CHARACTERIZATION AND REDUCTION OF METASTABILITY ERRORS IN CMOS INTERFACE CIRCUITS

Clemenz Lenard Portmann

Technical Report No. CSL-TR-95-671

June 1995

Abstract

In synchronous digital logic systems, asynchronous external signals must be referenced to the system clock or synchronized. Synchronization of asynchronous signals, however, inevitably leads to metastability errors. Metastability error rates can increase by orders of magnitude as clock frequencies increase in high performance designs, and supply voltages decrease in low-power designs. This research focuses on the characterization of metastability parameters and error reduction with no penalty in circuit performance. Two applications, high-speed flash analog-to-digital conversion and synchronization of asynchronous binary signals in application-specific integrated circuits have been investigated.

Applications such as telecommunications and instrumentation for time-domain analysis require analog-to-digital converters with metastability error probabilities on the order of 10^{-10} errors/cycle, achievable in high performance designs only through the use of dedicated circuitry for error reduction. A power and area efficient externally pipelined metastability error reduction technique for flash converters has been developed. Unresolved comparator outputs are held valid, causing the encode logic to fail benignly in the presence of metastability. In an n bit converter, errors are passed as a single unsettled bit to the converter output and are reduced with an external pipeline of only n latches per stage rather than an internal pipeline of $2^{n}-1$ latches per stage.

An 80-MHz, externally pipelined, 7-bit flash analog-to-digital converter was fabricated in 1.2-μm CMOS. Measured error rates were less than 10^{-12} errors/cycle. Using internal pipelining with two levels of 127 latches to achieve equivalent performance would require 3.48 times more power for the error reduction circuitry with a Nyquist frequency input. This corresponds to a reduction in the total power for the implemented converter of 1.24 times compared with the internally pipelined converter.

In synchronizers and arbiters, general purpose applications require mean time between failures on the order of one per year or tens of years. Comparison of previous designs has been difficult due to varying technologies, test setups, and test conditions. To address this problem, a test circuit for synchronizers was implemented in 2-µm and 1.2-µm CMOS technologies. Using the test setup, the evaluation and comparison of synchronizer performance in varying environments and technologies is possible. The effects of loading, output buffering, supply scaling, supply noise, and technology scaling on synchronizer performance are discussed.

Acknowledgments

I would like to convey my sincere appreciation to my adviser, Professor Teresa Meng. Without her guidance, encouragement, and patience, this work would never have been possible. Additionally, I would like to thank Prof. Robert Dutton, my associate advisor, who encouraged me to attend Stanford and got me started in what would eventually become this dissertation. I would also like to acknowledge Prof. Bruce Wooley for providing direction in this research, Prof. Gene Franklin for serving as my orals chair, and Prof. Kunle Olukotun for reading my dissertation. Finally, I would like to thank Prof. Michael Smith of the University of Hawaii for encouraging me to pursue a Ph. D.

Due to health problems, much of this work has been done with the aid of other students, many of whom did the repetitive tasks of mask layout, data entry, and document typing. I would like to thank Phil Acosta, Chris Giles, Larry Tanikawa, and Edwin deAngel who helped me in this respect. These students were funded partially by the Stanford Disability Resource Center. I would also like to acknowledge Donald Ramsey, Dr. John Maneatis, and Dr. Thomas Chanak. They performed the final assembly, debugging, and submission of an integrated circuit while I was hospitalized. I believe they have created the first IC that was debugged over the telephone. Dr. John Maneatis also provided access to the software routines he wrote, speeding up the simulation portion of this work.

The advice of Dr. David Su, Dr. Marc Loinaz, and Prof. Tom Lee was invaluable in implementing the test boards, creating test setups, and tracking down problems in the lab. Dr. Behzad Razavi provided insight into the design and testing of A/D converters. Prof. Mark Horowitz and Rambus graciously allowed me to probe my chip with their laser and locate problems at a critical juncture in the research. Ben Gordon and Ken Yang were also helpful in locating the problem.

I would like to thank my colleagues in CIS and AEL for providing a rewarding and interesting experience both on the professional and personal level. While I could not possibly hope to list them all here, I would like to especially thank the past and present fellow students of Prof. Meng's research group, as well as the students from Prof. Wooley's and Prof. Horowitz's research groups, and my office mates over the years. Special mention is also due to Lilian Betters for helping me navigate the paper mill while here at Stanford.

Finally, I want to extend my appreciation to my parents, brothers, and sister who supported me throughout my education in Washington, Hawaii, Japan, and here at Stanford. In the largely agricultural, Swiss-American community of the Pacific Northwest, there are not too many children pursuing advanced degrees. Many extended family members and acquaintances could not understand how someone could stay so long in college. My mother often joked, to thwart their inevitable questions, that I was a slow learner. While this may or may not be the case, I have finished at last.

Table of Contents

A/D Converter Test Board

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Recent advances in complementary metal oxide semiconductor (CMOS) technology have led to unparalleled levels of integration in digital logic systems. By and large, these digital logic systems require a clock to synchronize signals and ensure proper operation. However, in practical applications, digital systems must communicate with the outside world or other systems. This requires external signals asynchronous to the clocked system to be converted to signals the clocked digital system can understand. The requirement for synchronous systems is that all circuits required to respond to an input uniformly agree on the binary value of the input. In many applications, however, due to the nature of one or more of the external inputs, the input requirements for the synchronous circuit will be violated, and errors will result. Depending upon the application, the errors are described by a number of different terms including "synchronization failure," "arbitration error," and "metastability error." The underlying mechanism for all of these problems is the same, and of these terms, "metastability error" is the most general because it describes the failure of the element within the circuit and not the application.

Metastability is a widespread phenomenon and errors may occur in any synchronous circuit where an input signal can change randomly with respect to a reference signal [1]-[4]. The reference signal may be either a voltage based reference, such as a bias voltage, or a time based reference, such as a clock signal. Circuits in which metastability can occur include analog-to-digital converters [5]-[9], memories [10]-[12], time digitizers [13], [14], and bus controllers [15]-[17]. All of these circuits have a characteristic error probability, P_E, or mean time between failures, *MTBF*, associated with the circuit. Depending upon the
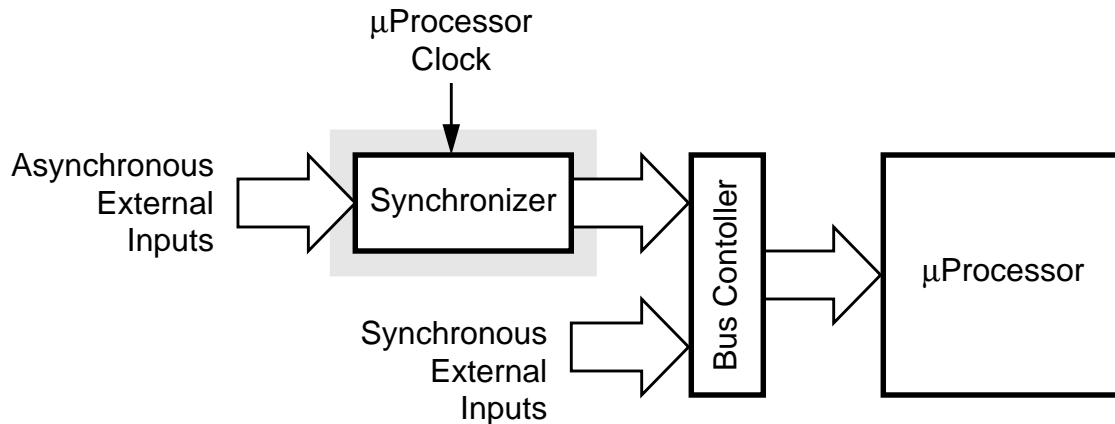
Figure 1.1: Microprocessor bus controller.

application, error requirements for the system can vary widely. For instance, applications with uncompressed digitized speech signals can tolerate relatively high error rates due to the resilience of human hearing perception. For speech applications, symbol error probabilities on the order of 10^{-3} are tolerable. However, other applications require error rates many orders of magnitude lower. These applications can include purely digital applications or mixed-signal applications.

A purely digital application where metastability is a problem is in bus controllers. Figure 1.1 shows a block diagram of a bus controller for a microprocessor. External asynchronous signals are synchronized with a synchronizer cell which will be discussed in detail in Chapter 2. An asynchronous signal is asynchronous to the system clock. This can include signals which are not synchronous to any clock or signals which are synchronous to a non-harmonically related clock. If the external signal is a relatively low frequency signal such as the user on a keyboard, the error rate will be relatively low. However, if the external input is another signal generated from a high-speed digital IC such as a math co-processor with a different clock frequency, the error rate with the same synchronizer will increase by many orders of magnitude. Since the microprocessor is a general purpose machine, very high metastability performance is desirable. Mean time between failures on

the order of one error per year or one error per tens of years for the worst case high-speed signal are usually implemented.

A mixed-signal application requiring low error rates is high-speed digitizing oscilloscopes. Figure 1.2 shows a block diagram for a digitizing oscilloscope. The analog input signal is conditioned with analog signal processing as determined by the user from the scope display. The analog signal is then quantized by a high-speed analog-to-digital converter. In this application, high sampling rates are desirable to allow high frequency input waveforms to be captured. However, metastability error rates for A/D converters are exponentially related to the sampling frequency. High error rates in the A/D converter will cause the waveform to trigger improperly on the screen and corrupt the display image. Error probability requirements are generally on the order of 10^{-9} to 10^{-10}. At high sampling rates, reducing the error rate for the A/D converter can require considerable circuitry, consuming power and area.

In general, applications which require low error rates fall into two categories: applications where at any particular time a single piece of information carries the pertinent information or applications requiring large amounts of processing where all bits of information are pertinent. Examples of the former are the microprocessor and digitizing oscilloscopes shown. Examples of the latter would be automated testers for manufacturing or telecommunications receivers for compressed high definition video. For an automated
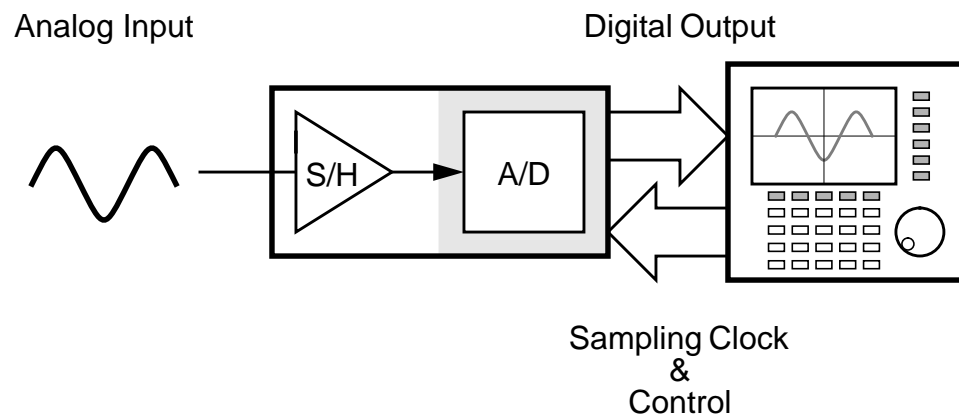
Figure 1.2: Digital oscilloscope.

tester searching for faulty pixel in charge-coupled device (CCD) imagers, for high definition screens, the error probability of the sampling A/D converter in the tester would have to be low enough to keep the false hit rate acceptable.

1.2 Organization

This thesis focuses on the characterization and reduction of metastability errors in CMOS circuits. While this research focuses on CMOS circuits explicitly, the techniques for error reduction and characterization are applicable to other technologies as well. The results obtained in this thesis are relevant to CMOS circuit designers who must interface with signals from other systems and deal with two areas where metastability errors are often problematic, ASIC standard cell design and analog-to-digital conversion.

A large number of synchronous digital designs are created using computer aided design (CAD) tools and are synthesized using standard cell or gate array methodologies. In a standard cell or gate array design environment, synchronizer cells from cell libraries are used repeatedly in a variety of chips. Over the lifetime of the library, the cells are used in a number of loading conditions, technologies, and supply voltages. For synchronous designs with asynchronous external inputs, factors other than the error rate will ultimately determine the environment the synchronizer will be placed in. The application will determine the supply voltage, technology, and target error rate. The specific implementation will determine the synchronizer loading and the supply noise. Power dissipation and chip cost will impact the library selection. The performance of synchronizer cells in a variety of environments has not been well studied, as prior research has been in full custom synchronizer designs and board-level synchronizer solutions.

Chapter 2 discusses the fundamentals of regenerative circuits and a brief overview of the signal conditions which cause regenerative circuits to enter into the metastable state. The equation to determine the exponential time constant of a regenerative circuit is derived. From this relationship, the equations for the error probability in comparators and the mean time between failures in synchronizer and arbiter applications are generated. Additionally, the optimum device sizes for synchronizer/arbiter performance are derived.

Chapter 3 characterizes the effects of loading, power supply, technology scaling, and supply noise on the metastable parameters of CMOS latches used as synchronizers. Results are presented from test-chips designed to measure the performance of buffered and unbuffered CMOS latches and simulations. A normalization scheme to compare latches with different supplies and technologies is presented with data from test-chips, and the minimum supply voltage for scalable metastable performance is investigated. This chapter also discusses the test setup and measurements to characterize the effects of supply noise on synchronizer performance.

Chapter 4 discusses architectures for high-speed A/D conversion. The origins of metastability errors in high-speed flash converters and the measurement of error probabilities for A/D converters are discussed. Measured error rates for a 6-bit flash A/D converter with no error reduction circuitry are shown, and conventional techniques to reduce metastability error rates are discussed.

Chapter 5 presents an area and power efficient method of reducing metastability errors in high-speed flash A/D converters. The design of a 7-bit, 80-MHz, flash A/D converter in 1.2-µm CMOS as a test vehicle for the architecture is discussed. Performance measurements for fabricated parts are shown, and the design is compared to an internally pipelined converter.

Chapter 6 contains the conclusion and suggestions for further work. Appendix A contains information about the test setups for the measurements of metastability errors in synchronizers, and the setup to characterize these parameters in the presence of noise. Appendix B contains information on the test board for the A/D converter and the measurement setups for the performance of the A/D converter.

Chapter 2

Metastability

2.1 Introduction

Many of the most widely used CMOS static circuits are formed from regenerative circuits. Static memory cells, latches, flip-flops, sense amplifiers, comparators, arbiters, and synchronizers are all implemented with circuits which have potential metastability problems. The regenerative circuit is used because it can provide logic information, storage, and/or amplification. In many applications however, the inputs to the regenerative circuit cannot be carefully controlled, leading the circuit to remain in an intermediate, or metastable, state. An element in a metastable state can propagate errors to the rest of the circuit, causing the system to generate invalid outputs. The amount of time the circuit remains in the metastable state is unbounded, so the creation of an error free environment is not possible. Ultimately, the circuit designer must define an acceptable level of errors and design the circuit accordingly.

This chapter looks at three basic circuit blocks which use regenerative circuits; arbiters, synchronizers, and comparators. Applications are described where timing problems occur with regenerative circuits. The exponential delay of a regenerative circuit with inputs insufficient to generate a valid output is then derived. From the equation for delay, the equations to describe the performance metrics for synchronizers, arbiters, and comparators are derived. Most applications can be modeled by one of these three circuits, and error rates can be determined with minimal changes in the performance metric equations. In addition, the optimum device sizes for unloaded synchronizer/arbiter circuits is shown, and the main method of reducing error rates, pipelining, is discussed.

2.2   Regenerative Circuits

A regenerative circuit is a circuit with two stable states. The stable states are arbitrarily called one and zero. The regenerative circuit can be forced into either of these two states provided the circuit inputs meet some timing and voltage requirements. If the inputs are not sufficient to unambiguously define either a one or zero, the circuit will become metastable, and the final state is not deterministic. The circuits required to respond to a metastable regenerative circuit may not agree on the value, leading to intermittent errors posing reliability problems.

Figure 2.1 shows a conceptual view of the metastability problem in regenerative circuits. The simplest regenerative circuit is shown in (a), two cross-coupled inverting gain stages shown here as inverters. Possible operating points for the circuit are
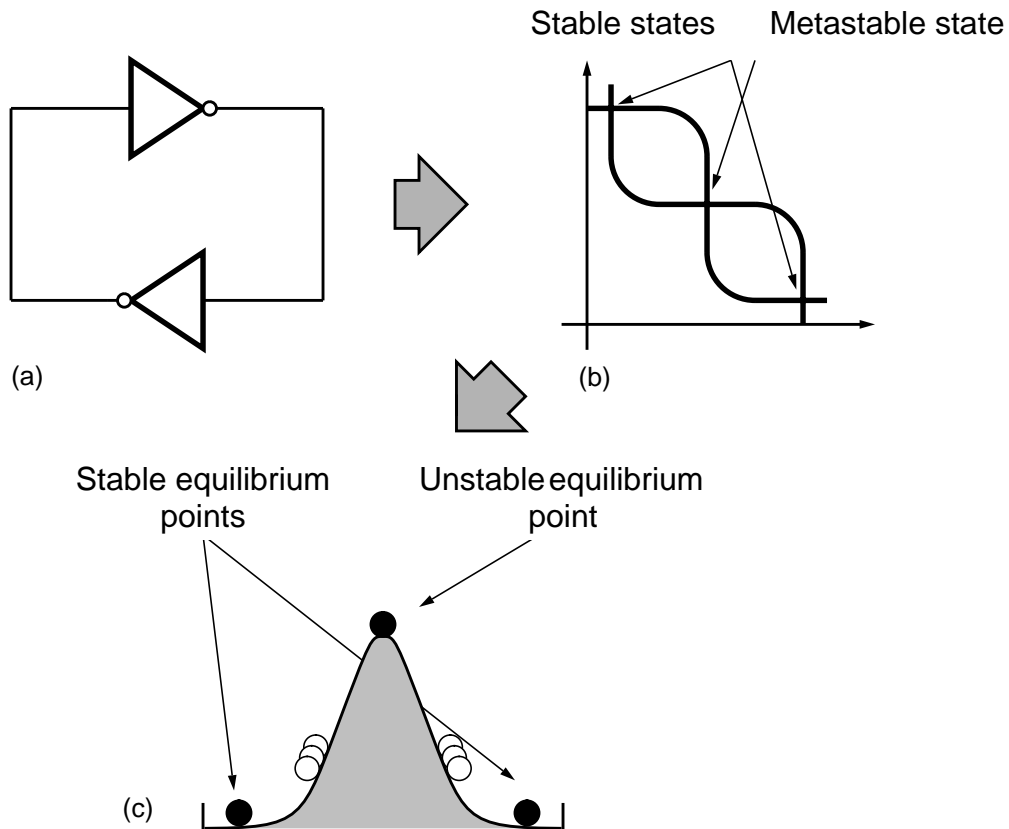
Figure 2.1: Mechanical analogy for metastable state. (a) Regenerative circuit; (b) load line representation; (c) mechanical analogy.

represented with the load line intersection points in (b) on the right hand side of the figure. The center intersection point in the load line plot is the circuit's operating point which corresponds to the ball on the hill's energy maximum and is an unstable operating point. However, in passing from one state to another, there is inevitably an energy maximum. The two stable states are near the positive and negative supply. Eventually the circuit will exit the unstable operating point due to noise, but the amount of time spent in the state is not well controlled.

The circuit remaining in the metastable state can be modeled conceptually as a ball on a hill as shown in Figure 2.1(c). The two stable states of the system correspond to the valleys, or stable equilibrium points, on either side of the hill [18]. If the ball is placed in either location, it will remain there indefinitely. However, if the ball is placed at the very top of the hill, it will remain there indefinitely, even though this is not an energy minimum. This is termed an unstable equilibrium point. If the ball is displaced to the left or right, it will quickly find one of the stable states. A metastable circuit operates in much the same manner. However, to characterize the circuit, the rate of entering the metastable state and the escape mechanism must be characterized. In addition, the distance that the ball must be displaced from the energy maximum before the location can be unequivocally declared in the left hand valley or the right hand valley must be ascertained.

2.2.1 Arbiters and Synchronizers

One operation in which metastability errors can occur is arbitration. An arbiter is a circuit that decides which of a number of signals arrives first. Figure 2.2 shows the circuit diagram and timing diagram for a two-input arbiter. The function of a two-input arbiter is to lock out one of the request inputs while the other is active. In this figure REQ_A activates before REQ_B, so REQ_B is locked out until REQ_A is completed. Arbiters are commonly used in multiport memories [10]. In a multiport memory, the arbiter must ensure that more than one of the ports does not attempt to access the memory core at the same time.

Two-input arbiters are implemented using an *RS* (reset/set) style latch. For the logic shown in Figure 2.2, both inputs low will clear the latch. One of the inputs going high will either set or reset the latch, and both inputs high will store the state of the first input pulled high. As long as REQ_A and REQ_B do not arrive at the same time, the timing restrictions for
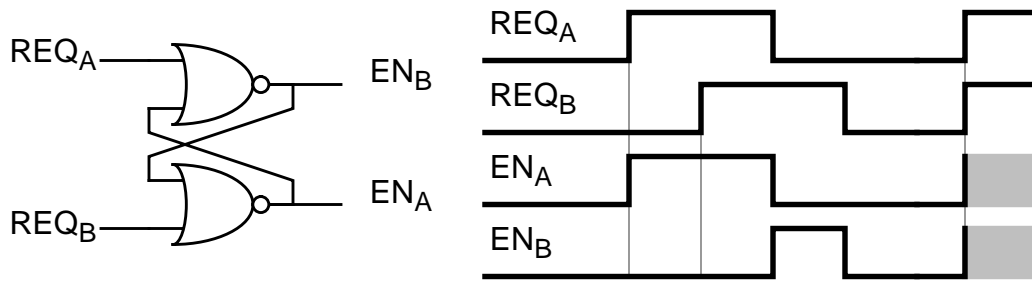
Figure 2.2: Circuit and timing diagram for two-input arbiter.
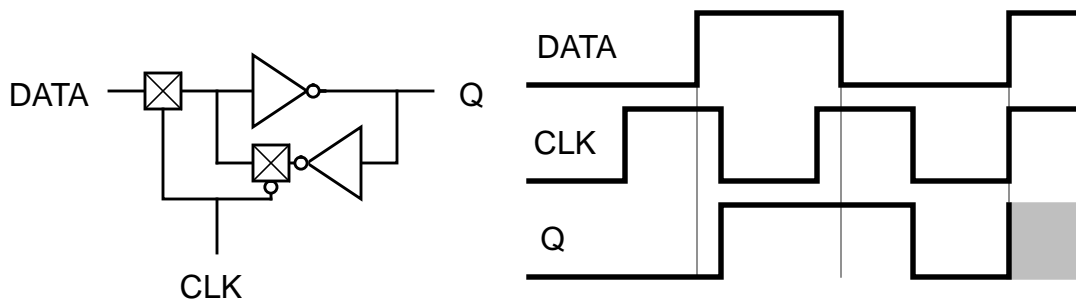
Figure 2.3: Synchronizer block and timing diagram.

the latch will not be violated. However, if REQ_A and REQ_B arrive simultaneously, full rail signals will not be transferred to the output of the latch. Both outputs, Q_A and Q_B, will languish in the intermediate state for an unbounded amount of time, leading to errors in the circuits that Q_A and Q_B drive. This is shown figuratively on the right hand side of the timing diagram with the undefined outputs of EN_A and EN_B.

Another application similar to an arbiter is a synchronizer. Chapter 1 described the most common application for synchronizers, bus controllers. Figure 2.3 shows the circuit diagram and timing for a synchronizer circuit. The function of the synchronizer is to maintain valid signals in Q, the output, during the low phase of the clock. This is often implemented with a D latch or flip-flop. If the *DATA* signal is asynchronous or synchronous with a clock frequency other than that of the receiving system's clock, *CLK*, the synchronizer will fail when *DATA* transitions near the clock rising edge. This is shown on the right hand side of the timing diagram with the undefined output in Q.

Due to the regenerative nature of the *RS* and *D* latches, the longer the application waits before accessing the outputs, the lower the probability that the outputs will be undefined when they are accessed. The relation is in fact an exponential one as section 2.3 will show. From the exponential relationship between the resolution time and error rates, section 2.4.1 will derive equations for the failure rates of arbiter and synchronizer circuits.

2.2.2 Comparators

The last circuit that will be discussed in which metastability errors occur is the latched comparator. Although the mechanism for metastability is the same as that of arbiters and synchronizers, the input waveforms are different. Latched comparators are used to amplify small initial voltage differences to valid logic levels, and to provide a sampling moment. The amplification is often done with a regenerative circuit because nonlinear regenerative amplification is the most rapid [19]. Figure 2.4 shows a circuit diagram and timing for a regenerative latched comparator. On the rising edge of the clock, the voltage difference between V_A and V_{REF} is amplified by the two inverting gain stages to valid logic levels. On the falling edge of the clock, the comparator outputs reset, and the process begins anew on the next clock rising edge. Although not shown explicitly here, there is additional circuitry to ensure the reset state will be a valid high.
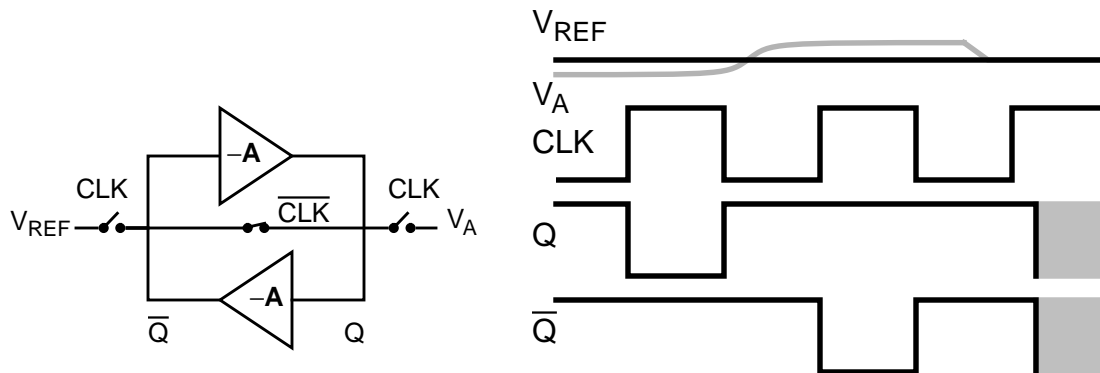
Figure 2.4: Timing diagram for comparator.

There will be some range of input voltages, however, for which the comparator will not be able to generate valid logic outputs in the time allotted for amplification. Invalid signals will propagate to the circuitry connected to the outputs, leading to metastability errors. Section 2.4.2 will calculate the range of input voltages which will lead to errors for a given resolution time and extend the analysis to calculate the probability of errors in a comparator.

2.3 Regeneration

In all of the previously described applications, metastability errors are caused by insufficient inputs to achieve a valid output in the time allotted to the regenerative circuit. This section describes the amplification of regenerative circuits and derives the exponential relationship which describes voltage levels attained in the time allowed to amplify initial conditions. A comparison of the regeneration time constant is made with the gain-bandwidth product of one of the inverting gain stages in the regenerative circuit. Finally, the optimum device sizes for the minimum regeneration time constant in an unloaded synchronizer/arbiter circuit is shown.

2.3.1 Regeneration Time Constants

While the actual implementation for the circuit elements discussed in section 2.2 can change depending upon circuit requirements, most regenerative circuits can be modeled as two identical back-to-back inverting gain stages driving equal loads [20]-[22]. Figure 2.5(a) shows a schematic for an idealized regenerative circuit. To model the regenerative amplification properties from the metastable state, initial voltage are placed on the outputs V_1 and V_2, and at time $t=0$ the switch is closed. For small initial voltage differences, linearized elements can be used to model the operation of the circuit in the region of interest, as shown in Figure 2.5(b). The node voltages v_1 and v_2 are referenced to the metastability voltage, which is equal to the switching point of the inverting gain stage. v_1 and v_2 satisfy the following two differential equations,

$$g_m v_1 + 2C_M \frac{d}{dt}(v_2 - v_1) + g_{out} v_2 + C_{out} \frac{dv_2}{dt} = 0 \tag{2.1}$$

$$g_m v_2 + 2C_M \frac{d}{dt}(v_1 - v_2) + g_{out} v_1 + C_{out} \frac{dv_1}{dt} = 0 \tag{2.2}$$

where g_m is the transconductance, g_{out} is the output transconductance of one of the inverters, C_M is the Miller capacitance around the gain stage, and C_{out} is the lumped capacitive load at the gain stage's output. For a CMOS inverter implementation, the equation parameters are:

$$g_m = g_{mn} + g_{mp} \tag{2.3}$$

$$g_o = g_{on} + g_{op} \tag{2.4}$$

$$C_M = C_{gdn} + C_{gdp} \tag{2.5}$$

$$C_{out} = C_{gsn} + C_{gsp} + C_L + C_{jn} + C_{jp}. \tag{2.6}$$

In the above equations, the subscript n denotes an n-channel device, and the subscript p denotes a p-channel device. C_{gs} and C_{gd} are the gate-source and gate-drain capacitance of an MOS device, respectively. C_j is the junction capacitance of a drain region of an MOS device. All of the parameters are calculated about the inverter switching point.
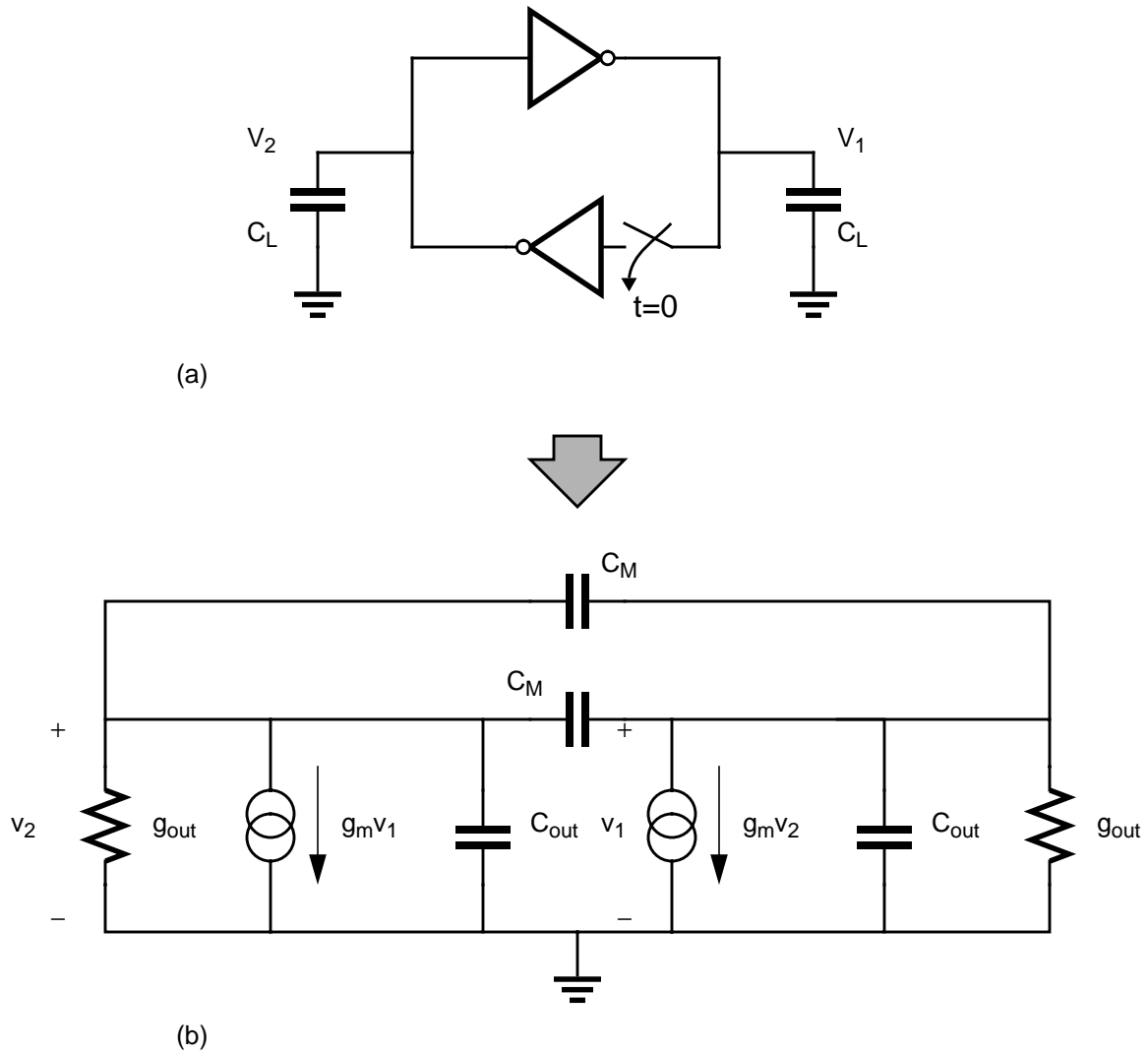
Figure 2.5: Regeneration in regenerative circuits. (a) Simplified schematic; (b) small-signal model.

Equations (2.1) and (2.2) can be solved to yield equations for the node voltages with respect to time. However it is much more intuitive to discuss the common-mode and differential-mode voltages at the two nodes because the polarity and magnitude of the differential voltage will ultimately determine whether the output can be determined as a valid 'one' or 'zero'. The common-mode, V_{comm}, and differential-mode, V_{diff}, voltages are equal to

$$V_{comm}(t) = \frac{v_1(t) + v_2(t)}{2} \tag{2.7}$$

$$V_{diff}(t) = \frac{v_1(t) - v_2(t)}{2} . \tag{2.8}$$

From equations (2.1) and (2.2), the solutions for the differential and common-mode voltages can be shown to be

$$V_{comm}(t) = V_{comm}(0) \, e^{-t(g_m + g_{out})/C_{out}} \tag{2.9}$$

$$V_{diff}(t) = V_{diff}(0) \, e^{t(g_m + g_{out})/(C_{out} + 4C_M)} . \tag{2.10}$$

The common mode voltages with the negative exponential can generally be ignored after several time constants. From equation (2.10) it can be seen that the differential-mode voltage will grow exponentially with time. The equation is usually represented in the form

$$V_{diff}(t) = V_{diff}(0) \, e^{t/\tau} \tag{2.11}$$

where

$$\tau = (C_{out} + 4C_M)/(g_m + g_{out}) . \tag{2.12}$$

τ is called the regeneration time constant for the circuit, and will determine the amount of time required for an initial voltage difference to reach the required output level for processing by successive circuits. For very small differential initial voltages, the time to generate a valid output voltage can be many time constants. Theoretically, for a perfectly balanced input, the circuit will remain balanced at the metastable point indefinitely.

In most practical implementations g_m is greater than g_{out}, and C_{out} is greater than C_M, so the exponential time constant can be approximated by

$$\tau = C_{out}/g_m. \tag{2.13}$$

The parameter τ is often related to the gain-bandwidth product of one of the gain stages. However, strictly speaking, this is incorrect. As will be shown in section 2.3.2, while the gain-bandwidth product is closely related to τ, there is by no means an exact correspondence.

2.3.2 Regeneration Time Constant and Gain-Bandwidth Product

The parameter τ is often related to the gain-bandwidth product of one of the gain stages because the two quantities are equal to the first order [20], [23]. This section will show the calculation of the gain-bandwidth product for an inverter biased to the point $V_{out} = V_{in}$. Figure 2.6 shows a schematic of the regenerative circuit of Figure 2.5 with the feedback loop opened. If V_1 has a dc bias equal to the switching point of the inverter, both transistors in the inverter will be in the saturation region, and the small-signal model can be generated for the circuit as shown in Figure 2.6(b). All parameters in the schematic are equal to those of equations (2.3) - (2.6).

For small-signal perturbations, the low frequency gain, v_2/v_1, for the circuit is

$$A = g_m/g_{out}. \tag{2.14}$$

Because the two stages are identical, the Miller capacitance around the load stage can be modeled by $(1-A)\,C_M$. Lumping the Miller capacitance with the load capacitance, the circuit of Figure 2.5(b) can be modeled as a one pole system with a pole at

$$p_1 = (C_{out} + (1-A)\,C_M)/g_{out}. \tag{2.15}$$

The gain-bandwidth for the inverter is then

$$GB = \frac{g_m}{(C_{out} + (1-A)\,C_M)}. \tag{2.16}$$

It should be noted that equations (2.12) and (2.16), though similar, differ in the multiplying factor on the term C_M. If C_{out} is large compared to the term $(1-A)\,C_M$, then the gain-bandwidth product can be approximated by g_m/C_{out}, which is the inverse of the first order approximation for the regeneration time constant, τ. Because of this, the
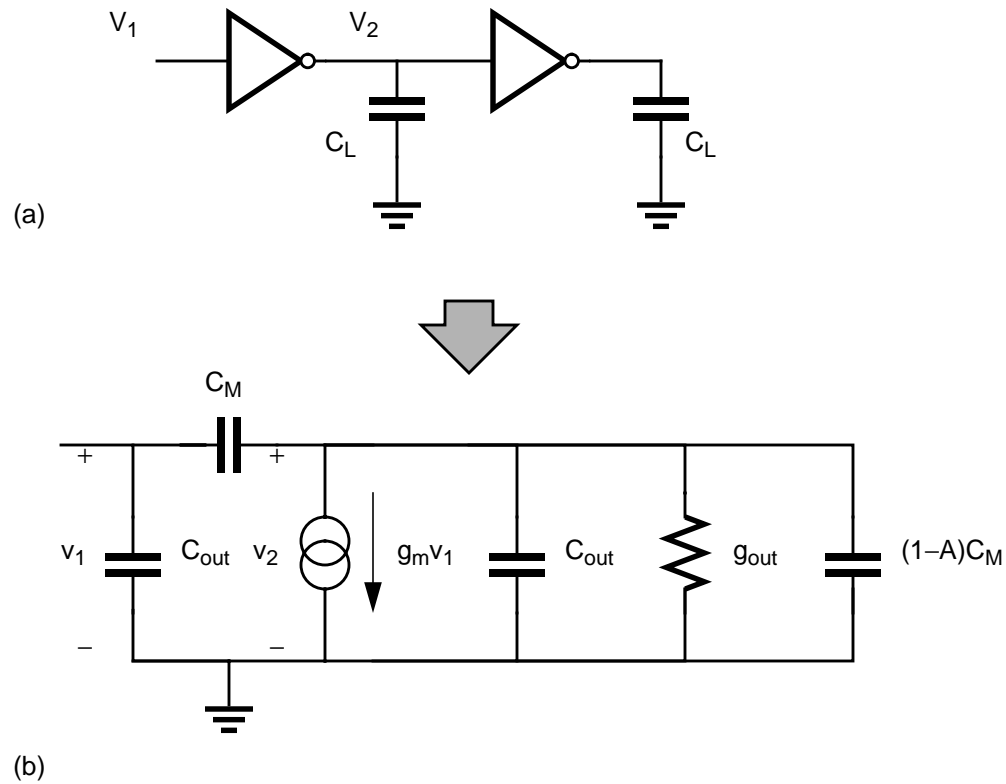
Figure 2.6: Gain-bandwidth product in inverters. (a) Simplified schematic; (b) small-signal model.

regeneration time constant τ can be approximated using ac simulations [23]. Using ac simulations greatly reduces the simulation time, because the timing point for metastability does not need to be found. However, as will be seen in the section 2.4.1, there are two parameters needed to fully describe the metastability performance of a regenerative circuit in an application, the regeneration time constant τ and a pre-exponential constant T_o. To fully characterize a regenerative circuit in latches, transient simulations must be used to find the timing point for metastability with data and clock signals to trigger the element into metastability.

2.3.3 Minimizing Latch Regeneration Time Constants

Using current equations for the MOS devices in the inverters of Figure 2.5, an expression for the regeneration time constant be derived, and device sizes for the minimum regeneration time constant can be calculated. Because other constraints also impact device sizes in high-speed, precision comparators, the equations derived in this section are inadequate to fully quantify a comparator's performance. However for synchronizers and arbiters, the small-signal model of Figure 2.5(b) is sufficient, and optimum device sizes for the regeneration time constant can be determined [21], [22].

As shown in section 2.3.1, the regeneration time constant, τ, can be approximated by the quantity C_{out}/g_m to the first order. If C_{out} is assumed to be some multiple of the input capacitance,

$$C_{out} = K_\tau C_{ox} (W_p + W_n) L, \tag{2.17}$$

where K_τ is the ratio of output capacitance to input capacitance, C_{ox} is the gate capacitance per unit area, L is the device length, and W_n and W_p are the channel widths for the n and p-channel devices, respectively. Device currents can be calculated using the following equations,

$$I_n = K_n \mu_n W_n (V_{gs} - V_{thn})^{x_n} \tag{2.18}$$

and

$$|I_p| = K_p \mu_p W_p (V_{DD} - V_{gs} - |V_{thp}|)^{x_p} \tag{2.19}$$

where μ_n and μ_p are the electron and hole mobility, respectively, and x_n and x_p are constants between one and two. For square law devices with $x_n = x_p = 2$, the constants K_n and K_p are equal to $C_{ox}/2L$. V_{DD} is the power supply voltage.

Assuming $x_n = x_p = x$, and $V_{thn} = |V_{thp}| = V_{th}$, the metastable voltage, V_m, can be solved for by setting I_n equal to I_p and is equal to

$$V_m = \frac{V_{DD} + V_{th}\left(\left(\frac{\mu_n W_n}{\mu_p W_p}\right)^{1/x} - 1\right)}{\left(\frac{\mu_n W_n}{\mu_p W_p}\right)^{1/x} + 1}. \tag{2.20}$$

The transconductance, dI/dV_{in}, of one of the gain elements in Figure 2.5(a) is

$$g_m = xK_n\mu_n W_n\left(\frac{V_{DD}-2V_{th}}{R^{1/x}+1}\right)^{x-1} + xK_p\mu_p W_p R^{\frac{(x-1)}{x}}\left(\frac{V_{DD}-2V_{th}}{R^{1/x}+1}\right)^{x-1} \tag{2.21}$$

where R is the ratio of the device widths and carrier mobilities, and is equal to

$$R = \frac{\mu_n W_n}{\mu_p W_p}. \tag{2.22}$$

The regeneration time constant, C_{out}/g_m, can be shown to be

$$\tau = \frac{K_\tau C_{ox}(W_p + W_n)L}{xK_n\mu_n W_n\left(\frac{V_{DD}-2V_{th}}{R^{1/x}+1}\right)^{x-1} + xK_p\mu_p W_p R^{\frac{(x-1)}{x}}\left(\frac{V_{DD}-2V_{th}}{R^{1/x}+1}\right)^{x-1}}. \tag{2.23}$$

For square law devices with $x = 2$, τ can be simplified to

$$\tau = \frac{K_\tau L^2(W_n + W_p)}{(V_{DD}-2V_{th})\sqrt{\mu_n W_n}\sqrt{\mu_p W_p}}. \tag{2.24}$$

By taking the derivative of equation (2.24) with respect to W_p, the device sizes for minimum τ can be shown to be at the point

$$W_n = W_p. \tag{2.25}$$

For arbitrary values of x, the derivative of equation (2.23) becomes unwieldy, and it is easier to represent equation (2.23) in graphical form. Figure 2.7 shows a plot of τ normalized to the minimum τ versus the ratio of device widths for several values of x. As can be seen from the plotted data, as the exponent of the current drive equations increases the ratio, W_n/W_p, for minimum τ values increases. For the extreme point of $x = 1$, the optimum ratio of W_n/W_p is equal to infinity.
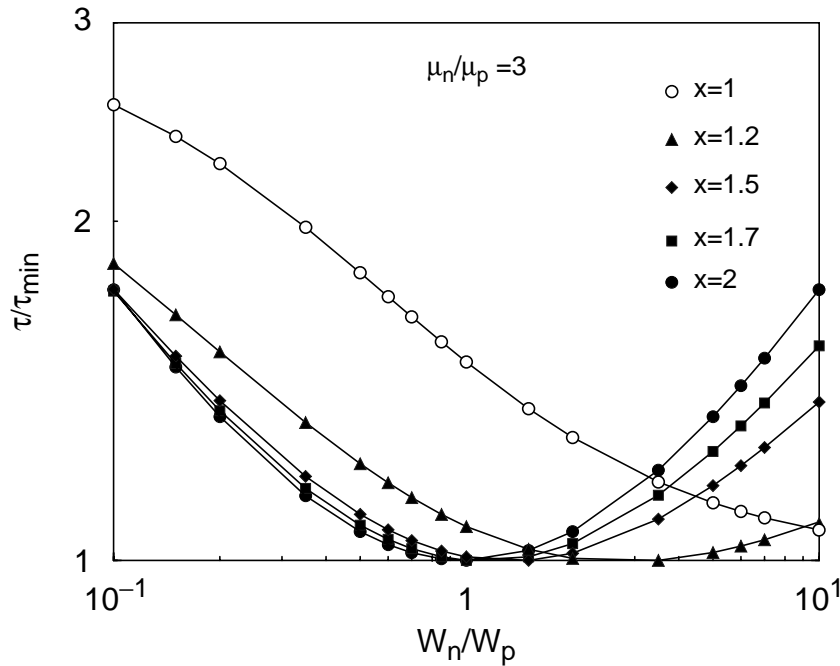
Figure 2.7: Normalized latch regeneration time constants for various device size ratios.

2.4 Performance Metrics

The exponential relationship between the differential output and the resolution time for regenerative circuits derived in equation (2.11) can be extended to calculate the performance metrics for systems using regenerative circuits. The mean time between failures, *MTBF*, is normally used for digital applications using synchronizers and arbiters. However, the probability of errors, P_E, is normally used for analog applications using comparators, namely high-speed flash A/D converters. Historically, the metastability characterizations of analog and digital systems have evolved independently so the performance metrics are different. However, the metrics are more or less equivalent as will be shown in section 2.4.2.

2.4.1 Synchronizers and Arbiters

In synchronizers and arbiters, metastability occurs when the setup/hold window for the circuit is violated. As an example, the synchronizer of Figure 2.3 is used. If the data input of the synchronizer is swept over time with respect to the clock as shown in Figure 2.8(a), a plot similar to that of Figure 2.8(b) can be obtained [24]. A similar plot can be generated for an arbiter by fixing one *REQ* input arrival time and sweeping the other with respect to it. On the far left of Figure 2.8(a), *DATA* arrives while the latch is open, so the output is valid after the propagation delay of the latch. On the far right, *DATA* arrives after the latch closes, so there is no data transferred to the output node and there is no delay. In the center region, the *DATA* transition occurs just as the latch closes. Because the internal circuit has partial inputs when it is switched to the regenerative mode, the latch requires extra time to achieve a valid output. Typically, this region of additional delay is in the hundreds of picoseconds for 2-μm technology and tens of picoseconds for 1.2-μm technology. When data arrives at time t_{meta}, the latch requires the maximum time to resolve. Theoretically, the time required to resolve the output would be infinite because the regenerative circuit would be perfectly balanced. The location of t_{meta} can be arbitrarily placed with respect to
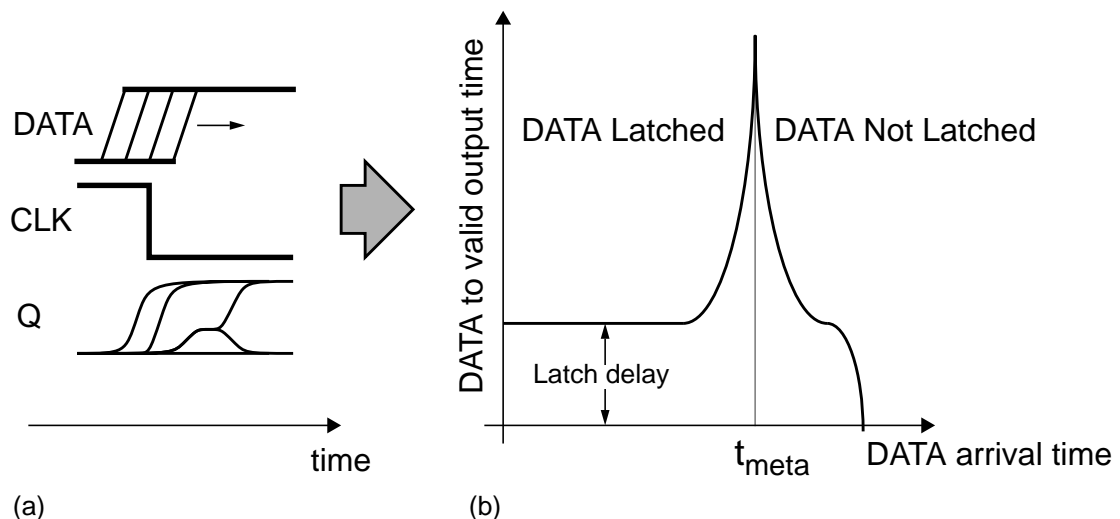
Figure 2.8: Metastability in latch. (a) DATA arrival time is swept with respect to CLK edge; (b) time to valid output versus DATA arrival time.
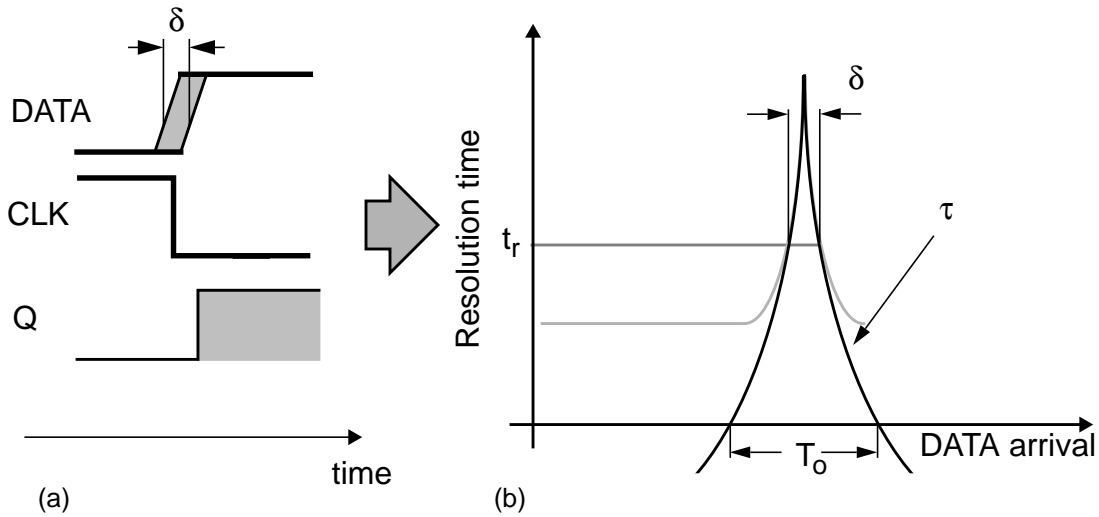
Figure 2.9: Detail of fitted exponentials showing metastability window, regeneration time constant, and pre-exponential constant.

the clock edge because the data and clock signals have different signal paths within the latch. As described in sections 2.3.1 and 2.3.2, the delay in the region close to t_{meta} is exponential in nature, where the exponential constant is, to the first order, the inverse of the gain-bandwidth product of the feedback element within the latch [20]. Figure 2.9 shows the plot of Figure 2.8(b) with two exponentials to model the behavior of the delay in the region of t_{meta}. A window called the metastability window, δ, can be defined such that data which transitions within δ will not be resolved within a given resolution time, t_r. The metastability window's width can be calculated for a given resolution time with the equation,

$$\delta = T_o e^{-t_r/\tau} \tag{2.26}$$

where T_o is the asymptotic width of the window with no resolution time, and τ is the regeneration time constant. The pre-exponential constant, T_o, is related to the setup time and minimum voltage required for a valid output. By specifying T_o and τ for a latch with a

given supply, loading condition, and supply noise characteristic, the metastability performance of the latch in the application can be determined.

If data transitions randomly with respect to the clock, the probability that the data will transition within δ is $f_D\delta$, where f_D is the average frequency of the data. The mean time between failures is described by the equation

$$MTBF = \frac{1}{f_D f_{CLK} T_o e^{-t_r/\tau}} \tag{2.27}$$

where f_{CLK} is the frequency of the clock.

If the *MTBF* for a synchronizer in an application is unacceptable, another synchronizer with better performance can be selected, or pipelined synchronizers can be used to reduce the *MTBF* performance to acceptable levels. The characterization and comparison of the τ and T_o parameters for various designs is covered in chapter 3. Pipelining to improve performance is discussed in section 2.4.3.

2.4.2 Comparators

In comparators, metastability also causes problems because the outputs are often processed by digital logic or digital control loops. In high-speed analog-to-digital conversion with large numbers of parallel comparators, metastable errors can become prohibitively large. However, due to the nature of the input, the calculation for the error performance is voltage based rather than time based.

In a regenerative comparator, the comparator is presented with a voltage difference which is amplified nonlinearly up to a full-rail signal when the comparator is strobed. The regeneration process provides amplification, but also keeps the outputs valid for processing for the remainder of the clock cycle. Figure 2.10 is a sketch of the response time for a regenerative comparator with various input differences. As can be seen, as the voltage difference grows small, the time required for amplification increases. The behavior is in fact the exponential relationship derived in equation (2.11). From this relationship, and the input voltage behavior, the probability of error or P_E for the comparator can be calculated.
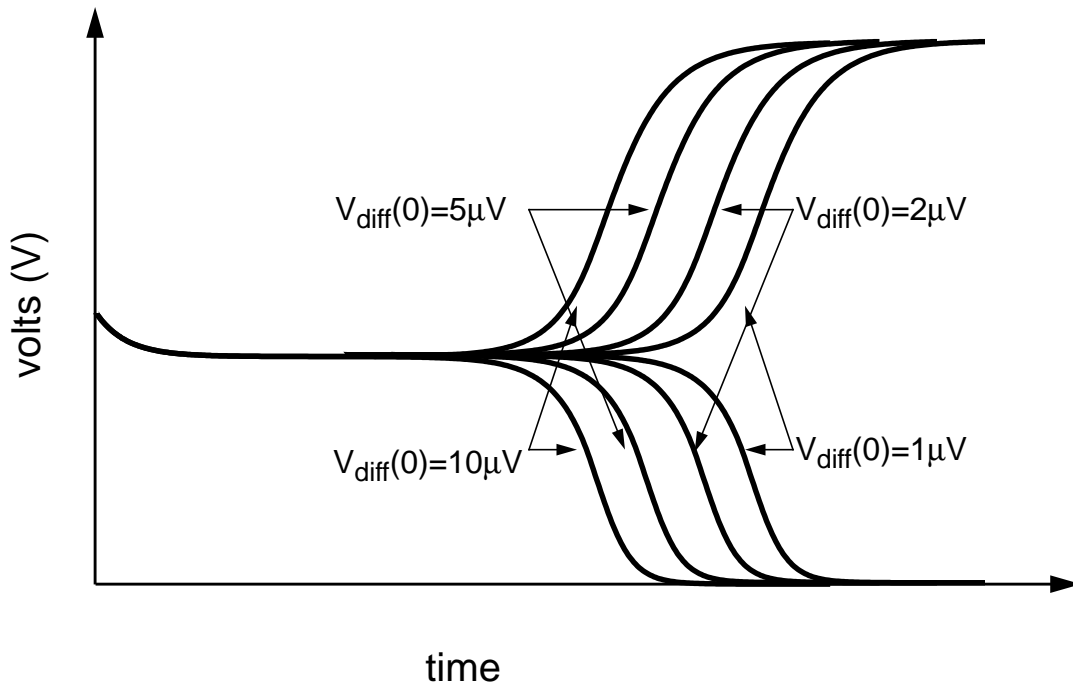
Figure 2.10: Comparator delay with different input voltages.

Assuming the latched comparator consists of a low offset preamp followed by a regenerative latch, the input voltage range which will be unable to resolve in the time slot satisfies the equation

$$\left| V_{REF\,(i)} - V_{in} \right| \leq \frac{V_o}{A} e^{-t_r/\tau} \tag{2.28}$$

where V_o is the output voltage swing required at the output for valid logic levels, A is the preamplifier gain and the linear gain of the latch in the open state, τ is the regeneration time constant for the latch, and t_r is the resolution time of the latch. The term $A e^{t_r/\tau}$ can be viewed as the effective gain for the comparator over the time period.

For a uniformly distributed input, the probability that the input is within the range of (2.28) is simply this range divided by the full scale range of V_{in}. The error probability for the comparator then is

$$P_E = 2\frac{V_o}{V_R A}e^{-t_r/\tau} \tag{2.29}$$

where V_R is the input range. It is interesting to note that the error probability is independent of the frequency of the analog input to the first order. From equation (2.29) it can be seen that the most effective means of reducing a comparator's P_E is by manipulating the terms in the exponential. Section 2.4.3 discusses pipelining to improve comparator metastability performance.

Typically, comparator metastability error performance has been described by the error probability rather than the *MTBF*, but the *MTBF* can be calculated to show the similarity between the error rates for comparators and synchronizers/arbiters. Multiplying the error probability by the clock frequency, the *MTBF* for a comparator in an application is

$$MTBF = \frac{1}{\frac{2V_o}{AV_R}f_{CLK}e^{-t_r/\tau}} \quad . \tag{2.30}$$

Equation (2.30) is similar to equation (2.27) with $f_D T_o$ replacing the term $2V_o/AV_R$. In both cases these are dimensionless quantities to represent the probability that the input signal is within a specific range. Synchronizers and arbiters use a time base probability while comparators use a voltage based probability.

2.4.3 Pipelining

In both synchronizer/arbiter and comparator applications, the most effective way of improving the metastable performance is by pipelining regenerative circuits. Although τ can be improved up to a point with device sizing, it is ultimately limited by the technology. In this section we generate equations for synchronizer and comparator applications to show the performance improvements with pipelining.
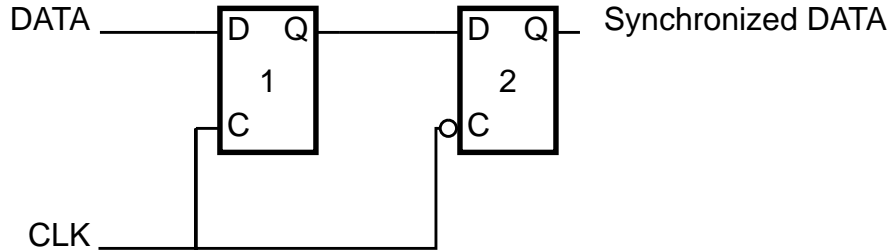
Figure 2.11: Pipelined synchronizers.

Figure 2.11 shows a schematic for a pipelined synchronizer. To determine the mean time between failures, the probability that the output of the first synchronizer is unsettled when the second synchronizer samples its output must be determined. From equation (2.26) we know that the second synchronizer will fail if data falls within the second window, δ_2. The only way data can fall within the second window is if the data is still unresolved from the first synchronizer's window, δ_1. The frequency of data unresolved at the output in the first window is $f_{CLK}f_D\delta_1$. Thus, the frequency of data unresolved in the second window is $f_Df_{CLK}\delta_1f_{CLK}\delta_2$. The mean time between failures is described by the equation

$$MTBF = \frac{1}{f_Df_{CLK}\delta_1f_{CLK}\delta_2}.$$ (2.31)

In general, the quantity δf_{CLK} must be much less than one for the circuit to operate, so pipelining improves performance by effectively increasing the resolution time.

For comparators, pipelining increases the effective gain of a comparator, so the error probability with a pipeline latch at the output is

$$P_e = \frac{2^n V_o}{V_R A A_l} e^{-(t_r - t_{su})/\tau} e^{-t_l/\tau_l}$$ (2.32)

where A_l, t_l, and τ_l are the gain of the latch in the transparent state, the resolution time for the latch, and the regeneration time constant for the latch, respectively, t_{su} is the setup time

for the latch, and *n* is the number of pipeline stages [7]. Errors can be reduced to arbitrarily low levels using pipelining, but there is a sacrifice of power, area, and latency.

2.5 Summary

In this chapter, applications with regenerative circuits where metastability errors are a concern have been discussed. Simple timing diagrams for an arbiter, synchronizer, and comparator have been shown with a discussion of the timing problems. The equation for the exponential behavior of a regenerative circuit with inputs that place the circuit into the metastable state has been derived. A discussion of this exponential time constant and its relationship to the gain-bandwidth product for one of the gain elements in the regenerative circuit is included. The equations for the optimum device sizes for synchronizer/arbiter performance have derived. The equations for failure rates in synchronizer and comparator application have been calculated. Additionally, the primary circuit technique for improving synchronizer performance, pipelining, was discussed, and the equations for synchronizer and comparator performance with pipeline latches at the output have been shown.

Chapter 3

Characterization of Standard Cell Synchronizers

3.1 Introduction

As device dimensions shrink and die sizes grow, the number of transistors implemented in very large scale integrated (VLSI) circuits becomes increasingly larger. In addition, time-to-market concerns push for faster and faster design cycles. These two factors have combined to push digital designers to use high-level computer aided design (CAD) tools to make designs manageable and reduce design times. High-level CAD tools generally simplify the design space over full custom designs, and integrated circuit designs are synthesized using a predesigned library often provided by a vendor. Simplified characteristics of the cells in the library are provided for the modeling of the circuit. From this information, the designer generally creates a netlist using a set of tools to allow for schematic capture and/or behavioral description. The design is then created using placement and routing tools. From the synthesized design, a back-annotated netlist is created to allow simulation of the final implementation. Along the way, the designer provides input to optimize the design to meet design criteria such as speed, power dissipation, and die size.

Prior research in metastability has concentrated on clocking and circuit schemes to increase the synchronizer's resolution time [15], [17], minimizing the regeneration time constant, τ, for unloaded latches [21]-[23], and integrated measurement techniques for the metastability window, δ [25]. However, the behavior of latches placed in circuits with respect to loading has not been investigated. Without such information, the results of

previous papers, while important, represent only discrete points in the design space. In application specific integrated circuit (ASIC) design, values for the metastable parameters, T_o and τ, for different loading conditions are necessary to predict metastable performance in the variety of loading conditions synchronizer/arbiter cells can be placed. In addition, to make design trade-offs in a variety of applications, the behavior of these parameters with respect to buffering, gate delay, power supply, and technology is helpful.

This chapter characterizes the metastability behavior of CMOS standard cell latches under different loading conditions, discusses latch metastable performance with respect to power supply and technology scaling, and investigates the effects of power supply noise. A discussion of simulating metastability is included showing the procedures needed for the extraction of metastable parameters from simulation. The assumptions as well as the circuits and models used are discussed in detail. The concept of standardized loads is modeled with gate fanouts and is used to characterize the latches with respect to loading. The design of a standard circuit to characterize metastability behavior of CMOS library latches is presented, which was implemented in 2-μm and 1.2-μm CMOS technologies. Measured results for latches versus loading, supply voltage, technology, and noise are presented. All the simulations and measurements shown in this chapter are for *D* latches used as synchronizers, but the results hold for *RS* latches used in arbiters as well.

Both theoretical and measured results show that both τ and T_o for an unbuffered latch vary linearly with loading. For a buffered latch, τ does not vary with loading to the first order, but T_o is exponentially increased by loading. A formula to calculate T_o for buffered latches and the effects of further buffering are presented. A normalization scheme for metastable parameters is developed to predict the parameters versus the supply voltage. The effects of power supply reduction are largely normalized out with proper definition of the parameters of interest for a wide range of supply voltages. However, metastable performance is shown to become markedly worse as the supply voltage approaches the sum of the magnitude of the device thresholds. The issue of technology scaling is addressed with the same normalization scheme, and the performance of a scaled circuit is shown to be influenced by a number of technology dependent factors. In contrast to prior work, supply noise is shown to degrade performance by degrading the T_o parameter. The results are useful to an ASIC designer with a standard cell or gate array library, or a custom designer wishing to design a multipurpose synchronizer or arbiter cell.

3.2 Latch Selection

While T_o and τ will vary for synchronizing elements of different styles and sizes, the behavior of these parameters for synchronizers placed in a circuit falls into two distinct categories: buffered and unbuffered. As such, synchronizing elements, regardless of style, can be broadly grouped into these two categories. Figure 3.1 shows schematics for the unbuffered and buffered latches used in this study. This latch style was chosen because it is one of the most commonly used latches in CMOS cell libraries. The parameters for flip-flops based on this latch style can be easily calculated from our results with the appropriate loading choices for the master and slave latches from equation (2.31). In addition, other latch styles exhibit similar behavior over loading, supply, and technology. No device size optimization has been performed with respect to metastable parameters, but a simple sizing scheme more in line with a standard cell or gate array library has been used. In all cases inverter transistors are four times the minimum length, and pass-gate transistors are two times the minimum length. For the buffered latch, the negative output, $\overline{Q}$, is used to eliminate any changes in T_o and τ due to internal loading changes of the latch. This allows easy comparison and parameter extraction between the latches.

3.3 Simulating Metastability

As discussed in section 2.3.2, because of the close correlation between τ and the gain-bandwidth product, the parameter τ can be approximated using ac simulations, greatly speeding up simulation time [26]. Additionally, the latch can be forced into the metastable state using ideal switches to find τ, but again the parameter T_o must be approximated or found by other means with this approach [15]. To fully describe the latch with respect to metastability, both parameters T_o and τ are needed. T_o can only be obtained with transient simulations. Transient simulations are performed with HSPICE [26] using clock and data signals to trigger the latch into metastable behavior.
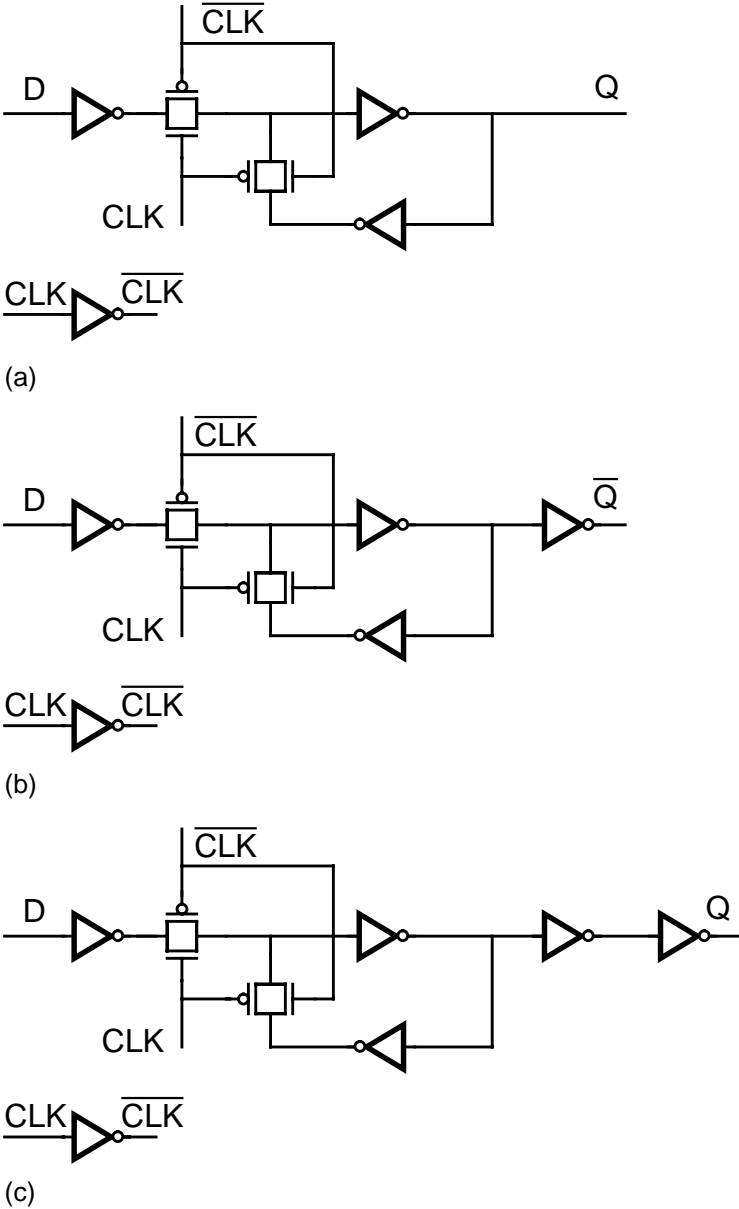
Figure 3.1: Latch schematics. (a) Unbuffered latch; (b) buffered latch; (c) double-buffered latch.

3.3.1 Simulation Environment

For the simulations throughout this chapter, latches are used as loads. To model transistor mismatch, a voltage source is included before the final buffer of the buffered latch. If the voltage source is not included, the latch's performance appears better with additional buffers by appearing to resolve finer and finer voltages. The value of this voltage source is estimated from transistor mismatch using parameters from [27] scaled for the various technologies. A charge conservation model is used, and the channel charge is split equally between the drain and source [28]. Simulations are run with a single 5-V supply, and all signals are measured from the inverter switching point rather than an arbitrary reference such as 50% of the power supply. This reduces edge rate effects that alter T_o. Additionally, since the data and clock inputs are buffered internally, increasing the edge rates of the external signals has little effect on the parameters of the latch.

3.3.2 Finding the Metastable Point

Since the latch delay is maximum at t_{meta}, an optimization program is used to find t_{meta} to speed up the simulation time. Performing transient sweeps by displacing the data arrival time linearly would be very time consuming due to the large number or simulations that would have to be performed. The optimization routine is a one dimensional minimization algorithm based on Brent's method [29] which maximizes the latch's time to achieve a valid output to a given precision. The data input signal is varied relative to the clock signal within a specified range, with the variation determined from the results of the previous runs. The output voltage vectors from each simulation are searched from the end of the simulation back in time to eliminate any spurious trip points caused by nonmonotonic behavior of the outputs in the metastable state. Charge, current, and voltage tolerances are reset to achieve consistent behavior near t_{meta}. These tolerances are set to 10^{-16} C, 1 fA, and 0.1 nV, respectively.

Figure 3.2(a) shows simulation results for a fanout of four buffered latch in 2-μm technology. The simulation models used are *n*-well CMOS models from the MOS Implementation System (MOSIS) [30], [31]. The use of the optimization routine allows easy automation in finding t_{meta} for the various styles and loading conditions. Once t_{meta} is determined, the data arrival time is displaced logarithmically to one side of t_{meta}. If the
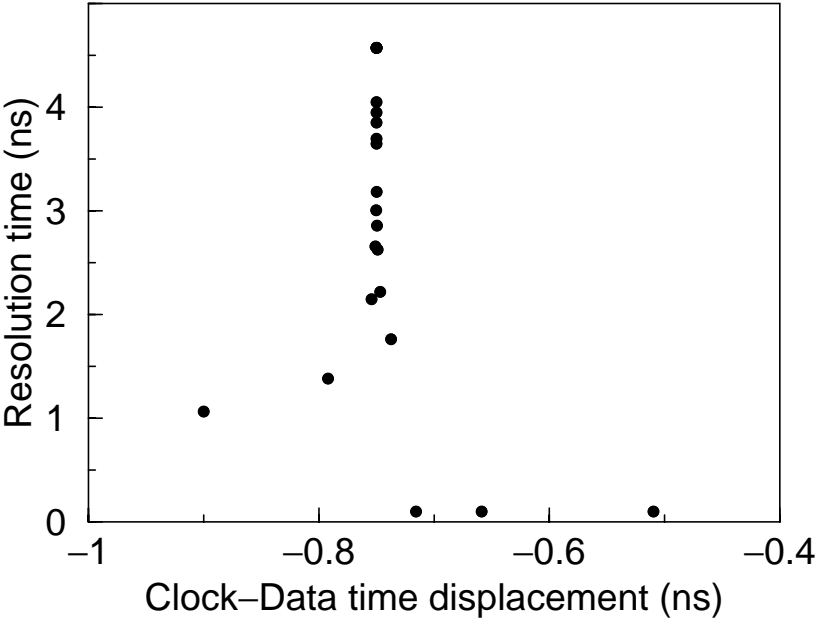
(a)
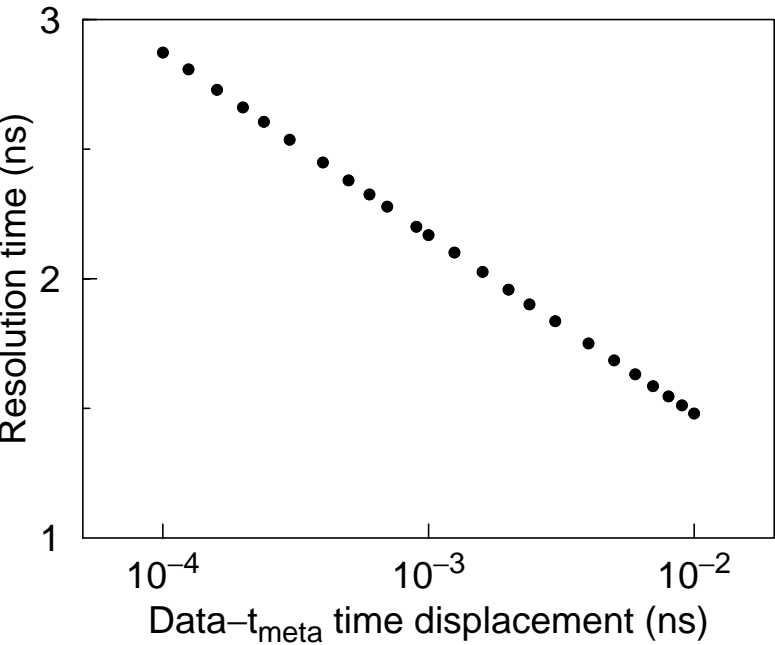
(b)

Figure 3.2: Metastable latch behavior. (a) Optimizer results; (b) data near metastable point.

resolution time is plotted versus the displacement of the input on a natural log-linear graph, the slope of the curve is τ, and the intercept is $\ln(T_O/2)$. Figure 3.2(b) shows the results of these simulations on a semilog plot. As can be seen, the results follow the exponential model quite well. Similar results were achieved in 1.2-μm and 0.8-μm simulations.

3.4 Measuring Metastability Parameters

In addition to simulating metastability it is important to measure the characteristics of a synchronizer or arbiter. This section describes an integrated test measurement technique for measuring T_O and τ. The technique is based on a previously reported circuit, but includes the ability to vary the synchronizer's load [25]. Additionally, the implemented circuit is specifically designed to operate in a wide variety of CMOS technologies and a wide range of supply voltages. The specifics of the test board and instrument configuration are included in Appendix A.

3.4.1 Test Circuit

A total of three test-chips were designed in CMOS technology to measure τ and T_O of simple latches used as synchronizers. Figure 3.3 shows a schematic of the measurement technique used. The latch being tested is given clock and data signals of non-harmonically related frequencies, so there is a finite probability of metastability. The output of the latch under test is connected to two latches which are enabled by delayed versions of the clock. One latch is latched with a long constant delay, the other with a shorter, variable delay. The variable delay line allows changes in t_r for the latch under test and latches the upper compare-latch while the output signal of the latch under test is still unresolved. The lower compare-latch is clocked with a long enough delay that the probability of the output not settling is orders of magnitude lower. The long delay models the case of infinite clock skew between the two compare-latches. Exclusive-ORing the outputs of the two compare-latches determines if the test latch output was unresolved at time t_r. By keeping track of the number of metastable events and the length of the test, the *MTBF* can be measured. Knowing the test clock and data frequencies, the metastability window for the latch can be calculated. Dummy loads are included to characterize the latches with respect to loading.
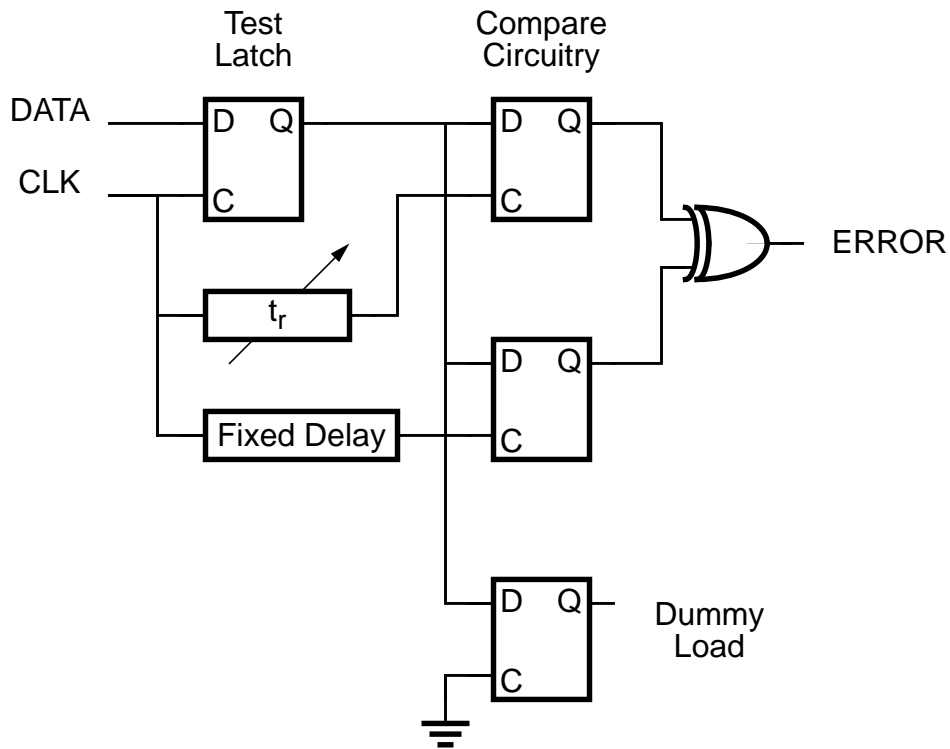
Figure 3.3: Metastability error rate test circuitry.

The 2-μm test-chip has modules to characterize the top two latches of Figure 3.1 under three different loading conditions. Figure 3.4 shows a die photograph of the 2-μm test-chip. The die area is 2.2×2.2 mm^2. The active area is 1.8×1.9 mm^2. Input capacitance, clock loading, and data edge rates are equal for all test configurations. A replication scheme is used to create the various loading conditions. An additional module is included to measure the delay of the variable delay elements, implemented here as a tapped inverter chain with a multiplexer to select the appropriate tap. Figure 3.5 shows the schematic for the variable delay element. The multiplexer is designed with each input having an input loading of a fanout of one, independent of delay selection. In addition, all inputs have equal delay to the output node.

Since the loading for the test latch in Figure 3.3 is implemented with gates, and the delay elements are implemented on-chip, the test structure is easily scalable to smaller technologies. A second, 1.2-μm chip was implemented to allow measurement of
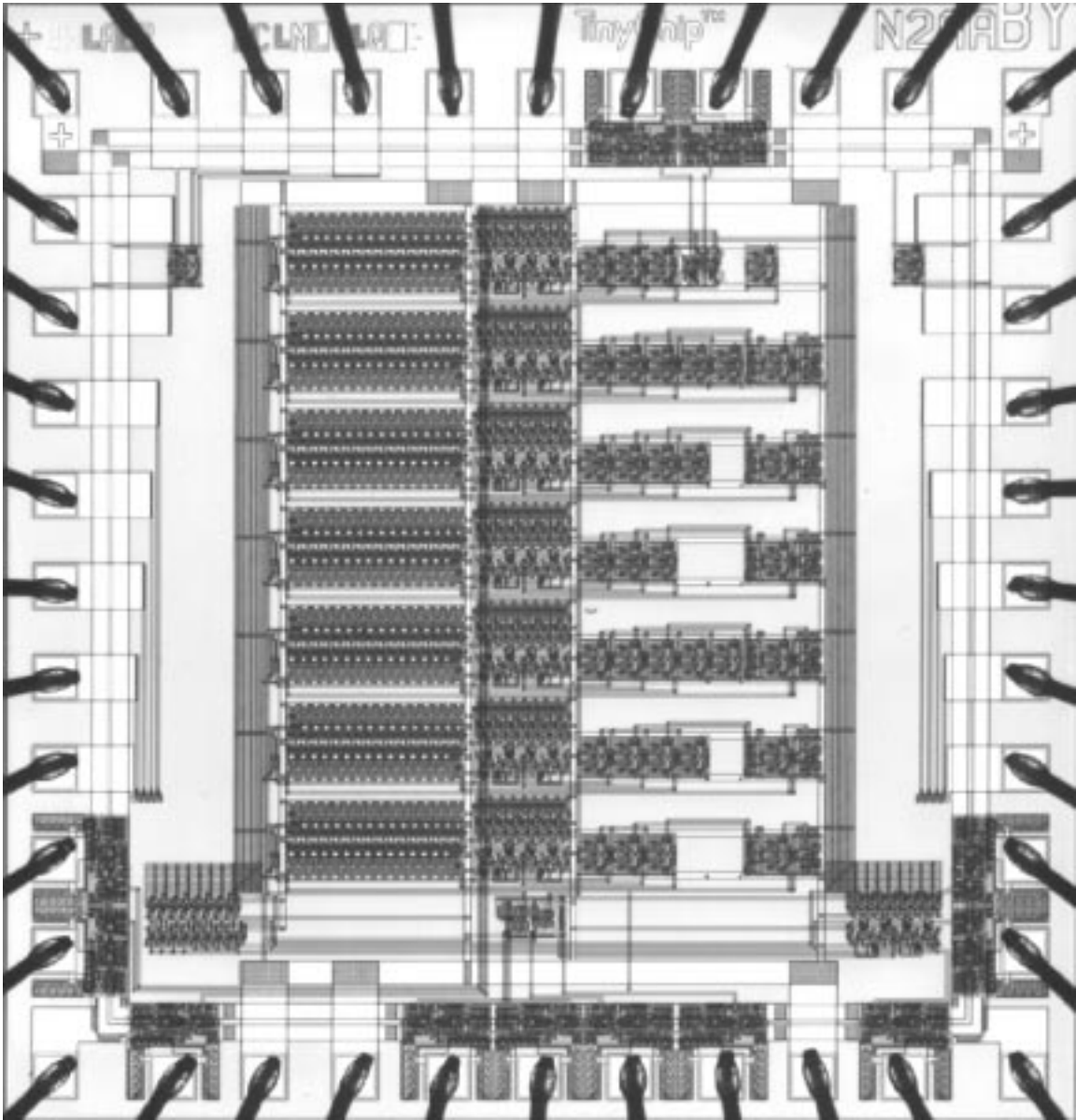
Figure 3.4: Die photograph of synchronizer test-chip.
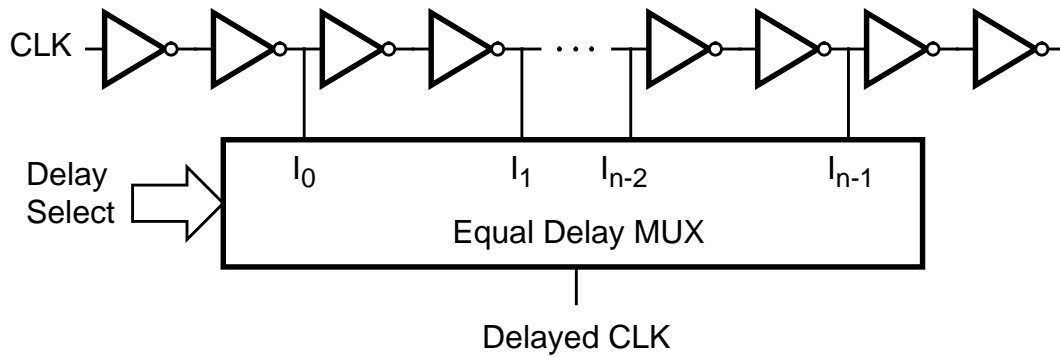
Figure 3.5: Programmable delay element.

technology scaling effects on metastable parameters. A final 2-μm test-chip was designed to quantify the effects of noise on synchronizer performance. The specifics of the noise test-chip are described in section 3.7.

3.5   Loading

3.5.1 Unbuffered Latch Simulation Results

Figure 3.6 shows simulation results for an unbuffered latch in 2-μm technology. Table 3.1 is a listing of selected parameters for *n* and *p*-channel transistors in the various technologies discussed in this paper [32], [33]. The 1.2-μm and 2-μm parameter values are measured from the fabricated test-chips. Figure 3.6(a) shows T_o derived from simulations under different loading conditions with the data input rising. A similar set of curves can be generated for the input falling. On the horizontal axis is the output swing required for valid output referenced to switching point of an inverter. The results show a linear relationship between T_o and the output voltage required for a valid 'one' or 'zero' to be sensed by the following stage. The voltage required for a valid output is a function of the transistor mismatch in addition to any logic threshold mismatch between the latch and the gates wired to the output of the latch. Logic threshold mismatch is usually dominant in standard cell and gate array libraries as the optimum device sizes for metastable performance are not the
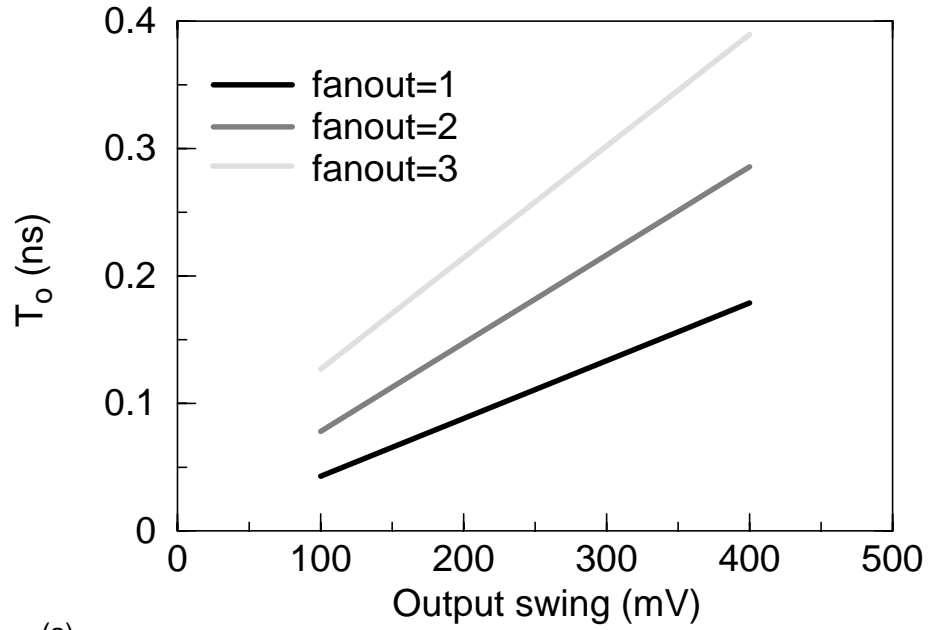
(a)
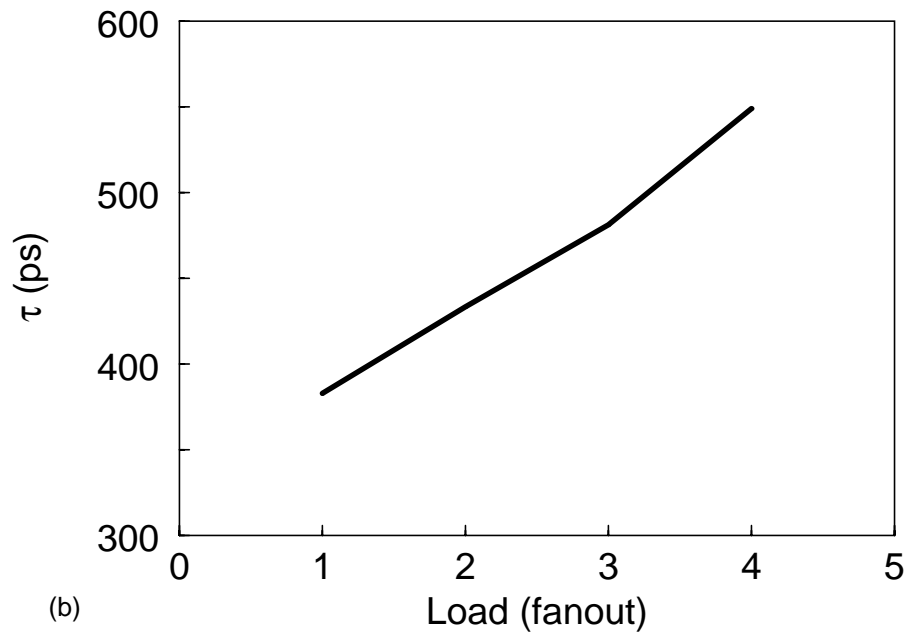
(b)

Figure 3.6: Unbuffered latch simulation results. (a) T_o versus output swing and loading; (b) τ versus loading.

Table 3.1: Selected SPICE parameters for various processes.

Parameter	nMOS 2 µm	pMOS 2 µm	nMOS 1.2 µm	pMOS 1.2 µm	nMOS 0.8 µm	pMOS 0.8 µm
V_{th} (V)	0.66	–0.79	0.79	–0.88	0.71	–0.90
μC_{ox} (µA/V^2)	50.24	24.42	102.3	30.58	110.9	34.45
t_{ox} (Å)	392	392	208	208	178	178
C_j (F/m^2)	$9.3\ 10^{-5}$	$2.4\ 10^{-4}$	$3.9\ 10^{-4}$	$4.7\ 10^{-4}$	$1.4\ 10^{-4}$	$5.9\ 10^{-4}$
C_{jsw} (F/m)	$4.6\ 10^{-10}$	$2.6\ 10^{-10}$	$1.4\ 10^{-10}$	$1.5\ 10^{-10}$	$4.1\ 10^{-10}$	$8.9\ 10^{-11}$
C_{gso} (F/m)	$5.1\ 10^{-10}$	$5.3\ 10^{-10}$	$4.5\ 10^{-10}$	$1.2\ 10^{-10}$	$4.2\ 10^{-10}$	$3.0\ 10^{-10}$

same as those for optimum delay in gates, and the logic following the latch may contain stacked and/or parallel devices. In addition, supply noise will affect the minimum voltage required for a valid output.

The value for the slope of the lines is a complicated function, but a first order approximation is the output voltage divided by a weighted sum of the signal rates at the resolving and intermediate nodes. The weights are determined by the latch structure and the capacitances at the intermediate and resolving nodes. For increasing loads, the slope of the curves increases linearly. This tracks well with the first order model because the signal edge rate at the output node decreases linearly with increased loading. For any given output voltage swing, T_o increases linearly with loading. These results agree well with previous results for nMOS synchronizers [34]. However, here we have included the effects of different loads.

Figure 3.6(b) is a plot of τ versus loading. The results show a linear relationship since the regeneration time constant of the system can be approximated by C_{out}/g_m where C_{out} is the capacitance at the output and g_m is the transconductance of the resolving element. Any additional loading is added directly into C_{out} for an unbuffered latch of the circuit style of Figure 3.1(a).

Other latches will have different behavior with respect to loading. Section 3.5.2 discusses the behavior of the buffered latch in Figure 3.1. In an unbuffered *RS* type latch, the decision to place the input transistor on the top or bottom of the stacked transistors that

implement the NAND operation is influenced by the process parameters and the exact loading condition [23], [35]. For latches with large loads, the cascode effect gained by placing the input transistor on the top of the stack is beneficial to reduce the τ dependence on C_{out}. However, unbuffered *RS* latches with relatively small loads will have smaller τ by placing the feedback transistor on the top of the stack.

3.5.2 Buffered Latch Simulation Results

Figure 3.7(a) shows the T_o extracted from simulation results for a buffered latch in 2-µm technology. The results show that T_o has an exponential relationship with loading, which means that the *MTBF* will decrease exponentially with increasing loading. If a buffer is inserted in the unbuffered latch, there is some delay associated with that buffer, t_d. If the load is varied, t_d varies linearly with loading, so T_o varies exponentially with loading. The shape of the curves matches the shape of delay versus loading on a linear plot. The input rising curve falls off for lower fanouts due to the overdrive of the *n*-channel transistor. With sufficient buffer gain, we can approximate T_o for a buffered latch as

$$T_o = T_{o\,(unbuffered)}\, e^{\,t_d/\tau} \tag{3.1}$$

where $T_{o(unbuffered)}$ is the T_o of an unbuffered latch with a fanout of one, and t_d is the time required for the voltage level to propagate from the internal resolving node to the output node.

Figure 3.7(b) shows the τ versus loading results. The buffer isolates the resolving nodes from the load, so τ is pinned at the τ with a fanout of one, and is relatively independent of loading. τ is essentially constant, but the slight decline is due to the Miller effect of the capacitance across the output buffer. As loading increases, the output signal transition becomes slower, causing the capacitance across the buffer to look smaller. Due to the open loop nature of the gain of the buffers and transistor mismatches, additional buffering does not increase the resolving capabilities of the latch. From equation (3.1) it can be seen that additional buffers can be added to the latch to optimize t_d and T_o for very heavy loads. For very light loads the addition of the buffer will increase T_o and τ needlessly, so an unbuffered latch should be used. Our simulation results presented here agree well with measured results presented in the next section.
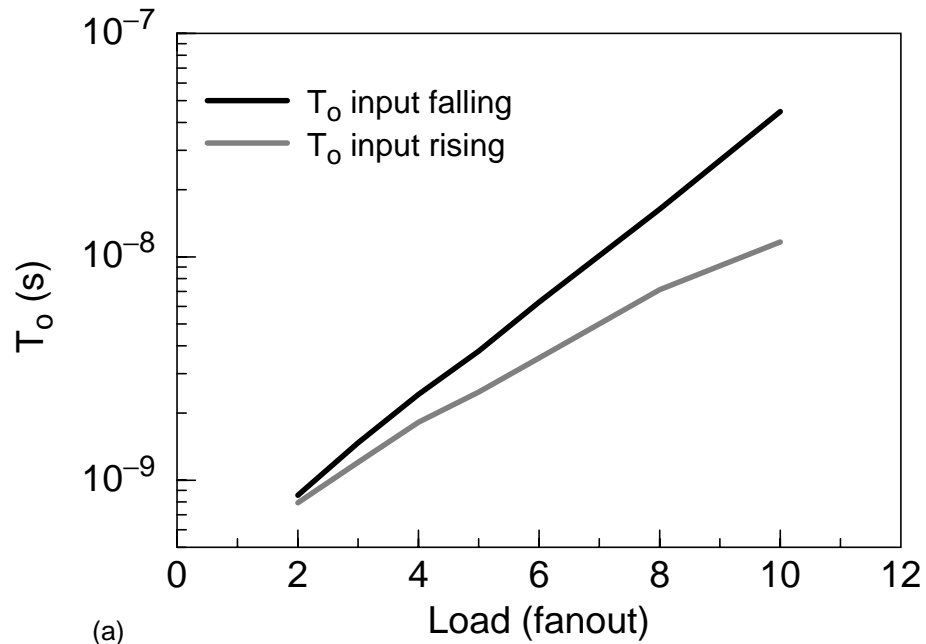
(a)
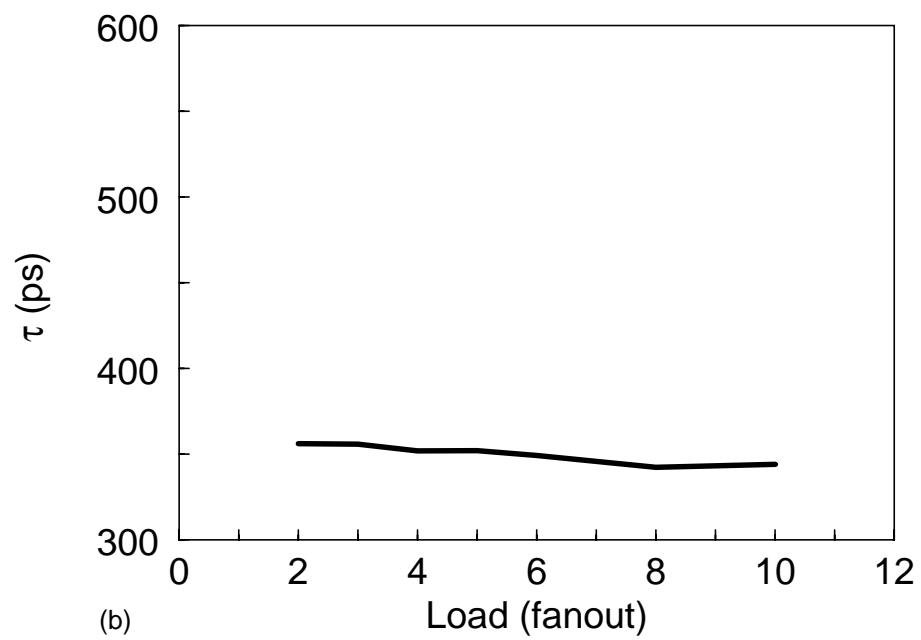
(b)

Figure 3.7: Buffered latch simulation results. (a) T_o versus loading; (b) τ versus loading.

3.5.3 Measured Results

Figure 3.8 shows measured results for unbuffered and buffered latches in 2-μm technology measured at room temperature with a 5-V power supply. The metastability window rather than the *MTBF* is plotted because *MTBF* is a function of the test's clock and data frequencies. For the unbuffered latch, both the intercept, T_o, and the slope of the curves, τ, increase with increased loading, showing the degrading performance as the latch is loaded.

For the buffered latch, τ is unchanged versus loading, but T_o changes. Although T_o is larger for the buffered latches, buffered latches are superior due to their greater τ term after some t_r. As loading decreases, the crossover t_r moves toward infinity. The theoretical limit is at a fanout of one, where the unbuffered latch becomes superior for all resolution times. From our measured results, τ for the buffered latch is slightly higher than that extrapolated from the unbuffered results due to the proximity of the output buffer,
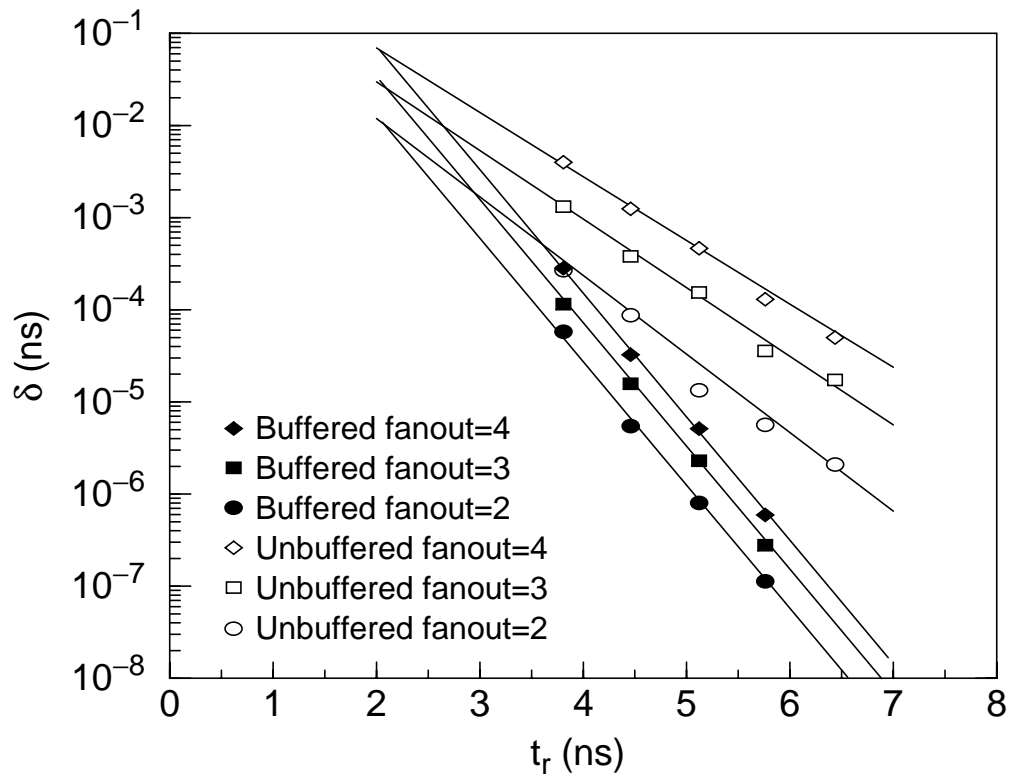
Figure 3.8: Measured metastability window for different loading conditions.

eliminating much of the wiring capacitance. In addition, the measured T_O parameter is influenced by the setup time of the compare-latches, t_{su}, so an additional term, $e^{t_{su}/\tau}$, is entered into the measured T_O values. While the performance of buffered latches is superior for all fanouts greater than one with adequate resolution time, the performance still degrades exponentially with loading. By increasing the loading from a fanout of two to a fanout of four, the metastability window is increased by almost an order of magnitude. This corresponds to almost an order of magnitude drop in the *MTBF* for the latch.

3.6   Supply and Technology Scaling

While metastability concerns are not the driving force for supply and technology scaling, it is useful to know how a latch's parameters vary with supply and technology scaling to predict behavior of library cells in different applications. Cell libraries are often used for a variety of chips, with various technologies and supplies. Low power design and reliability problems have driven the push toward lower supply voltages, and technology scaling has been a matter of course in the semiconductor industry since its inception. For a system with a fixed clock frequency and resolution time, *MTBF* performance can be improved by operating at higher supply and smaller technologies [22], [23], [25]. However, reduced supply usually implies reduced clock frequencies, and smaller technologies, higher clock frequencies. It has been noted that changes in on-chip delays help to compensate for changes in τ [22], [25]. Here we develop a normalization scheme to predict metastable performance over supply scaling and establish a method of directly comparing latches in different technologies.

3.6.1 Metastability Parameter Estimation

As shown in section 2.3.3 first order equations for synchronizers can be derived and τ can be shown for square law devices to be [21]

$$\tau = \frac{K_\tau L^2 \left(W_n + W_p\right)}{\left(V_{DD} - 2V_{th}\right)\sqrt{\mu_n W_n}\sqrt{\mu_p W_p}} \ . \tag{3.2}$$

K_τ is the ratio of output capacitance to input capacitance, C_{ox} is the gate capacitance per unit area, L is the device length, and W_n and W_p are the channel widths for the n and p-channel devices, respectively. μ_n and μ_p are the mobility of holes and electrons, respectively, and the magnitude of the threshold voltage, V_{th}, is equal for both p-channel and n-channel devices. It can be seen from equation (3.2) that decreasing the power supply will degrade τ, and as V_{DD} approaches $2V_{th}$, τ will increase markedly.

3.6.2 Parameter Normalization with Supply and Technology Scaling

To evaluate a library cell in different environments for design trade-offs, it is necessary to make some assumptions about the application's system clocking. To account for changes in the clock frequency with supply and technology, we assume the receiving system's clock is a constant multiple of an on-chip reference delay [36]. Thus,

$$T_{CLK} = N_{g1}t_{ref} \tag{3.3}$$

where T_{CLK} is the period of the clock, N_{g1} is the number of standard delays in the clock cycle, and t_{ref} is the standard delay, defined as a fixed on-chip gate delay. The resolution time, t_r, is usually generated on-chip with some fraction or multiple of the clock period, so an equation similar to equation (3.3) can be established. Substituting (3.3) into (2.27) and replacing t_r with equivalent reference delays, the following equation for the system's *MTBF* is obtained,

$$MTBF = \frac{N_{g1}}{f_D \dfrac{T_o}{t_{ref}} e^{-N_{g2}\frac{t_{ref}}{\tau}}} \tag{3.4}$$

where N_{g2} is the number of standard delays in the resolution time.

Comparing equations (3.4) and (2.27), it can be seen that in addition to τ and T_o, the parameters τ/t_{ref} and T_o/t_{ref} should be used to describe the latch's metastable behavior for systems where the clock period and resolution time are fixed multiples of an on-chip reference delay. Making assumptions similar to those for equation (3.2), the reference delay can be approximated by

$$t_{ref} = \frac{C_L V_{DD}}{I} = \frac{K_{ref}(W_n + W_p) L^2 V_{DD} (\mu_n W_n + \mu_p W_p)}{\mu_n \mu_p W_n W_p (V_{DD} - V_{th})^2} \tag{3.5}$$

where K_{ref} is the ratio of output capacitance to input capacitance. Combining (3.2) with (3.5), the equation

$$\frac{\tau}{t_{ref}} = \frac{K_\tau (V_{DD} - V_{th})^2 \sqrt{\mu_n \mu_p W_n W_p}}{K_{ref} V_{DD} (V_{DD} - 2V_{th}) (W_n \mu_n + W_p \mu_p)} \tag{3.6}$$

can be generated to show the relationship of τ with respect to t_{ref}.

It can be seen from (3.6) that if V_{DD} is much greater than V_{th}, τ/t_{ref} is a constant. At lower supply voltages, both t_{ref} and τ increase dramatically, but τ increases faster because the latch's metastable voltage biases the devices near $V_{DD}/2$, whereas delay elements use the full supply swing. This causes the τ/t_{ref} ratio to increase, degrading performance of the cell at lower supply voltages. In low voltage design, a low threshold device is often desirable to reduce the allowable supply voltage from a delay standpoint [37]. Reducing the threshold voltages will improve τ as well [23], and from equation (3.6) it can be seen that this will reduce the minimum supply for constant τ/t_{ref} operation. However, it will have little effect on τ/t_{ref} for V_{DD} much greater than V_{th}. For changes in technology, τ/t_{ref} is largely dependent upon changes in μ_n, μ_p, and the differences in K_τ and K_{ref}. The constants K_τ and K_{ref} are heavily dependent upon technology because the relative values of diffusion capacitance, gate capacitance, and wiring capacitance do not scale equally with technology.

It is difficult to make quantitative expressions for T_o that hold over process and supply changes because changes in T_o can be dominated by second order effects in the exponent of (3.1). In addition, measurement introduces a second exponential term due to the setup time of the measurement latch. $T_{o(unbuffered)}$ is proportional to the output voltage and signal edge rates as stated earlier, and inversely proportional to the switching current. For an unbuffered latch T_o is expected to increase with reduced supply because the available current is reduced. With reduced technology, T_o will decrease because to the first order the current is the same, but C_{out} is reduced. For buffered latches, however, second order changes in the exponential term in equation (3.1) due to the buffer insertion can dominate

these effects. The quantity T_o/t_{ref} to the first order is a constant over supply for both buffered and unbuffered latches.

3.6.3 Supply Scaling Measured Results

Figure 3.9 shows plots of the 2-μm buffered latch with a fanout of four test results versus power supply from 2 V to 5 V. Figure 3.9(a) is the δ versus t_r over supply on an absolute scale. The metastable performance is dramatically worse at lower supply voltages with τ changing from 0.275 ns at 5 V to 1.048 ns at 2 V. The measured worst case $τ/t_{ref}$ variation is approximately 21% between 2 V and 5 V, but only 3% between 2.5 V and 5 V. Figure 3.9(b) shows normalized values with t_r plotted in reference delays and the measured δ scaled by a reference delay. The reference delay here is a fanout of one inverter followed by a fanout of two inverter in the variable delay line.

Table 3.2 shows a summary of the latch performance for normalized and absolute parameters. With respect to the $τ/t_{ref}$ parameter, the circuit performance is relatively independent of supply down to about 2.5 V. The T_o/t_{ref} variation over the supply voltages measured is less than an order of magnitude. Similar results are obtained for the 1.2-μm fanout of four latch, but the performance drops off at a higher supply voltages due to the higher threshold voltages in the 1.2-μm process. Thus, latches can be used in the same circuit with no penalty over a wide range of supply voltages, but in extremely low supply environments performance drops off rapidly. Simulations show that for a 1.5-V supply,

Table 3.2: Metastable parameters versus power supply.

V_{DD} (V)	t_{ref} (ns)	τ (ns)	$τ/t_{ref}$	T_o (s)	T_o/t_{ref}
2.0	1.790	1.048	0.586	1.819×10^{-6}	1.016×10^3
2.5	1.274	0.608	0.477	4.125×10^{-6}	3.238×10^3
3.0	1.010	0.470	0.465	2.258×10^{-6}	2.236×10^3
4.0	0.741	0.344	0.464	1.310×10^{-6}	1.768×10^3
5.0	0.594	0.275	0.463	4.451×10^{-6}	7.493×10^3
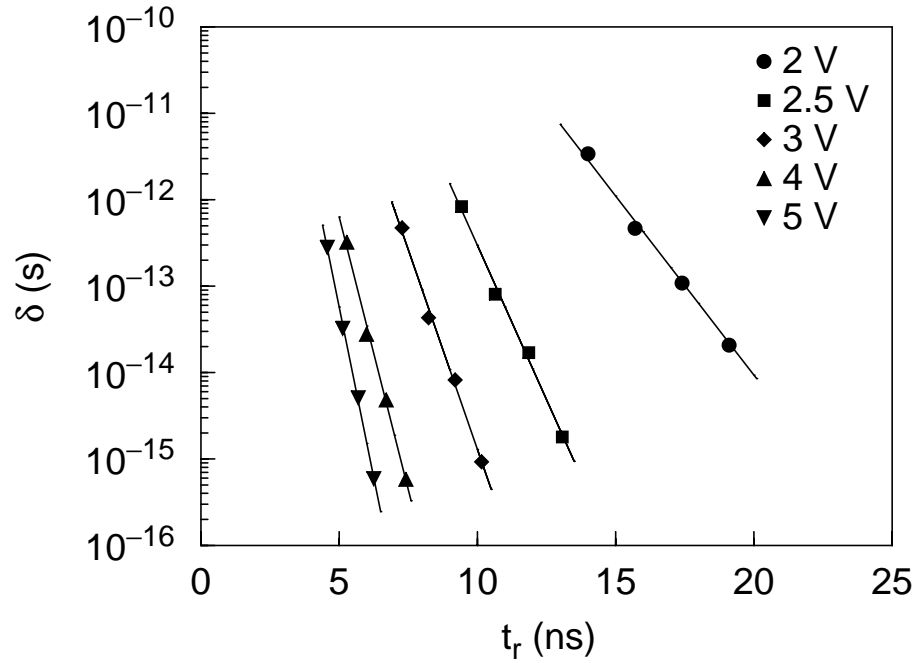
(a)
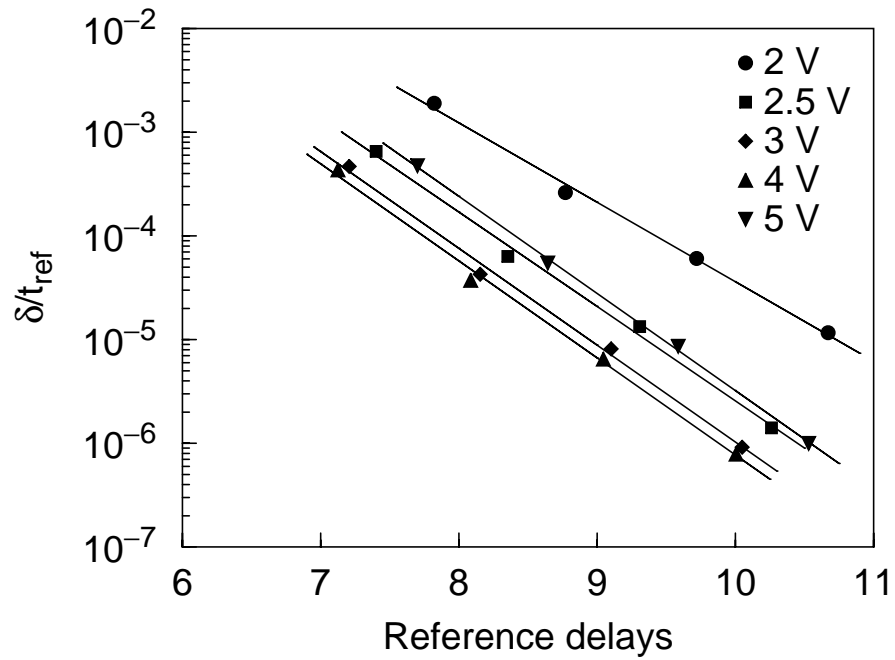
(b)

Figure 3.9: Measured results versus supply. (a) Unnormalized results; (b) normalized results.

τ/t_{ref} is greater than 3 in 1.2-µm technology. This represents a severe degradation in performance if supply voltages are scaled down and the number of gate delays per clock cycle is held constant. We were not able to measure the parameters at 1.5 V. With a 1.5-V supply, τ has degraded so badly much longer delays than those we can generate on-chip are needed to obtain exponential behavior of δ versus t_r.

3.6.4   Technology Scaling Measured Results

Figure 3.10 shows the performance of buffered latches with a fanout of two and four with respect to technology scaling at a 5-V power supply. The absolute plot shows a significant improvement in the value of the τ parameter with smaller gate lengths. The normalized plot, however, shows the scaling of t_{ref} and τ is not equal with respect to technology. Table 3.3 shows the extracted parameters for the two measured latches. Our measured results show an improvement in the relative parameters as the technology is shrunk from 2-µm to 1.2-µm, but performance worsens as the technology is shrunk to 0.8-µm. Further reductions in technology will undoubtedly yield improvements of metastable parameters in the absolute case, but the relative performance must be evaluated case by case.

Table 3.3: Metastable parameters versus technology.

tech. & latch	t_{ref} (ps)	τ (ps)	τ/t_{ref}	T_o (s)	T_o/t_{ref}
2µ fanout=2	594	272	0.458	$1.034\ 10^{-6}$	$1.743\ 10^3$
2µ fanout=4	594	275	0.463	$4.451\ 10^{-6}$	$7.493\ 10^3$
1.2µ fanout=2	464	172	0.371	$2.682\ 10^{-8}$	$5.779\ 10^1$
1.2µ fanout=4	464	155	0.334	$1.618\ 10^{-6}$	$3.488\ 10^3$
0.8µ fanout=2[*]	291	127	0.437	$1.221\ 10^{-7}$	$4.193\ 10^3$
0.8µ fanout=4[*]	291	124	0.424	$6.031\ 10^{-7}$	$2.070\ 10^3$

[*] simulated results
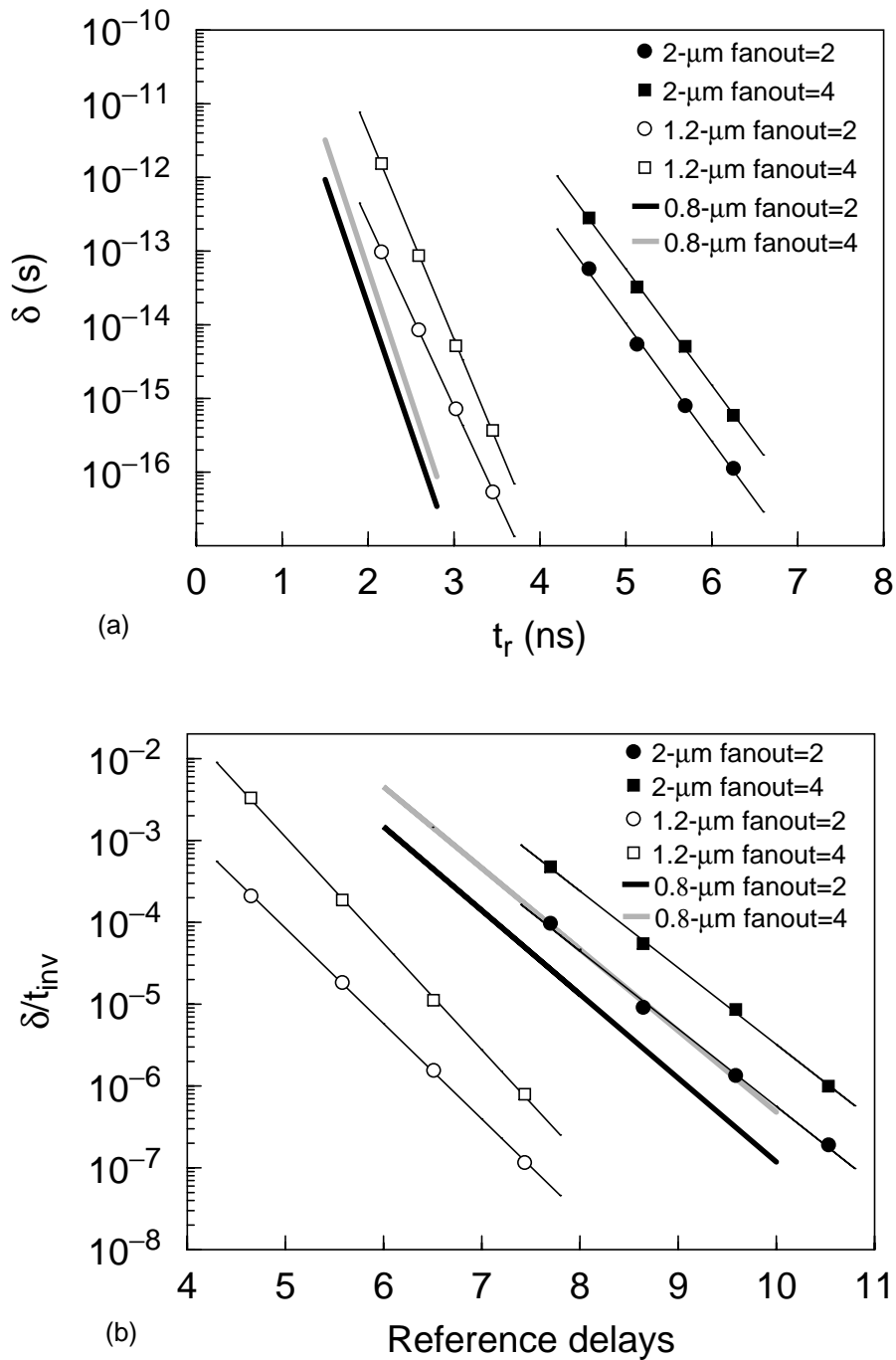
Figure 3.10: Measured results versus technology. (a) Unnormalized results; (b) normalized results.

3.7 Supply Noise

As seen earlier in the chapter, characterization of the metastability parameters, particularly T_O, is difficult and time consuming. Simplifications must be made to reduce the amount of measurements or constrain the simulation environment. One simplification is the supply environment. Previous work stated that circuit noise in CMOS synchronizers and arbiters does not effect the metastability error rate of the circuit [3], [20], [34]. Supply noise in an integrated circuit, however, varies depending upon the power distribution networks, chip sizes, and circuit elements, and an increase *MTBFs* for the same chip was observed as the board technology and design improved. Additionally, the question of whether output buffering would improve a synchronizer's noise immunity could not be ascertained from published work.

This section discusses the test methodology and presents measured results from a 2-μm CMOS test-chip to demonstrate that supply noise does in fact alter the error rate of synchronizers by degrading the parameter T_O. The degradation shows a linear relationship to the root mean square (RMS) value of noise on the supply. The simulation of this behavior is discussed. In addition, our measured results show that buffering the synchronizer does not appear to improve the synchronizer's noise immunity.

3.7.1 Measurement Setup

A test-chip was designed in 2-μm, *n*-well CMOS process to measure the effects of supply noise on τ and T_O of latches used as synchronizers. The measurement circuit is the same as that described in section 3.4 and the latches measured are those of Figure 3.1. However, in the implemented chip for noise measurement, the supplies of different parts of the circuit are pinned out separately to allow injection of noise into selected parts of the circuit as shown in Figure 3.11. Since the *p*-channel devices for each portion of the circuit have separate wells, the effects of varying the supply are contained to devices connected to the noisy supply only.

A two layer printed circuit board was designed with a clean power supply and a noisy power supply. Shorting bars allow wiring of the synchronizer or the compare-latch to either the noisy or clean supply. The noisy supply is created by capacitively coupling ac
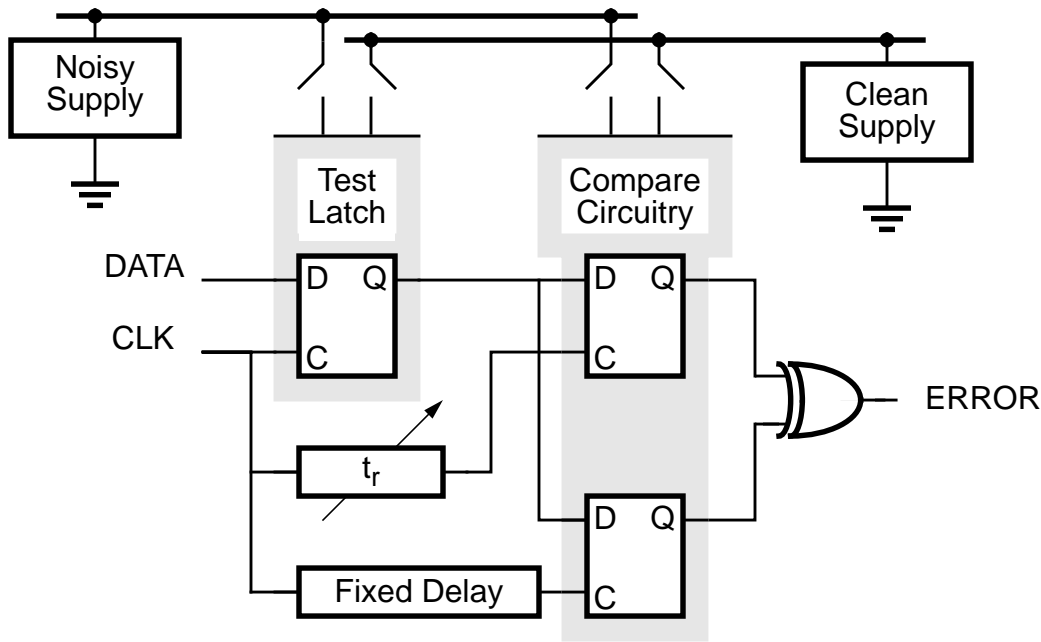
Figure 3.11: Synchronizer performance versus supply noise test circuit.

signals into the supply from an external signal source. A digital oscilloscope is used to continually measure the root mean squared (RMS) value of the noise on the supply at the pin of the package. By increasing the amplitude of the external function generator, the RMS value of the supply noise can be freely controlled. Appendix A discussed the specifics of the board implementation and instrument configuration.

3.7.2 Measured Results

Figure 3.12 shows a plot of δ versus the resolving time, t_r, for a double buffered synchronizer with 250-mV and 0-mV RMS noise in the supply for both the synchronizer and compare-latches of Figure 3.11. The supply voltages are set to 5 V. To approximate white noise, a 50-MHz signal is injected into the supply with a clock frequency of 6.25 MHz and a data frequency of 5.99 MHz. The measurements were performed with a sinusoidal signal coupled into the supply, but measurements with square waves were well
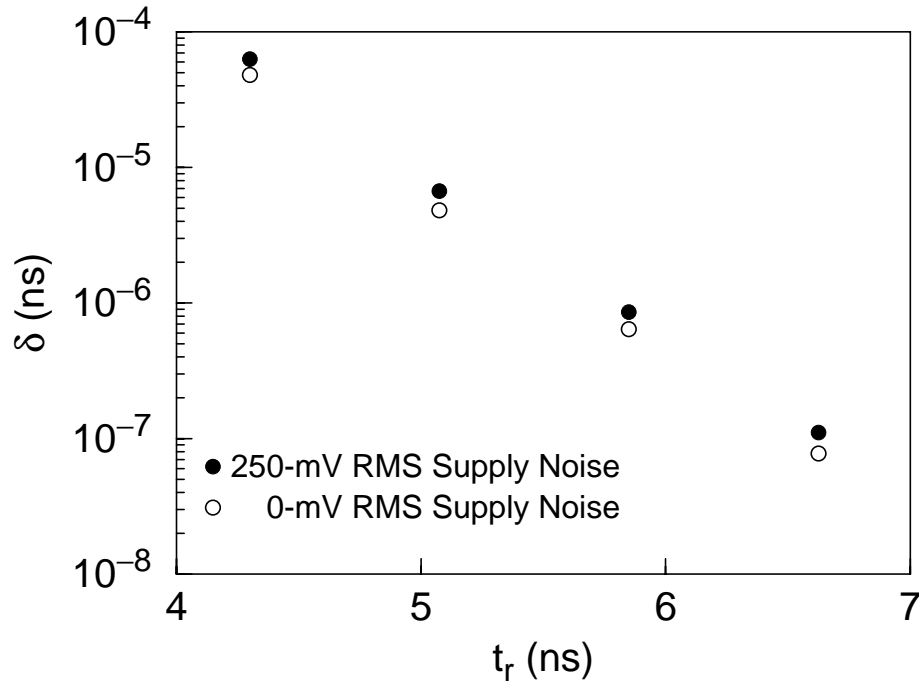
Figure 3.12: δ versus synchronizer resolving time. Noise frequency is 50 MHz. Noise RMS value is 250 mV.

within measurement error limits. The slope of the lines is the same, so the parameter τ is unchanged with respect to supply noise. The intercept, T_o, however is increased due to the injected noise, degrading the synchronizer's performance. Here the metastability window is increased in size by about 30% for each point along the line. Because τ is not affected by supply noise, the measurement of δ can be performed at only one delay setting, the shortest, to speed up data acquisition times. Figure 3.13 shows a plot of the δ/δ_{avg} for a double buffered synchronizer with respect to the frequency of the noise. The RMS supply noise is 250 mV. The variation in the metastability window is less than $+/-6.5\%$ showing the window is relatively independent of the noise frequency. The remaining tests were performed with a noise frequency of 50 MHz.

Figure 3.14(a) shows the measured results for double and single buffered synchronizers of δ versus the RMS value of supply noise with noise injected into the compare-latches' supply. Each point represents the average of 1500 synchronization
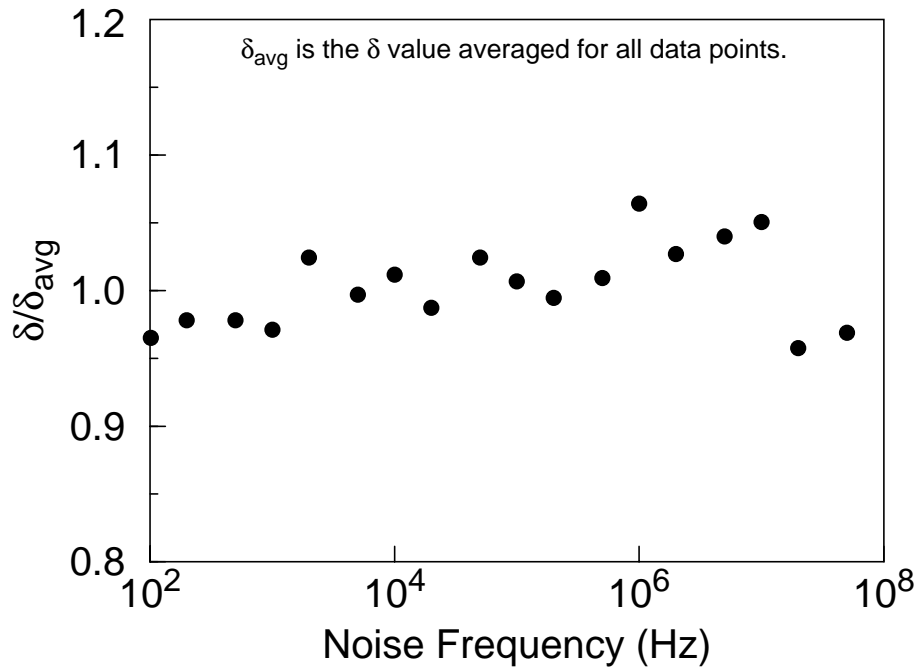
Figure 3.13: δ versus frequency for 250-mV RMS sinusoidal noise signal.

failure times. δ/δ_{avg} rather than δ is plotted to allow comparison between the two synchronizers. The measured error is less than +/−5%. The plot of Figure 3.14(a) shows that noise in the compare-latches has little effect on the metastability window because supply noise in the compare-latches is shielded from the synchronizer's internal nodes by the output buffers. Over the noise range measured, a single level of buffering is adequate to remove the effects of noise at the synchronizer output.

Figure 3.14(b) shows a plot of δ/δ_{avg} versus RMS noise on the synchronizer supply. In this case, RMS noise increases the metastability window linearly. Additionally, both the single and double buffered synchronizers have essentially the same behavior, so additional output buffering does not improve the synchronizer's noise immunity. In the absolute measurements, single buffered synchronizers have metastability windows about two times smaller than double buffered synchronizers, so buffering in an attempt to reduce noise sensitivity would increase the error rate by a factor of two with no improvement. The

(a)

(b)
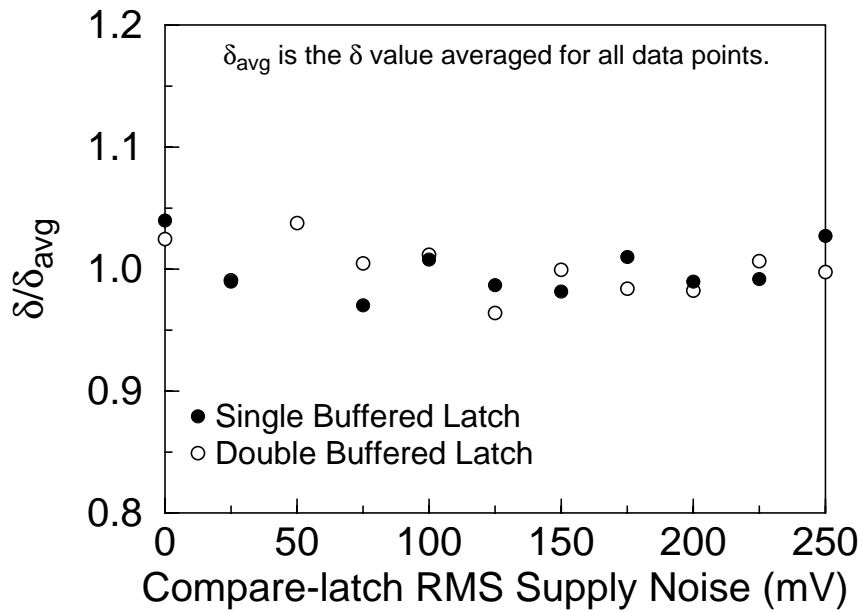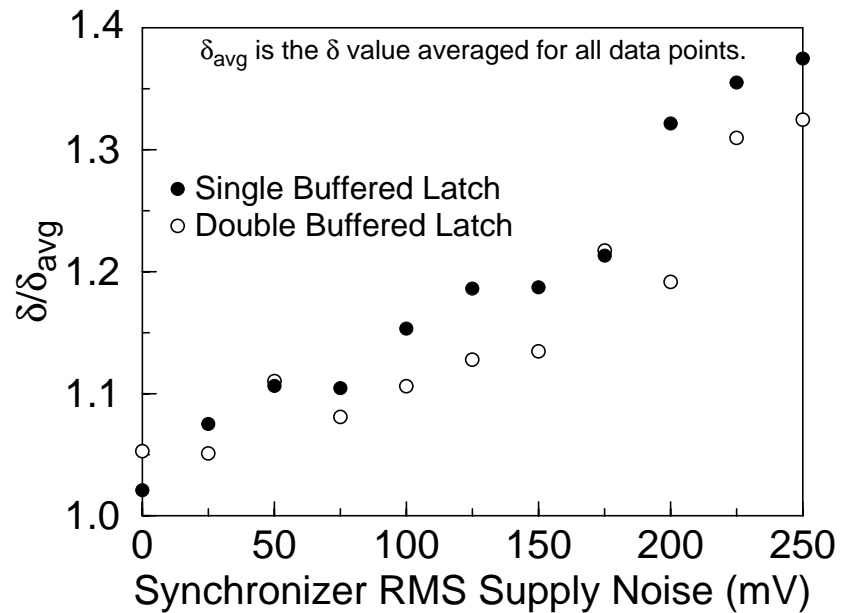
Figure 3.14: δ versus RMS noise amplitude. (a) Noise on compare-latches' supply; (b) noise on synchronizer's supply.

behavior exhibited in Figure 3.14(b) can be modeled by simulating τ with the average supply voltage, and T_O with average supply voltage less an offset voltage to model the noise. A conservative choice for this offset voltage would be the peak noise amplitude.

3.8 Summary

This chapter discussed the characterization of CMOS cell library elements with respect to metastability. The behavior of buffered and unbuffered latches versus loading from a metastable performance viewpoint has been discussed. For an unbuffered latch, both τ and T_O vary with loading. For a buffered latch only T_O varies with loading, but the variation is exponential. A formula to determine T_O for a buffered latch from an unbuffered one has been described. The optimum synchronizer in any given technology is unbuffered with very small load, but for any appreciable load the buffered latch is superior. Additional buffering past the optimal delay was shown to degrade the parameter T_O needlessly.

A normalization scheme to compare latches with different supply voltages and technologies has been presented. The degradation of metastable parameters for reduced supply are largely cancelled when the effects of longer on-chip delays is factored in. However, the performance for the normalized circuit parameters degrades sharply as the supply approaches $2V_{th}$. With reduced channel lengths, the parameter's behavior is much more complex, and performance can improve or degrade when the scaling of on-chip delays is factored in.

Measured results from a 2-μm CMOS test-chip to characterize the effects of supply noise on synchronizer error rates were presented. The *MTBF* of synchronizers is reduced with power supply noise by increasing the parameter T_O. This *MTBF* reduction can be measured by injecting noise into the supply of a test-chip, or modeled using an offset voltage on the supply in the T_O simulations. Increasing the number of buffers at the synchronizer output does not improve the synchronizer's noise immunity, and decreases the *MTBF* needlessly.

Chapter 4

Metastability Errors in A/D Converters

4.1 Introduction

As seen in Chapter 1, high-speed analog-to-digital converters are used in applications where metastability error rates must be kept low. For digitizing oscilloscopes, automated test equipment, and telecommunications receivers, A/D converters require resolutions of 6 to 8 bits with sampling speeds as high as possible. High-speed A/D conversion has been dominated in past years by research in bipolar designs due to bipolar junction transistor's higher performance. However because of increased levels of integration, lower cost, and improvements in CMOS performance, the need for high-speed CMOS A/D converters has increased. In bipolar designs, high-speed A/D conversion with 6 to 8 bits of resolution has been dominated by flash, and flash-based architectures [38]-[44] and folding techniques have become increasingly popular [45]-[48]. However, resolution in full flash implementations in CMOS is limited by the matching characteristics of the MOS devices [27] to approximately seven bits [9], [49]. The feasibility of CMOS folding architectures has just begun to be demonstrated [50], [51]. Two-step, pipelined architectures are popular choices for resolutions of 8 bits or more. However, the speed is ultimately somewhat less than that achievable by full flash or folding implementations [52]-[55].

This chapter reviews the basic flash and flash-based architectures. In addition, folding converters and interleaved converters are discussed. The origins of metastability errors and the equation to estimate the error rate for a flash A/D converter is described. A discussion of the test method to measure metastability error rates in A/D converters is

included. Previous methods for error probability reduction and their limitations are also discussed.

4.2   High-speed A/D Conversion

Flash converters or fully-parallel converters are conceptually the simplest architectures and the most commonly used high-speed converters in the 6 to 8 bit range. In addition, other high-speed architectures are either based on or make use of flash stages, so an understanding of flash converters is vital to understanding other converters. However, due to the high input capacitance, large element count and area requirements, and power dissipation of full flash converters, it is often desirable to use other high-speed A/D architectures to meet performance requirements. In this section, flash, interpolating, two-step, and folding A/D converter architectures are discussed. A brief discussion of parallel converters is also included.

4.2.1 Flash Converters

In an n-bit flash converter, the analog input, V_A, is compared to 2^n-1 reference voltages, ideally separated by one least significant bit (LSB) in voltage, as shown in Figure 4.1(a). Comparators with V_A above their reference voltage generate a low output, and comparators with V_A below their reference voltages generate a high output. The outputs of the comparators form what is known as a thermometer code, where the high comparator outputs track the analog input. In read only memory (ROM) based encoding schemes, the thermometer code is converted to a 1-of-n code which can be used to turn on a word line in a ROM code book as shown in Figure 4.1(b). Thermometer code to 1-of-n code conversion is usually accomplished with two or three input logic gates comparing a comparator output to its nearest neighbors. The words in the ROM are generally binary or Gray codes with the code corresponding to the location of the analog input.

Advantages of flash architectures include modular design and high-speed operation. Additionally, the flash architecture does not require a sample-and-hold circuit because the conversion is performed at only one point in time, when the comparators are strobed [40]. Disadvantages, however, are a large number of devices, high input capacitance, stringent

V_{REF} V_A

$2^n - 1$ Comparators

$V_{REF(N-1)}$

$V_{REF(N-2)}$

$V_{REF(2)}$

$V_{REF(1)}$

Thermometer Code Detector
&
Encode ROM

n-bit output

(a)

Comparator
outputs

Thermometer
code detector
output

ROM
codes

0
0
0
1
1
1
1

0
0
0
1
0
0
0
0

1 1 1
1 1 0
1 0 1
1 0 0
0 1 1
0 1 0
0 0 1
0 0 0

Converter
output

Thermometer
code

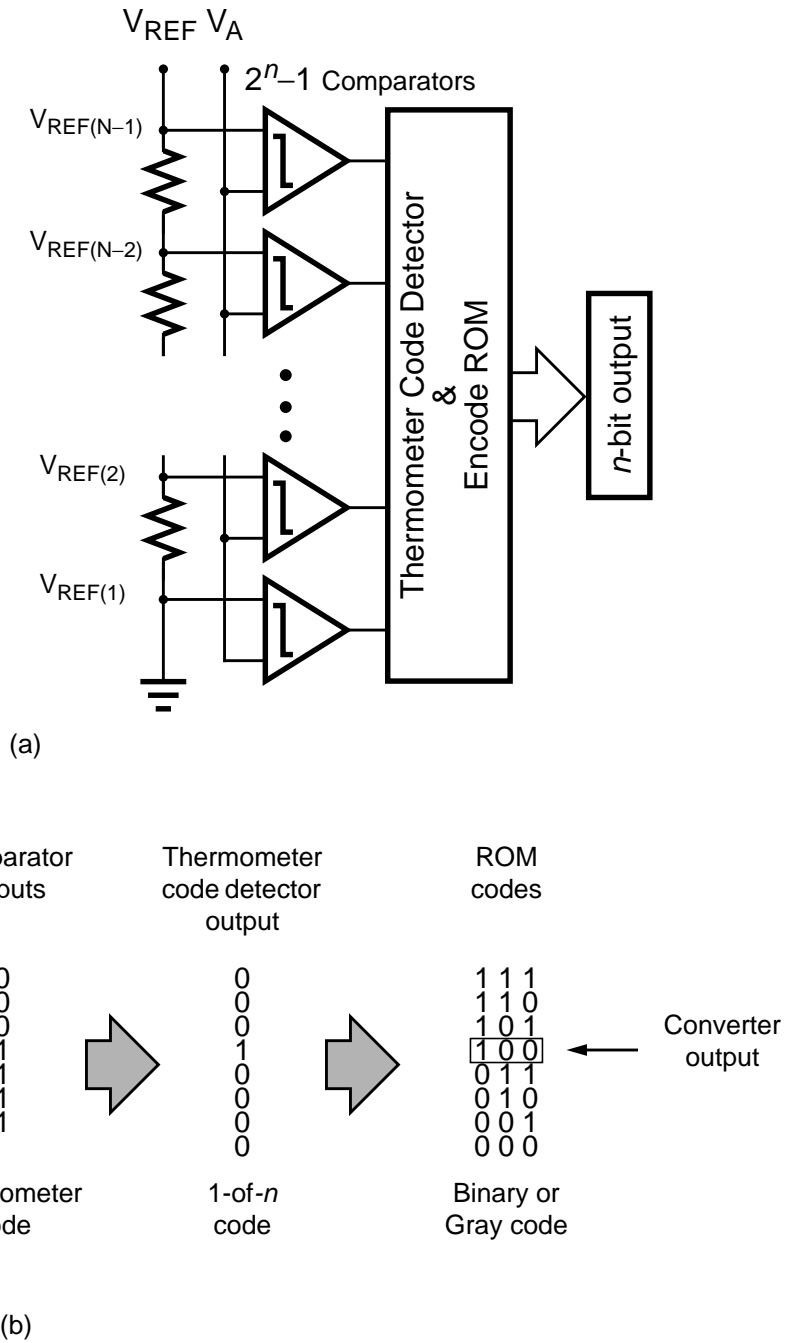1-of-n
code

Binary or
Gray code

(b)

Figure 4.1: Flash A/D converter with ROM-based encoding. (a) Schematic of
converter; (b) operation with a 3-bit example.

timing requirements, and variable comparator delay. The stringent time requirements and comparator delay requirements can be relaxed by using a front end sample-and-hold amplifier. However, the front end sample-and-hold amplifier then becomes the design bottleneck. Additionally, the resolution of CMOS implementations is limited due to the poor matching of MOS devices.

4.2.2 Interpolation

To reduce input capacitance, area, and power with respect to full flash architectures, interpolating architectures can be used [43], [44]. Interpolating architectures consist of $2^n/k$ input stages where k is a multiple of 2. The $k-1$ intermediate reference levels between the input stages are linearly interpolated. Figure 4.2(a) shows a simple schematic for interpolation of a single intermediate reference voltage. Preamplifiers are used to amplify the difference between the analog input voltage and the two reference voltages, $V_{REF(j)}$ and $V_{REF(j-1)}$. The transfer functions for the two preamplifier's outputs versus the analog input voltage are shown in Figure 4.2(b). By inserting an additional differential latch between the preamplifiers, an intermediate reference voltage can be interpolated. In the figure shown, the outputs of the preamplifiers, V_{j-1} and $\overline{V}_j$, are used to drive the intermediate latch and generate the additional bit of resolution, and the reference level

$$V_{REF(interpolated)} = \frac{V_{REF(j)} + V_{REF(j+1)}}{2} \tag{4.4}$$

is interpolated.

For proper operation, the differential latch must reject the common mode signal and generate logic output levels based on the differential voltage placed at the inputs. Additionally, the transfer characteristic of the preamplifier must by symmetric to place the interpolated reference level in the center of the two reference levels. In practice, more than one additional reference level can be interpolated, but this increases the requirements of the preamplifier and the number of additional circuitry needed. The thermometer code detection and digital encoding proceed exactly as that of a full flash converter.

In practice, interpolating converters are slower than full flash converters due to the increased output capacitance on the preamplifiers and the additional linearity requirements imposed on the preamplifiers. However, the power and area requirements for the converter
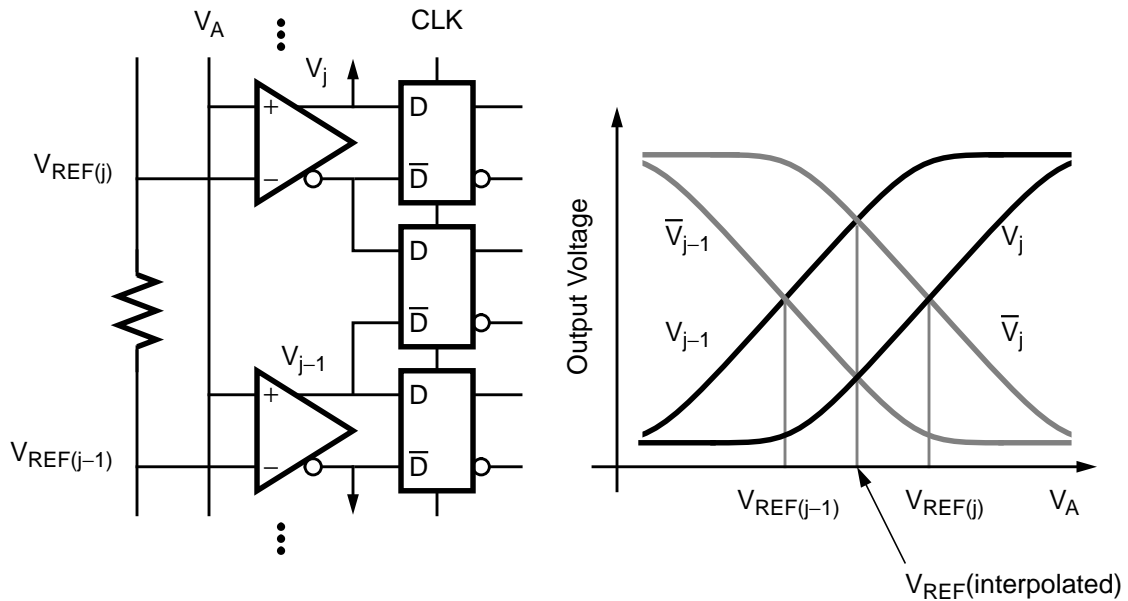
Figure 4.2: Interpolating A/D converter operation. (a) Schematic of section; (b) interpolation of intermediate level.

are reduced since the number of preamplifiers is reduced, and the effects of comparator offset are reduced. The preamplifiers are usually quite large in area to reduce the input-referred offset voltages, so the power and area savings can be substantial. Interpolation is possible with both voltage-based schemes, like the one shown in Figure 4.2, and current-based schemes. In addition, interpolation is commonly used in other converter architectures such as folding converters.

4.2.3 Two-step Pipelined Converters

Another alternative to full flash implementations to reduce power and area is the pipelined, two-step converter [52]-[55]. High-speed two-step converters are comprised of two flash stages which encode the analog signal in successive clock cycles. Figure 4.3
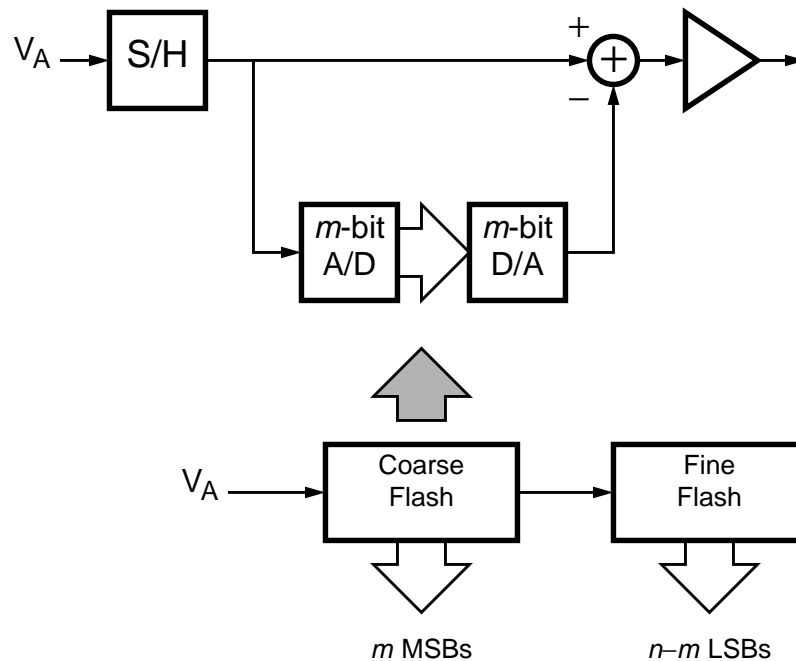
Figure 4.3: Two-step pipelined A/D converter.

shows a block diagram for a two-step pipelined A/D converter. A sample-and-hold circuit is used to sample the analog input, and an m-bit encoding of the analog input is subtracted off the analog input leaving a "residue". The residue is passed to the second stage for encoding of the $n-m$ lower order bits.

In principle, more than two stages of pipelining are possible, but in low resolution converters, two-step architectures dominate. Additionally, two-step converters do not require pipelining for proper operation. However, to operate at high speeds, pipelining is necessary. With a two-step architecture, the number of comparators can be reduced substantially. For an 8-bit converter, instead of the 255 comparators required for a full flash implementation, the converter can be implemented with a minimum of 30 comparators if two 4-bit flash stages are used. In practice, two-step converters often include overlap to allow for any offsets between the two stages. This doubles the number of converters in the second stage, increasing the area and power. In addition to the flash

stages, the two-step, pipelined architecture requires a subtracter, an *m*-bit DAC, and a sample-and-hold circuit. Typically, the speed bottleneck in such designs is the path through the coarse stage, the DAC, and the subtracter.

4.2.4 Folding Converters

The final architecture in high-speed, low resolution converters is the folding architecture. In folding converters, the residue voltage is generated in parallel to the coarse conversion by analog preprocessing. By generating the residue in parallel, the need for subtracter, D/A, and sample-and-hold circuit is eliminated. As these are the speed bottlenecks in the two-step architecture, folding converters can obtain very high speed operation. Folding A/D converters have been used for a number of years in bipolar implementations. CMOS implementations, however, have only recently been introduced.

Figure 4.4(a) shows the block diagram for a folding A/D converter. The analog input is split into two paths. In one path the circuit performs a coarse A/D conversion to obtain the higher order output bits. In the other path, the analog input voltage is processed with a continuous time folding circuit. The folding circuit converts a full-scale signal to a signal which represents the input signal less some voltage, depending upon which region the input voltage is in. Figure 4.4(b) shows the output for the folding circuit for a ramp input. The dotted line represent the analog input, and the solid line represents the folded output. The folded output can now be encoded with $2^n/k$ comparators, where *k* is the number of times the signal is folded. Due to the different delays in parallel paths of the input, a bit synchronization is required to synchronize the lower order bits to the higher order bits.

In principle, folding can reduce the number of circuit elements required in a similar manner to the two-step architecture over flash converters. However, due to the rounding of signal peaks in the residue, overlap is required to allow conversion on the central portion of the residue voltage. Because of this overlap, a large number of folding circuits is required. For this reason folding schemes must be used in conjunction with interpolation to reduce the number of circuit elements. A drawback to the folding architecture however, is the increase in the frequency of the analog input voltage. The analog input frequency is increased by the number of times the input is folded, ultimately limiting the analog input bandwidth possible for folding converters.
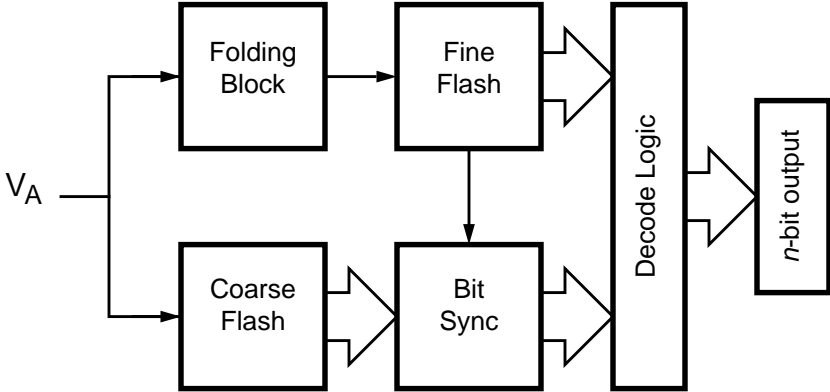
(a)
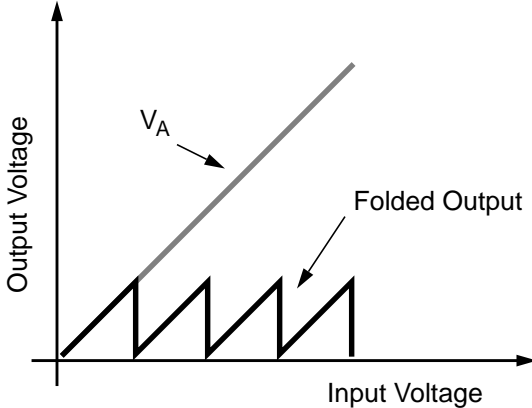
(b)

Figure 4.4: Folding A/D converter. (a) Block diagram; (b) folding operation.

4.2.5 Interleaved Converters

A final technique for high-speed conversion is parallelism. By operating two (or more) converters in an interleaved fashion, high-speed A/D converters have been designed [56], [57]. Figure 4.5 shows a diagram of two interleaved flash A/D converters. The implementations of the converters themselves need not be flash [57], [58]. In monolithic implementations, the bias circuitry and reference generators can be shared, reducing power and area. However, the input capacitance for such a system is increased. Additionally, mismatches in the converters will appear as spurious tones in the output spectrum, degrading the performance [58]. For high-speed operation, a sample-and-hold
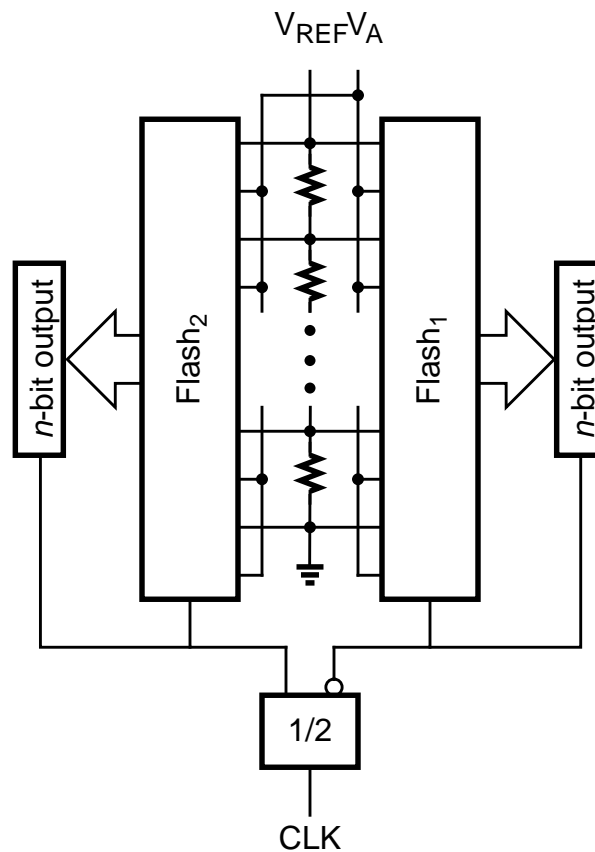

Figure 4.5: Interleaved flash A/D converters.

amplifier is often required to sample the input at the full sampling speed of the system which may limit the ultimate performance of the system [57].

4.3 Metastability Errors in Flash A/D Converters

A/D converters with low error rates have to date been implemented with flash architectures due to the simplicity of error reduction. In this section we show the source of metastability errors and the calculation for the error probability in a flash converter.

In flash converters, all comparators are presented with a voltage difference in parallel, and latched every clock cycle as shown in Figure 4.1(a). Metastability errors manifest themselves when undefined comparator outputs pass to the converter output bits. As shown in the example in Figure 4.6, under normal operating conditions the comparator outputs generate a thermometer code. The encoder will select the appropriate output code for the thermometer code representation. However, if V_A is near the reference voltage for a comparator, the comparator output may be undefined at the end of the evaluation time. V_A will only be close to the reference voltage for the comparator which is at the transition point of the thermometer code from '0' outputs to '1'. Since V_A is continuous in nature, there is always a finite probability that the voltage difference cannot be amplified sufficiently in the time allotted for comparison. If the output is not sufficiently amplified, nonbinary signals are propagated to the encoder logic, leading to errors.

If the comparator consists of a preamplifier followed by a regenerative latch, with a uniformly distributed analog input waveform, the metastability error probability for an n-bit converter with 2^n-1 parallel comparators can be calculated from equation (2.29) and is equal to

$$P_E = \frac{2\,(2^n-1)\,V_o}{V_R A}e^{-t_r/\tau}. \tag{4.1}$$

V_R is the analog input range and V_o is the output voltage swing required for valid logic levels to the thermometer code circuitry. A is the combined gain of the preamplifier and the latch's gain in the transparent state, τ is the regeneration time constant for the latch, and t_r is the resolution time of the latch.

Thermometer code with a
metastable comparator output.

$$
\begin{array}{cc}
0 & 0 \\
0 & 0 \\
0 & 0 \\
V_{REF(i)}<V_A<V_{REF(i+1)} \longrightarrow \quad 1 & \mathbf{X} \longleftarrow V_A{\sim}V_{REF(i)} \\
1 & 1 \\
1 & 1 \\
1 & 1
\end{array}
$$

Thermometer code with valid
comparator outputs.
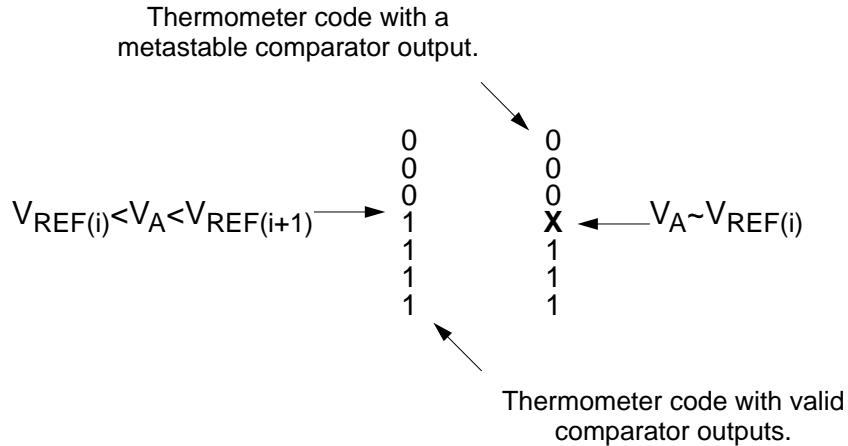
Figure 4.6: Metastable comparator output in thermometer code.

Several of the parameters in equation (4.1), however, are not known with sufficient accuracy to determine the error rate to more than an order of magnitude or two. In particular, the parameters V_o and t_r are difficult to predict over process skew and device mismatch. For single-ended implementations, V_o is a strong function of the supply noise in the circuit. V_o is also influenced by the comparator's metastable voltage and the logic following the comparator. Nominally, t_r is half of the clock cycle, however t_r is influenced by the propagation delay of the thermometer code circuitry and the setup time of any latches placed after the comparators. Additionally, comparators may have varying t_r and τ parameters due to different loading conditions caused by the converter's floor plan. For an accurate determination of the metastability error rate of a converter, the performance often must be measured.

4.4 Measuring Metastability Errors in A/D Converters

To characterize metastability error rates in A/D converters, the measurement technique of Figure 4.7(a) is used [6]. The circuit compares the current sample with a previous sample plus an offset, Δ. If the A/D converter input waveform is a low frequency, full-scale
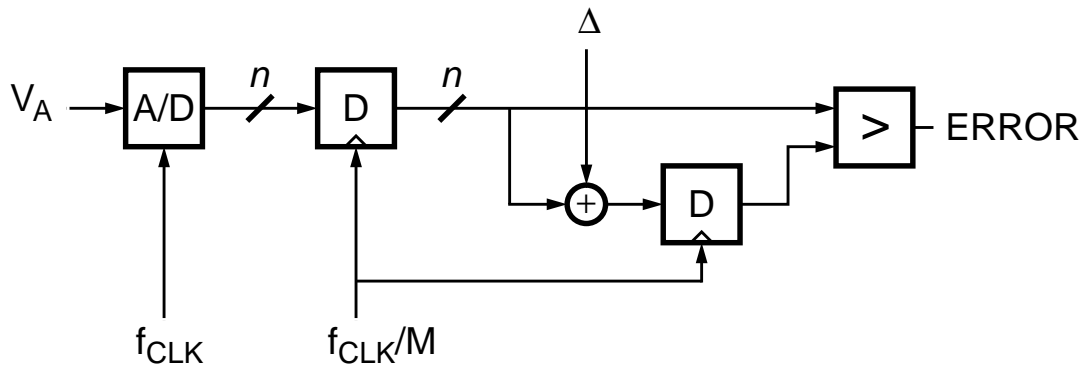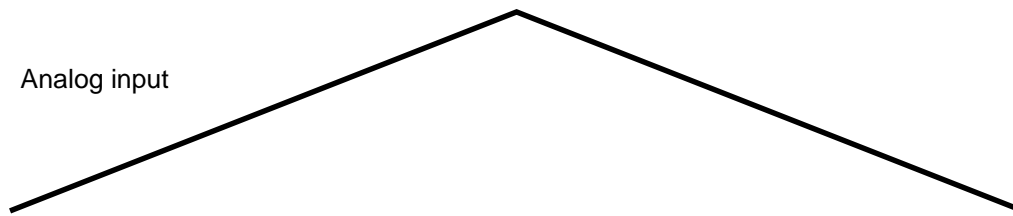
Figure 4.7: Measurement circuit for metastability errors. (a) Circuit schematic; (b) analog input; (c) digitized output.

waveform, successive samples should differ by one bit or less. A metastability error will appear as a large code variation and will be flagged by the greater than circuit block. Errors which glitch low will be flagged when the data returns to the input levels as shown in Figure 4.7(b). The number of errors can be accumulated, and the error probability calculated.

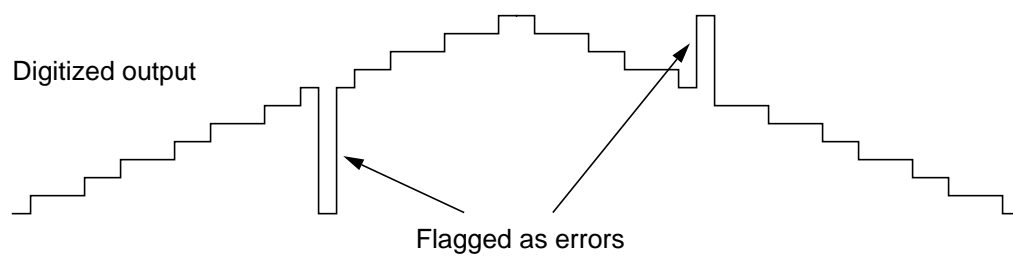If the output is sub-sampled by dividing the sampling clock down by a factor *M* as shown in the figure, there are two benefits. Because of the longer cycle time, sub-sampling allows the circuit following the converter to be relatively low-speed, so long as the set-up and hold window for the *D* registers is small. Additionally, the converter can be tested with high frequency analog inputs because the sub-sampled output for high frequency inputs will be aliased to low frequencies. The maximum frequency of operation for a triangle wave input is

$$f_A < \frac{f_S}{M2^{n+1}} \tag{4.2}$$

where f_A is the analog input frequency, f_S is the sampling frequency, n is the number of bits in the converter, and *M* is the division factor for the clock. For a sinusoidal input, the maximum analog input frequency is

$$f_A < \frac{f_S}{\pi M2^n}. \tag{4.3}$$

Additionally, for *i*=2 to *M*, the aliased sinusoidal inputs which satisfy the equation

$$\left(\frac{f_S}{i} - \frac{f_S}{\pi M2^n}\right) < f_A < \frac{f_S}{i} \tag{4.4}$$

can be used.

Figure 4.8(a) shows measured error rates versus sampling frequency for a 6-bit converter implemented in 1.2-μm CMOS based on the architecture of [49]. No explicit metastability error correction is implemented in the converter. The supply voltage is 5 V. The analog input waveform is a full-scale, 1-kHz triangle wave. Because t_r is derived from the clock signal, the error rate displays an exponential relationship with the sampling frequency. Although there is some variation, τ for the circuit is approximately 240 ps. For a 70-MHz sampling frequency, the converter has a measured error rate of approximately

10^{-7} errors/cycle, corresponding to about seven errors per second. This level of errors is unacceptable in many applications, and to achieve acceptable performance, the converter must be specifically designed for low error rates.

Figure 4.8(b) shows a plot of the error rate dependence on the analog input frequency. The analog input is a full-scale sinusoidal input, to avoid problems with the higher harmonics in sub-sampled triangle waves. The sampling frequency is 70 MHz. The error rate is weakly related to the analog input frequency, but the variation over almost five orders of magnitude of input frequency is less than one order of magnitude. The measurement for a 1-kHz full-scale sinusoid differs slightly from that of the 1-kHz triangle wave due to the different distribution functions for the two waveforms [7]. The exact mechanism of the error probability variation with respect to analog input frequency was not investigated due to the weak dependence. Additionally, error rates generally can be lowered by orders of magnitude with the techniques described in the next section.

4.5   Metastability Error Reduction in Flash Converters

From equation (4.1), it is apparent that the metastability error rate in flash A/D converters is most readily impacted by the comparator regeneration time constant, τ, and the comparator settling time, t_r. Since the relationship is exponential, error rates can increase dramatically as designs are pushed to higher and higher speeds. Additionally, reducing the supply voltage can impact τ, and unless t_r is scaled appropriately, the error rates will increase. High-speed comparators must meet a variety of design criteria including input-referred offset and overdrive recovery. These design criteria may be at odds with decreasing the regeneration time constant. Because of this, increasing the resolution time available to comparators is generally used for low-error rate converters. This can be accomplished through parallelism or pipelining. The following sections discuss parallelism and two methods of pipelining: encoder pipelining and internal pipelining.

Although this discussion focuses on the flash architecture, other converter architectures, such as folding and interpolating, can use the error reduction schemes discussed in this section and chapter 5. In two-step, pipelined converters however, error reduction is difficult because the conversion is done serially. This serialization requires the
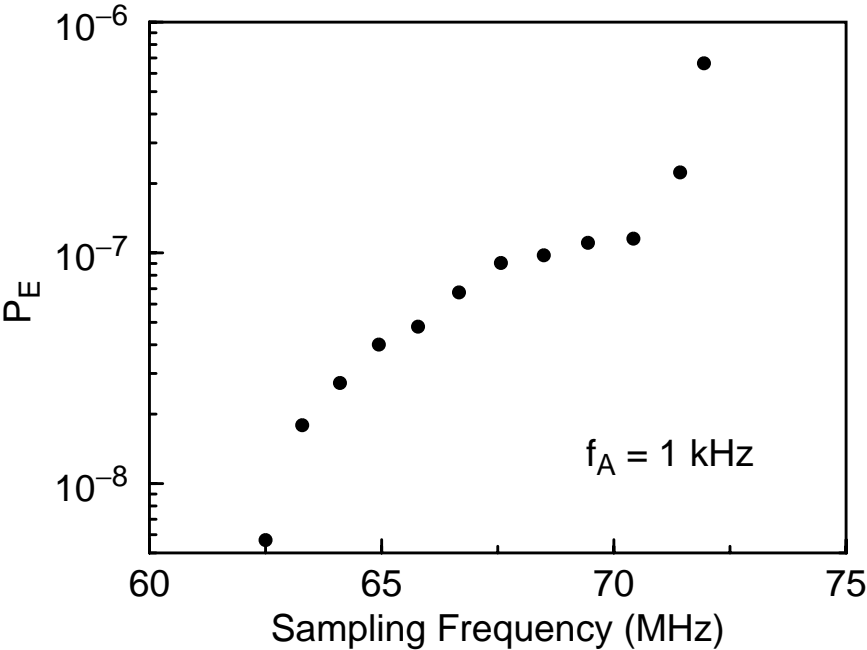
(a)
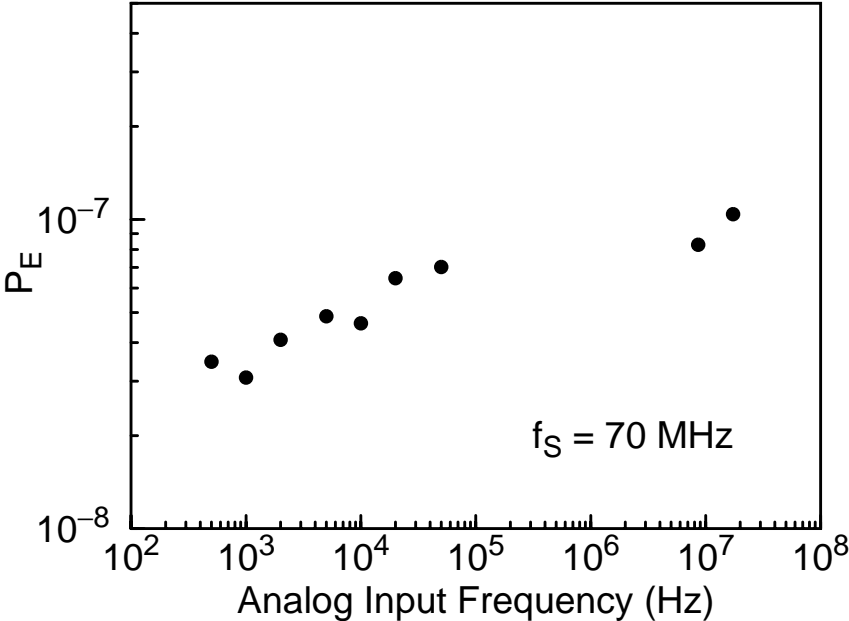
(b)

Figure 4.8: Measured error rates. (a) Versus sampling frequency; (b) versus analog input frequency.

analog input be pipelined as well as the digital output. In addition, the DAC structure is often merged directly with the thermometer code, requiring large numbers of latches to pipeline the inputs to the DAC. These problems make pipelined, multistep converters a poor choice for architectures in applications requiring low metastability error rates [59].

4.5.1 Parallelism

As shown in section 4.2.5, high-speed implementations based on parallel converters is possible. However, this technique can also be used to reduce the conversion system's metastability error rate. Additionally, the parallelism can be accomplished with discrete parts and not just in a monolithic implementation. By placing two (or more) converters in parallel and interleaving them, the clock frequency for each converter can be reduced while maintaining the system throughput. By reducing the clock frequency, the resolution time, t_r, available to each converter is increased. For example, if the clock frequency for each converter is reduced by a factor of two, and the resolution time is assumed to double, this will result in an error probability reduction of $e^{-t_r/\tau}$. However, implementing two converters in parallel requires a large amount of area and doubles the input capacitance of the system. In addition, mismatch in the converters can limit performance. In monolithic implementations of low error rate converters, pipelining is more common.

4.5.2 Encoder Pipelining

In contrast to using a ROM-based encoding scheme, it is possible to use a coding scheme which is more tolerant to an undefined comparator output. By ensuring the encoding occurs without the output from a single comparator affecting more than one bit in the output code, metastability error rates can be reduced with pipelining within the encoder. This can be accomplished using a Gray code with a gate-based encoding scheme rather than ROM-based encoding. Figure 4.9 shows the logic required for a 3-bit flash converter from the thermometer code to the converter output. The comparator outputs which generate the thermometer code are represented with the notation T_i where i is a number from 1 to $n-1$. Because the thermometer code outputs are used only once, latches can be inserted between the logic stages to reduce the probability of an unsettled output bit.
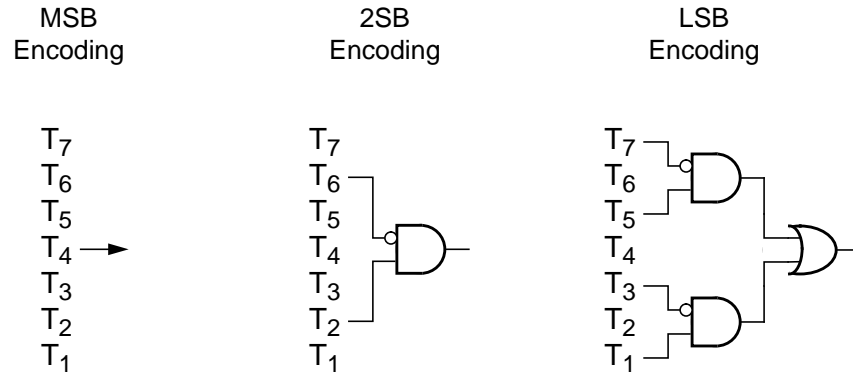
Figure 4.9: Logic-based Gray encoding.

Additionally, the number of latches needed per level of pipelining is reduced by roughly a power of two with each level of encoding.

Gate-based encoding however, requires that signals from different areas of the chip be logically combined. In implementations with six to eight bits of resolution, the wiring capacitance of these signals becomes substantial. In the example of Figure 4.9, the furthest signals are only four comparators apart. However, for an 8-bit flash A/D converter, the signals T_{64} and T_{192} must be logically combined. Additionally, the encoder would require large area, and the encoder structure would have a strong influence on the converter's floor plan.

4.5.3 Internal Pipelining

Because of the regularity of the encoder structure and the need for only local connections to perform thermometer to 1-of-n code conversion, ROM-based encoding schemes are used extensively in flash A/D converters. Previous high-speed flash converters have increased t_r to improve metastability error rates by inserting pipeline latches at the comparator outputs internal to the A/D converter. One benefit of internal pipelining is that the metastability error reduction is independent of the encoding scheme used in the ROM,
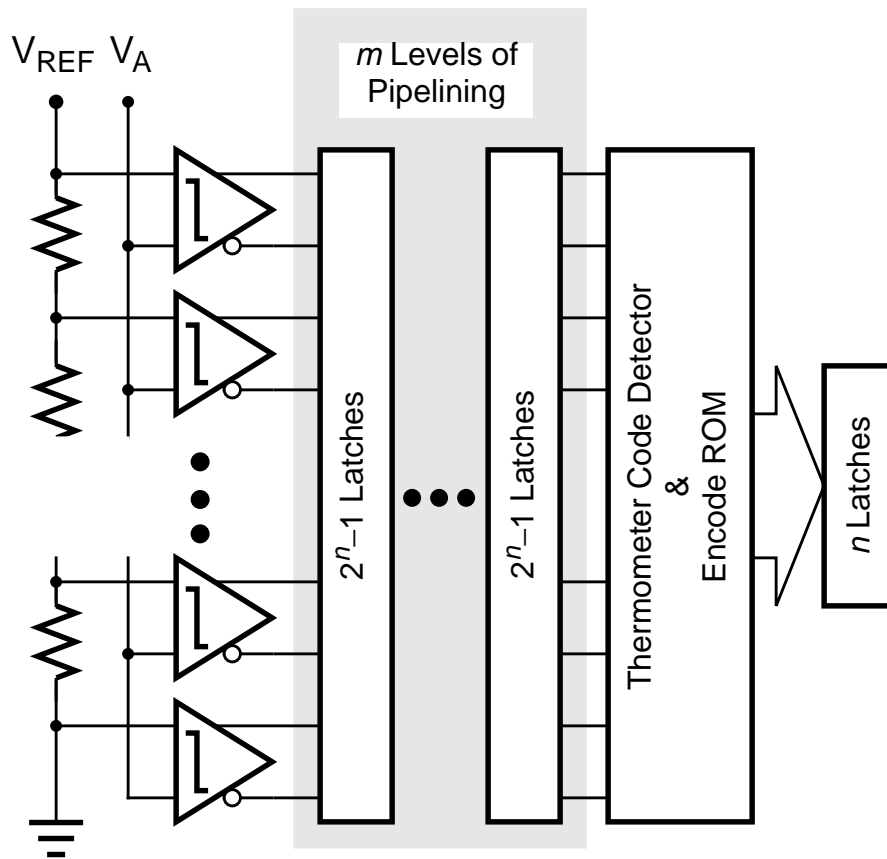
Figure 4.10: Internal pipelining to reduce metastability errors.

and allows for more sophisticated coding schemes to remove bubbles from the thermometer code. However, an n-bit flash A/D converter with m levels of pipelining requires $m(2^n-1)$ latches as shown in Figure 4.10. Previous 8-bit high-speed bipolar flash designs have included as many as three levels of pipelining. Clock loading can be increased substantially as a clock signal must be distributed to each of these latches. Additionally, with a Nyquist frequency input, all of the latches in the pipeline toggle at the sampling frequency, dissipating power. To reduce error rates with less power and area, the number of latches must be reduced and the clock loading decreased.

4.6 Summary

A brief description of the fundamentals of high-speed A/D conversion has been presented. The main architectures for high-speed A/D conversion have been surveyed and the suitability of flash converters for low error operation has been discussed. The mechanisms of metastability errors in flash A/D converters have been presented, and the equation to estimate a converter's error probability has been generated. The measurement technique to characterize A/D converters was shown, as long with measured results from a 6-bit CMOS flash A/D conversion with no error reduction circuitry. A discussion of previous techniques to reduce high-speed error rates has been included as well.

Chapter 5

Power-Efficient Metastability Error Reduction in Flash A/D Converters

5.1 Introduction

As seen in chapter 4, previously reported methods of reducing metastability error rates in A/D converters have either required considerable area and power overhead, or will affect the implementation of the converter negatively. In this chapter an external pipeline technique for reducing metastability error rates in high-speed flash A/D converters with minimal power and area overhead is described. In addition, this chapter describes the implementation of a 7-bit flash converter in 1.2-μm CMOS using external pipelining. Measured results from the implemented chip are included, and a power comparison to an internally pipelined implementation is performed.

5.2 Error Reduction with External Pipelining

To reduce the error rate in A/D converters it is necessary to increase the effective resolution time available to the comparators in the converter. In order to increase the resolution time, t_r, in equation (4.1) without inserting stages of latches at the outputs of the comparators, the error propagation from a metastable comparator must be contained. This section investigates the methods of error propagation in ROM based encoders, and

presents a method for constraining the error propagation to allow pipelining at the converter output to reduce metastability errors.

5.2.1 Error Propagation

Figure 5.1 shows the error propagation for a flash A/D converter with a binary encoded, n-channel pull-down ROM. Zeros in the ROM are implemented with n-channel pull-down transistors, and ones are implemented as opens. The undefined comparator outputs can propagate to the pull-down transistors in the ROM connected to the undefined word lines. For an undefined comparator output propagating to the two output words '100' and '011'
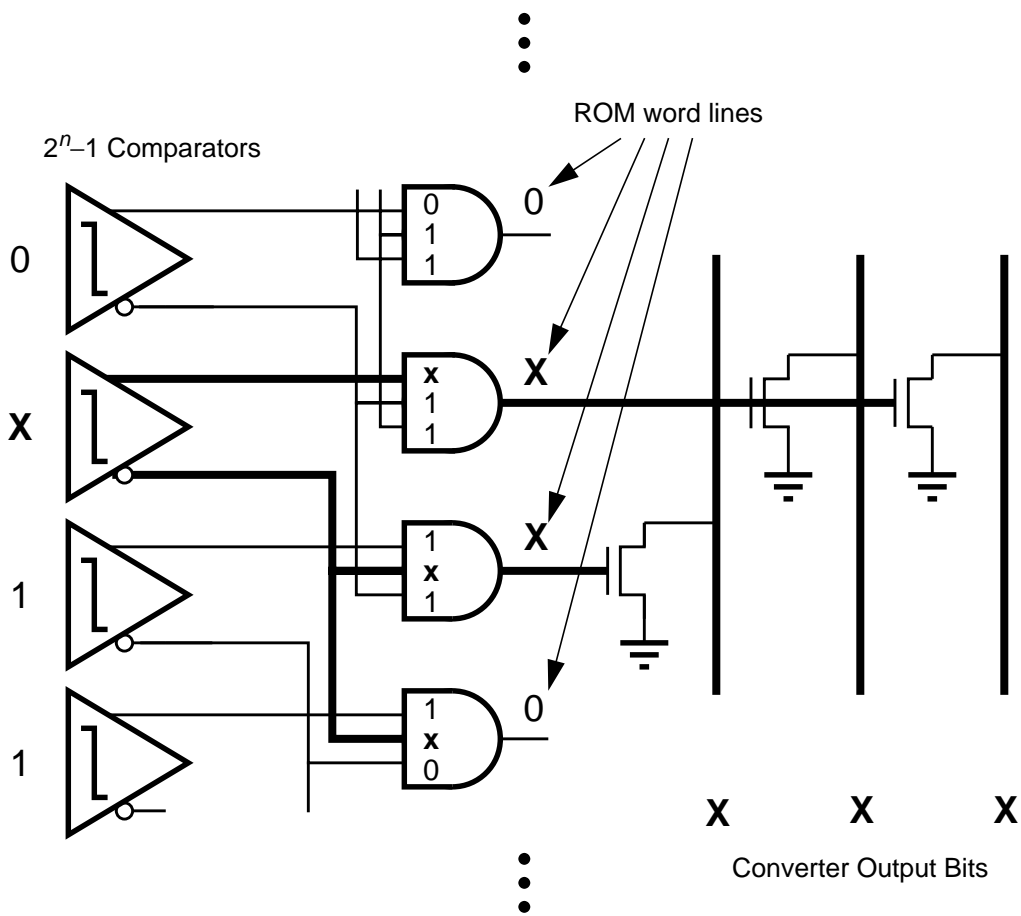

Figure 5.1: Error propagation in encode ROM words and output bits.

as shown in the figure, undefined values are propagated to all of the output bits. The final output is determined by random mismatches in transistors and/or process skew. Since the proper output encoding is either '100' or '011', but the bits will settle randomly, glitches in the output waveform are possible with very large deviations from the desired encoding. As an extreme example, assume all bits settle to zero. The encoding '000' is approximately half of the full-scale voltage away from the desired encoding. A converter with a Gray-encoded ROM has similar error propagation.

5.2.2 Containing Errors

To constrain error propagation from metastable comparator outputs, it is necessary to ensure that the metastable voltage level for comparators is interpreted as a valid logic level and use a code in the ROM which can handle the case of a metastable comparator. Figure 5.2 shows a current-mode implementation of a comparator in the latched state and a thermometer code logic gate. In this case, the logic gate is single-ended and biased to a reference voltage, V_{R1}. If V_{R1} is near V_M, the outputs of the gate are unpredictable or at worst undefined with a metastable comparator. If V_{R1} is higher than the metastable voltage level for the comparator, the logic gate will interpret both $V_{out(i)}$ and $\overline{V}_{out(i)}$ as valid "0" signals for metastable comparator. If V_{R1} is lower than the metastable voltage level for the comparator, the logic gate will interpret both $V_{out(i)}$ and $\overline{V}_{out(i)}$ as valid '1' signals if the comparator is metastable. For this case, both comparator outputs will appear as a valid '1' until the comparator resolves, but only one output will transition when the comparator resolves. Thus, only one output can be undefined at any one time for the comparator. By ensuring only one undefined output at a time and the other output a valid high, errors resulting from comparator metastability can be gracefully handled by the encode ROM.

For full-rail logic families such as static CMOS logic, the reference level V_{R1} is implicit within the device, and the voltage level is determined by the device sizes within the comparator and the encode logic. However, the implied voltage level can vary considerably over process and supply skew.

Error propagation from the comparator outputs proceeds to the ROM as follows. If both comparator outputs are high when the encode ROM is clocked, two adjacent word lines will be turned on in the ROM causing large errors in a ROM with binary words.
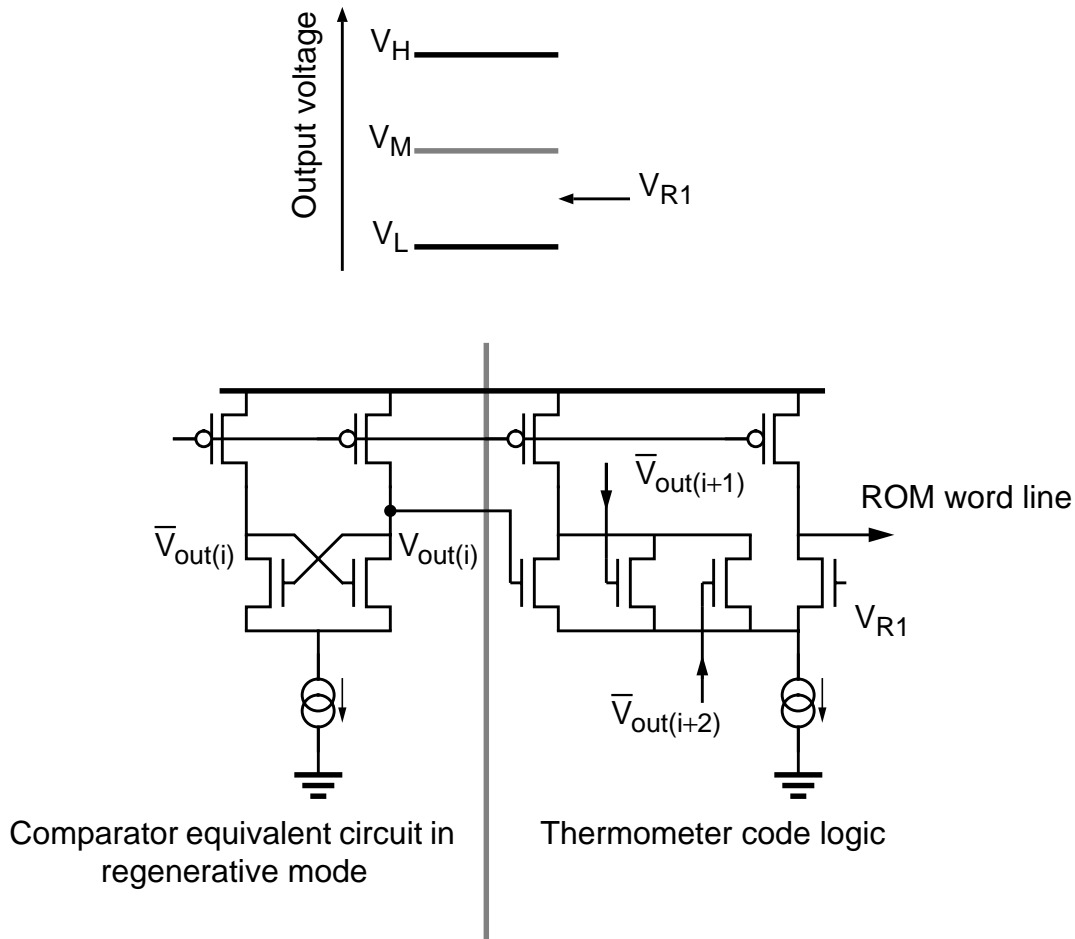
Figure 5.2: Implementation of valid high outputs for metastable comparators with current steering logic.

However, if a Gray-encoded ROM is used, the output bits will be the logical AND of the two words. This causes no problems because the words differ by only one bit. Since metastability errors only occur with the analog input at a boundary between two output words, either code can be considered correct. The worst case scenario for the encode ROM now becomes the single comparator output settling as the ROM is clocked. The undefined comparator output is now passed to the ROM as a word line high and an adjacent line undefined as shown in Figure 5.3. Due to the Gray code, the logically ANDed output words leave only one output bit unsettled. The probability of the single bit
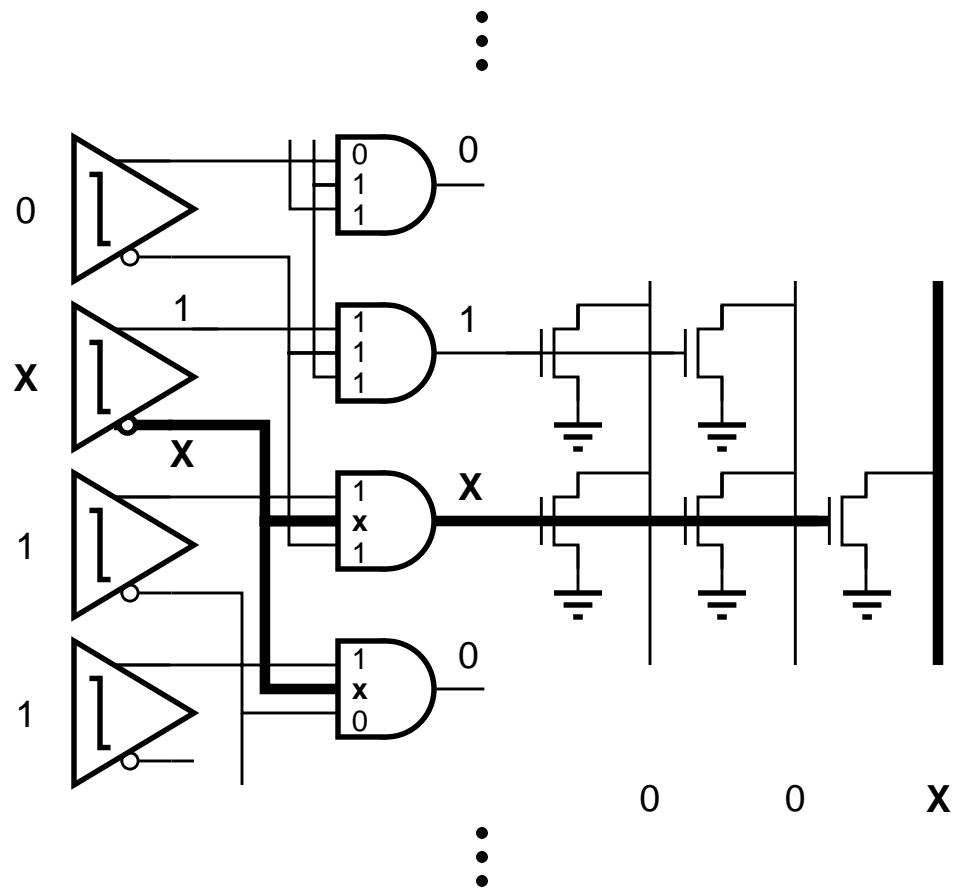
Figure 5.3: Error propagation with comparator outputs valid high and Gray-encoded ROM.

remaining unresolved can be reduced with pipeline latches at the converter output, but only n latches are required per stage, for an n-bit converter. A converter normally has at least one set of latches here to resynchronize the output so there is no overhead for a single pipeline stage.

With AND gates used as the thermometer decode logic, error reduction at the output is possible only with the use of valid outputs for unsettled comparators and the Gray code in the coded ROM. Figure 5.4 shows a table with a binary code rather than a Gray code being used, as well as the outputs with a Gray code and unresolved outputs undefined. As can be seen from the chart, only the case using both techniques, comparator error containment

	Binary ROM			Gray ROM		
	$1XX$	ROM_4		$XX1$	ROM_1	Unresolved comparator undefined.
AND	$X11$	ROM_3	AND	XXX	ROM_0	
	XXX			XXX		
	$1XX$	ROM_4		001	ROM_1	Unresolved comparator outputs held high.
AND	011	ROM_3	AND	XXX	ROM_0	
	$0XX$			$00X$		

Figure 5.4: Errors in binary and Gray code ROMs with unresolved comparator outputs undefined and unresolved comparator outputs held high.

and Gray encoding, leaves only one output bit unresolved. The other schemes shown in the figure will result in more than one bit remaining unsettled, resulting in large errors. Other encoding schemes for ROMs can be evaluated using similar figures.

5.2.3 Other A/D Converters

The circuit techniques described in section 5.2.1 can be used to reduce metastability error probabilities in other high-speed converter architectures as well. Interpolating converters can implement the circuits as described. The techniques used here to constrain error propagation in thermometer code detection should be directly applicable to cyclical code detection, allowing low-power reduction of error rates in folding A/D converters, by implementing both the fine and coarse stages of Figure 4.4, for a low error rate folding A/D converter. Two-step architectures can use the techniques to reduce the severity of errors in the course flash stage, but this would require a pipeline be inserted in the analog input path to synchronize the analog input with the pipelined coarse stage output. In practice, this would introduce nonlinearity and noise with each additional sample-and-hold and may prove unfeasible.

5.3   A 7-bit, 80-MHz Flash A/D Converter

A high-speed flash converter was designed in 1.2-μm CMOS as a test vehicle for the architecture described in section 5.2. To facilitate a simple high-speed design, no offset cancellation was used in the comparator. For the process available [32] this limited resolution to 7-bits. The converter is designed to have the highest sampling frequency possible and an error probability of less than 10^{-10}. The implementation is similar to previously published CMOS flash converters, but the external pipeline scheme described in section 5.2 is used to minimize metastability errors [9], [49].

5.3.1   Comparator

The comparator consists of a differential preamplifier followed by a drain-strobed latch. Figure 5.5(a) shows the differential preamplifier design. The diode connected loads of M5 and M6 effectively clamp the output signal swings, decreasing the reset time for the preamplifier and reducing the kickback noise on to the reference ladder. The negative resistance loads of M3 and M4 cancel the low input transconductances, g_{m5}, of M5 and M6, increasing the small-signal gain of the preamplifier. A shorting switch, M8, clears the preamp during ϕ_1. The fully symmetric circuit reduces the reset time of the preamplifier. The preamp has no offset cancelation to allow high-speed performance, but device sizes have been chosen to give an estimated input-referred offset of 1/2 LSB for a 7-bit converter with a common-mode input range of 2 V. The operating current is 80-μA, and the input bandwidth is approximately 80-MHz.

Figure 5.5(b) shows a schematic of the low offset latch used in the comparators. When ϕ_2 is pulled high, the regenerative amplifiers M5 and M6 amplify the voltage difference of V_p and $\overline{V}_p$ to near full rail. The circuit is designed for maximum frequency of operation in the overdrive recovery test, not low error rates. The circuit is similar to regenerative latches in other CMOS comparators, but the cross-coupled *n*-channel transistors are removed [49]. Removing the cross-coupled *n*-channel devices decreases the latch reset time and increases the maximum operating frequency in the overdrive recovery test [60]. However, this increases the regeneration time constant, τ, increasing the metastability error rate. With a 5-V supply and typical device parameters, τ is simulated as 385 ps.
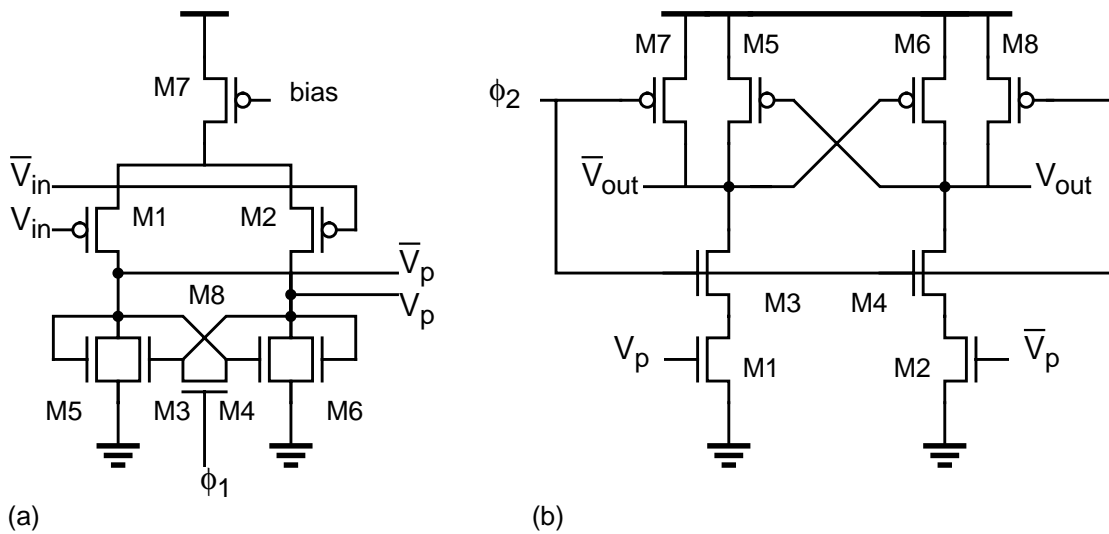
Figure 5.5: Comparator circuitry. (a) Preamplifier; (b) regenerative latch.

5.3.2 Completion Detection Circuit

As discussed in section 5.2, to enable external pipelining at the output of the converter, the metastable voltage for the comparator of Figure 5.5 must be interpreted as a valid logic level. With transistor sizing of the thermometer code logic gates, it is difficult to ensure valid logic outputs for metastable comparator voltages over process and supply skew. Additionally, device sizes for the logic gates become prohibitively large and load the comparator excessively. Figure 5.6 shows a completion detection circuit used to ensure valid high outputs for metastable inputs. All of the devices sizes are minimum. Devices M3 and M4 are on and hold the outputs, V_{out} and $\overline{V}_{out}$, high when the inputs from the comparator are at intermediate and equal. After the input voltages separate by more than one threshold voltage, one of the outputs is pulled low by the input. In contrast to a simple gain stage, the circuit of Figure 5.6 guarantees at least one valid output. Additionally, the circuit operates properly over a wide range of common mode input voltages. Additionally, the circuit does not require a clock signal. Although the delay through the completion detection circuitry reduces the settling time for the comparator during the clock phase when the comparator is strobed, this time is more than compensated for in an increase in
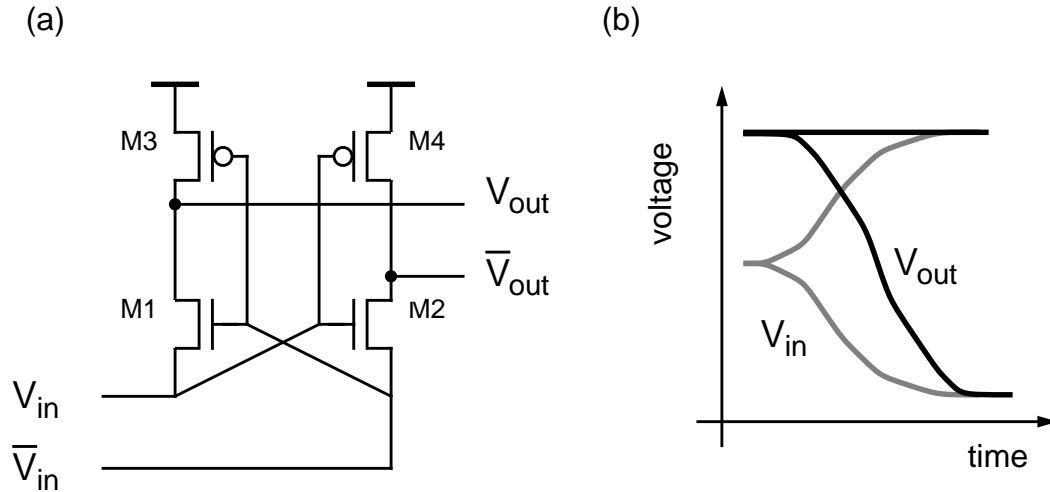
Figure 5.6: Completion detection circuit. (a) Circuit schematic; (b) input-output voltage relationship.

the settling time created by the pipeline latches at the output. The cycle time for the comparator is set by the overdrive recovery condition and not the forward delay, so the insertion of the completion detection circuitry does not impact the converter's cycle time.

5.3.3 A/D Block Diagram

Figure 5.7 shows the block diagram for an A/D converter with external pipelining to reduce metastability errors. The encode ROM is implemented as a dynamic ROM with *n*-channel pull-down transistors. The completion detection circuitry of Figure 5.6 replaces the pipeline latches at the outputs of the comparators, and the pipeline stage is now placed at the A/D converter output. For minimum size latches, each level of pipelining is estimated to reduce the metastability error probability by about six orders of magnitude. For a 7-bit, 80-MHz converter, simulations show two levels of pipelining are needed to reliably reduce the error rate to below the 10^{-10} errors/cycle.

To reduce timing skew, the clock and analog input are both routed from the top of the design to the bottom. The pins to bring the signals on chip, however, are physically
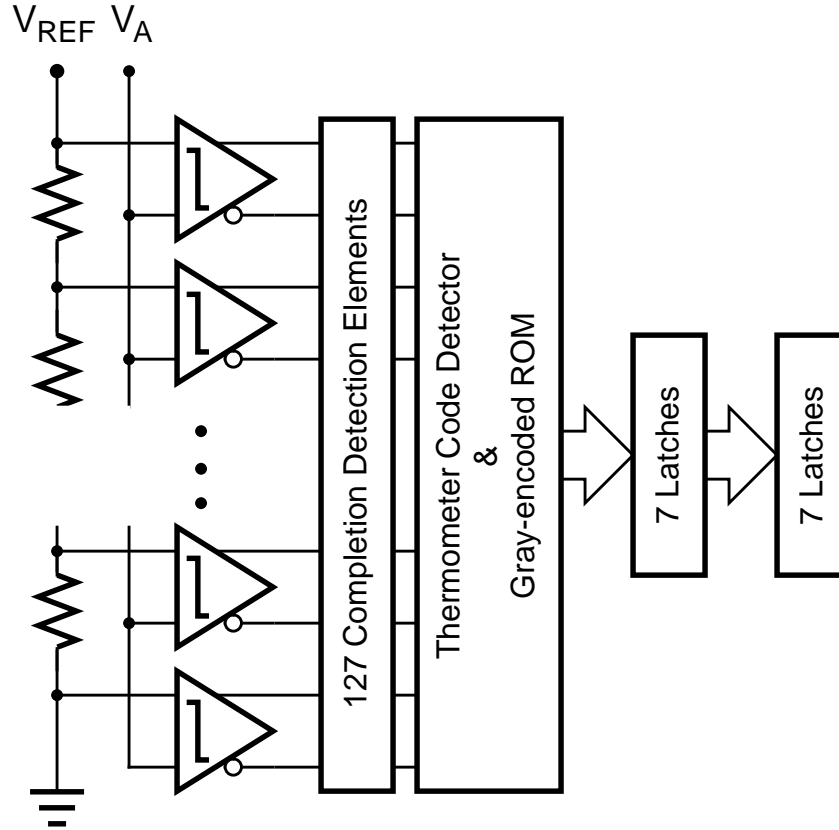
Figure 5.7: External pipelining to reduce metastability error rates.

separated by four ground and power pins to reduce capacitive coupling between the signals. In addition, the design is folded to improve the aspect ratio of the layout. By folding the design, the output coding can be implemented with 5-bit ROMs due to bit redundancy in columns. This also reduces loading on the word line drivers, increasing the speed of the encode circuitry.

5.4 Power Comparison

Simulations were performed on the extracted netlist of the converter of Figure 5.7 with a switch-level simulator to compare the power dissipation of the implemented A/D

Figure 5.8: Power comparison. (a) Error reduction circuitry; (b) total power dissipation.

converter to a converter with two levels of 127 internal pipeline latches. To allow the switch-level simulator to evaluate the circuit, the clamping transistors, M5 and M6, in the preamplif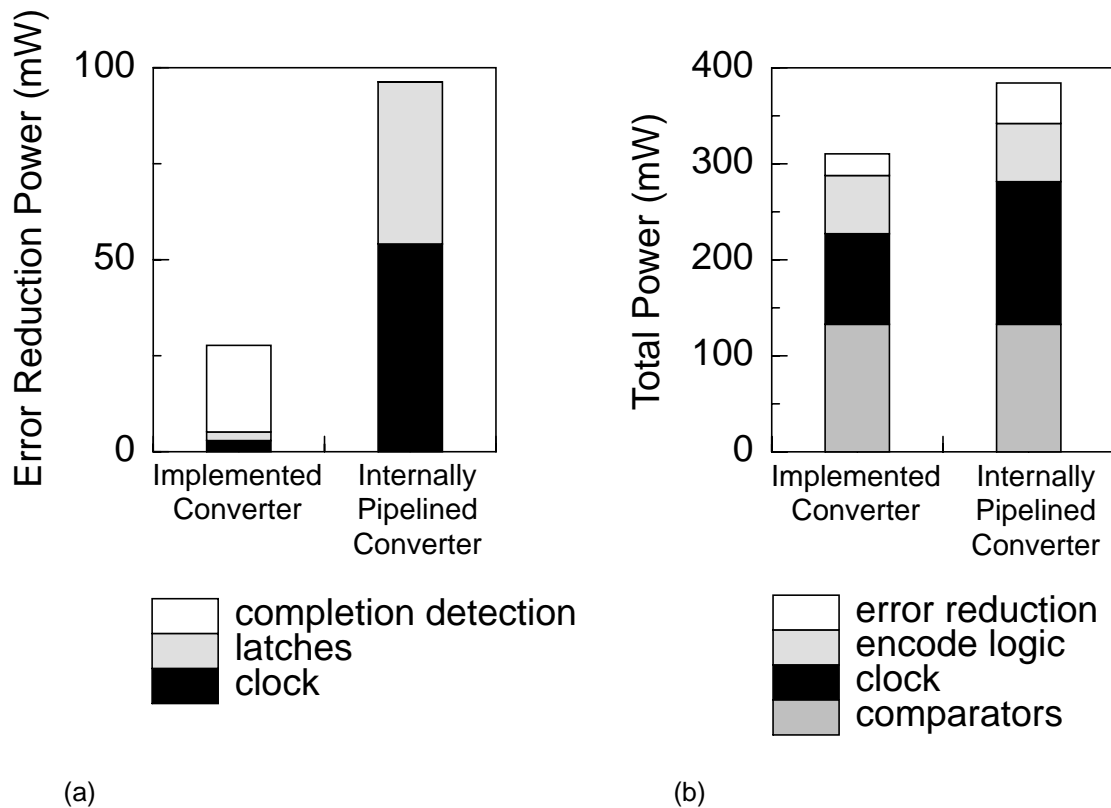ier must be removed, and the analog input simulated by setting the 127 comparator inputs to binary values at each clock cycle. Figure 5.8(a) shows the breakdown of simulated power dissipation for the error reduction circuitry in both converters with a Nyquist frequency, full-scale sinusoidal input. The clock frequency is 80 MHz. The internally pipelined converter requires 3.48 times more power for the error reduction circuitry, with most of the increase in the power for the clock signal required for the internal latches.

Figure 5.8(b) shows the total power breakdown for the implemented converter and the internally pipelined converter with 254 internal pipeline latches. Due to the static current paths in the comparators and the removal of the clamping transistors, the comparators'

measured values are used. The simulated power dissipation for the digital portion of the circuit is within 5% of the measured values for the implemented converter. The total increase in total power for the internally pipelined converter is 1.24 times. Other converter designs, with lighter clock loads in the comparators, higher resolution, or deeper latch pipelines, will have even larger improvements.

5.5 Measured Results

Figure 5.9 shows a die photograph of the 7-bit flash A/D converter of Figure 5.7 implemented in 1.2-µm CMOS. The die area is 3452×2424 µm^2, and the active area for the converter is 3037×1585 µm^2. Power supplies are split to reduce noise from the digital portions of the chip injecting into the sensitive analog portions. Three power supplies are brought out to the board, V_{DDA}, V_{DDQ}, and V_{DDD}. These are the analog supply for the preamplifiers and biasing circuitry, the quasi-analog supply for the regenerative latches, and the digital supply for the encode logic, respectively. There are also corresponding ground pins. The substrate is tied to the analog ground on chip. Analog and quasi-analog power and ground are routed from the bottom of the chip, and digital power and ground is routed from the top of the chip. The bias lines and reference voltages are routed from the bottom of the chip, and the digital outputs are brought out at the top of the chip. A Gray-

Table 5.1: Measured A/D performance summary.

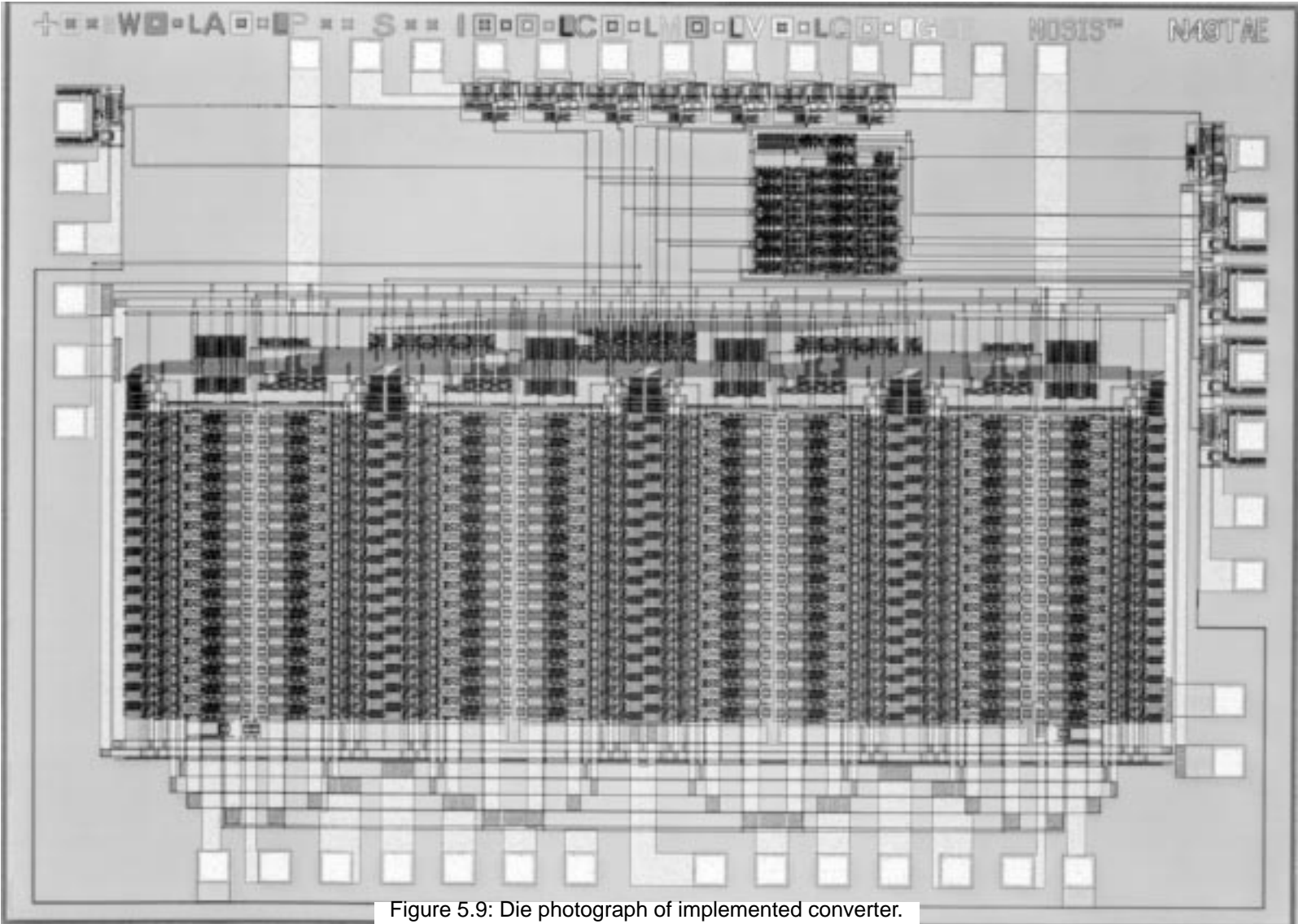|	
Process	1.2-µm CMOS
Converter Area	1585×3037 µm^2
Resolution	7 bits
Max Sampling Frequency	80 MHz
Input Range	2 V (single-ended)
SNDR	39.8 dB
Error Rate	$< 10^{-12}$ errors/cycle
Power $(f_A=11\ MHz)$	307.2 mW

Figure 5.9: Die photograph of implemented converter.

to-binary converter and test circuitry to measure the error rates is included on-chip to facilitate testing.

Table 5.1 shows a performance summary of the implemented converter. Measured performance for the converter is shown in Figure 5.10. The low frequency signal-to-noise plus distortion ration (SNDR) is 39.8 dB. This corresponds to 6.34 effective bits. The distortion level is due to the integral nonlinearity in the circuit. This can be seen in the second, third, and fifth harmonic tones in the FFT output plot of Figure 5.11. However, for the level shown, the integral nonlinearity is less than 1 LSB because all tones are less than 42 dB below the fundamental frequency's power level [5]. The harmonic distortion increases for input frequencies over 11 MHz, causing the degradation in performance for inputs seen in Figure 5.10.

The metastability test circuit described in section 4.4 was implemented on-chip along with the frequency division circuit and requires only $408 \times 432\ \mu m^2$. The registers for the test circuitry are implemented as edge-triggered flip-flops. The clock is divided by a factor of four, easing the cycle time requirements for the test circuit to 50 ns. The adder is a
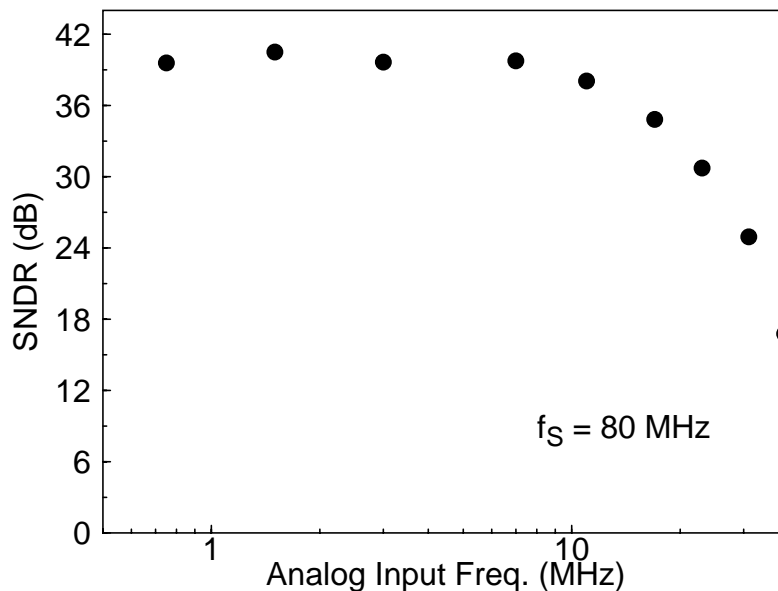
Figure 5.10: Signal-to-noise plus distortion versus input frequency.
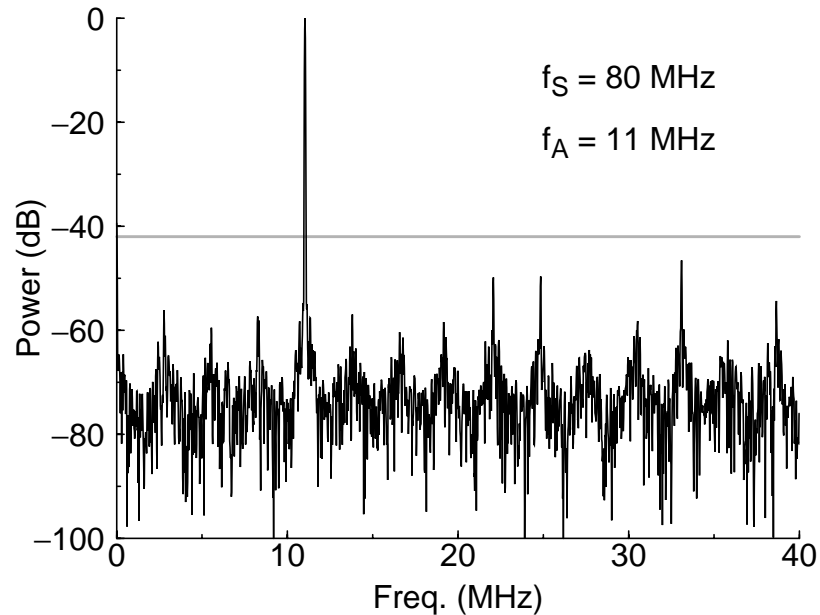
Figure 5.11: 2048 point FFT with 11-MHz sinusoidal input and 80-MHz sampling rate.

simple 8-bit ripple adder to handle overflow. The greater-than function is implemented with the adder. The carry-in connection is wired high, and one of the input words complemented. The measured error probability is less than 10^{-12} with a clock frequency of 80 MHz. The analog input was a 1-kHz, full-scale triangle wave.

5.6 Summary

An architecture to reduce metastability error rates in high-speed flash A/D converters was introduced, and an 80-MHz, 7-bit A/D converter implemented in 1.2-μm CMOS has measured error probability of less than 10^{-12}. By using external pipelines with n latches instead of internal pipelines with 2^n latches, clock loading and overall power is reduced. The converter requires two levels of pipelining to achieve this error rate. Similar error

reduction circuitry with $2(2^n-1)$ internal pipeline latches would require 3.48 times more power to implement. An internally pipelined converter with the same metastability error performance, resolution, and speed would require 1.24 times more total power.

Chapter 6

Conclusions

6.1 Conclusion

Synchronous applications with asynchronous inputs requiring low error rates must deal with metastability errors. The reduction of metastability error rates to acceptable levels in CMOS circuits requires a thorough understanding of the phenomenon and the parameters to characterize circuits. This thesis has investigated the characterization of synchronizer performance in a variety of environments commonly encountered in a standard cell based design environment. Additionally the use of architectural changes to minimize metastability errors with minimal power and area overhead in flash A/D converters has been investigated.

Following a derivation of the formulae to predict the exponential relationship between the differential output voltage versus the resolution time available in a regenerative circuit, the performance metrics for synchronizers/arbiters and comparators were calculated.

A standard test circuit for characterization of CMOS standard cell and gate array synchronizers' metastability parameters was created. The test circuit takes advantage of the concept of standardized loads or fanouts and standard delays implemented in the technology with the synchronizer to allow scalability to finer geometry processes with no changes to the circuit. This test circuit was implemented in both 2-μm and 1.2-μm CMOS processes to measure both buffered and unbuffered latches.

The results from the two synchronizer test chips showed that buffered synchronizers are superior to unbuffered synchronizers when loading is greater than a fanout of one. However, the performance is still exponentially related to loading. Additionally, buffering adds delay into the system without increasing resolution, so increased buffering past the optimum delay chain for a given load will negatively affect performance. The results also showed that reduced supply voltage negatively impacts synchronizer performance as expected, but by normalizing the parameters of interest to account for lower clock and data frequencies, the relative performance is unchanged over a wide range of supply voltages. The relative performance, however, degrades badly as the supply voltage approaches $2V_{th}$. Technology scaling, however, was shown to not always improve relative performance because τ may not scale proportionally to the gate delay for the process shrink.

In addition, a test methodology for measuring the impact of supply noise on synchronizer performance was developed, and a 2-μm test chip implemented. Supply noise was shown, in contrast to previous work, to negatively impact synchronizer performance, degrading the performance linearly with the RMS value of supply noise.

A survey of high-speed A/D converter architectures was performed, and the use of flash converters for high-speed, low error architectures was discussed. The technique was implemented in a 7-bit, 80-MHz flash A/D converter in 1.2-μm CMOS. The measured error probability was less than 10^{-12}.

6.2   Recommendations for Future Investigation

In the course of this investigation, several sections deserving further investigation have been brought to light. In this section, some of those areas are discussed.

6.2.1   Synchronizers

In section 3.6.2, the performance drop off as supply voltages approach the sum of the threshold devices in the synchronizer was calculated. As the supply voltage approaches $2V_{th}$, both devices in the synchronizer begin to operate in the subthreshold regime and performance suffers significantly. Previous work on device size optimization has focused

exclusively on optimization for devices operating above threshold voltages. It is unclear what device sizes are optimum for the regeneration time constant, τ, in the subthreshold regime.

Another area of investigation is in the creation of a noise-immune synchronizer. A fully differential implementation should reduce some of the effects of noise. However, the absolute value of τ may suffer. It is necessary to implement a noise immune synchronizer with the minimum possible τ.

6.2.2 A/D Converters

In A/D converters, several areas of investigation are possible. Initially, the error reduction technique of chapter 5 could be implemented in a folding converter for a low-power converter with very low error probabilities. Additionally, there is a need for higher resolutions in the 10 to 12 bit range with 50 to 100 MHz sampling. Such converters would be used in automated testers and require low error rates as well. Increasing the number of bits to 12 bits introduces significant design complexities which must be overcome to achieve low metastability error rates. As discussed in section 4.5, in two-step (or multi-step) pipelined converters, it is difficult to reduce the error rates with pipelining because the analog input would have to be pipelined as well. In 10 to 12 bit converters however, two-step architectures dominate because of power, area, and input capacitance requirements. A suitable architecture for high resolution, high-speed, low metastability error converters is needed.

Appendix A

Synchronizer Test Setups

This appendix discusses the board design and test methodology for the metastability parameter test chips. The boards are two-level etched kepro^{TM} boards. The power supply was bypassed with 470-μF electrolytic, 22-μF tantalum, and 1-μF surface-mount ceramic capacitors where the supply pin enters the board. Additionally, each supply pin on the chip was bypassed with a 0.1-μF ceramic surface mount capacitor at the package edge.

A.1 Measurement Setup

Figure A.1 depicts the configuration of test equipment for the mean time between failure measurements for the synchronizer performance measurements presented in sections 3.5.3, 3.6.3, and 3.6.4. The board includes DIP switches to allow for the selection of the appropriate test structure. The counter to monitor the number of errors was implemented on the board with 74ACT161 4-bit synchronous counters to count the number of failures. The outputs of the counters drive 7-segment LED displays for a digital readout of the number of failures. The counters include an asynchronous clear to allow the reset of the test. In principle, a universal counter such as the Fluke 7261A could be used to monitor the number of errors. However, this counter was not available due to equipment availability in the laboratory. The timer used was a stop-watch because of the exponential nature of the phenomenon. With the wide range of times needed to be measured, other types of timers based on the system clock would overflow before the longer tests could be run.
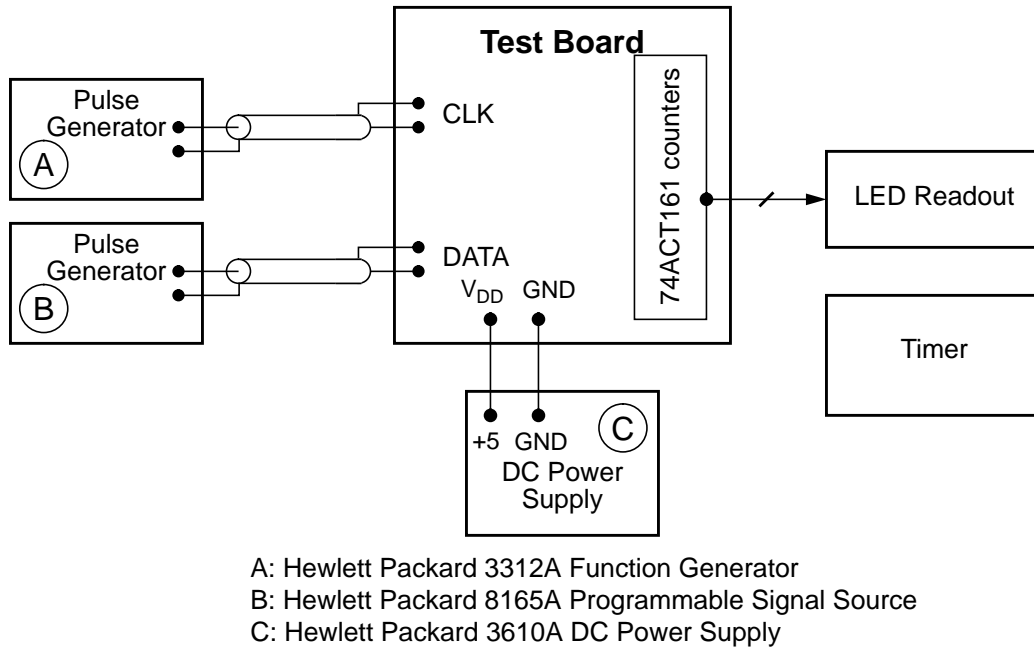
Figure A.1: Measurement setup for synchronizer performance.

A.2 Noise Measurement Setup

Figure A.2 depicts the configuration of test equipment for the mean time between failure measurements in the presence of supply noise data presented in section 3.7.2. The clean supply was bypassed as discussed above, but the bypassing at the package pins was not included. the waveform from the noise source was coupled through a 1-μF ceramic capacitor onto the noisy supply. Care was taken to ensure the measurements with the noise source off were within experimental error to the measurements with all pins connected to the clean supply. The noisy supply was monitored with a connector soldered to the board as close to the chip as possible. The oscilloscope monitor was continuously connected to ensure the capacitive loading on the supply was always the same.
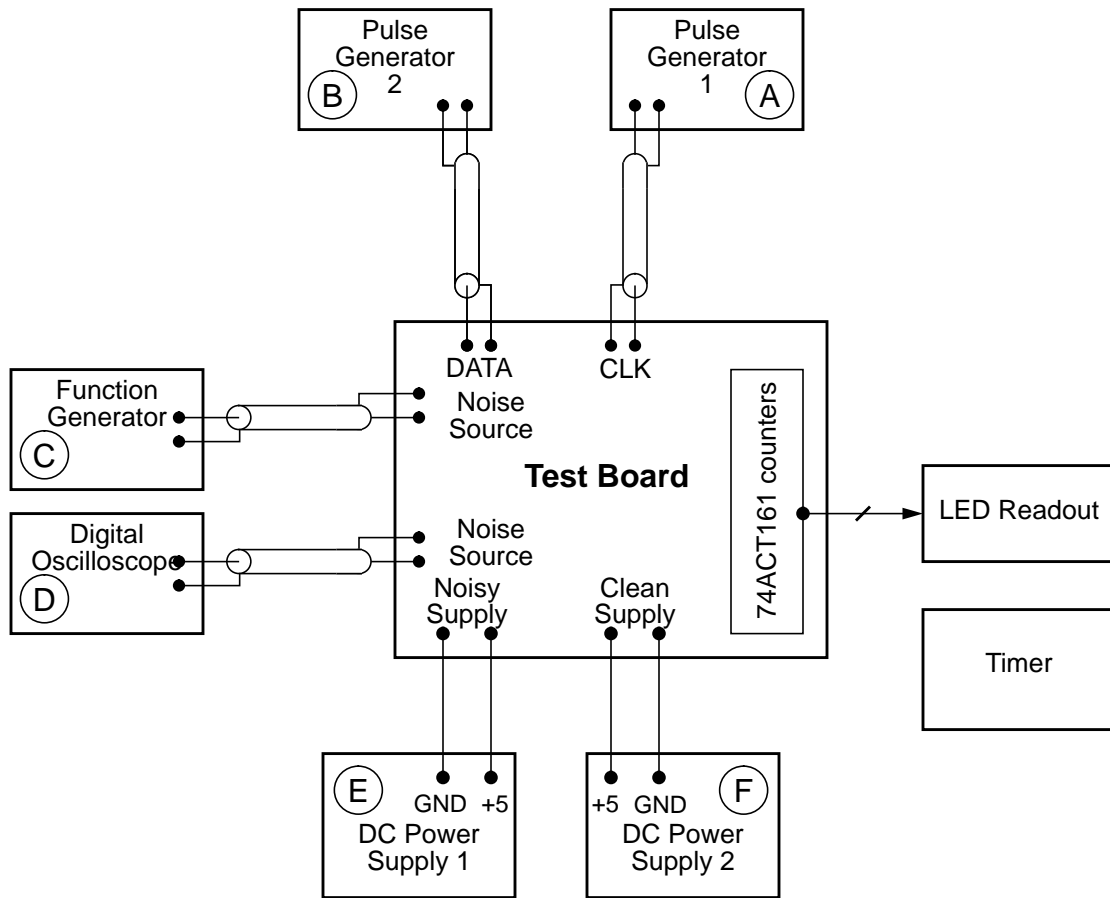
A: Hewlett Packard 8131A Pulse Generator
B: Hewlett Packard 3312A Function Generator
C: Hewlett Packard 8165A Programmable Signal Source
D: Hewlett Packard 54601A Digital Oscilloscope
E: Hewlett Packard 3610A DC Power Supply
F: Hewlett Packard 3610A DC Power Supply

Figure A.2: Measurement setup for synchronizer performance in the presence of supply noise.

Appendix B

A/D Converter Test Board

This appendix discusses the test boards for measurement of the performance of the A/D converter. All boards are custom two-level etched kepro^{TM} board. The digital supply was bypassed with a 470-μF electrolytic, a 22-μF tantalum, and 1-μF surface-mount ceramic capacitors where the supply enters the board. The package was a 52 pin J-lead PLCC package to allow bypassing with surface mount capacitors under the chip. All supply pins and bias pins are bypassed with 0.1-μF capacitors under the chip. The analog supply and quasi-analog supply are generated on-chip with Analog Devices Inc. REF3 precision voltage supplies.

B.1 SNDR Measurement Setup

Figure B.1 depicts the configuration of test equipment for the signal to noise plus distortion (SNDR) presented in section Figure B.5.5. The analog input was generated from a spectrally pure RF source to minimize harmonic distortion in the input. The input from the RF signal source was capacitively coupled through a 1-μF ceramic capacitor to a resistor divider network as shown in Figure B.2. The resistor values are used 100Ω. The parallel combination of the two resistors 50Ω terminates the signal generator. The output of the coupling circuit was inspected with a spectrum analyzer and the harmonic frequency components were low enough to allow testing of the 7-bit converters. The clock signal was unterminated to reduce coupling of clock related currents into the ground plane of the test board. In addition, the clock signal was brought in with a 2-V swing to reduce coupling on the board and amplified up on chip. The data buffered was triggered with the
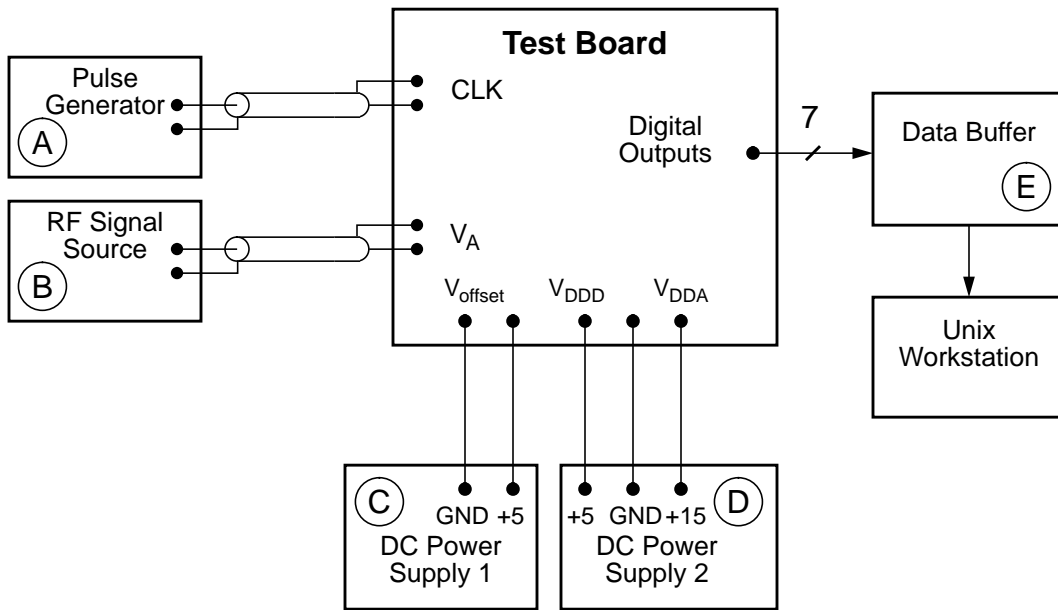
A: Hewlett Packard 8131A Pulse Generator
B: Hewlett Packard 8640B Signal Generator
C: Hewlett Packard 3610A DC Power Supply
D: Hewlett Packard 6236B Triple DC Power Supply
E: Tektronix DAS9200 Digital Analysis System with 92A16 Acquisition Module

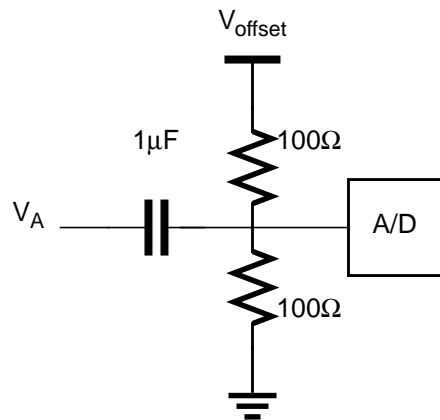Figure B.1: Measurement setup for SNDR measurement.

Figure B.2: Resistor divider.

other channel from the clock source to allow skewing between the A/D clock and the sampling clock to identify the location of the clock edge for maximum SNDR.

B.2 Error Probability Measurement Setup

The test equipment configuration for the measurement of the A/D converter error rate was shown in Figure B.3. The RF signal source was replaced with a function generator to
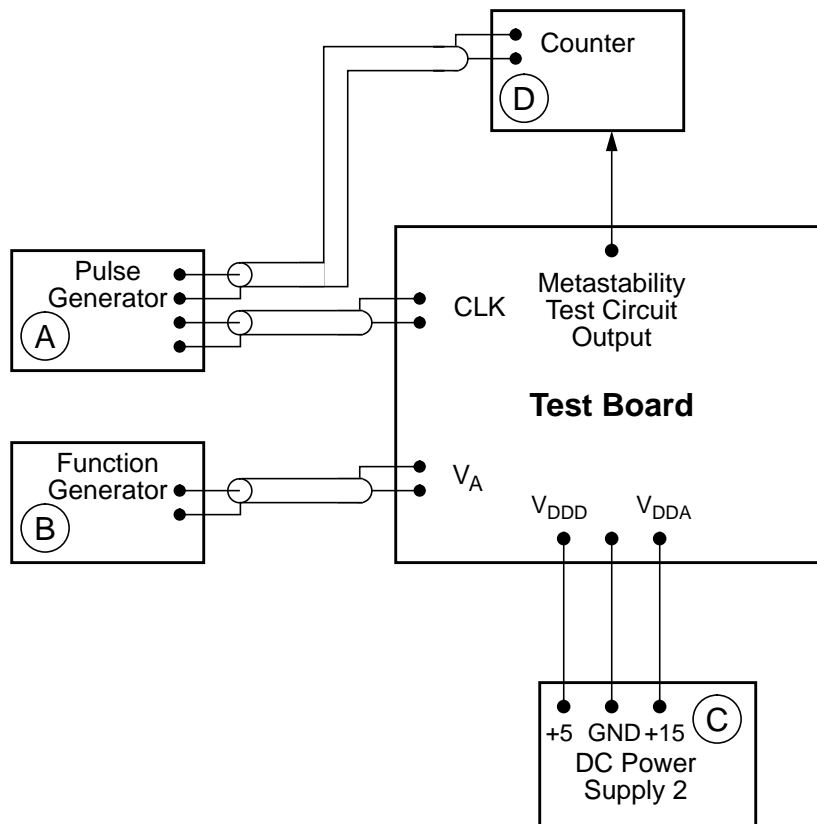
A: Hewlett Packard 8131A Pulse Generator
B: Hewlett Packard 8165A Programmable Signal Source
C: Hewlett Packard 6236B Triple DC Power Supply
D: Fluke 7261A Universal Counter

Figure B.3: Measurement setup for A/D converter error probability.

allow testing with non-sinusoidal signals. The timer again was a stopwatch due to the exponential behavior of the measurements. Δ in Figure B.7 was programmed with DIP switches and was set to 3 LSB. The extra channel from the clock generator was used to clock the error measurement circuitry.

Bibliography

[1] T. J. Chaney and C. E. Molnar, "Anomalous behavior of synchronizer and arbiter circuits," *IEEE Trans. Computers*, vol. C-22, no. 4 pp. 421-422, Apr. 1973.

[2] L. R. Marino, "General theory of metastable operation," *IEEE Trans. Computers*, vol. C-30, pp. 107-115, Feb. 1981.

[3] G. R. Couranz and D. F. Wann, "Theoretical and experimental behavior of synchronizers operating in the metastable region," *IEEE Trans. Computers*, vol. C-24, no. 6, pp. 604-616, June 1975

[4] M. Pechoucek, "Anomalous response times of input synchronizers," *IEEE Trans. Computers*, vol. C-25, pp. 133-139, Feb. 1976.

[5] B. Zojer, R. Petschacher, and W. A. Luschnig, "A 6-bit/200-MHz full Nyquist A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-20, no. 3, pp. 780-786, June 1985.

[6] C. W. Mangelsdorf, "Improving bandwidth and error rate in flash converters," *1989 Symposium on VLSI Circuits Digest of Technical Papers,* pp. 53-54.

[7] C. W. Mangelsdorf, "A 400-MHz input flash converter with error correction," *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 184-191, Feb. 1990.

[8] H. Reyhani and P. Quinlan, "A 5-V, 6-b, 80 Ms/s BiCMOS flash ADC," *IEEE J. Solid-State Circuits*, vol. 29, no. 8, pp. 873-878, Aug. 1994.

[9] A. K. Joy, R. J. Killips, and P. H. Saul, "An inherently monotonic 7-bit CMOS ADC for video applications," *IEEE J. Solid-State Circuits*, vol. SC-21, No. 3, pp. 436-440, June 1986.

[10] F. E. Barber, et al, "A 2K x 9 dual port memory," *ISSCC Digest of Technical Papers*, pp. 44-45, Feb. 1985.

[11] E. A. Reese, D. W. Spaderna, S. T. Flannagan, and F. Tsang, "A 4K x 8 dynamic RAM with self-refresh," *IEEE J. Solid-State Circuits*, vol. SC-16, no. 5, pp. 479-487, Oct. 1981.

[12] K. Nogami, et al., "A 1-Mbit virtually static RAM," *IEEE J. Solid-State Circuits*, vol. SC-21, no. 5, pp. 662-669, Oct. 1986.

[13] T. A. Knotts, D. Chu, and J. Sommer, "A 500MHz time digitizer IC with 15.625ps resolution," *ISSCC Digest of Technical Papers*, pp. 58-59, Feb. 1994.

[14] M. Loinaz and B. A. Wooley, "A CMOS multi-channel IC for pulse-timing measurements with 1mV sensitivity", *ISSCC Digest of Technical Papers*, pp., Feb. 1995.

[15] B. Joshi, et al., "A BiCMOS 50Mhz cache controller for a superscalar microprocessor," *ISSCC Digest of Technical Papers*, pp. 110-111, Feb. 1992.

[16] J. Beaston and R. S. Tetrick, "Designers confront metastability in boards and buses," *Computer Design,* Vol. 25, No. 5, pp. 67-71, Mar. 1, 1986.

[17] H. H. Chao, S. Ong, J. Y. Tang, and F. Shih, "On the synchronization of a microprocessor," *IEEE 1986 Custom Integrated Circuits Conference Proceedings*, pp. 447-450, May 1986.

[18] D. Halliday and R. Resnick, *Fundamentals of Physics,* pp. 117-118, Second Edition, John Wiley & Sons, New York, 1981.

[19] J. T. Wu and B. A. Wooley, "A 100-MHz pipelined CMOS comparator," *IEEE J. Solid-State Circuits*, vol. 23, no. 6, pp. 1379-1385, Dec. 1988.

[20] H. J. M. Veendrick, "The behavior of flip-flops used as synchronizers and prediction of their failure rate," *IEEE J. Solid-State Circuits*, vol. SC-15, no. 2, pp. 169-176, Apr. 1980.

[21] S. T. Flanagan, "Synchronization reliability in CMOS technology," *IEEE J. Solid-State Circuits*, vol. SC-20, no. 4, pp. 880-882, Aug. 1985.

[22] T. Sakurai, "Optimization of CMOS arbiter and synchronizer circuits with submicrometer MOSFET's," *IEEE J. Solid-State Circuits*, vol. 23, no. 4. pp. 901-906, Aug. 1988.

[23] L.-S. Kim and R. W. Dutton, "Metastability of CMOS latch/flip-flop," *IEEE J. Solid-State Circuits*, vol. 25, no. 4, pp. 942-951, Aug. 1990.

[24] P. A. Stoll, "How to avoid synchronization problems," *VLSI Design*, pp. 56-59 Nov./Dec. 1982.

[25] J. U. Horstmann, H. W. Eichel, and R. L. Coates, "Metastability behavior of CMOS ASIC flip-flops in theory and test," *IEEE J. Solid-State Circuits*, vol. 24, no. 1, pp. 146-157, Feb. 1989.

[26] HSPICE manual, Metasoftware Inc., 1991.

[27] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433-1439, Oct. 1989.

[28] D. E. Ward and R. W. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE J. Solid-State Circuits*, vol. SC-13, no. 5, pp. 703-708, Oct. 1978.

[29] W. H. Press, et. al., *Numerical Recipes in C*, Cambridge University Press, New York, 1992.

[30] The MOSIS Service, USC/ISI, Marina del Rey, CA.

[31] VLSI Technology Inc. *n*-well 2-μm CMOS process.

[32] Hewlett-Packard *n*-well 1.2-μm CMOS process.

[33] Hewlett-Packard *n*-well 0.8-μm CMOS process.

[34] J. H. Hohl, W. R. Larsen, and L. C. Schooley, "Prediction of error probabilities for integrated digital synchronizers," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 2, pp. 236-244, Apr. 1984.

[35] F. U. Rosenberger and C. E. Molnar, "Comments on 'Metastability of CMOS latch/-flip-flop,' "*IEEE J. Solid-State Circuits*, vol. 27, no. 1, pp. 128-130, Jan. 1992.

[36] M. Horowitz, "Clocking strategies in high performance processors," *1992 Symposium on VLSI Circuits Digest of Technical Papers,* pp. 50-53, June 1992.

[37] A. P. Chandrakasan, S. Sheng, and R. W. Broderson, "Low power CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 27, no. 4, pp. 473-484, Apr. 1992

[38] T. Wakimoto, Y. Akazawa, and S. Konaka, "Si bipolar 2-GS/s 6-b flash A/D conversion LSI," *1988 ISSCC Digest of Technical Papers*, pp. 232-233, Feb. 1988.

[39] Y. Yoshii, et. al., "An 8-Bit, 100-ms/s flash ADC," *IEEE J. Solid-State Circuits*, vol SC-19, no.6, pp. 842-846, Dec. 1984.

[40] B. Peetz, B. D. Hamilton, and J. Kang, "An 8-bit 250 megasample per second analog-to-digital converter: Operation without a sample and hold," *IEEE J. Solid-State Circuits*, vol. SC-21, no.6, pp. 997-1002, Dec. 1986.

[41] Y. Akazawa, et. al., "A 400-MSPS 8-b flash AD conversion LSI," *1987 ISSCC Digest of Technical Papers*, pp. 98-99, Feb. 1987.

[42] Y. Gendai, et. al., "An 8-b 500-MHz ADC," *1991 ISSCC Digest of Technical Papers*, pp. 172-173, Feb. 1991.

[43] C. Lane, "A 10-Bit 60 MSPS flash ADC," *1989 Bipolar Circuits and Technology Meeting*, pp. 44-47, 1989.

[44] M. Steyaert, R. Roovers, and J. Craninckx, "A 100 MHz 8 bit CMOS interpolation A/D converter," *1993 Custom Integrated Circuits Conference*, pp. 28.1.1-28.1.4, 1993.

[45] J. J. Corcoran, et. al., "A 400MHz 6b ADC," *1984 ISSCC Digest of Technical Papers*, pp. 294-295, Feb. 1984.

[46] R. E. J. Van De Grift, I. W. J. M. Rutten, and M. Van Der Veen, "An 8-bit video ADC incorporating folding and interpolation techniques," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 6, pp. 944-953, Dec. 1987.

[47] R. J. van de Plassche, and P. Baltus, "An 8-bit 100-MHz full Nyquist analog to digital converter," *IEEE J. Solid-State Circuits*, vol. 23, no. 6, pp. 1334-1344, Dec. 1988.

[48] J. van Valburg and R. J. van de Plassche, "An 8-b 650-MHz folding ADC," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1662-1666, Dec. 1992.

[49] A. Yukawa, "A CMOS 8-bit high-speed A/D converter IC," *IEEE J. Solid-State Circuits*, vol. SC-20, No. 3, pp. 775-779, June 1985.

[50] M. P. Flynn and D. J. Allstot, "CMOS folding ADCs with current-mode interpolation," *1995 ISSCC Digest of Technical Papers*, pp. 274-275, Feb. 1995.

[51] B. Nauta and A. G. W. Venes, "A 70MSample/s 110mW 8b CMOS folding interpolating A/D converter," *1995 ISSCC Digest of Technical Papers*, pp. 276-277, Feb. 1995.

[52] N. Fukushima, et. al., "A CMOS 40-MHz 8-b 105-mW two-step ADC," *1989 ISSCC Digest of Technical Papers*, pp. 14-15, Feb. 1989.

[53] T. Matsuura, et. al. "An 8-b 50-MHz 225-mW submicron CMOS ADC using saturation eliminated comparators," *1990 Custom Integrated Circuits Conference*, pp. 6.4.1-6.4.4, 1990.

[54] M. Yotusyanagi, T. Etoh, and K. Hirata, "A 10-Bit 50-Msample/sec pipelined CMOS A/D converter with S/H," *1992 Custom Integrated Circuits Conference*, pp. 16.7.1-16.7.4, 1992.

[55] M. Ishikawa and T. Tsukahara, "An 8-bit 50-MHz CMOS subranging A/D converter with pipelined wide-band S/H," *IEEE J. Solid-State Circuits*, vol. 24, no.6, pp. 1485-1491, Dec. 1989.

[56] A. Matsuzawa, et. al., "A 6-b 1-GHz dual-parallel A/D converter," *1991 ISSCC Digest of Technical Papers*, pp. 174-175, Feb. 1991.

[57] K. Poulton, J. J. Corcoran, and T. Hornak, "A 1-GHz 6-bit ADC system," *IEEE J. Solid-State Circuits*, vol. SC-22, no.6, pp. 962-969, Dec. 1987.

[58] C. S. G. Conroy, D. W. Cline, and P. R. Gray, "An 8-b 85-MS/s parallel pipeline A/D converter in 1-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 28, no. 4, pp. 447-454, Apr. 1993.

[59] B. Razavi, *Principles of Data Conversion System Design,* IEEE Press, New York, 1995.

[60] P. J. Lim and B. A. Wooley, "An 8-bit 200-MHz BiCMOS comparator," *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 192-199, Feb. 1990.

[61] J. Doernberg, H. S. Lee, and D. A. Hodges, "Full-speed testing of A/D converters," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 6, pp. 820-827, Dec. 1984.