

Introduction

The history of semiconductor devices began in the 1930s, when Lilienfeld and Heil [1,2] first proposed the Metal Oxide Semiconductor (MOS) Field-Effect Transistor (FET). However, it took 30 years before this idea was applied to functioning devices to be used in practical applications [3], and, up to the late 1970s, bipolar devices were the mainstream digital technology. Around 1980, this trend took a turn when MOS technology caught up and there was a crossover between bipolar and MOS shares. Complementary-MOS (CMOS) was finding more widespread use due to its low power dissipation, high packing density, and simple design, such that, by 1990, CMOS covered more than 90% of the total MOS sales, and the relation between MOS and bipolar sales was two to one.

In digital circuit applications, there was a performance gap between CMOS and bipolar logic. The existence of this gap, as shown in Figure 1.1, implies that neither CMOS nor bipolar had the flexibility required to cover the full delay-power space. This flexibility was achieved with the emergence of bipolar-compatible CMOS (BiCMOS) technology, the objective of which is to combine bipolar and CMOS so as to exploit the advantages of both at the circuit and system levels.

In 1983, a bipolar-compatible process based on CMOS technology was developed, and BiCMOS technology with both the MOS and bipolar devices fabricated on the same chip was developed and studied [4–6]. The principal

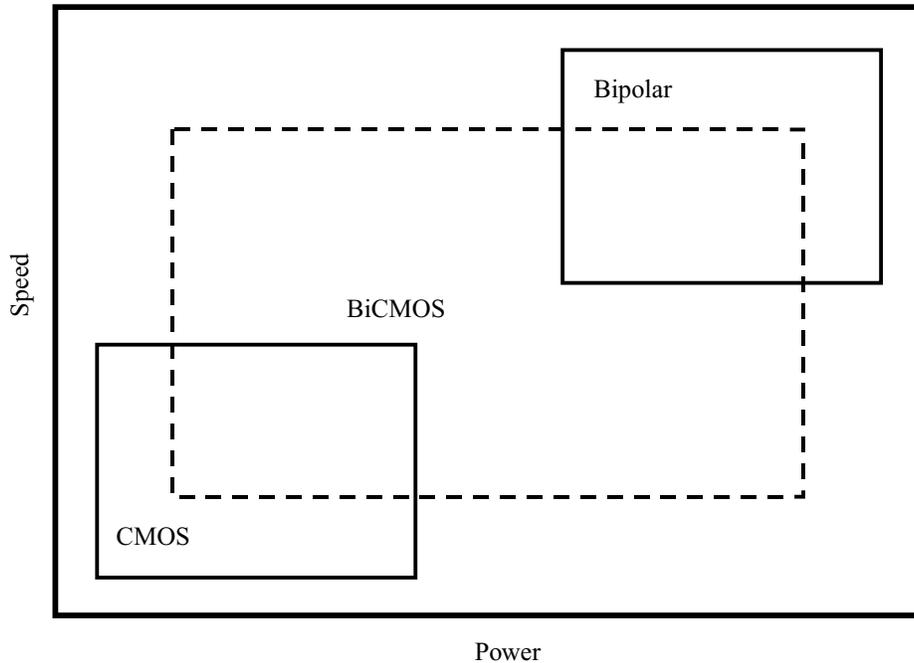


Fig. 1.1 A comparison of CMOS, bipolar, and BiCMOS technologies in terms of speed and power.

BiCMOS circuit in the early days was the BiCMOS totem-pole gate [7] as shown in Figure 1.2. This circuit was proposed by Lin et al. and is one of the earliest versions to be used in practice. It is commonly referred to as the conventional BiCMOS circuit. Since 1985, BiCMOS technologies developed beyond initial experimentation to become widespread production processes. The state-of-the-art bipolar and CMOS structures have been converging. Today, BiCMOS has become one of the dominant technologies used for high-speed, low-power, and highly functional Very-Large-Scale-Integration (VLSI) circuits [8–11], especially when the BiCMOS process has been enhanced and integrated into the CMOS process without any additional steps [12]. Because the process steps required for both CMOS and bipolar are similar, these steps can be shared for both of them.

The concern for power consumption has been part of the design process since the early 1970s. At that time, however, the main design focuses were providing for high-speed operation and a design with minimum area; design tools were all geared toward achieving these two goals. Since the early 1990s, the semiconductor industry has witnessed an explosive growth in the demand and supply of portable systems in the consumer electronics market. High-performance portable products, ranging from small hand-held personal communication devices, such as pagers and cellular phones, to larger and more



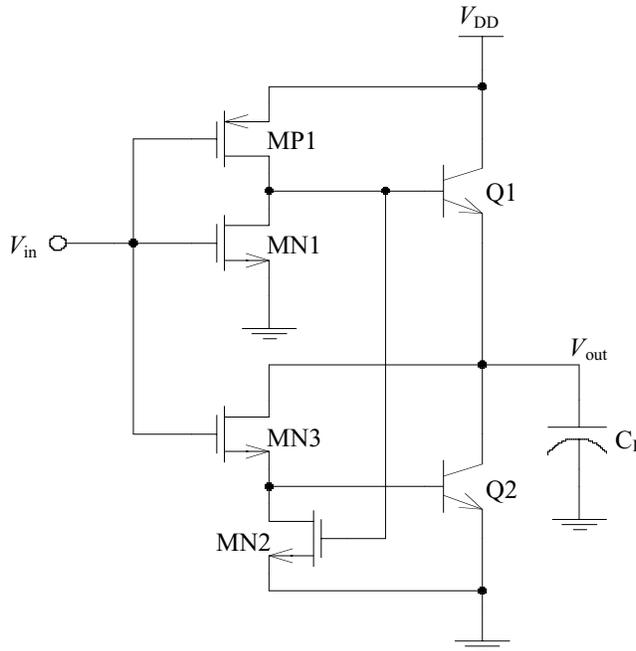


Fig. 1.2 Conventional BiCMOS inverter [43] (Reprinted by permission of Pearson Education, Inc.).

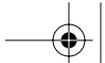
sophisticated products that support multimedia applications, such as lap-top and palm-top computers, have enjoyed considerable success among consumers. Indeed, we anticipate that, in the near future, almost half of the consumer electronics market will be in portable systems. Even though the performance, support features, and cost of a portable product are important to the consumers, its portability is often a key differentiator in a user's purchase decision.

1.1 LOW-POWER DESIGN: AN OVERVIEW

In the past, due to a high degree of process complexity and the exorbitant costs involved, low-power circuit design and applications involving CMOS and BiCMOS technologies were used only in applications where very low power dissipation was absolutely essential, such as wrist watches, pocket calculators, pacemakers, and some integrated sensors. However, low-power design is becoming the norm for all high-performance applications, as power is the most important single design constraint. Although designers have different reasons for lowering power consumption, depending on the target application, minimizing the overall power dissipation in a system has become a high priority.

One of the most important reasons for this trend is the advent of portable systems. As the "on the move with anyone, anytime, and anywhere" era





becomes a reality, portability becomes an essential feature of the electronic systems interfacing with nonelectronic systems, emphasizing efficient use of energy as a major design objective.

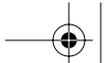
The considerations for portability are due to numerous factors. First, the size and weight of the battery pack is fundamental. A portable system that has an unreasonably heavy battery pack is not practical and restricts the amount of battery power that can be loaded at any one time. Second, the convenience of using a portable system relies heavily on its recharging interval. A system that requires frequent recharging is inconvenient and hence limits the user's overall satisfaction in using the product.

Although the battery technology has improved over the years, its capacity has only managed to increase by a factor of two to four in the last 30 years or so; the computational power of digital integrated circuits has increased by more than four orders of magnitude. To illustrate the importance of low-power design, or the lack of it in portable systems, consider a future portable multimedia terminal that supports high-bandwidth wireless communication; bidirectional motion video; high-quality audio, speech, and pen-based input; and full texts/graphics. The power of such a terminal—when implemented using off-the-shelf components not designed for low power—is projected to reach approximately 40 W. Based on the current Nickel-Cadmium (NiCd) battery technology, which offers a capacity of 20 W-hour/pound, a 20-pound battery pack is required to stretch the recharge interval to 10 hours. Even with new battery technologies, such as the rechargeable lithium or polymers, battery capacity is not expected to improve by more than 30 to 40% over the next 5 years. Hence, in the absence of low-power design techniques, future portable products will have either unreasonably heavy battery packs or a very short battery life.

The issue of power also embraces reliability and the cost of manufacturing nonportable high-end applications. The rapidly increasing packing density, clock frequency, and computational power of microprocessors have inevitably resulted in rising power dissipation. The trends relating to the power consumption of microprocessors indicate that power has increased almost linearly with area-frequency product over the years. For example, the DEC21164, which has a die area of 3 cm² and runs on a 300-MHz clock frequency, dissipates as much as 50 W of power. Such high power consumption requires expensive packaging and cooling techniques given that insufficient cooling leads to high operating temperatures, which tend to exacerbate several silicon failure mechanisms. To maintain the reliability of their products, and avoid expensive packaging and cooling techniques, manufacturers are now under strong pressure to control, if not reduce, the power dissipation of their products.

Finally, due to the increasing percentage of electrical energy usage for computing and communication in the modern workplace, low-power design is in line with the increasing global awareness of environmental concerns. As a result, power has emerged as one of the most important design and performance parameters for integrated circuits. Only a few years ago, the power





dissipation of a circuit was of secondary importance to such design issues as performance and area. The performance of a digital system is usually measured only in terms of the number of instructions it can carry out in a given amount of time, that is, its throughput. The area required to implement a circuit is also important as it is directly related to the fabrication cost of the chip. Larger die areas lead to more expensive packaging and lower fabrication yield. Both effects translate to higher cost. Because the performance of a system is usually improved at the expense of silicon area, a major task for integrated chip (IC) designers in the past was to achieve an optimal balance between these two often-conflicting objectives. Now, with the rising importance of power, this balance is no longer sufficient. Today, IC designers must design circuits with low-power dissipation without severely compromising the circuits' performance.

Clearly, power has become a major consideration in VLSI and giga-scale-integration (GSI) engineering due to portability, reliability, cost, and environmental concerns. The BiCMOS technology that combines the low-power dissipation and high packing density of CMOS with the high-speed and high-output drive of bipolar devices has proven to be an excellent workhorse for portable as well as nonportable applications. For many years to come, device miniaturization together with the search for even lower power and lower voltage requirements will continue. To cater to such an ever-increasing demand, the CMOS/BiCMOS technology shall be the answer.



1.2 LOW-VOLTAGE, LOW-POWER DESIGN LIMITATIONS

1.2.1 Power Supply Voltage

From the device designer's viewpoint, it has been said, "the lower the supply voltage, the better." Even though the dynamic power is largely dependent on the supply voltage, stray capacitance, and the frequency of operation, the overall supply voltage has the largest effect. Therefore, with overall supply voltage lowered, the power dissipation of the circuits can be largely reduced, without compromising the frequency of operation, or, in other words, the speed performance. However, there are various problems associated with lowering the voltage. In CMOS circuitry, the drivability of MOSFETs will decrease, signals will become smaller, and the threshold voltage variations will become more limiting. As shown in Figure 1.3, the increase of the gate delay time is serious when the operating voltage is reduced to 2 V or less, even when the device dimensions are scaled down. The supply voltage scaling in BiCMOS circuits puts even more serious constraints on the circuit performance. Although BiCMOS Ultra-Large-Scale-Integration (ULSI) systems realize the benefits of the low-power dissipation of CMOS and the high-output drive capability of bipolar devices, under low-power supply voltage conditions, the gate delay time significantly increases. This occurs because the effective voltage applied



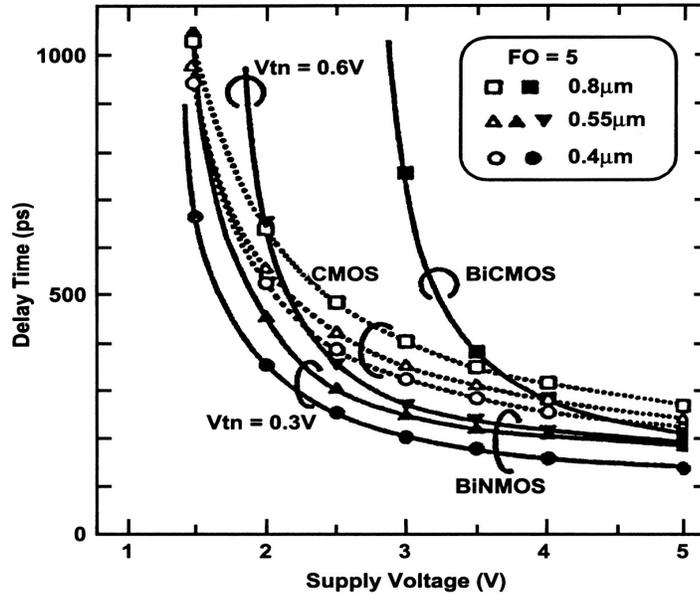


Fig. 1.3 Inverter time versus supply voltage [13] (©1995 IEEE).

to MOS devices is dropped by the inherent built-in voltage ($V_{BE} \sim 0.7\text{ V}$) of the bipolar devices in the conventional totem-pole type circuit. New methods, therefore, must be devised to overcome these obstacles to lowering the supply voltage.

1.2.2 Threshold Voltage

Another related issue of scaling down the power supply voltage is the threshold voltage restriction. At a low-power supply voltage, a low threshold voltage is preferable to maintain the performance trend. However, because the reduction of the threshold voltage causes a drastic increase in the cut-off current, the lower limit of the threshold voltage should be carefully considered by taking into account the stability of the circuit operation and the power dissipation. Furthermore, the threshold voltage dispersion must be suppressed proportional to the supply voltage. The dispersion of threshold voltage affects the noise margin, the standby power dissipation, and the transient power dissipation. Because the worst case critical path restricts LSI performance, it is influenced by the threshold voltage dispersion. Therefore, suppressing the threshold voltage is strongly recommended for low-power large-scale integration (LSI) from the process control and the circuit design point of view [14].

Figure 1.4 shows the V_{th}/V_{DD} dependence of the gate delay time of the CMOS inverter [15]. When the threshold voltage approaches $V_{DD}/2$, the delay time increases rapidly causing a drastic reduction of the MOSFET current and



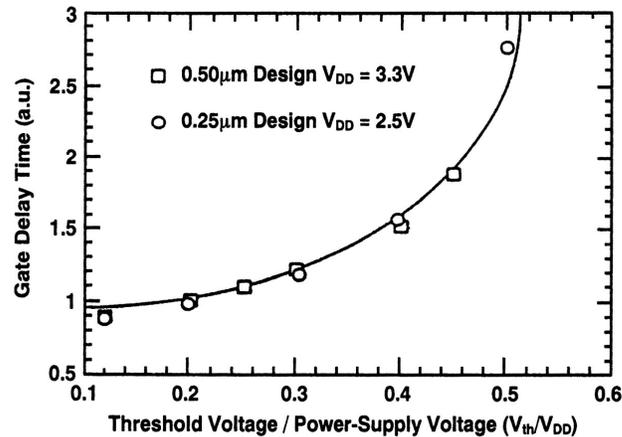


Fig. 1.4 Gate delay time of CMOS inverter versus threshold voltage/power supply voltage [43] (Reprinted by permission of Pearson Education, Inc.).

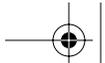
a corresponding increase in the CMOS inverter threshold. On the other hand, lowering the threshold voltage drastically improves the gate delay time. Therefore, a V_{th}/V_{DD} ratio of 0.2 and below is required for high-speed operation, and it is necessary to reduce the threshold voltage to as low as possible when lowering the power supply voltage. However, because the subthreshold swing is almost constant in any device generation, reduction of the threshold voltage sharply increases the MOSFET cut-off current and degrades its ON/OFF ratio. Moreover, the threshold voltage reduction increases the power dissipation due to the switching transient current. At high threshold voltages, the transient power dissipation is negligible as compared to the total power dissipation. On the other hand, at low threshold voltage, the transient power greatly increases with the transient current. Thus, a compromise needs to be found for the V_{th}/V_{DD} ratio to have both low-power and high-speed operation.

1.2.3 Scaling

As the demand for high-speed, low-power consumption and high packing density continues to grow each year, there is a need to scale the device to smaller dimensions. As the market trend moves toward a greater scale of integration, the move toward a reduced supply voltage also has the advantage of improving the reliability of IC components of ever-reducing dimensions. This change can be easily understood if one recalls that IC components with smaller dimensions have more of a tendency to break down at high voltages. It has already been accepted that scaled-down CMOS devices even at 2.5 V do not sacrifice device performance as they maintain device reliability [16].

Scaling the supply voltage for digital circuits has historically been the most effective way to lower the power dissipation because it reduces all components of





power and is felt globally across the entire system. The 1997 National Technology Roadmap for Semiconductors (NTRS) [17] projects the supply voltage of future gigascale integrated systems to scale from 2.5 V in 1997 to 0.5 V in 2012 primarily to reduce power dissipation and power density, increases of which are projected to be driven by higher clock rates, higher overall capacitance, and larger chip sizes.

Scaling brings about the following benefits:

1. Improved device characteristics for low-voltage operation due to the improvement in the current driving capabilities
2. Reduced capacitance through small geometries and junction capacitances
3. Improved interconnect technology
4. Availability of multiple and variable threshold devices, which results in good management of active and standby power trade-off
5. Higher density of integration (It has been shown that the integration of a whole system into a single chip provides orders of magnitude in power savings.)

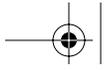
However, during the scaling process, the supply voltage would have to decrease to limit the field strength in the insulator of the CMOS and relax the electric field from the reliability point of view. This decrease leads to a tremendous increase in the propagation delay of the BiCMOS gates, especially if the supply voltage is scaled below 3 V [18]. Also, scaling down the supply voltage causes the output voltage swing of the BiCMOS circuits to decrease [19,20]. Moreover, external noise does not scale down as the device features' size reduces, giving rise to adverse effects on the circuit performance and reliability.

The major device problem associated with the simple scaling lies in the increase of the threshold voltage and the decrease of the carrier surface mobility, when the substrate doping concentration is increased to prevent punch-through. To sustain the low threshold voltage with a high carrier surface mobility and a high immunity to punch-through simultaneously, substrate engineering will be a prerequisite.

1.2.4 Interconnect Wires

In the deep submicron era, interconnect wires are responsible for an increasing fraction of the power consumption of an integrated circuit. Most of this increase is attributed to global wires, such as busses, clocks, and timing signals. D. Liu et al. [21] found that, for gate array and cell library-based designs, the power consumption of wires and clock signals can be up to 40 and 50% of the total on-chip power consumption, respectively. The influence of this interconnect is even more significant for reconfigurable circuits. It has also been reported that, over a wide range of applications, more than 90% of the power dissipation of traditional Field Programmable Gate Array (FPGA) devices has





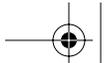
been attributable to the interconnect [22]. Therefore, it is both advantageous and desirable to adopt techniques that can help to reduce these ratios. For chip-to-chip interconnects, wires are treated as transmission lines, and many low-power Input/Output (I/O) schemes were proposed at both the circuit level [23] and the coding level [24]. One of the effective techniques to reduce the power consumption of on-chip interconnects is to reduce the voltage swing of the signal on the wire. A few reduced-swing interconnect schemes have been proposed in the literature [25–30]. These schemes present a wide range of potential energy reductions, but other considerations such as complexity, reliability, and performance play an important role as well. Nakagome et al. [26] proposed a static driver with a reduced power supply. This driver requires two extra power rails to limit the interconnect swing and uses special low-threshold devices (~ 0.1 V) to compensate for the current-drive loss due to the lower supply voltages. Differential signaling, proposed and analyzed by Burd [31], achieves great energy savings by using a very low-voltage supply. The driver uses nMOS transistors for both pull-up and pull-down, and the receiver is a clocked unbalanced current-latch sense amplifier. The receiver overhead may hence be dominant for short interconnect wires with small capacitive loads. The main disadvantage of the differential approach is the doubling of the number of wires, which certainly presents a major concern in most designs. Another class of circuits comes under the category of Dynamically Enabled Drivers. The idea behind this family of circuits is to control the charging and discharging times of the drivers so that a desired swing on the interconnect is obtained. This concept has been widely applied in memory designs. However, it only works well in cases when the capacitive loads are well known beforehand.

Another scheme called the Reduced-Swing Driver–Voltage-Sense Translator (RSD–VST [29]) also uses a dynamically enabled driver, with an embedded copy of the receiver circuit, called voltage-sense translator (VST), to sense the interconnect swing and provide a feedback signal to control the driver. Inherent in the scheme is the drawback of the mismatch of the switching threshold voltage between the two VSTs. The charge intershared bus (CISB) [27] and charge-recycling bus (CRB) [28] are two schemes that reduce the interconnect swing by utilizing charge sharing between multiple data bit lines of a bus. The CRB scheme uses differential signaling, and the CISB scheme is single ended with references. Both schemes reduce the interconnect swing by a factor of n , where n is the number of bits.

1.3 SILICON-ON-INSULATOR (SOI)

SOI is one of the leading technologies that offer high packing density as well as high-speed and low-power operations. SOI has a distinct characteristic—the delay improvement in the circuit depends very much on the circuit topology and its design [32]. Generally, the more complex the circuit is, the more





impact SOI provides. This understanding is especially applicable to circuits using stacked transistors since the body voltage is rarely negative with respect to the source. The performance improvement of SOI over bulk varies according to the type of circuit as well, namely static, dynamic, or array. SOI also has an impact on the path delay. Since the improvement of individual circuits in SOI depends on their topologies, the improvement in overall cycle time is the net effect of the improvement of all the individual circuits composing the frequency-limiting critical paths.

The advantages rendered by SOI do not come by without circuit design challenges. These concerns are mainly caused by the uncertainty in the potential of the FET body. The potential of the body with respect to ground is a function of many factors, including the circuit topology and switching history. This “history effect” makes the delay through a particular circuit difficult to predict without full knowledge of the prior states and transitions of the circuit. This effect on delays varies according to the circuit topology, environment, and other factors. Another point of concern is the parasitic bipolar current, which puts dynamic circuits and some static circuit families at risk. Even though it is not a serious problem in fully restoring circuits, it becomes alarming when it comes to topologies, which combine a large number of parallel devices, such as wide muxes and OR gates. The floating body in SOI transistors also leads to uncertainty in threshold voltages, which in turn means lower noise margins for dynamic circuits. A low noise figure requires changes in dynamic circuit design and redesigning circuits that were originally intended for a bulk technology. Several design techniques are commonly used to improve the noise figure without adversely affecting the circuit delay. These include cross-connecting inputs to stacked devices, precharging intermediate nodes, and remapping logic. Device self-heating can be problematic with SOI. This phenomenon is attributed to the thermal resistance of the buried oxide layer. Vulnerable devices are those in a high current state for a significant portion of the clock cycle, such as some off-chip drivers.

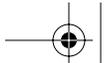


1.4 FROM DEVICES TO CIRCUITS

The key features of the current generation of silicon CMOS devices include the use of self-aligned ion implantation to dope the source, drain, and gate; metal silicides on silicon surfaces for reducing the sheet resistance and to improve upon the ohmic contact; shallow trench isolation (STI) to separate FETs to save silicon area; and the employment of nonuniform (retrograde) channel doping that is coupled with halo implants to control short-channel effects [33].

According to the exponential projections of the Semiconductor Industry Association (SIA) roadmap for the year 2012, dynamic random access memories (DRAMs) are expected to have a capacity of 256-gigabit (Gb), microprocessors with 1.3×10^9 logic FETs, a gate lithography of 35 nm (channel lengths at or below 20 nm), across-chip clocks of 3 gigahertz (GHz), a power





supply voltage of 0.5 V, and an equivalent gate oxide thickness of less than 1.0 nm will emerge. However, realizing these goals will necessitate finding new technologies.

The ongoing shrinking of silicon MOSFETs complicates device behavior. To maintain desirable transfer characteristics, more specialized doping or more complex structures are required. The lateral dimensions are constrained by gate lithography and lateral doping profiles, whereas the vertical dimensions are restricted by gate insulator tunneling considerations, vertical doping profile abruptness, and, for bulk or partially depleted SOI designs, maximum body doping constraints due to body-to-drain tunneling.

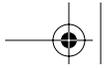
In the recent published work on possible 25-nm bulk design, many of these issues surfaced [34]. The first design consideration is the requirement for the SiO_2 gate insulator to go below 1.0 nm, which could lead to high tunneling leakage through the SiO_2 . A compromise scheme such as the employment of a thicker oxide/nitride insulator with equivalent oxide thickness of 1.5 nm gives tunneling leakage of $\sim 1 \text{ A/cm}^2$. This is possibly the thinnest usable oxide/nitride insulator because thinner insulators will have excessive standby power dissipation thereby restricting the use of dynamic logic.

The reduction of supply voltage to minimize reliability problems [35] and power dissipation to, say, 1 V for high-performance designs and perhaps as low as 0.6 V for low-power circuits [36] necessitates a low threshold voltage. However, the threshold voltage must be high enough to prevent the off-state current from exceeding the power budget. The threshold roll-off can be compensated by using super-halo implants (see chapter 2), but the compensation can lead to performance degradation. The implementation of thin high-permittivity insulators to conventional FETs may not help to achieve significant device scaling because the body-doping constraints will still limit the depletion depth. A possible measure is to change the device structure so that a gate below the channel replaces the body [33]. Other feasible variations exist: The two gates may be either separate or connected, the work functions may be different or the same, and the current flow may be in the x , y , or z direction [33]. Theoretically, the three-dimensional (3D) generalization using a cylindrical gate type is the most scaleable. These double- or surround-gated structures have the potential for more scaling than the conventional FETs. The details of these new features are discussed in section 2.8.3.

The electromigration (EM) current threshold for aluminum (Al) is $2 \times 10^5 \text{ A/cm}^2$. Beyond the 0.25- μm device generation, the current densities could reach a level that could induce EM failure for traditionally doped-aluminum conductors. Copper (Cu), on the other hand, offers an EM current threshold of $5 \times 10^6 \text{ A/cm}^2$, thereby providing a good substitute for Al metallization and overcoming the EM limitation. Cu strong resistance to EM is primarily due to its high melting point of 1082°C, whereas the melting point for Al is only 660°C [37].

The continual shrinking of the interconnect cross section leads to higher line resistance. The smaller pitch also results in elevated line-to-line capaci-





tance. Bohr reported that a 0.25- μm line-width Al-metal that is longer than 436 μm can contribute more delay than a 0.25- μm gate [36]. Using copper will certainly help to lower the interconnect delay and provide further shrinkage to the upper interconnect levels, thereby increasing the wiring density and reducing the number of metal layers. The use of copper should also help to drive down processing costs. Because copper cannot be dry-etched easily, the damascene (in-laid) approach is used to deposit copper. This approach gives an additional advantage because the dual damascene process can fabricate both the line and via levels concurrently, which results in approximately 30% fewer steps (and hence lower cost) than the single damascene or subtractive patterning method [38].

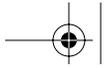
As chip complexities increase, design problems, such as layout of a chip and simulation for the circuit, all correspondingly escalate. Today, circuit designers are often required to design large, complex circuits. Generally, this task is becoming increasingly difficult because designers are now facing not only circuit problems but also process- and device-related issues. Furthermore, they must juggle different design requirements and balance conflicting constraints. First come the multiple levels of abstraction from the specification of a chip function to a layout. This process needs much work because it covers both the “front-end” and “back-end” design activities. Generally, front-end design activities include system definition, functional design and simulation, logic design and simulation, and circuit design and simulation. The back-end design activities involve those physical design details that require little creative work but instead the mechanical translation of the design into semiconductor (e.g., silicon). Another important constraint is the cost, where one must strike a balance between the performance and the price paid to achieve it. This constraint, together with the generally short design time, makes integrated circuit design a challenging process. For the physical implementation of complex circuits, three layout methodologies are commonly used: full-custom design, gate array, and standard-cell approach (semicustom design).

Full-custom design is the design methodology whereby the layout of each function and transistor is fully optimized. Though this approach is capable of attaining the objective of minimizing the power consumption of circuits, it can lead to low design productivity. Therefore, this approach is not recommended for Application-Specific Integrated Circuits (ASIC) and processors.

The second methodology is the gate array approach. Gate arrays consist of already-implemented cells and thus require only personalization steps. The design of the logic gates includes the wiring of different transistors from the continuous array of nMOS and pMOS transistors in the internal cell array using metallization and contact. This methodology allows the reduction of design cost at the expense of some other constraints such as the area, power, and performance.

The third methodology is the standard-cell approach. With the creation of digital library cells, several logic gates and functions can be created and compiled in the library. Generally, in the library cell approach, two layout





styles exist. The first is to optimize the cell area to reduce the silicon space. The second is to optimize the cell performance, usually resulting in high speed and requiring more space. This methodology provides lower cost and higher productivity in speeding up the design process.

1.4.1 Latches and Flip-Flops

Latches and flip-flops are basic sequential elements commonly used to store logic values and are always associated with the use of clocks and clocking networks. The clocking network with its 20 to 40% contribution to the overall power dissipation is a major obstacle in implementing high-performance systems [39–41]. This deterrent leads to a growing need to improve clocked structures. The analysis and previous research [14–16,42] suggest that the main focus for low-power design must be the off-chip power, the cache, latches, and flip-flops. The off-chip power cannot be reduced unless the off-chip circuits are optimized. The cache design styles, the size, and the cache access and its coherency maintenance algorithms dictate the power dissipation of the cache.

The direction taken by research in a field is a function of the prevailing design philosophy, the requirements imposed on that field by other disciplines, and the shortcoming of existing designs/systems. For latches and flip-flops, the story is no different. Latch and flip-flop designs at any point in time are natural outcomes of important design requirements of those times and the primary use they are put to. The quality measures of latches and flip-flops have also evolved along the same lines and have shaped their design theme. The main features of the theme are functionality, synchronous versus asynchronous, area optimization, performance, pipelining, and high-speed/low-power operation.

Apart from ensuring the functionality of the circuit, the design should be implemented using the least number of transistors to reduce the area. A large decrease in the number of transistors is possible by utilizing the bi-directionality feature of the MOS transistor—the pass transistor design style. The pass transistor version of the D flip-flop requires only 12 transistors. Despite the area optimization advantage, pass transistor designs do not provide full swings nor do they isolate the outputs from inputs. One of the concerns that is associated with clocked flip-flop designs and that affects their performance is the skew that develops between different clock phases. To address this concern, each phase should be routed in an identical manner to the others. However, two parallel long wires tend to suffer from crosstalk. Furthermore, when the chip's operation speed is above 100 MHz, it is difficult to generate nonoverlapping clocks and to control the clock skew properly in a VLSI chip due to the statistical variations of components in the clock distribution path [17]. Such a difficulty involved in routing many clock phases rekindles the interest, in many cases, in single-phase clock structures.

Pipelining improves throughput in combinatorial circuit design. Pipelines are constructed by breaking up a large circuit into many stages by inserting registers. The objective here is to develop a methodology by which the

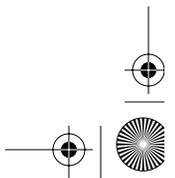
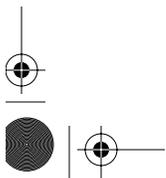


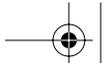


latches and flip-flops (see chapter 5) could be interspersed with logic to yield fast pipeline structures.

1.5 REFERENCES

1. J. E. Lilienfeld, U.S. Patent 1,745,175 (1930).
2. O. Heil, U.S. Patent 439,457 (1935).
3. D. Kahng and M. M. Atalla, "Silicon-Silicon Dioxide Field Induced Surface Devices," *IRE Solid-State Devices Res. Conf.*, Carnegie Institute of Technology, Pittsburgh, Pa., 1960.
4. H. Momose, H. Shibata, Y. Mizutani, K. Kanzaki, and S. Kohyama, "High Performance 1.0 μm N-Well CMOS/Bipolar Technology," *Symp. VLSI Technology, Tech. Dig.*, pp. 40–41, 1983.
5. I. Walezyk and J. Rubinstein, "A Merged CMOS/Bipolar VLSI Process," *IEEE IEDM, Tech. Dig.*, pp. 59–62, 1983.
6. J. Mayamoto, S. Saitoh, H. Homose, H. Shibata, K. Kanzaki, and S. Kohyama, "A 1.0 μm N-Well CMOS/Bipolar Technology for VLSI Circuits," *IEEE IEDM, Tech. Dig.*, pp. 63–66, 1983.
7. H. C. Lin, J. C. Ho, R. R. Iyer, and K. Kwong, "CMOS-Bipolar Transistor Structure," *IEEE Trans. Electron Devices*, Vol. ED-6, No. 11, pp. 945–951, 1969.
8. H. Higuchi, G. Kitsukawa, T. Ikeda, Y. Nishio, N. Sasaki, and K. Ogiue, "Performance and Structures of Scale-Down Bipolar Devices Merged with CMOSFETs," *IEEE IEDM, Tech. Dig.*, pp. 694–697, 1984.
9. A. R. Alvarez, P. Meller, and B. Tien, "2 μm Merged Bipolar-CMOS Technology," *IEEE IEDM, Tech. Dig.*, pp. 761–764, 1984.
10. T. Ikeda, T. Nagano, N. Momma, K. Miyata, H. Higuchi, M. Odaka, and K. Ogiue, "Advanced BiCMOS Technology for High Speed VLSI," *IEEE IEDM, Tech. Dig.*, pp. 408–411, 1986.
11. H. Iwai, G. Sasaki, Y. Unno, Y. Niitsu, M. Norishima, Y. Sugimoto, and K. Kanzaki, "0.8 μm BiCMOS Technology with f_T Ion-Implanted Emitter Bipolar Transistor," *IEEE IEDM, Tech. Dig.*, pp. 28–31, 1987.
12. J. Miyamoto, S. Saitoh, H. Momose, H. Shibata, K. Kanzaki, and T. Iizuka, "A 28 ns CMOS RAM with Bipolar Sense Amplifiers," *IEEE ISSCC, Tech. Dig.*, pp. 245–248, 1984.
13. M. Takada, K. Nakamura, and T. Yamazaki, "High Speed Submicron BiCMOS Memory," *IEEE Trans. Electron Devices*, Vol. 42, No. 3, pp. 497–505, March 1995.
14. T. Kobayashi and T. Sakurai, "Self-Adjusting Threshold Voltage Scheme (SATS) for Low Voltage High Speed Operation," *IEEE CICC Proc.*, pp. 271–274, 1994.
15. Y. Mii, S. Wind, Y. Taur, Y. Lii, D. Klaus, and J. Bucchigano, "An Ultra-Low Power 0.1 μm CMOS," *Symp. VLSI Technology, Tech. Dig.*, pp. 9–10, 1994.
16. W. H. Chang, B. Davaari, M. R. Wordeman, Y. Taur, C. C. H. Hsu, and D. Rodriguez, "A High Performance 0.25 μm CMOS Technology (I) & (II)," *IEEE Trans. Electron Devices*, Vol. 39, No. 4, pp. 959–967, 1992.
17. The National Technology Roadmap for Semiconductor, SIA Handbook, 1997.





18. S. Shukuri, A. Watanabe, R. Izawa, T. Nagano, and E. Takeda, "The Guiding Principle for BiCMOS Scaling in ULSIs," *Symp. VLSI Technology, Tech. Dig.*, pp. 53–54, 1989.
19. A. Bellaouar, S. H. K. Embabi, and M. I. Elmasry, "Scaling of Digital BiCMOS Circuits," *IEEE J. Solid-State Circuits*, Vol. 25, No. 4, pp. 932–941, 1990.
20. G. P. Rossel and R. W. Dutton, "Scaling Rules for Bipolar Transistors in BiCMOS Circuits," *IEEE IEDM, Tech. Dig.*, pp. 795–798, 1989.
21. D. Liu et al., "Power Consumption Estimation in CMOS VLSI Chips," *IEEE J. Solid State Circuits*, Vol. 29, No. 6, pp. 663–670, June 1994.
22. E. Kusse, "Analysis and Circuit Design for Low Power Programmable Logic Modules," M.S. thesis, Univ. Calif., Berkeley, 1997.
23. B. Gunning et al., "A CMOS Low-Voltage-Swing Transmission-Line Transceiver," *IEEE ISSCC Dig. Tech. Papers*, pp. 58–59, Feb. 1992.
24. E. Musoll et al., "Working-Zone Encoding for Reducing the Energy in Microprocessor Address Busses," *IEEE Trans. VLSI Syst.*, Vol. 6, No. 12, pp. 568–572, Dec. 1998.
25. H. Zhang and J. Rabaey, "Low-Swing Interconnect Interface Circuits," *Proc. 1998 Intl. Symp. Low-Power Electronic Devices*, Monterey, Calif., Aug. 1998, pp. 161–166.
26. Y. Nakagome et al., "Sub-1 V Swing Internal Bus Architecture for Future Low-Power USLI's," *IEEE J. Solid State Circuits*, Vol. 28, pp. 414–419, April 1993.
27. M. Hiraki et al., "Data Dependent Logic Swing Internal Bus Architecture for Ultra Low-Power LSI's," *IEEE J. Solid State Circuits*, Vol. 30, No. 4, pp. 397–402, April 1995.
28. H. Yamauchi et al., "An Asymptotically Zero Power Charge-Recycling Bus Architecture for Battery-Operated Ultra-High Data Rate ULSI's," *IEEE J. Solid-State Circuits*, Vol. 30, No. 4, pp. 423–431, April 1995.
29. R. Colsham and B. Jaroun, "A Novel Reduced Swing CMOS BUS Interface Circuit for High-Speed Low-Power VLSI Systems," *Proc. IEEE Intl. Symp. Circuits and Systems*, Vol. 4, pp. 351–354, May 1994.
30. J. Rabaey, *Digital Integrated Circuits*, Prentice Hall, Englewood Cliffs, N.J., 1996.
31. T. Burd, "Energy Efficient Processor System Design," Ph.D. dissertation, Univ. Calif., Berkeley, 1998.
32. G. Shahidi et al., "Partially-Depleted SOI Technology for Digital Logic," *Proc. IEEE Int. Solid-State Circuits Conf.*, pp. 426–427, Feb. 1999.
33. H. S. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, and J. J. Welser, "Nano-scale CMOS," *Proc. of IEEE*, Vol. 87, No. 4, pp. 537–570, April 1999.
34. Y. Taur, C. H. Wann, and D. J. Frank, "25-nm CMOS Design Considerations," *IEEE IEDM Tech. Dig.*, pp. 789–792, 1998.
35. J. H. Stathis and D. J. DiMaria, "Reliability Projection for Ultra-Thin Oxides at Low Voltage," *IEEE IEDM Tech. Dig.*, pp. 167–170, 1998.
36. M. T. Bohr, "Interconnect Scaling—The Real Limiter to High Performance ULSI," *IEEE IEDM Tech. Dig.*, pp. 241–244, Dec. 1995.
37. D. Pramanik, "Aluminum-Based Metallurgy for Global Interconnects," *MRS Bulletin, Metallization for Integrated Circuit Manufacturers*, Vol. XX, No. 11, pp. 57–60, Nov. 1995.





38. R. L. Jackson, E. Broadbent, T. Cacouris, A. Harrus, M. Biberger, E. Patton, and T. Welsh, "Processing and Integrations of Copper Interconnects," *Damascus Technical Paper, Novellus homepage (www.damascus.novellus.com/damascus/index.htm)*, Novellus Systems, San Jose. Calif.
39. R. Mehra et al., "A Partitioning Scheme for Optimizing Interconnect Power," *IEEE J. Solid-State Circuits*, Vol. 32, No. 3, pp. 433–443, March 1997.
40. E. De Man and M. Schobinger, "Power Dissipation in the Clock System of Highly Pipelined ULSI CMOS Circuits," *1994 International Workshop on Low Power Design*, pp. 133–138, April 1994.
41. D. Liu and C. Svensson, "Power Consumption Estimation in CMOS VLSI Chips," *IEEE J. Solid-State Circuits*, Vol. 29, No. 6, pp. 663–670, June 1994.
42. H. J. Chao and C. Johnston, "Behaviour Analysis of CMOS Flip-Flops," *IEEE J. Solid-State Circuits*, Vol. 24, No. 5, pp. 1454–1458, Oct. 1989.
43. S. S. Rofail and K. S. Yeo, *Low-Voltage, Low-Power Digital BiCMOS Circuits: Circuit Design, Comparative Study and Sensitivity Analysis*, Prentice Hall, Upper Saddle River, N.J., 2000.

