel architecture, assuming that the bus activity is 100%, while maintaining the same throughput as shown in Fig.3-11. This is because the parallel architecture can reduce to merely 0.36 of the SSL's power, even when scaling the supply voltage (Vcc) to 0.6 through doubling of the number of wires, as shown in Table 3-II. On the other hand, TM-CRB is able to reduce the power consumption to 1/31 even at the 4-times the number of data transition due to the time-multiplexed complementary data transitions at both of the 2-falling and 2-rising edges (Frq=4, shown in Table 3-II), resulting from using charge-recycling and reducing the signal-voltage swing to 1/8 of Vcc, when compared against SSL. Note

Table 3-II Power and Area Comparisons among various low power techniques

@Bus Activity = 100%

	SSL	Parallel	CRB	TMD	TM-CRB
Type	*Single*	*Single*	*Different.*	*Different.*	*Different.*
Frq	1	1/2	2	4	4
N• Cbus	1	2	1/8	1	1/8
Vbus	1	0.6	1/16	1/16	1/16
Vcc	1	0.6	1	1	1
Power	**1**	**0.36**	**0.016**	**0.26**	**0.032**
Area	1	2.2	< 2.1	< 1.2	< 1.2

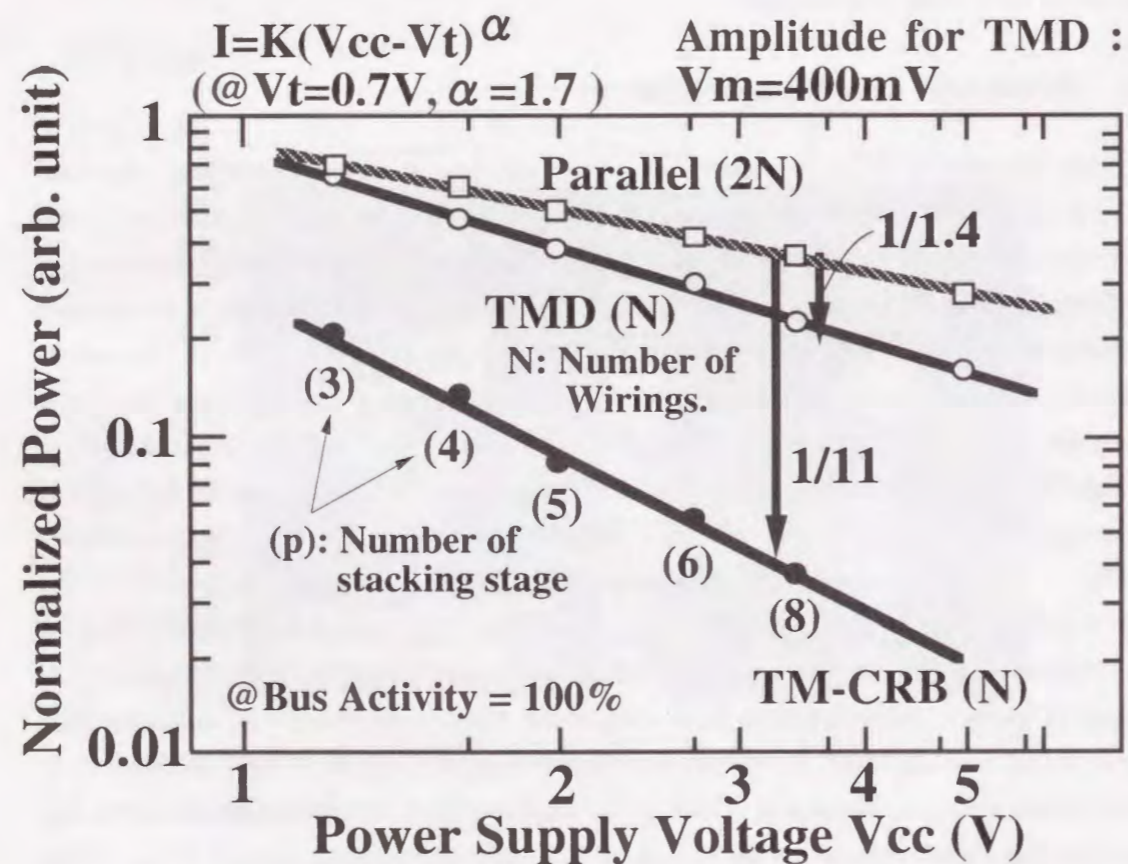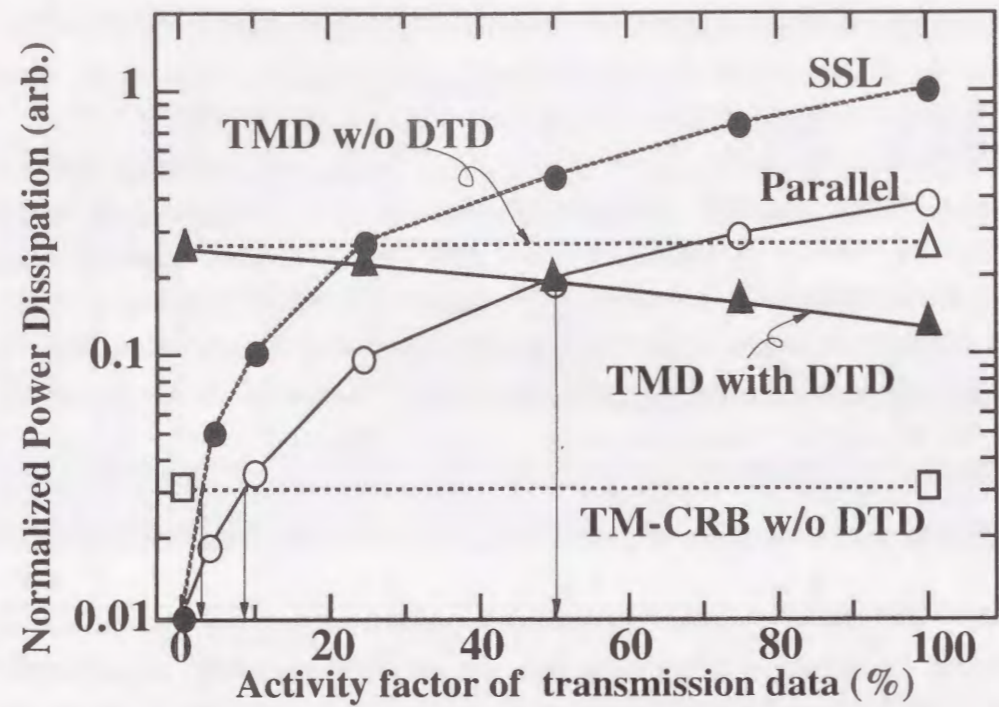☐ **Power ≈ Frq • N• Cbus • Vbus • Vcc**
@Vcc=3.3V, m=8, 2Vbus =Vm=Vcc/8

Fig. 3-11 Comparisons of power reducing capability compared to SSL
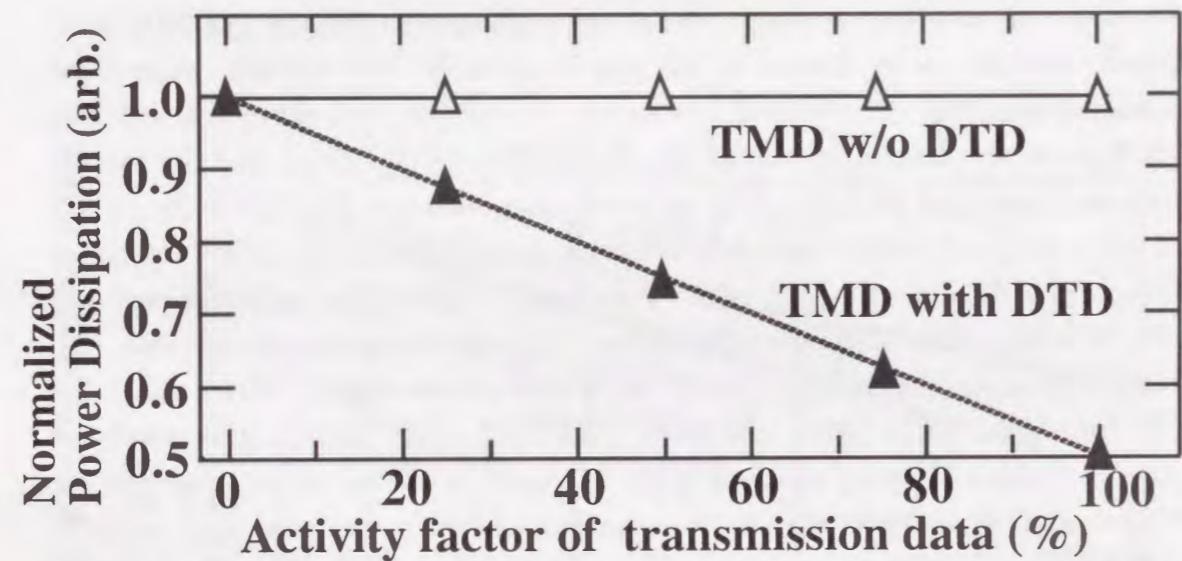
that N•Cbus=1/8 for CRB or TM-CRB compared with others (shown in Table 3-II) is caused by the charge-recycle data transmission, which no longer requires the charge-supplying for each bus pair from the power line, except for the top bus pair.

Another interesting point shown in Fig.3-11, is that the capability of power reduction in TMD, TM-CRB, and parallel architecture, are decreased as the Vcc approaches the sum of thresholds V_T in CMOS gates, or the suppressed signal amplitude Vm of 400mV. This is because the stacking number (p) allowable to maintain the Vm of 400mV is reduced to 3 at Vcc=1.2V, while p=8 at Vcc=3.3V for TM-CRB. For parallel architecture, the degradation of the scaling capability of the Vcc necessary to maintain the throughput, causes less power savings. However, even at Vcc of 1.2V, TMD and TM-CRB can still achieve greater power savings than parallel architecture, while reducing the number of signal wires by half, assuming that the bus activity is 100%.

Figure 3-12(a) shows the power dissipation versus activity factor of transmission data. When taking the activity of the busline (i.e. the transition probability of successively transferred data in Ai or Bi) into consideration, the power consumption in SSL is direct proportion to the activity of the busline. On the other hand, since the time-multiplexed bus spoils the statistical correlation, the power dissipation in TMD or TM-CRB never directly depends on the activity of the data in Ai or Bi. For example, even if the individual activity of Ai or Bi is "0", the activity of multiplexed bus between Ai and Bi is 100% when Ai is inverse of Bi (e.g. Ai= all-"0", Bi= all-"1"). Thus, assuming that the individual activity of Ai or Bi is 25%, the power consumption in TM-CRB is 1/8 and 1/3 of that in SSL and of that in parallel architecture, respectively, which are still significant improvement, as shown in Table 3-III. Note that the DTD scheme can not be employed into TM-CRB. This is because DTD scheme disturbs the charge-recycle operation. (i.e. when DTD works, the charge will not be transferred to lower stage.) On the other hand, TMD provides a greater power savings by using DTD, even when taking the activity of Ai or Bi into consideration, as shown in Fig.3-12(a). Assuming that the worst case for the time-multiplexed bus is when the Ai is inverse of Bi, the power consumption in TMD with DTD is reduced in inverse-proportion to the activity factor of Ai and Bi, as shown in Fig.3-12(b). When the bus activity is larger than 50%, the TMD with DTD has a more capability of power savings compared to the parallel scheme, as shown in Fig.3-12(a). Data invert coding scheme[6] seems to be necessary to avoid the above mentioned worst case (when the Ai is inverse of Bi). This scheme tends to prevent the the statistical correlation between successively transferred data from spoiling, resulting from data invert pre-decoding so as to have the same transmission data between Ai and Bi as much as possible. When using this scheme, the power consumption in TMD with DTD is reduced in proportion to the activity factor of Ai and Bi (no longer in
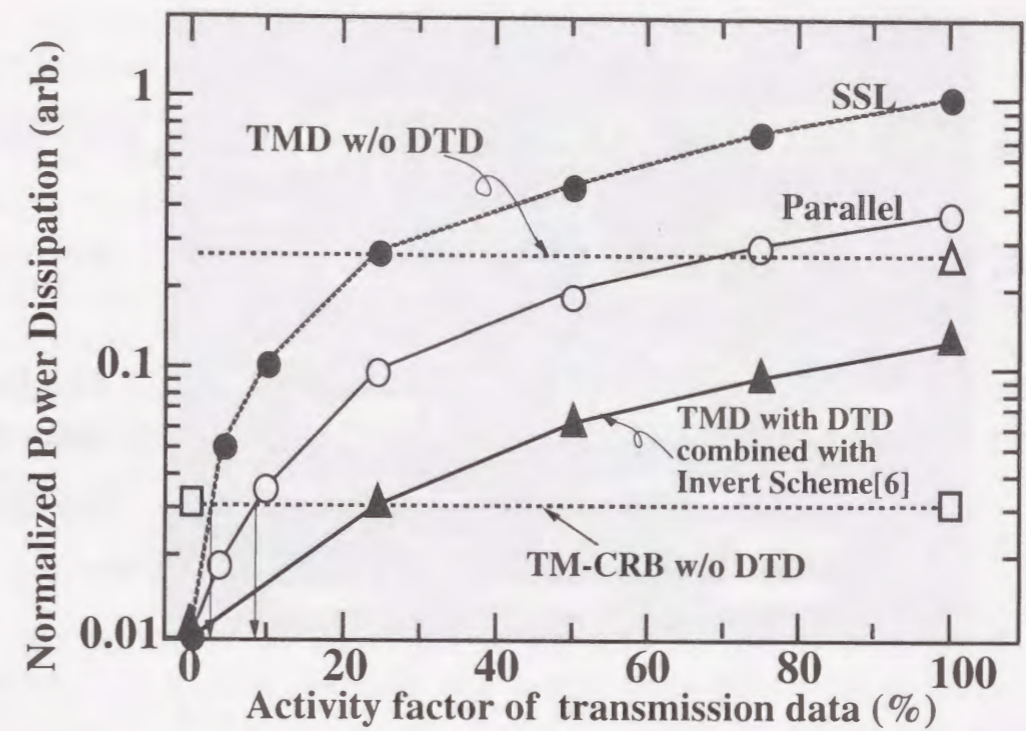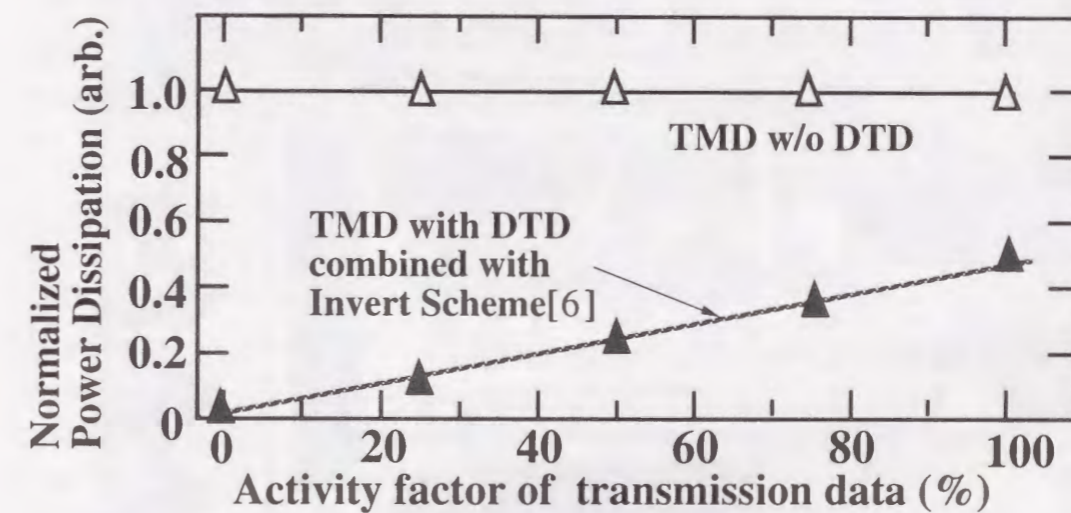
Figure 3-12 (left):

(a)

Normalized Power Dissipation (arb.)

1

SSL

TMD w/o DTD

0.1

Parallel

TMD with DTD

0.01

TM-CRB w/o DTD

0   20   40   60   80   100
Activity factor of transmission data (%)

(a)

Normalized Power Dissipation (arb.)

1.0

TMD w/o DTD

0.9

0.8

0.7

TMD with DTD

0.6

0.5

0   20   40   60   80   100
Activity factor of transmission data (%)

Ai 0 1 0 1 · · · Activity factor of Ai and Bi is 100%
                       No transition occurs every 2-bits
Bi 1 0 1 0 · · · when using DTD.

Ai 0 0 1 1 · · · Activity factor of Ai and Bi is 50%
                       No transition occurs every 4-bits
Bi 1 1 0 0 · · · when using DTD.

(b)

Fig.3-12 (a) Power savings versus activity factor and
          (b)Power dissipation comparison between with and without DTD.

92

Figure 3-13 (right):

Normalized Power Dissipation (arb.)

1

SSL

TMD w/o DTD

0.1

Parallel

TMD with DTD
combined with
Invert Scheme[6]

0.01

TM-CRB w/o DTD

0   20   40   60   80   100
Activity factor of transmission data (%)

(a)

Normalized Power Dissipation (arb.)

1.0

TMD w/o DTD

0.8

0.6

TMD with DTD
combined with
Invert Scheme[6]

0.4

0.2

0

0   20   40   60   80   100
Activity factor of transmission data (%)

Ai 0 1 0 1 · · · Activity factor of Ai and Bi is 100%
Bi─▷○─ B̄i 0 1 0 1 · · · No transition occurs every 2-bits
                       when using DTD.

Ai 0 0 1 1 · · · Activity factor of Ai and Bi is 50%
Bi─▷○─ B̄i 0 0 1 1 · · · Only one transition occurs every 4-bits
                       when using DTD.

(b)

Fig.3-13 (a) Power savings versus activity factor,
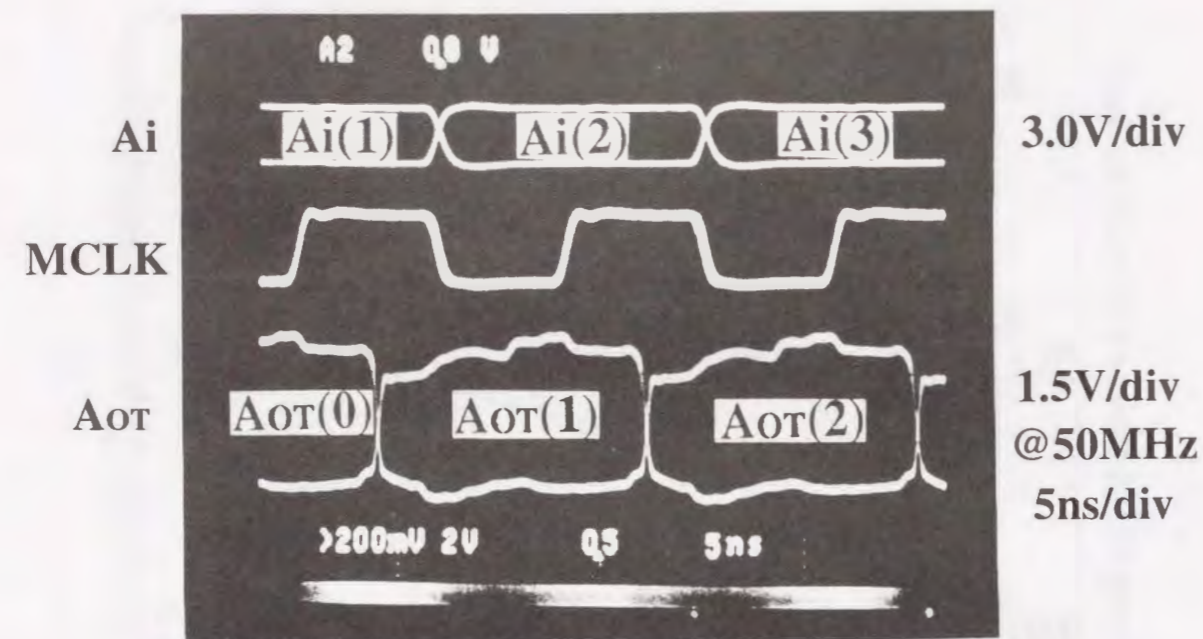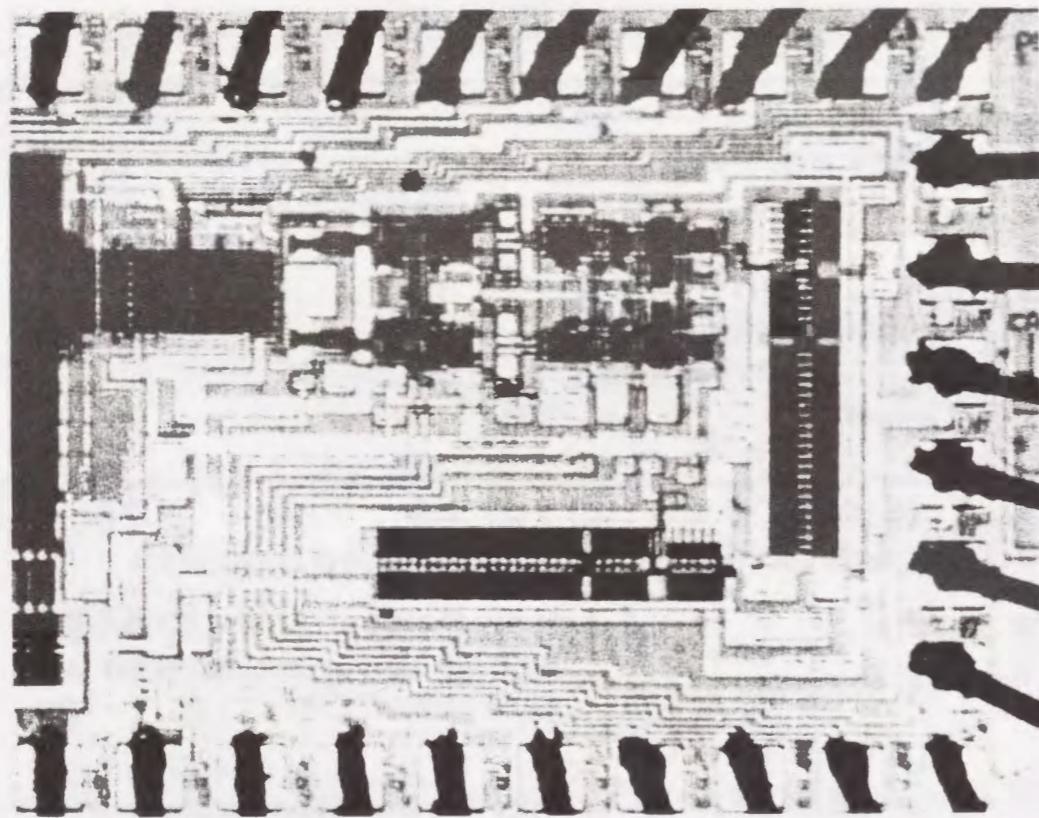   (b)Power dissipation comparison between with DTD
             combined invert scheme[6] and without DTD.

93

A_i $Ai(1)$ $Ai(2)$ $Ai(3)$ 3.0V/div

MCLK

A_{OT} $A_{OT}(0)$ $A_{OT}(1)$ $A_{OT}(2)$ 1.5V/div
@50MHz
5ns/div

(a)

(b)

Fig.3-14. (a) Typical operating waveforms, and
(b) Photomicrograph of test device.

inverse proportion to that), as shown in Fig.3-13(b). As a result, even when the bus activity is less than 50%, the TMD with DTD has a more capability of power savings compared to the parallel scheme, as shown in Fig.3-13(a).

We also verified that a TM-CRB operation at the operating frequency of 50MHz and Vcc=3.3V for 64bits with the bus capacitance=4pF/bit, consumes merely 1.3mA (including 0.8mA in the receiver circuits). This value corresponds to only 6% of the SSL according to the measured data. The measured internal operating waveforms of the signal of Ai, MCLK, and A_{OT} in TMD scheme are shown in Fig.3-14(a). A stable data transmission was observed at the operating frequency of 50MHz and Vcc=3.3V. The test device was fabricated by using 0.5μm CMOS technology and the photomicrograph of the test device for TMD scheme is shown in Fig.3-14(b).

3-6 Conclusion

Once the required throughput can be met in the devices, the proposed TM-CRB become more attractive candidates rather than parallel architecture for saving the data-transmission power on heavily loaded data lines(including chip to chip). This is because TM-CRB can reduce the bus power to 1/11 and 1/3 when the bus activity are 100% and 25%, respectively, compared to the parallel architecture, while reducing the number of signal wires by half. It is our feelings that an even greater power savings in a practical chip can be achieved by combining the strategies of TM-CRB (for heavily loaded data lines) and parallel architecture (for logic circuits).

References

[1] H. Yamauchi, et al "A Low Power Complete Charge-Recycling Bus Architecture for Ultra-High Data Rate ULSI's", in Symposium. on VLSI Circuits Digest of Technical Papers, pp. 21 - 22 , Jun. 1994.

[2] A. Chandrakasan, et al, "Low-Power CMOS Digital Design" IEEE Journal of Solid-State Circuits, vol. 27, No.4SC, pp. 473 - 484, Apr. 1992.

[3] H. Yamauchi, et al, "A 20ns Battery-Operated 16Mb CMOS DRAM" ,ISSCC Digest of Technical Papers , pp. 44-45, Feb. 1993.

[4] H. Yamauchi, et al, "A Low Power Signal-Swing Suppressing Strategy Using Time-Multiplexed Differential Data-Transfer(TMD) Scheme" , in Symposium on Low Power Electronics Digest of Technical Papers, Session No.6.2, Oct. 1995.

[5] T. Yoshihara, et al,"A Twisted Bit Line Technique for Multi-Mb DRAMs" ISSCC Digest of Technical Papers , pp. 238-239, Feb. 1988.

[6] M. R. Stan, et al, "Bus-Invert Coding for Low power I/O", IEEE Transactions on VLSI systems Vol.3, No.1, pp.49-58, March,1995.