UNIVERSITY OF PENNSYLVANIA
MOORE SCHOOL OF ELECTRICAL ENGINEERING
SCHOOL OF ENGINEERING AND APPLIED SCIENCE

DESIGN AND IMPLEMENTATION OF A TIME-TO-VOLTAGE CONVERTER/ANALOG MEMORY FOR COLLIDING BEAM DETECTORS

ANDREW E. STEVENS

PHILADELPHIA, PENNSYLVANIA

January, 1988

A thesis presented to the Faculty of Engineering and Applied Science of the University of Pennsylvania in partial fulfillment of the requirements for the degree of Master of Science in Engineering for graduate work in Electrical Engineering.

Dr. Jan Van der Spiegel
Thesis Supervisor

Dr. Sohrab Rabii
Graduate Group Chair

Abstract

This thesis describes a new CMOS integrated circuit designed to measure the time interval between two digital voltage pulses. The measurement is stored as an analog voltage on a capacitor for later digitization. The targeted range of measurable times is 5—20 nanoseconds, with a resolution of 0.5 nanoseconds. An additional feature of the circuit is a storage depth of 8 samples, i.e. 8 consecutive time measurements may be recorded individually. Hence, the chip is a combination of a *time-to-voltage converter* (TVC) and an *analog memory.*

The motivation behind the development of the TVC/Analog Memory chip is in the design of a new *wire track chamber*, one type of detector used in colliders for high-energy physics. The track chamber is used to determine the path of particles emerging from a collision. The TVC will measure the drift time of particles in the chamber, while the analog memory will allow many measurements to be made in a short period of time. By recording the drift times, the path of a particle traversing the track chamber can be reconstructed. Because of the high rate of events in a collider, an analog storage system provides a better dynamic range and a shorter dead-time than a totally digital system.

The TVC/Analog Memory chip was manufactured using the Hewlett-Packard 1.6 μm double-metal digital process available from MOSIS. The total area of the 8-channel system was 1100 μm $\times$ 1500 μm. Laboratory test results showed that the TVC had a resolution of 0.2 ns over an input time interval range of 7—25 ns. Linearity matching between the eight Analog Memory channels was better than 1%, and matching between different packages was better than 2%.

Acknowledgments

I would like to thank my thesis advisors, Prof. Jan Van der Spiegel and Prof. Hugh H. Williams, for taking the time and effort to help me with my research.

Thanks to the following people for technical assistance: Prof. Willy Sansen, Rick Van Berg, Prof. Ken Laker, Larry Singer, Prof. Charlie Sodini, Ben McCarroll, Kurt Ware.

Thanks for help with the computer resources at Penn: Rick Bowen, John Bradley, Al Broscius, Dave Feldman, Tom Trojak. Thanks for help in the laboratory testing of the chip: Rick Van Berg, Mark Clougherty.

I am grateful Joe Burgiel and Dave Watkins for the opportunity to participate in the graduate degree program at AT&T Bell Laboratories.

Finally, thanks to my parents, sister, and grandmother for their support.

CONTENTS

1. Introduction

1.1 On Colliders and Detectors

In most high-energy physics experiments, particles are collided together in order to observe the debris from the collision. These collisions are usually referred to as *events*. By finding the path and energy of the emerging particles, a physicist may infer the existence of new sub-atomic particles, new types of interactions, etc. One typical configuration of a collider is an accelerator ring (Figure 1.1). Different types of particles are collided in different types of colliders; proton-antiproton ($p\bar{p}$) and electron-positron (e^-e^+) colliders are most common.

A typical detector consists of several track chambers (also referred to as drift chambers) and calorimeters (Figure 1.2). The purpose of a track chamber is to measure the paths of emerging charged particles. The calorimeter surrounds the track chamber and detects the energy escaping from the chamber after the collision, including neutral and charged particles. With this information, particles such as electrons, muons, etc. can be identified.

1.2 Wire Track Chambers

1.2.1 Physical Description A wire track chamber is usually the center-most part of a sophisticated detector system (Figure 1.2). The chamber contains a large number of

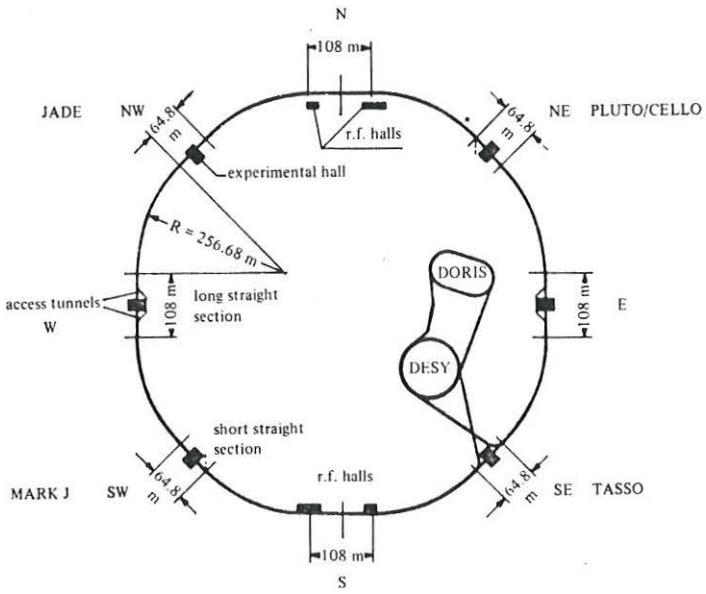


Figure 1.1. The layout of the PETRA e^-e^+ ring at DESY, W. Germany. Detectors are located in the NW, NE, SW, and SE corners of the ring. [2]

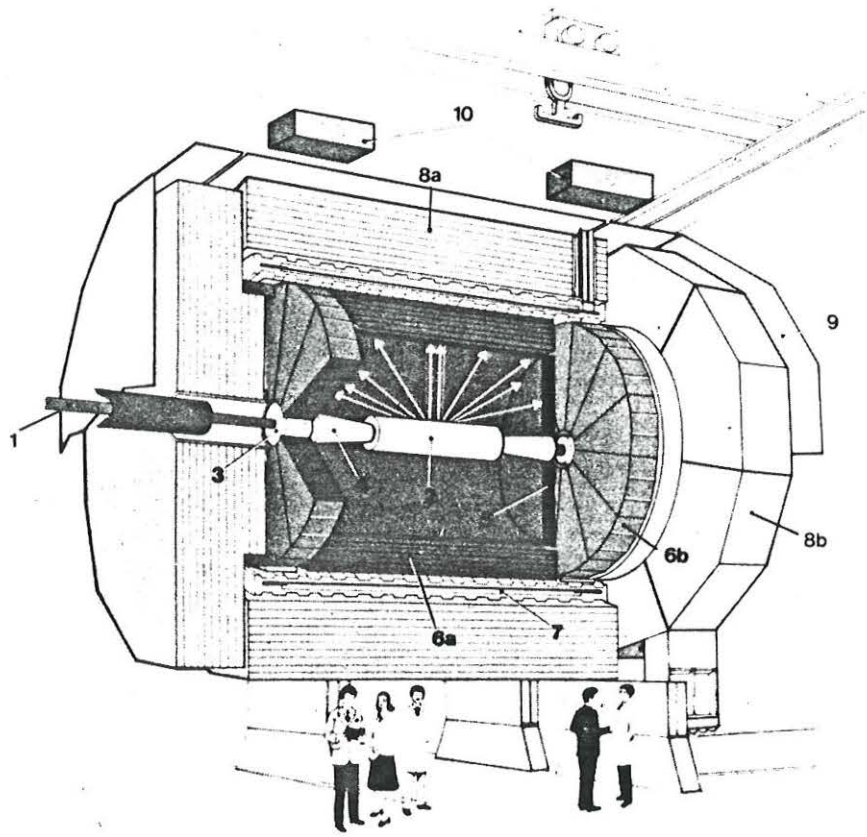


Figure 1.2. A schematic of the ALEPH detector at CERN, Switzerland. (1) Beam pipe, (2) Wire track chamber, (6a,6b,8a,8b) Calorimeters. [3]

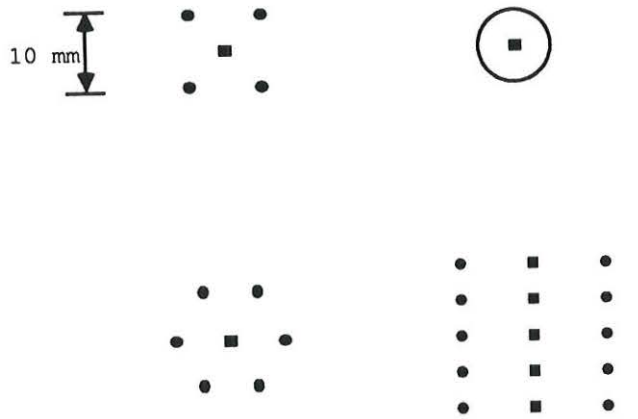
stretched wires held at fixed potentials (typically about 2000 V), thereby defining an electric field. The configuration of the wires varies from detector to detector, and is integral in the design; usually, the wires are arranged in groups of five to ten, referred to as *proportional cells.* Some typical proportional cells are shown in Figure 1.3a. Most configurations consist of a number of cathodes surrounding one or more anodes, thereby allowing the electric field to be shaped in a known pattern (Figure 1.3b).

The track chamber is also filled with a gas mixture such as argon/methane (typically 90% Ar and 10% CH_4 at atmospheric pressure). After a collision occurs, charged outgoing particles traverse the gas, leaving a path of ionized gas molecules and electrons. The positive ions will drift towards the cathodes, and the electrons will drift towards the anodes. As the electrons get close to the wires, the large increase in electric field strength will cause an avalanche effect; depending on the wire potential, this avalanche will cause a multiplication on the order of 10^4 electrons-per-electron, which will cause a measurable current pulse on the wire.

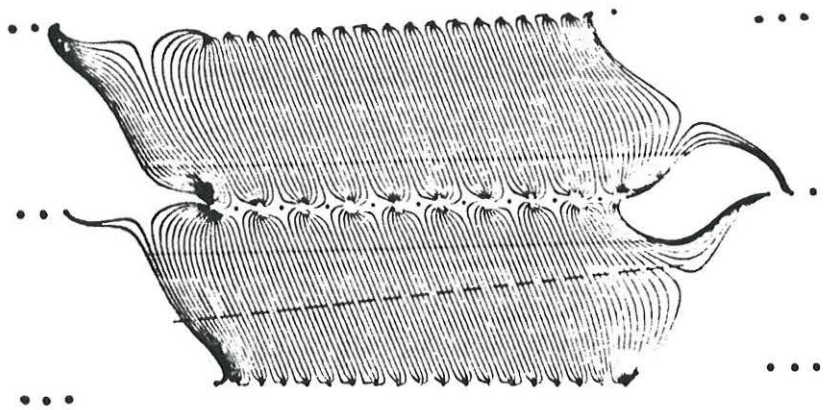
The physics of a typical proportional cell is shown in Figure 1.4 [1]. In the figure, a particle traverses the cell, creating ion/electron pairs along its path. For a constant potential between the anode and cathode, the magnitude of the electric field for this geometry is known from elementary physics to be

$$E_r(r) = \frac{V}{r \ln \frac{r_c}{r_a}} = E_a \frac{r_a}{r} \tag{1.1}$$

where $E_a = \dfrac{V}{r_a \ln \dfrac{r_c}{r_a}}$ is defined as the electric field at $r=r_a$. Using the equation for drift velocity $\mathbf{v}=\mu\mathbf{E}$, the velocity of a charge with respect to position will be



(a)



(b)

Figure 1.3. (a) Some typical arrangements for groups of wires in a track chamber. The circles represent the cathodes, and the squares represent the sense wires (anodes). A chamber may contain over 100,000 total wires. (b) A diagram of the electric field lines for the Mark II track chamber at Stanford. The ionization electrons produced along the trajectory of a charged particle (*dashed line*) drift to, and are collected by, the sense wires in the middle plane on which the field lines terminate. [3]

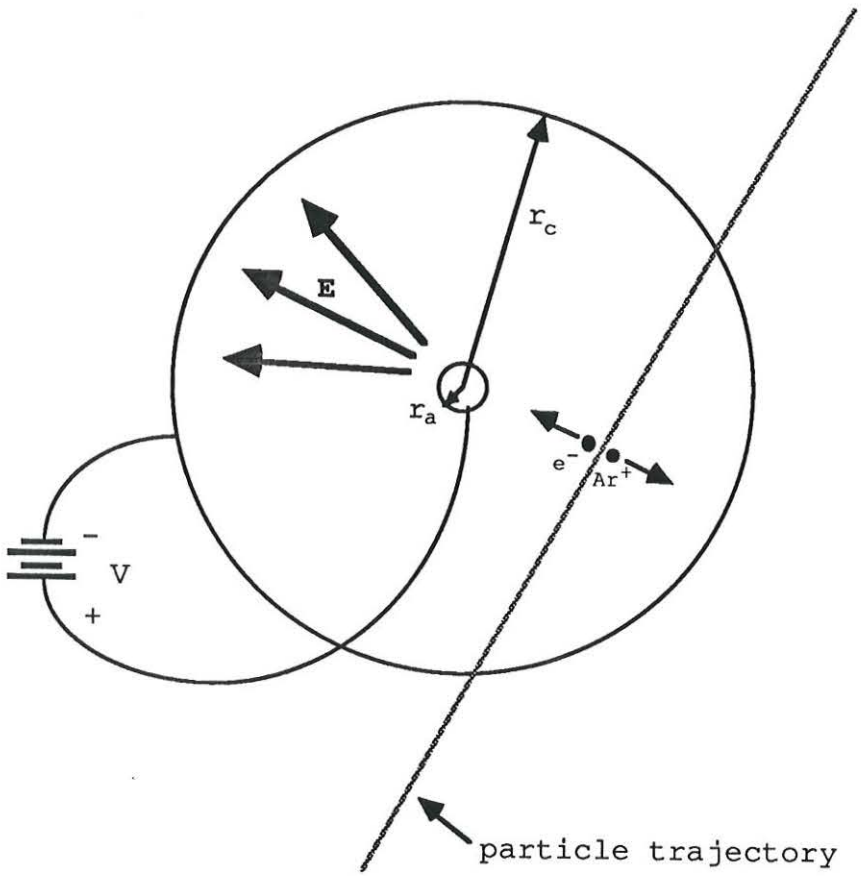


Figure 1.4. The geometry of a typical proportional cell. A typical particle trajectory is shown by the dotted line.

$$v_r(r) = \mu E_r = \mu E_a \frac{r_a}{r} \tag{1.2}$$

where μ is the mobility. Using $v = \frac{dr}{dt}$, we can integrate this and get

$$r^2 = r_a^2 + 2\mu E_a r_a t \tag{1.3}$$

Also, relating the power in the wire with the power in the moving charge,

$$Vi(r) = q\mathbf{E}\cdot\mathbf{v} = qV\frac{\mu E_a}{\ln\frac{r_c}{r_a}}\frac{r_a}{r^2} \tag{1.4}$$

Substituting (1.3) into (1.4), the result is a current pulse on the wire, having the form

$$i(t) = \frac{i_m}{1+\frac{t}{t_o}} \tag{1.5}$$

where $i_m = \frac{q}{2t_o \ln\frac{r_c}{r_a}}$ and $t_o = \frac{r_a}{2\mu E_a}$. The values of i_m and t_o are on the order of microamps and nanoseconds, respectively. The goal of the track chamber electronics is to measure the time for electrons to drift from the particle trajectory to the sense wire. Because the emerging particles move at light speed, they traverse the chamber almost instantaneously. Thus, by measuring the time interval between the collision and the small current pulse on the sense wire, we can determine the particle position. With information from all of the sense wires in the chamber, the path of the particle can be deduced (Figure 1.5). Since the drift velocity is on the order of 50 μm/ns, a measurement of the drift time to resolution of 0.5 ns allows a localization of the debris path to an accuracy of 25 μm.

1.2.2 Associated Electronics The electronics for the wire track chamber are divided into two main sections (Figure 1.6), a pulse detector and a time-to-voltage converter. The pulse detector consists of a low-noise preamplifier and a fast pulse-shaping amplifier followed by a comparator. The pulse detector converts the tiny current pulse

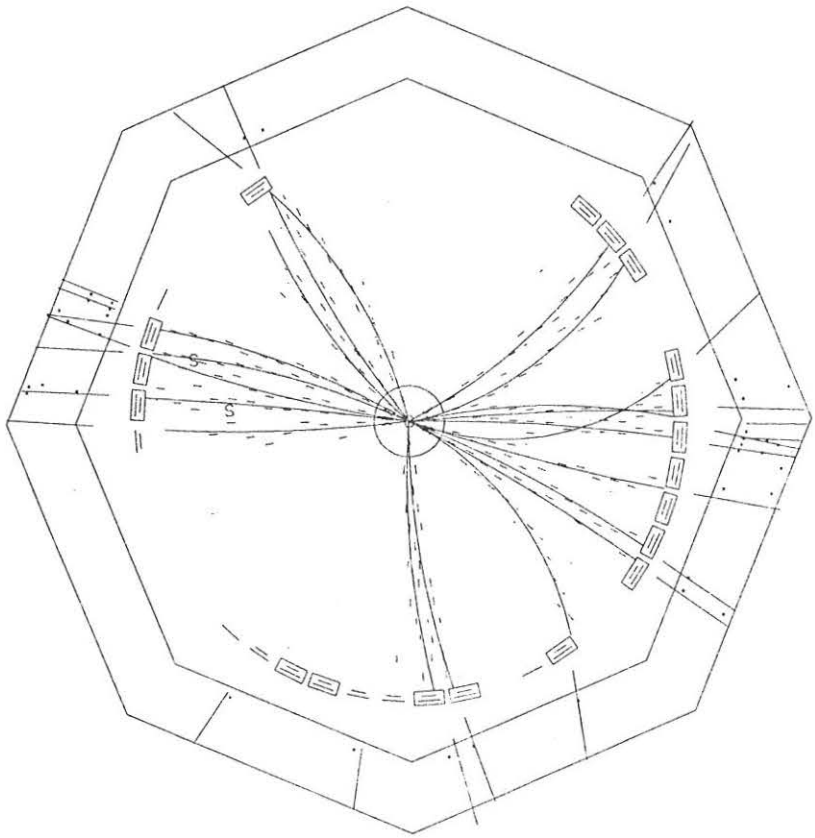


Figure 1.5. Recorded tracks for an event at the Mark II detector. Each tiny dashed line represents a proportional cell of 10 wires. Each solid line represents a particle emerging from the collision. [3]

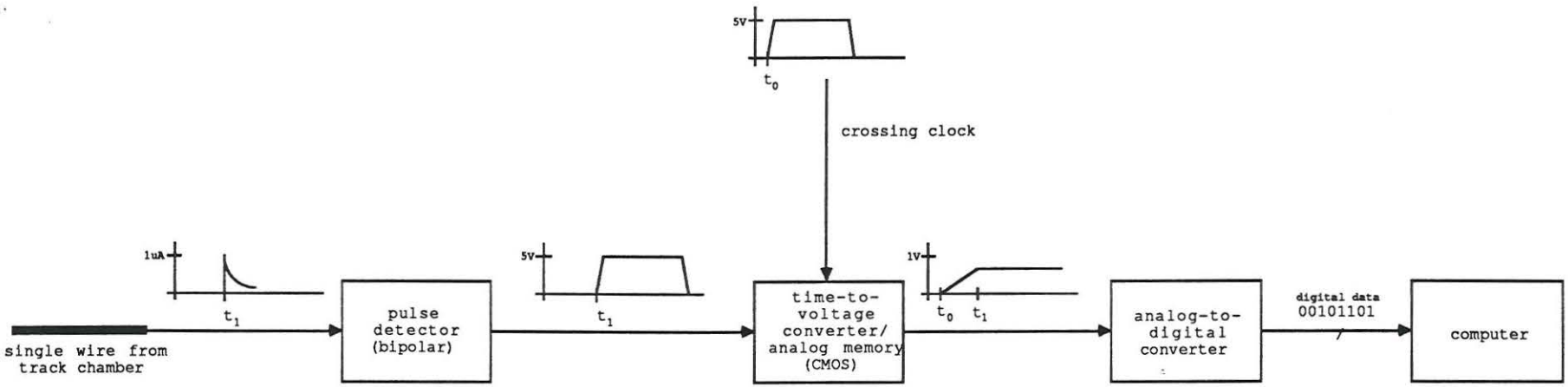


Figure 1.6. Electronics system for a wire track chamber.

on the wire into a fast digital-electronics-compatible signal with a rise time on the order of 1 ns. The time-to-voltage converter then measures the time between this digital signal and a collision-clock signal. This information is stored in the analog memory for later digitization and acquisition by a computer.

1.2.3 Motivation for Current Work Substantial work has already taken place to design and implement a wide variety of wire track chamber readout systems [3]. However, most of these systems are designed using old technologies employing discrete transistors. They also consume as much as one watt of power per wire. The proposed Superconducting Super Collider (SSC) will have a detector with a track chamber containing as many as 200,000 sense wires [4]. Clearly, a low-power, highly integrated design is desirable. Although work is currently in progress to meet these needs for other types of particle detectors [5-8], little or no effort has been made towards a monolithic system for drift chambers.

The design presented here is the latter portion of a wire track chamber electronics system. At the same time, a preamplifier/shaping-amplifier/comparator chip is currently under joint development at the University of Pennsylvania Department of Physics and Department of Electrical Engineering. It is intended that these two designs will form a state-of-the-art detector system for the SSC.

1.3 System Design

Because of the high frequency of events in the SSC ($\sim$ 100 MHz), the detector electronics must operate at exceedingly high speeds. Thus, the TVC must have a very fine resolution, and must have a very fast recovery time between events. The proposed scheme for the design of the TVC is shown in Figure 1.7. Referring to the figure, a flip flop accepts pulses from the crossing clock (*START*) and the pulse detector (*STOP*). The

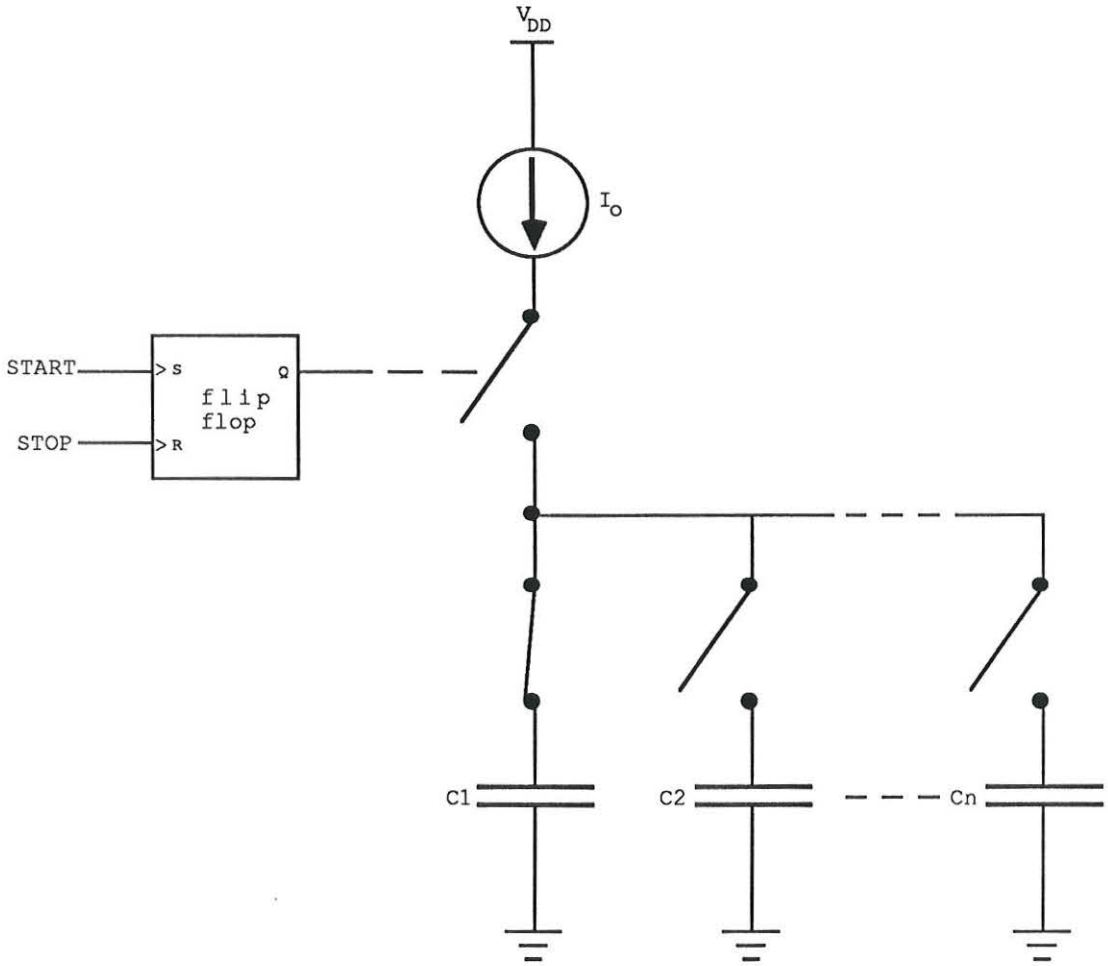


Figure 1.7. Top-level design of the Time-to-Voltage Converter/Analog Memory. n switches select one of the n capacitors, and one additional switch controls the current source.

resulting Q output is "on" for the length of time between the two rising edges of the input pulses. This signal is used to switch a current source which begins to charge a capacitor. Using the constituent relation for current through a capacitor $i_c = C\frac{dv_c}{dt}$, the final voltage on the capacitor will be

$$v_c = \frac{I_o}{C}t \tag{1.6}$$

Because of the relatively low speed of any data-acquisition system compared to the event rate, the data on the capacitor must be stored for later readout. Thus, the integrating capacitor also becomes a hold capacitor, and a bank of capacitors serves as the analog memory unit. In this method, data can be taken continuously, limited by the speed of the switching, and readout can occur at any time, limited by the storage depth of the analog memory.

1.4 Silicon Design

A TVC has been implemented with an analog memory size of 8 capacitors. A 1.6 micron CMOS double-metal digital process was used from the MOSIS System (Hewlett-Packard did the actual fabrication). A time resolution of .5 nanoseconds and a linear measurement range of 5–20 ns were the design goals. Special layout techniques were employed to ensure good matching between the 8 channels.

Test results of the prototype silicon were encouraging. A TVC resolution of 0.2 ns was achieved, with a linear measurement range of 7 to 25 ns. The analog memory achieved hold times of over 1 ms with negligible leakage on the hold capacitors at room temperature. Matching of linearity between channels had $<1\%$ error for adjacent channels and $<2\%$ error for different packages. However, DC offsets and chip-to-chip variations caused approximately 3 ns variations between different memory channels. Complete test results are related in Chapter 5.

2. Design of the TVC

2.1 Current Switching

The basic elements of the time-to-voltage converter are a current source, a capacitor, and two switches (Figure 2.1). Switch $S1$ controls the integration of the current I_o on the capacitor C, while $S2$ resets the capacitor voltage. Typical operation would initially close and open $S2$ to reset the capacitor (ϕ_{RESET}), and then close $S1$ for the time interval to be measured (ϕ_{TIME}). Nominal values for I_o and C are 50 μA and 1 pF respectively. These values were chosen to satisfy power requirements, output swings, etc.

2.1.1 Evolution of the Basic Design An obvious implementation would be to use MOS transistors for the switches. (Figure 2.2). However, this topology allows the current source to be connected to a floating node when $S1$ is off. This would allow the small parasitic capacitance at the floating node to be charged, resulting in a large voltage swing that would send the current source and switch out of their proper operating regions.

Thus, we must shunt the current somewhere while the capacitor is in its "hold" state (i.e. $S1$ off). This can be accomplished by adding a third switch ($S3$) which switches complementary to $S1$ and provides an escape path to ground for the current (Figure 2.3a). However, this introduces another problem; the clocking scheme for the two transistors $S1$

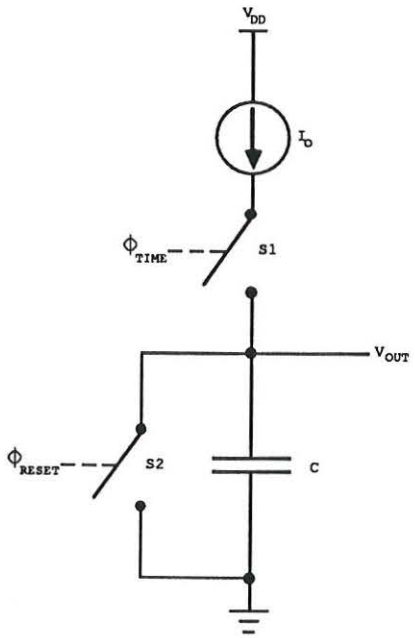


Figure 2.1. Basic TVC topology. Switch $S1$ controls the integration of the current I_o on the capacitor C. Switch $S2$ resets the capacitor voltage.

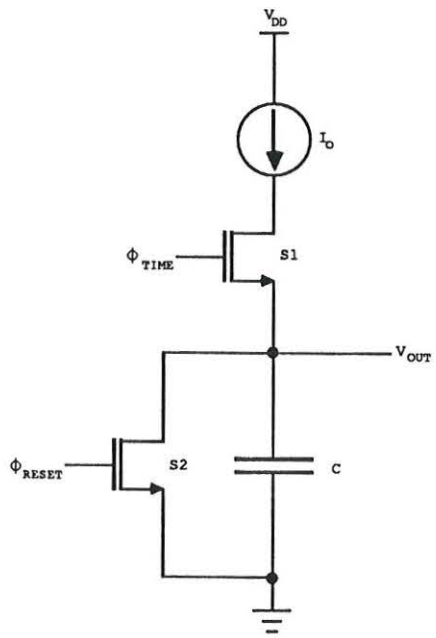


Figure 2.2. TVC implemented with MOS analog switching elements. This circuit will perform poorly because the current source will be connected to a floating node when $S1$ is off.

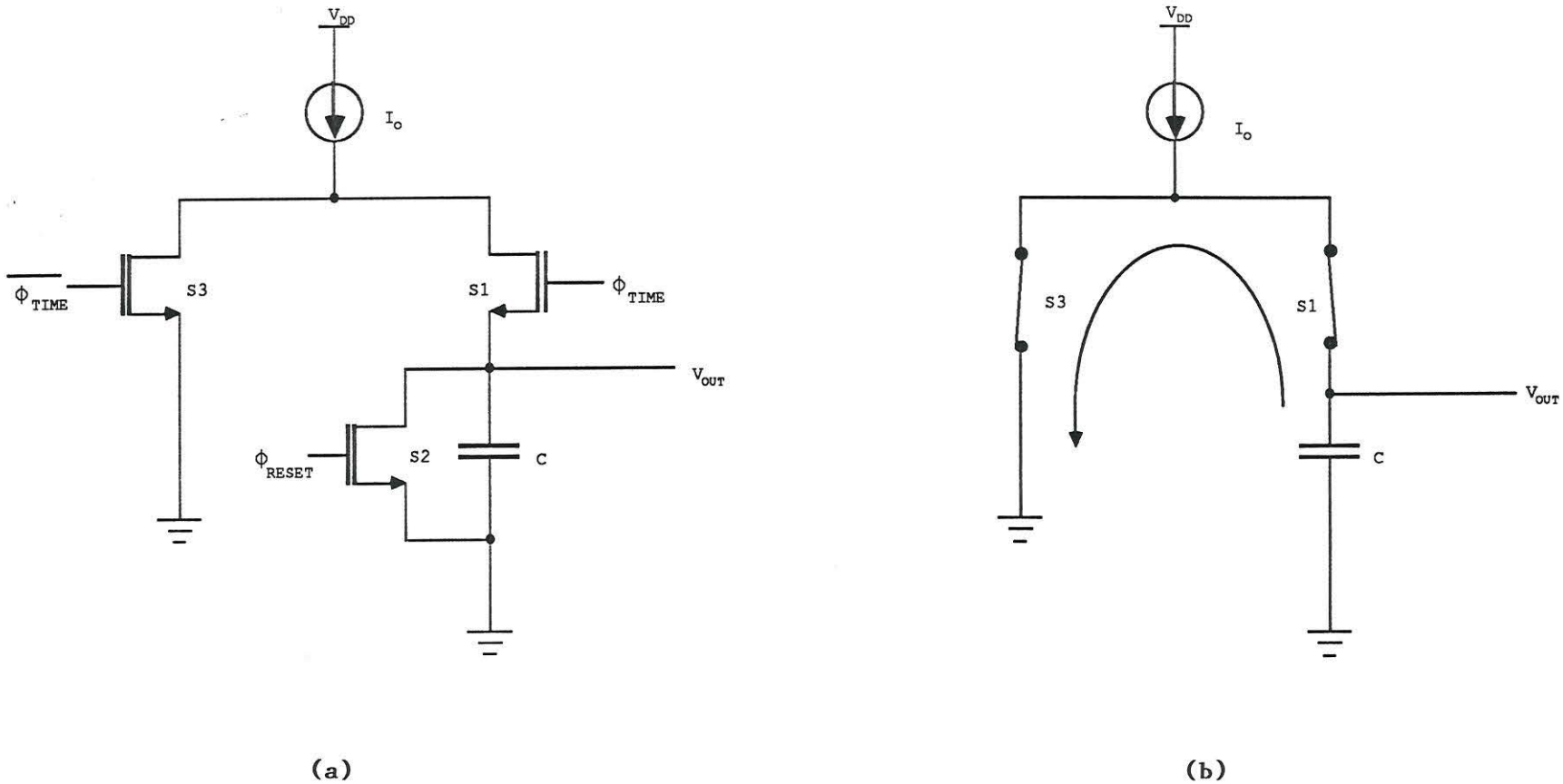


Figure 2.3. (a) $S3$ is added to provide an escape path for the current when $S1$ is off. (b) If $S1$ and $S3$ are closed simultaneously, charge will leak from the capacitor.

and $S3$ must be carefully controlled to guard against undesirable states. For example, if $S1$ and $S3$ are both off for an instant during normal switching, then the current source will be connected to a floating node. If $S1$ and $S3$ are both switched on for an instant, a direct path will exist from V_{OUT} to ground, which will bleed charge from the capacitor (Figure 2.3b). This would destroy any hold capability of the circuit. Because it is difficult if not impossible to ensure the precise simultaneous switching of $S1$ and $S3$, this method is unfeasible.

An alternate topology which is similar to that of Figure 2.3 is to operate $S1$ and $S3$ as MOS transistors in the saturation region rather than the linear region (Figure 2.4). This method has been used previously in implementations of single- and dual-slope analog-to-digital converters [9-12]. As shown in the figure, by using p-channel rather than n-channel transistors for $S1$ and $S3$, we can ensure that the transistors will operate in the saturated region because initially for both transistors, $\left|V_{DS}\right| > \left|V_{GS}-V_T\right|$. To emphasize the analog behavior of this circuit, the two switches $S1$ and $S3$ are now relabeled as MOS transistors $M2$ and $M1$ in the figure. The reset switch $S2$ remains unchanged, and is now omitted from the diagrams for clarity.

The actual operation of the circuit is shown in Figure 2.5. Initially, $M2$ is off, $M1$ is on, and the capacitor voltage V_c is reset to zero. Because $M1$ is saturated, it can be modeled as a dependent current source. In order to start the time measurement, $M2$ is turned on *before* $M1$ is turned off, and the two transistors share the current equally. This eliminates any leakage path to ground during switching as in Figure 2.3b. Then, $M1$ turns off and the capacitor charges at a uniform rate. At the end of the time interval, the process reverses.

A caveat of this circuit is the fact that the capacitor charges non-linearly while the

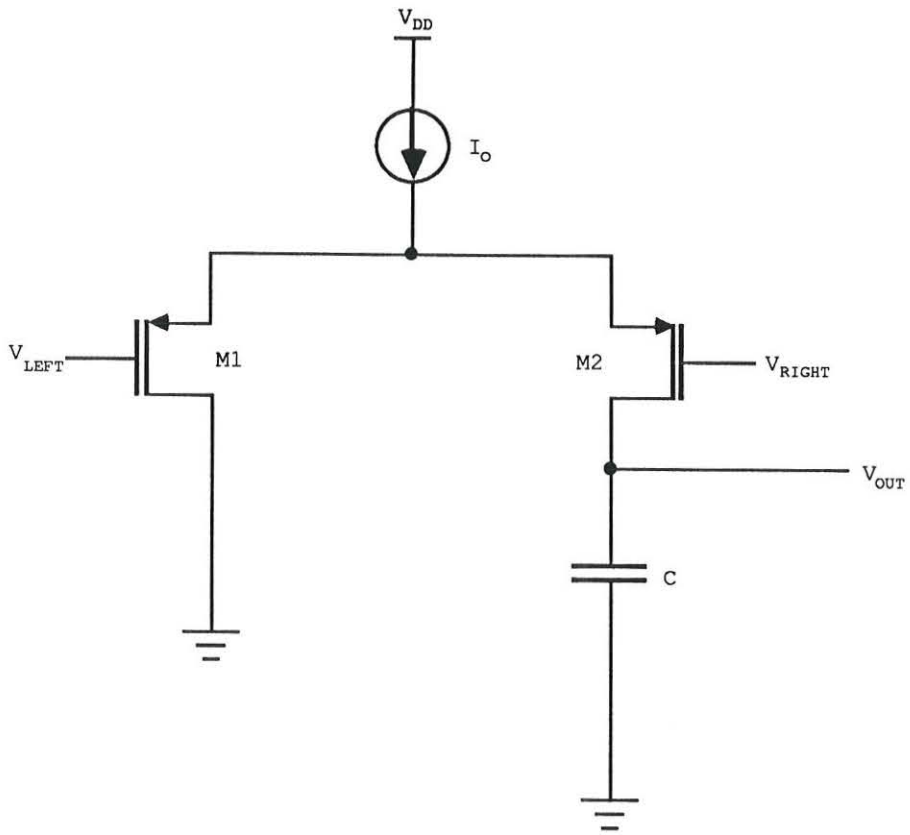


Figure 2.4. Current switching with transistors in saturation. This topology removes the leakage path to ground when both transistors are on.

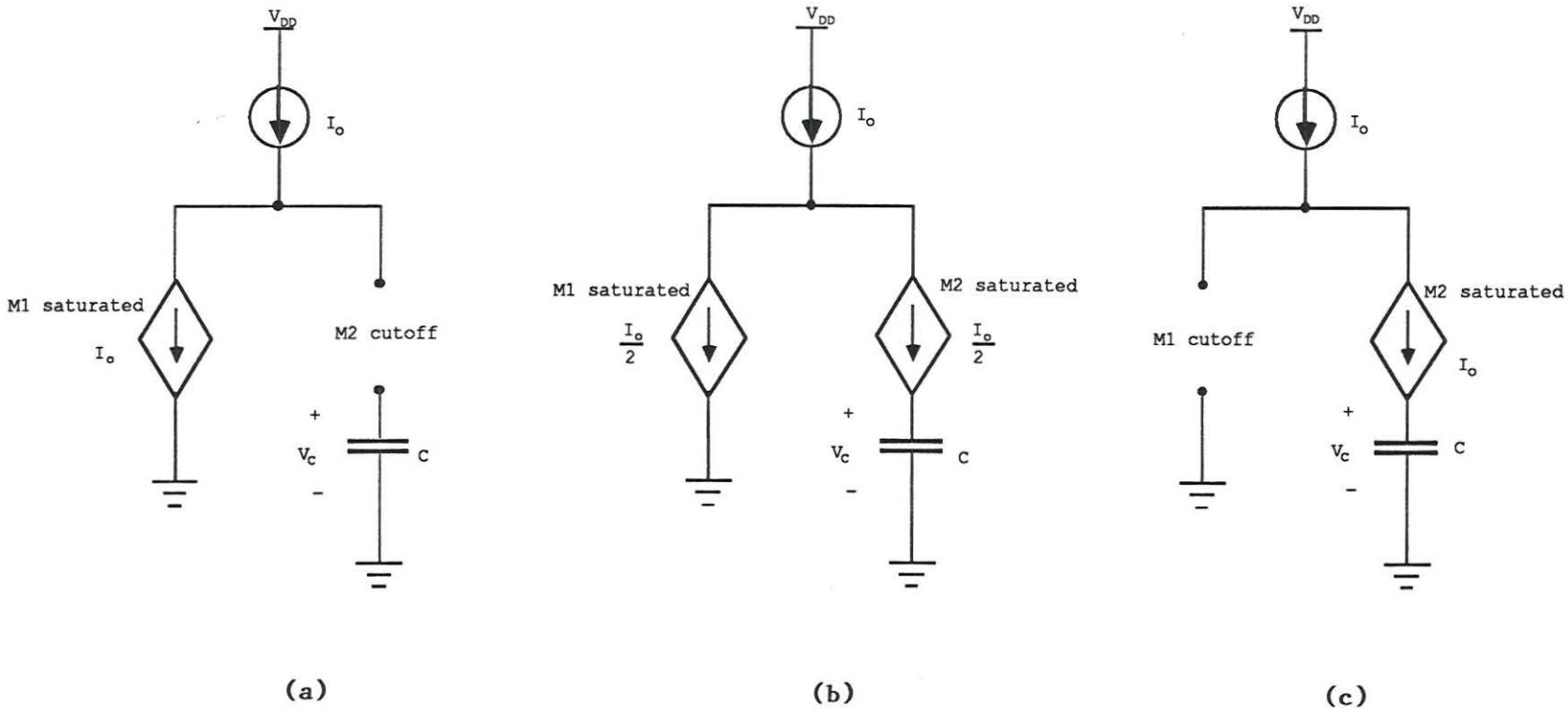


Figure 2.5. Idealized mechanism for saturation-region switching. (a) Initially, $M1$ is on and $M2$ is cutoff. (b) During switching, $M1$ and $M2$ will share the current equally. (c) After switching, $M2$ will carry all the current and capacitor charging will commence.

switching takes place (both transistors on) because the capacitor current is not constant. However, this effect should not be significant; clearly, some amount of non-linear charging will always take place due to the impossibility of moving the current instantaneously. By programming the proper switching sequence to occur very quickly (on the order of 0.5 ns), a minimal residual error voltage will result. This will be discussed in detail in section 4.1.

At this point, the entire circuit is inverted in order to push the limits of the technology as far as possible (Figure 2.6). Thus, all p-devices are changed to n-devices, and vice-versa. Since the carrier mobility for n-channel transistors is approximately three times larger than for p-channels, switching can take place faster when n-channels are used for $M1$ and $M2$.

2.1.2 Measurement Range and Output Swing The total measurement range of the TVC will depend on the maximum output swing of the circuit. Specifically, if $M2$ is left "on" for too long, the capacitor will get charged too much and send $M2$ out of saturation. The requirement on $M2$ is $V_{DS2} > V_{DSSAT2} = V_{GS2} - V_{T2}$. When $M2$ initially turns on, the capacitor voltage is zero and $V_{DS2} = V_{GS2}$. However, as the capacitor charges, V_{DS2} will drop toward V_{DSSAT2}; when V_{DS2} drops low enough, $M2$ will become linear. This means that the maximum capacitor voltage $V_{cmax} = V_{T2}$, and this relation will set the maximum output swing. Using Equation 1.6,

$$t_{\max} = \frac{C}{I_o} V_{T2} \tag{2.1}$$

For typical values of $C = 1$ pF, $I_o = 50$ μA, and threshold voltage $V_{T2} = 1$ V, $t_{\max} = 20$ ns.

2.2 Gate Clocks

As just described, an overlapping clock scheme is necessary to drive the transistor gates in the TVC. Two methods are considered, (1) single-ended clocking (Figure 2.7),

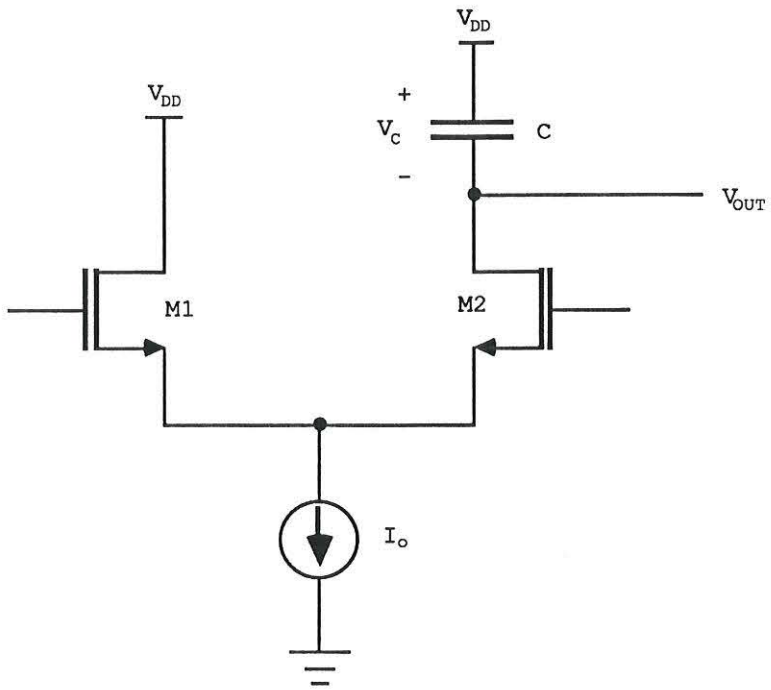
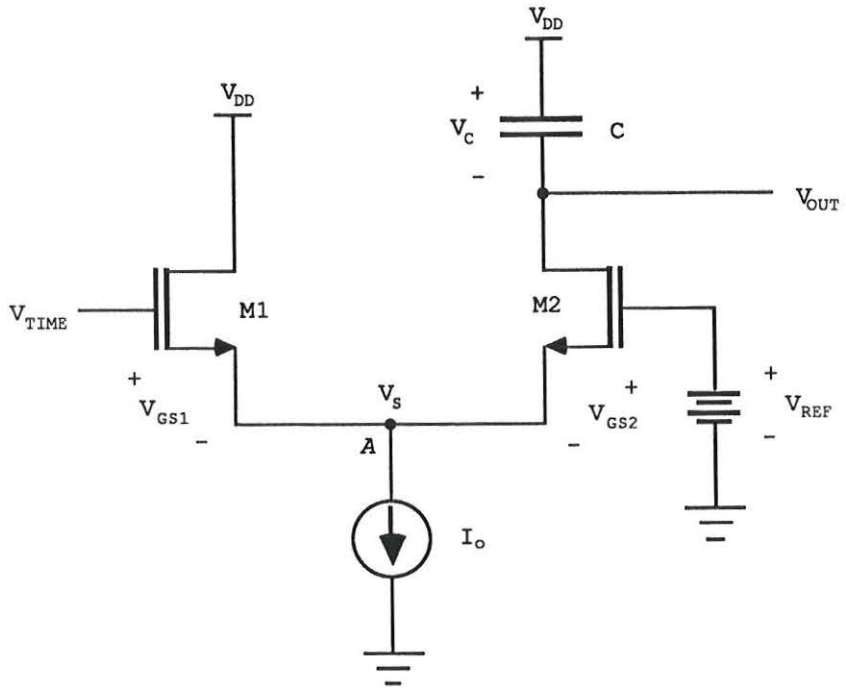


Figure 2.6. N-channel version of the TVC. This version should operate faster because the n-channel transistors are inherently faster devices.

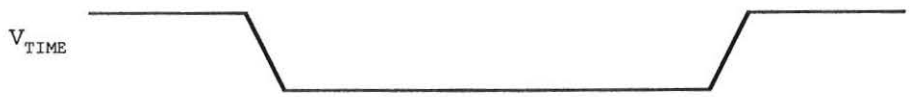
and (2) differential clocking (Figure 2.8). In single-ended clocking, the gate voltage at $M2$ is held constant while $M1$ is turned off and on for the desired time interval. In differential clocking, both gate signals are varied in a complementary fashion.

2.2.1 Single-ended Clocking This scheme is intended to resemble a bipolar ECL logic gate. The gate voltage at $M1$ (V_{TIME}) is a digital signal which is low for the time interval to be measured. V_{REF} is an intermediate voltage on the order of $\frac{V_{DD}}{2}$. Operation of the circuit is straightforward: initially, V_{TIME} is high, $M1$ is on and $M2$ is off. Since $M1$ is saturated at a current of $I_{D1}=I_o$, an imposed V_{GS1} will exist, creating a nominal value for the source voltage V_S (at node A in the figure). With proper device sizing and choice of V_{REF}, this voltage will be high enough to keep the current source in saturation and $M2$ off. At the start of the time measurement, V_{TIME} starts to fall; a corresponding fall will be seen in V_S in order to keep I_{D1} constant. Eventually, V_S will fall enough so that $V_{REF}-V_S > V_{T2}$, and $M2$ will turn on saturated. As V_{TIME} continues to fall, $M1$ will turn off and $M2$ will transmit the current to the capacitor.

The single-ended clocking scheme depends heavily on the voltage swing at node A. If V_{TIME} switches rail-to-rail (ground to V_{DD}), the size of $M1$ and $M2$ will determine the swing at A because W and L will determine V_{GS} for a fixed I_D. Ideally, we would like V_S to change approximately 0.5 to 1 volt to ensure a proper noise margin and to avoid subthreshold currents. A problem arises in the form of a parasitic capacitance between node A and ground. This capacitance is caused by the source-bulk diffusions of $M1$ and $M2$, the drain-bulk diffusion of the current source, and interconnect capacitance. As will be explained later, this capacitance will be on the order of 0.1 pF, and would cause any voltage change at A to be slow. This in turn would cause the wrong V_{GS} to be imposed on the transistors, which would then transmit the wrong amount of current. Since we



(a)



(b)

Figure 2.7. Single-ended clocking. This circuit will perform poorly if the parasitic capacitance at node A is too large. (a) Schematic. (b) Gate timing.

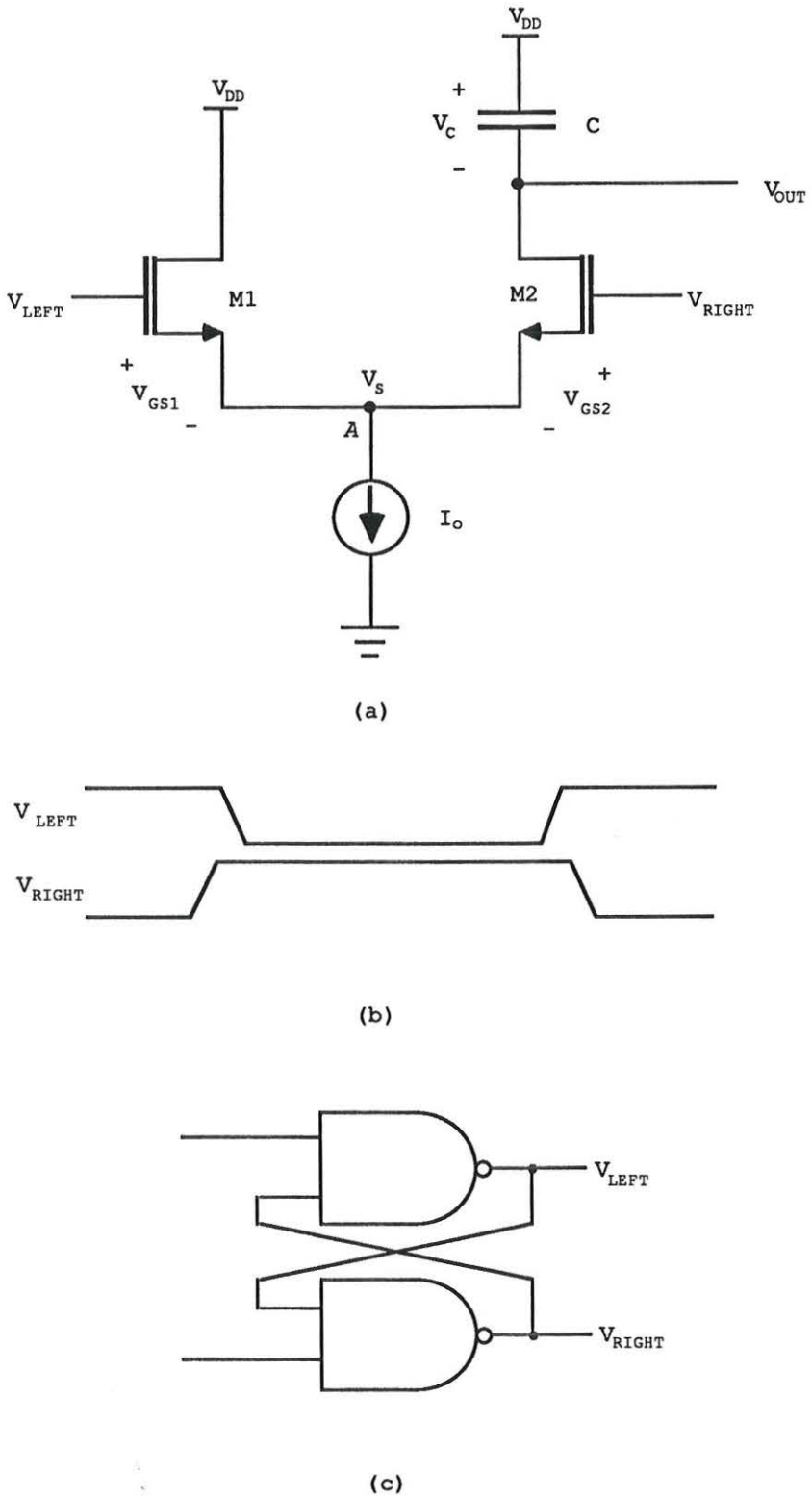


Figure 2.8. Differential clocking. This clocking scheme performs well due to the minimal variation in V_S. (a) Schematic. (b) Gate timing. (c) Gate voltage generator.

require subnanosecond switching times, single-ended clocking must be discarded.

2.2.2 Differential Clocking As shown in Figure 2.8a, differential clocking varies both gate voltages, with the timing shown in Figure 2.8b. This timing ensures that there will never be a state where both transistors are off. In addition, due to the symmetric nature of the circuit, the voltage at node A will never vary far from its nominal value, and clean current switching will occur.

Generation of the overlapping clocks is accomplished by using the outputs of a cross-coupled NAND pair (Figure 2.8c). This simple logic circuit automatically produces the desired timing as long as the loads are identical at each output. However, it is unclear whether the loading can be matched well enough to ensure that the proper sequence takes place. This issue can be minimized by designing the gates to drive an extra-large load, thereby making minor load variations tolerable.

2.3 Parasitic Effects in the TVC

Parasitic effects are introduced due to various capacitances in the circuit (Figure 2.9). These will affect performance in one of two ways, either by (1) injecting charge onto the capacitor or (2) varying the capacitor current. Intrinsic parasitics include the gate-source and gate-drain overlap capacitances (c_{gs} and c_{gd}), and the channel charge (Q_C). Extrinsic parasitics include the source-substrate and drain-substrate junction capacitances and the interconnect capacitance. The drain-substrate capacitance c_{db2} of $M2$ will add a small non-linear capacitor to the hold capacitor. The source-substrate junctions of $M1$ and $M2$, the drain-substrate junction of the current source, and the interconnect capacitance between $M1$ and $M2$ are summed together and are represented by the single capacitor C_P.

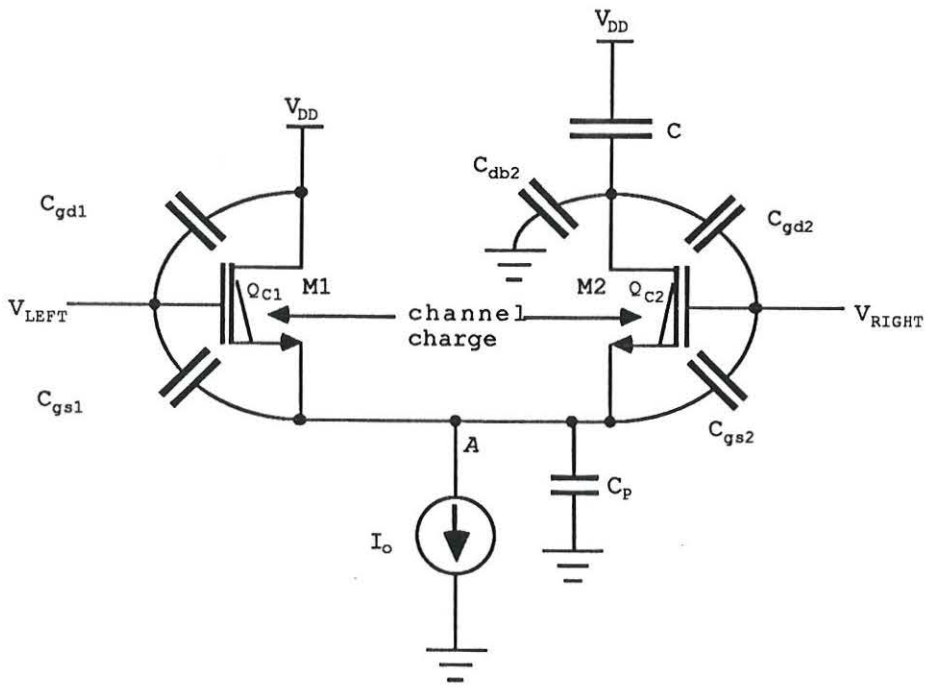


Figure 2.9. Parasitic elements in the TVC include the gate-drain and gate-source overlap capacitances, the drain-bulk and source-bulk junction capacitances, the interconnection capacitance, and the channel charge.

2.3.1 Charge Injection Effects Charge injection occurs due to two effects: (1) capacitive coupling from the gate clock, and (2) channel charge dissipation during transistor turn-off. In the first case, charge is coupled through c_{gd2} to the hold capacitor C. This charge will appear as a rise-time-dependent current of magnitude:

$$i_{feedthrough} = \frac{c_{gd2}C}{c_{gd2}+C}\frac{dV_{G2}}{dt} \approx c_{gd2}\frac{dV_{G2}}{dt} \tag{2.2}$$

For typical numbers of $c_{gd2} = 2$ fF and $\frac{dV_{G1}}{dt} = 5$ V/ns, the resulting $i_{feedthrough} = 10$ μA. This is not negligible for an I_o of 50 μA, and the resulting error voltage will be 10 mV. However, the turn-on charge injection from the rising clock edge will get canceled out by the subsequent falling clock edge at turn-off, making the net effect equal to zero. This point will be discussed in later SPICE simulations. It should be noted that the total error due to capacitive coupling will be independent of rise time because the total coupled *charge* will be the integral of the feedthrough current.

The second charge injection effect is due to the channel charge. This effect has been discussed widely in the literature [13-16]. However, these papers deal with MOS switches operating in the linear region with $V_{DS} \approx 0$. This is not really applicable to our topology of saturated transistors. Thus, an account of the movement of charge in a saturated switching transistor will be described here.

In our case, we are interested in the effect of excess charge from a saturated transistor ($M2$) onto a hold capacitor (Figure 2.10). As previously described, at the moment just before turn-off, a current of 50 μA flows through $M2$ and C. The channel of $M2$ is pinched off and negative charge is swept through the depletion region towards the drain due to the applied electric field. An approximation of the total channel charge is

$$\left|Q_C\right| = \frac{2}{3}WLC_{OX}(V_{GS}-V_T). \tag{2.3}$$

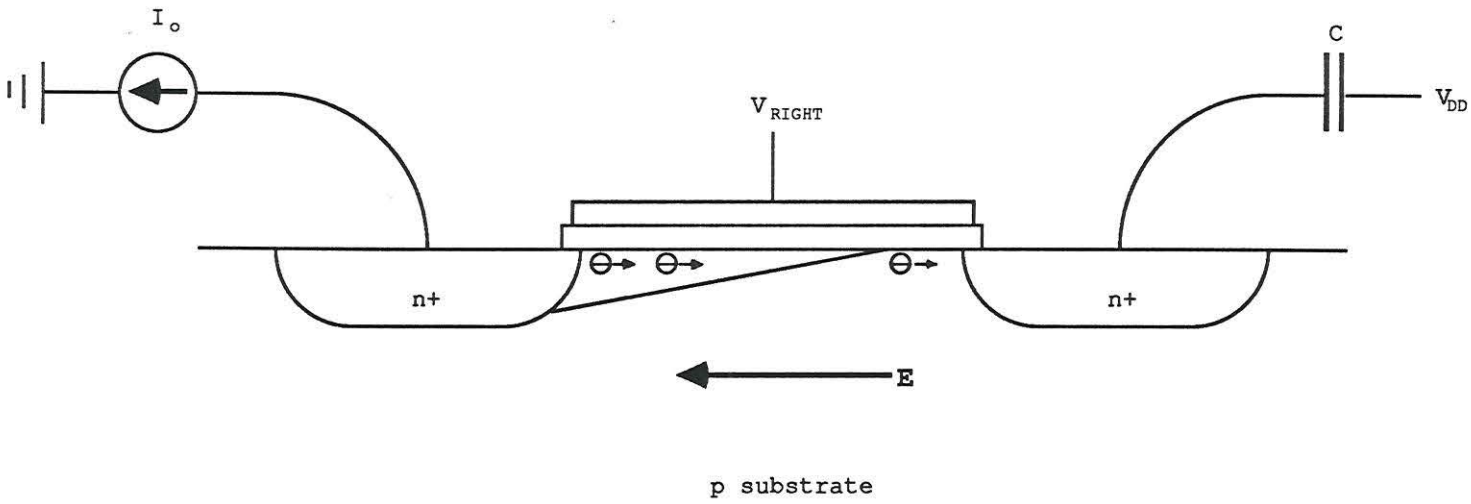


Figure 2.10. A look at the MOS transistor action as the capacitor is charged. Negative carriers will flow through the channel and then get swept across the depletion region by the electric field. As the gate voltage is decreased, the extra channel charge will drift into the drain.

For typical values of channel length $L = 1.6$ μm, channel width $W = 4.8$ μm, $V_{GS} = 2$ V, $V_T = 1$ V, and gate-oxide capacitance $C_{OX} = 1.4\times10^{-7}$ F/cm^2, $Q_C = 8$ femtocoulombs.

For an MOS transistor with the above characteristics, the carrier transit time

$$\tau_o = \frac{WLC_{OX}}{g_m} \tag{2.4}$$

is on the order of 100 ps. As the device is turned off, V_{GS2} falls slowly with respect to the transit time ($t_{fall} \approx 1$ ns), and the source current I_{S2} will also fall. However, in order for the drain current I_{D2} to decrease, extra channel charge must first drift into the drain. In this case, because the transit time is so short, this excess charge exits almost instantaneously, or at a fraction of the transit time. Thus, the net effect of the channel charge at this instant is to delay turn-off by a few tens of picoseconds. As the gate voltage continues to fall, the transit time of the transistor will rise (substituting for the transconductance $g_m = \mu_n C_{OX}\frac{W}{L}(V_{GS}-V_T)$ in (2.3), $\tau_o = \frac{L^2}{\mu_n(V_{GS}-V_T)}$, where μ_n is the mobility of the channel charge) and eventually reach the same order as the gate fall time. For example, in order for the transit time to be 1 ns, $V_{GS} = 1.3$ V, or 0.3 volts above turn-off. At this point, the channel charge will take a long time to drift out the drain, but will also be reduced to about 3 femtocoulombs. It is this charge which will cause a visible "dump" onto the hold capacitor. However, in reality, the entire 8 femtocoulombs channel charge will have been dumped onto the capacitor; the other 5 femtocoulombs will get there during the tens-of-picoseconds delay between V_{GS2} and I_{D2}.

Summarizing, the entire quiescent channel charge will end up on the hold capacitor. However, only a fraction of the charge will be an observable "charge feedthrough"; For most of the transition, the gate voltage is falling relatively slowly and the channel charge will decrease at the same rate. Only during the final part of turn-off, when $V_{GS} \approx V_T + 0.3$ does the transit time become long enough such that there might be

charge trapped in the channel after $V_{GS} < V_T$. This total charge, on the order of 3 femtocoulombs, will cause a visible dump of about 3 mV on the hold capacitor; however, the total error will be 8 mV due to the delay in the change of I_{D2}.

Circuit simulators such as SPICE do not accurately model the channel charge [13,15], but rather use an arbitrary ratio to do channel charge calculations (e.g. the channel charge splits 60-40 between the source and drain). However, by making the magnitude of the channel charge as small as possible and the gate rise time as fast as possible, we minimize any effect of the channel charge to an immeasurable level. This is accomplished by making $M1$ and $M2$ as small as possible in order to minimize the transit time and maximize the drive capability at the gate.

2.3.2 Current Varying Effects As previously mentioned, variation of the voltage V_S at the common source node (node A) will change the V_{GS} of $M1$ and $M2$. Some variation of this voltage is inevitable; the quiescent value is set by the imposed V_{GS} which exists when either $M1$ or $M2$ is on with a current of 50 μA. However, during switching, both transistors will be fully on (with gate voltages $V_{G1} = V_{G2} = V_{DD}$) and the imposed V_{GS} will be smaller due to the smaller current through each device. For a typical g_m of 0.1 mmho, a current change of 25 μA translates into a 250 mV change at A. If C_P is too big, then the voltage cannot change easily, and the transistor currents will take a long time to settle to their correct values. This will result in non-linear charging of the hold capacitor. The effect of C_P will be made evident in the SPICE simulations.

Another effect of varying the voltage at A is a possible change in the current through the current source. If the output resistance of the current source is too low, the voltage shift will vary the current from its nominal value of I_o. For example, for a current source with an output resistance of 100 kΩ, a 250 mV change in voltage will cause the current to

change by 2.5 μA. Also, the current source must have a good frequency response and settling time in order to return to its quiescent value after switching transients. This means that devices must be small in order to minimize parasitics. However, small transistors make it difficult to achieve high output resistance. Clearly, a tradeoff is necessary. This will be discussed shortly.

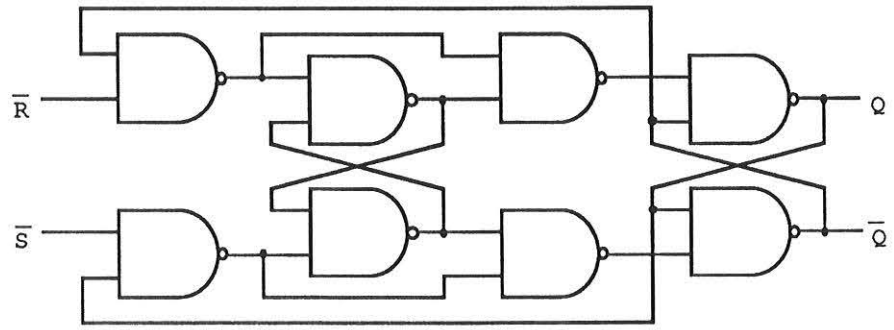
2.4 Flip-Flop Design

A flip-flop is necessary to translate the *START* and *STOP* signals into voltages to drive the gates of the TVC transistors. This function could be accomplished by a simple RS latch, i.e. a cross-coupled NAND pair. However, a cross-coupled NAND configuration produces an undefined output when both inputs are low; this places an unreasonable restraint on the input waveforms by requiring that both signals never be asserted simultaneously. An alternate configuration, albeit more complicated, is the edge-triggered RS latch shown in Figure 2.11a. This logic circuit will sense the falling *edges* of the two signals and generate Q and $\overline{Q}$ outputs which are "on" for the corresponding time interval (Figure 2.11b). Note that the output stage of the flip-flop is a cross-coupled NAND pair which ensures the properly overlapping signals required to implement differential clocking. This flip-flop design is adapted from the phase detector of a digital phase-locked loop [17].

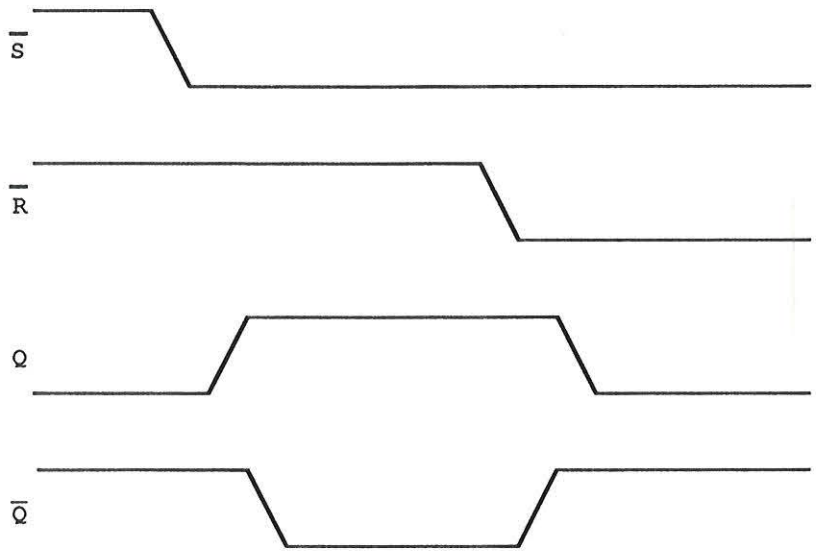
Relevant issues in the design of this flip-flop include the propagation delay, the hold time, and the minimum time interval resolution. Ideally, we would like to minimize all of these quantities; however, most of them will be technology dependent. Thus, the speed of the flip-flop will most likely set the maximum operation speed of the entire design.

2.5 Current Source Design

As previously mentioned, the current source must have a high output resistance and a



(a)



(b)

Figure 2.11. Edge-triggered RS flip-flop. (a) Logic diagram. (b) Signal timing.

good frequency response. Three different types of MOS current sources are considered: (1) simple current mirror, (2) cascode current mirror, and (3) modified cascode current mirror. The three circuits are shown in Figure 2.12 and have been discussed in many textbooks [18], so the details will not be discussed here. In the simple current mirror, the reference current is mirrored from $M1$ to $M2$, and the output resistance of the source will be equal to the output resistance of $M2$. In the cascode current source, the output resistance is increased by approximately an order of magnitude by adding the cascode devices. However, the usable voltage range is decreased by an additional V_T. This problem is fixed by using the modified cascode source, which uses a level shifter (source follower) to increase the voltage swing to that of a simple current mirror.

The frequency response of the current source is defined as the change in the current as a result of a small change in an imposed output voltage. For the simple current mirror, it is easy to derive a small signal model (Figure 2.13a) and subsequently find the frequency response by hand analysis. In this case, the response will be:

$$\frac{i_{OUT}(s)}{v_{OUT}(s)} = \frac{1}{r_{o2}} + c_{gd2}\frac{s^2(c_{gs1}+c_{gs2}) + s(2g_m)}{s(c_{gd2}+c_{gs1}+c_{gs2}) + g_m} \tag{2.5}$$

The small signal model for the modified cascode source is shown in Figure 2.13b. Six node equations are required to solve this circuit, so we resort to computer simulation to find the frequency response. The result of a SPICE simulation of the simple and modified cascode current sources is shown in Figure 2.14. The sources are simulated with typical device sizes and an output current of 50 μA. As is evident in the plot, the simple source will have a higher break frequency than the modified cascode, by a margin of one decade. However, it is also apparent that the modified cascode source will always have a smaller current variation that the simple source due to its higher output resistance.

The advantages and disadvantages of the different current sources are summarized in

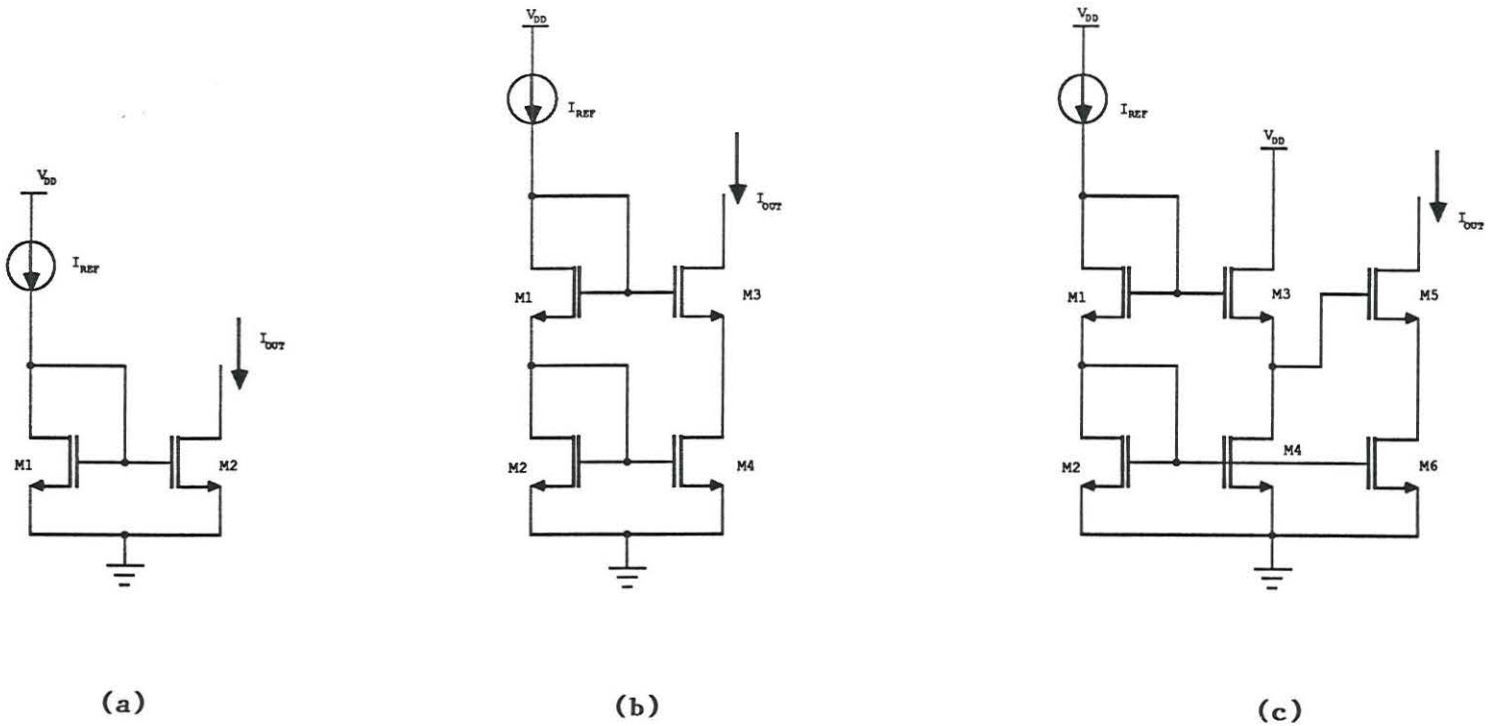
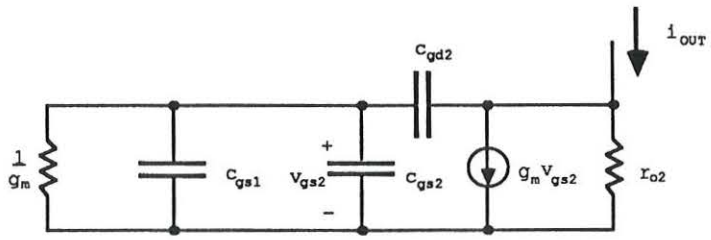
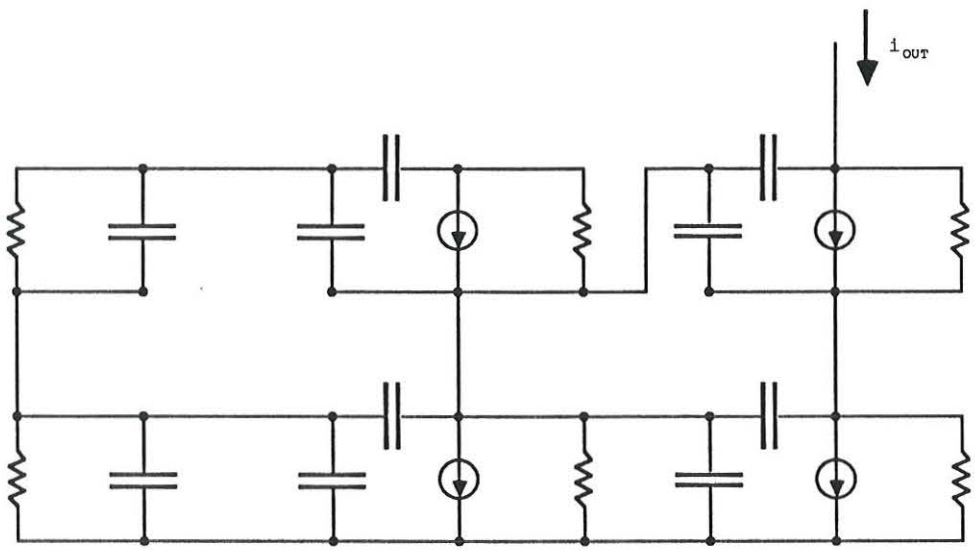


Figure 2.12. MOS current sources. (a) Simple current mirror. (b) Cascode current mirror. (c) Modified cascode current mirror.

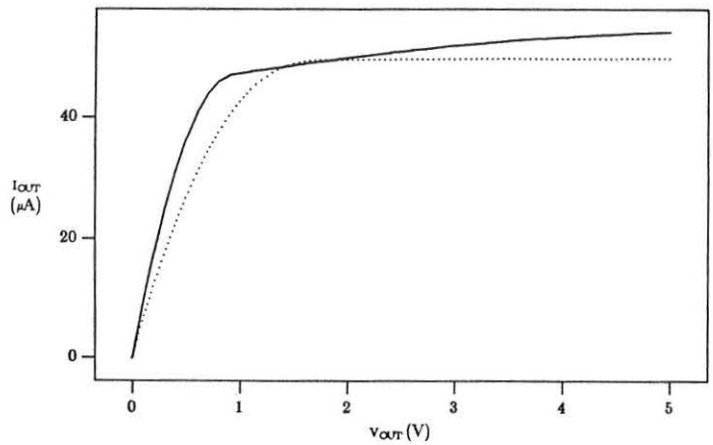


(a)

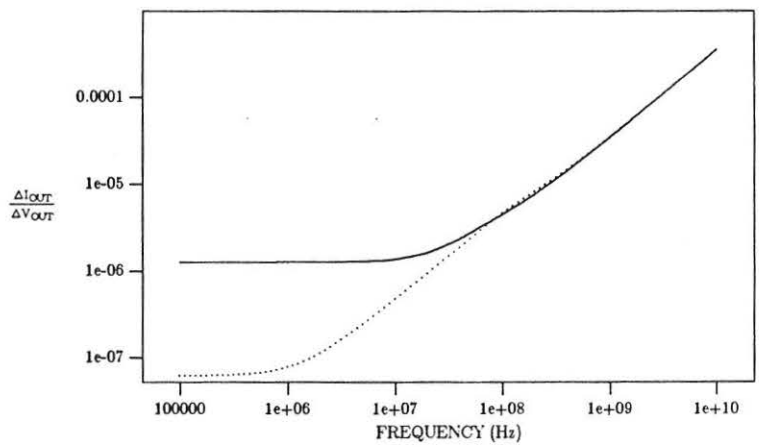


(b)

Figure 2.13. MOS current source small signal models. (a) Simple current mirror. (b) Modified cascode current mirror.



(a)



(b)

Figure 2.14. SPICE simulation comparing simple (*solid line*) and modified cascode (*dotted line*) current sources. (a) DC characteristics. (b) Frequency response.

Table 2.1. For the purposes of the TVC, the cascode source is eliminated because of its poor output swing. Thus, the choice falls between the simple and modified cascode sources. The requirements of the TVC are such that the voltage swing across the current source is small (for differential clocking, $\Delta V_S \approx 250$ mV), and the switching speed is high. Thus, the simple current mirror provides good performance for the lowest price, and was chosen for use in this design. Clearly, however, either source could have been chosen.

2.6 Output Buffer

An output buffer is required in order to perform a non-destructive read on the TVC hold capacitor. Ideally, the buffer should be linear with a high frequency response and a gain of one. This would allow a direct reading of the capacitor voltage for test purposes, allowing observation of charge feedthroughs, etc. Unfortunately, a high frequency response and slew rate are difficult to achieve due to the high current and therefore large transistors required to drive the high capacitance off chip. However, a large amount of information can be surmised as long as the gain of the buffer is uniform, allowing reliable observation of the final voltage on the hold capacitor. Thus, a simple implementation would be a low-power op amp in a unity-gain configuration. However, because of time limitations in the design sequence, a simple source follower was utilized (Figure 2.15).

The source follower consists of an n-channel transistor biased by a current source. Because the output of the TVC ranges from V_{DD} to about $V_{DD}-V_T$, an n-channel design was required. The bias current of 1 mA was arbitrarily chosen to provide sufficient drive off the chip.

In typical CMOS source follower designs, the well of $M1$ is tied to the source in order to cancel the body effect. In our case, this was not done in order to maintain a generic layout; because the type of well was not known during the generation of the layout, gen-

Type	Advantages	Disadvantages
simple current mirror	good frequency response good output swing	low output resistance
cascode current mirror	high output resistance	limited output swing
modified cascode	high output resistance good output swing	more transistors

TABLE 2.1. Summary of MOS current sources.

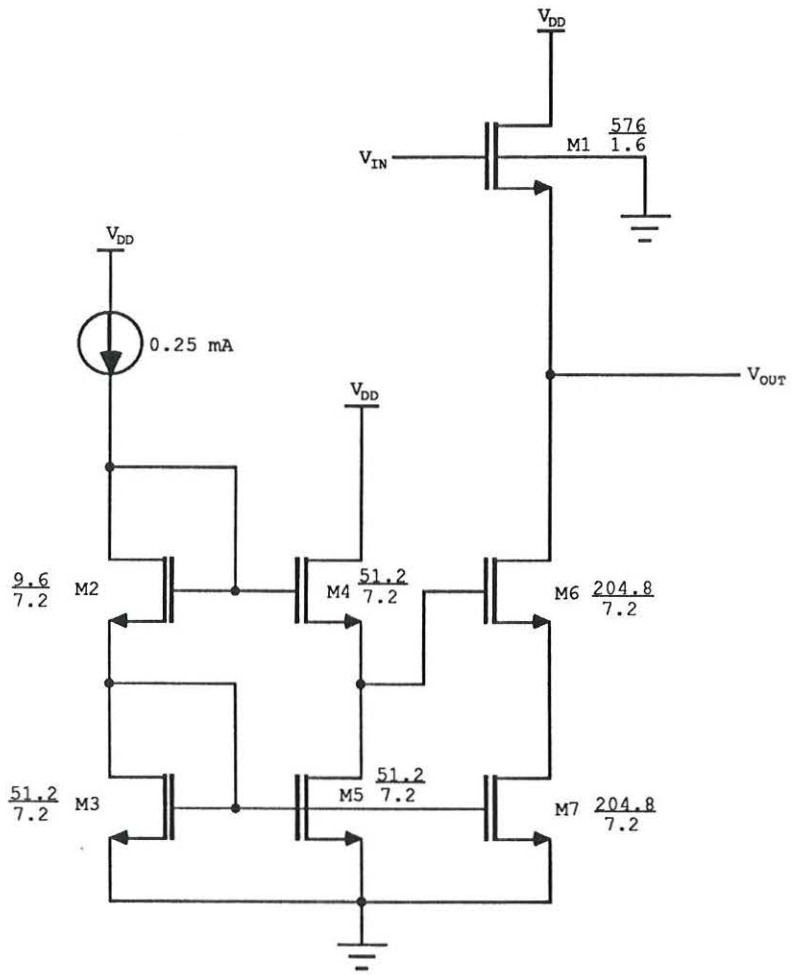


Figure 2.15. Source follower. Note that the bulk of $M1$ was tied to ground because of the n-well process.

eric design rules allowed fabrication of the chip in the next run, regardless of the well type. Thus, the transfer characteristic of the source follower is [18]

$$A_o = \frac{1}{1+\frac{g_{mb}}{g_m}} = \frac{1}{1+\frac{\gamma}{2\sqrt{2\phi_f + V_{SB}}}} \tag{2.6}$$

where g_{mb} is the backgate transconductance and ϕ_f is the Fermi potential of the bulk. Fortunately, the estimated bulk threshold parameter γ for the HP process was relatively small at 0.25 $\sqrt{\text{V}}$. Also, because the output range of the TVC was limited to between 5 and 4 V, the source follower remained linear despite any body effect; the only negative result was a gain of less than one. Assuming $\phi_f = 0.3$ V and $V_{SB} = 3.5 \pm 0.5$ V, the source follower gain $A_o = 0.94$.

The current source for the source follower was a modified cascode current mirror. The most important issue in the current source design was a high DC output resistance; any variations in the current with respect to the output voltage would change the gain and the offset of the source follower. The current source was designed to have a DC output resistance of 1 MΩ, which allowed a maximum current variation of 0.2% for an output swing of 2 V.

3. Design of the Analog Memory

3.1 Extension of the TVC to Include Analog Memory

As specified in Chapter 1, the TVC must have the capability of taking many consecutive time measurements. A straightforward method of increasing the throughput would be to run many parallel TVC's with some logic to choose between the individual circuits. A question arises, however, on exactly which system components must be paralleled. One approach would be to replicate everything, including the flip-flop, current source, differential pair, and capacitor. Alternately, one could repeat the minimum number of components, e.g. use a number of hold capacitors and some analog switching to select any single capacitor. Obviously, area, matching, and power requirements point toward repeating the fewest number of components; the goal then becomes one of determining the easiest method of implementing the selection logic.

This goal is accomplished by simply adding more transistors to the "differential pair" (Figure 3.1. Note the renumbering of the left transistor to $M0$). By performing a simple AND function at the gates of the transistors $M1$ through Mn, only one of the n channels will be active at any time; the rest of the transistors will be in cut-off. The total number of channels is limited by the added capacitance at node A; too many channels will make C_P large and make fast switching impossible. The amount of repeated circuitry is minimal, consisting of a capacitor, a saturated switching transistor, and a logic gate.

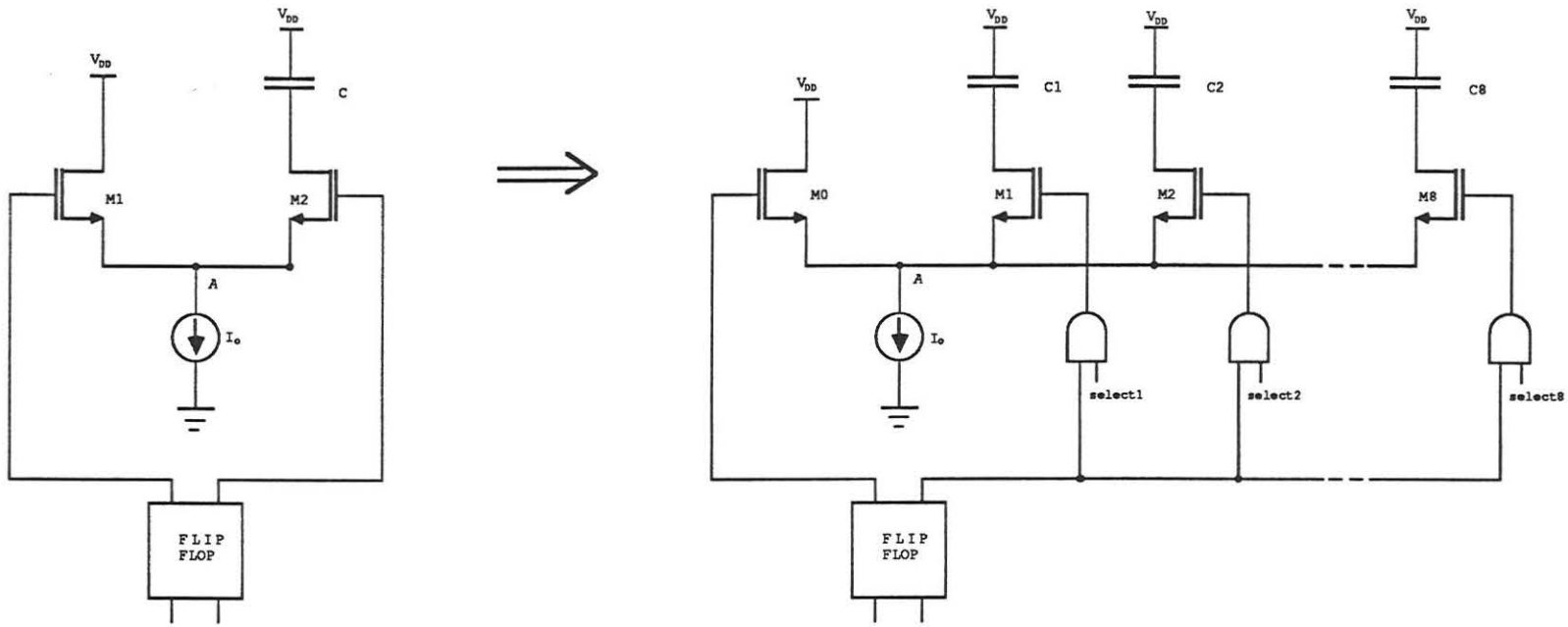


Figure 3.1. Extension of the single-channel TVC to an n-channel TVC/Analog Memory. The added circuitry for each channel is one switching transistor, one capacitor, and one logic gate. Note the renumbering of the left transistor to $M0$.

Some additional logic is not shown on the diagram; this includes an n-bit shift register to select the individual channels.

3.2 Component Matching

Matching of components between channels is important in order to maintain a constant level of performance across the system. However, certain components are more important to match than others. In particular, the hold capacitors must be matched as closely as possible because variations in the absolute value of the capacitor will have a first order effect on the output (remember Equation 1.6, $v_c = \frac{I_o}{C}t$). This matching is accomplished through use of various layout techniques, including common-centroid geometry. This scheme will be outlined in Chapter 4.

Matching of the other channel components is less critical because the effects will be second order. In the case of the switching transistor, process variations will cause changes in V_T from channel to channel. This will cause the quiescent voltage at node A to vary because V_{GS} is determined by

$$V_{GS} = V_G - V_S = V_{DD} - V_S = \left(\frac{I_D}{\frac{\mu_n C_{OX}}{2}\frac{W}{L}} \right)^{1/2} + V_T \tag{3.1}$$

Thus, variations in V_T will have a first order effect on the voltage at A; this in turn could affect the current I_o by varying the voltage at the output of the current source. However, V_T variations for adjacent devices are typically on the order of a few millivolts; a reasonable current source design should easily reject this effect. Therefore, the matching of the switching transistors is not critical.

Threshold variations can also affect the logic at the gates of the switching transistors. These variations appear as an uncertainty of the logic threshold level of the gate. This

means that we can never be sure of the precise instant when the gate will switch from high to low. In order to get around this uncertainty, the logic which drives the current switch is always overdriven so that the rise times will be as short as possible. This reduces any logic skew to a minimum.

3.3 8-Channel Scheme

An 8-channel analog memory was used in the layout of the prototype chip. This number was arrived at for two reasons. First, as previously mentioned, for a good transient response and settling time, the capacitance at node A had to be minimized. The second reason concerned complexity of the layout; due to the stringent matching requirement, most of the circuitry for the 8 channels had to be laid out in one piece; it was not possible to design one channel and then repeat it any number of times. Thus, a total of 8 channels provided a good measure of circuit performance while also keeping the layout reasonable.

The capacitance at node A had three components: the source-substrate junctions of the switching transistors, the interconnect capacitance, and the drain-substrate junction of the current source. A quick hand estimation can be made from the process parameters from MOSIS and dimensions from the layout. Assuming that the parasitic in the current source is negligible (this is true for a compact layout), the magnitude of this capacitance will be:

$$C_P = 9\times\left[A_{source}\frac{C_j}{\left(1+\frac{V_{SB}}{\psi_o}\right)^{.5}} + P_{source}\frac{C_{jw}}{\left(1+\frac{V_{SB}}{\psi_o}\right)^{.33}}\right] + A_{interconnect}\times C_{metal-substrate}$$

Assuming source area $A_{source} = 2.4\ \mu\text{m} \times 2.4\ \mu\text{m}$, source perimeter $P_{source} = 4.8\ \mu\text{m} + 4.8\ \mu\text{m}$, interconnect area $A_{interconnect} = 2.4\ \mu\text{m} \times 1500\ \mu\text{m}$, zero-bias junction capacitance $C_j = 2\times10^{-4}$ pF/μm^2, zero-bias sidewall capacitance $C_{jw} = 6\times10^{-4}$ pF/μm,

$C_{metal-substrate} = 2\times10^{-5}$ pF/μm^2, $V_{SB} \approx 3$ V, junction potential $\psi_o = 0.6$ V, the resulting $C_P = 0.1$ pF. As subsequent simulations will show, the magnitude of C_P will have an important effect on the operation of the TVC.

3.4 Selection Logic

3.4.1 Fast Gate Drive In order to assure uniform operation of the 8 TVC channels, the rise and fall times at the transistor gates needed to be as fast as possible. This was accomplished by using a dedicated CMOS inverter at each transistor gate. The logic for each channel was implemented as shown in Figure 3.2, and consists of an inverter and a NAND gate. The inverters were laid out with $\left(\frac{W}{L}\right)_P = 3\left(\frac{W}{L}\right)_N$. This geometry equalized the rise and fall times by accounting for the lower mobility of the p-channel transistors ($\mu_n \approx 3\mu_p$).

Another important issue was the drive capability of the flip-flop. Due to the addition of the seven more channels, the flip-flop needed to drive 8 CMOS loads on one output and one CMOS load on the other (Figure 3.3). The 8 loads also had a large interconnection load which was required to physically connect the logic gates together; in fact, it was this load which dominated the total load capacitance. In order to preserve the overlapping nature of the flip-flop outputs, the load at Q and $\overline{Q}$ needed to be matched. This was accomplished by using two techniques. First, an inverter chain was added at the outputs of the flip-flop in order to boost its drive capability. As seen in Figure 3.4, two inverters were added at each output; each subsequent inverter had $\frac{W}{L}$ larger than the previous stage by a factor of e. This provided the best method of increasing drive capability without sacrificing speed [19].

The second method to equalize the behavior at the two outputs was to match the

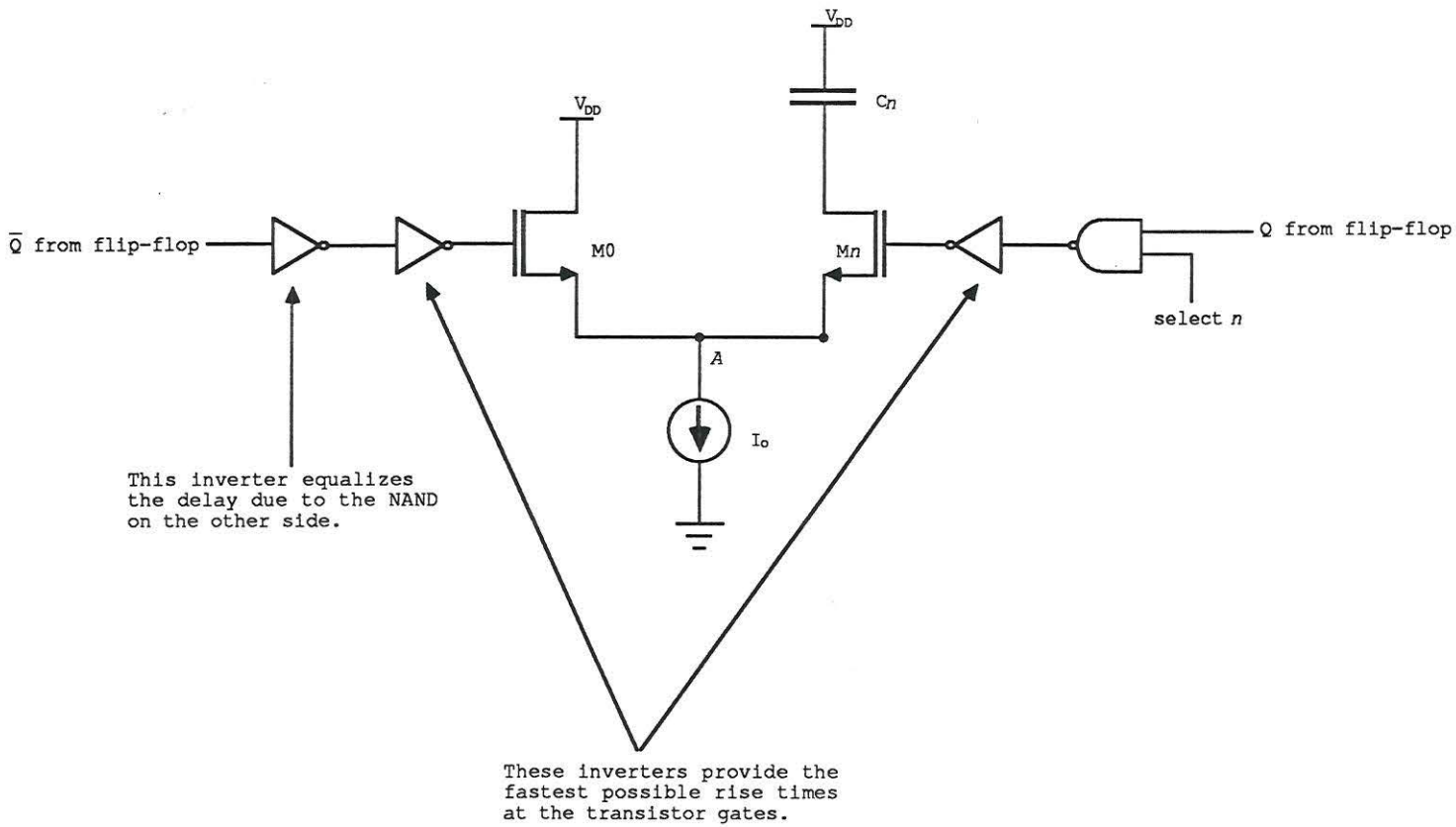
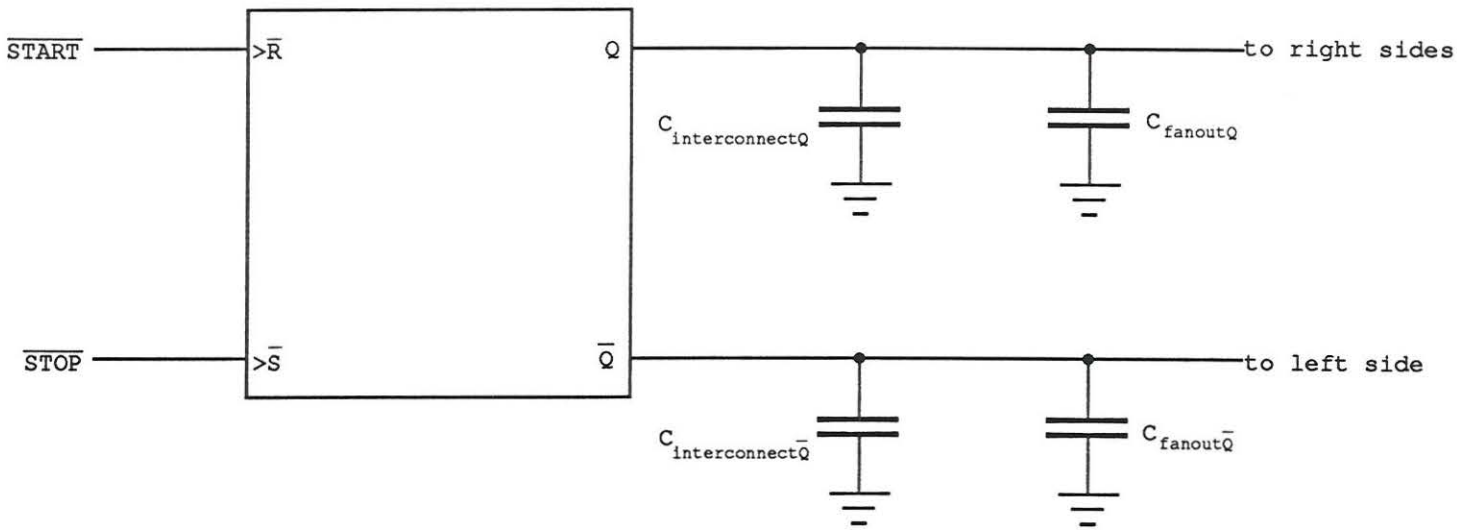


Figure 3.2. Fast differential gate drive with added selection logic.



$$C_{fanoutQ} = 8 \times C_{fanout\overline{Q}}$$

Figure 3.3. The outputs of the flip-flop will drive unequal loads because of the different fanouts of the right and left sides. The interconnection capacitances were approximately matched.

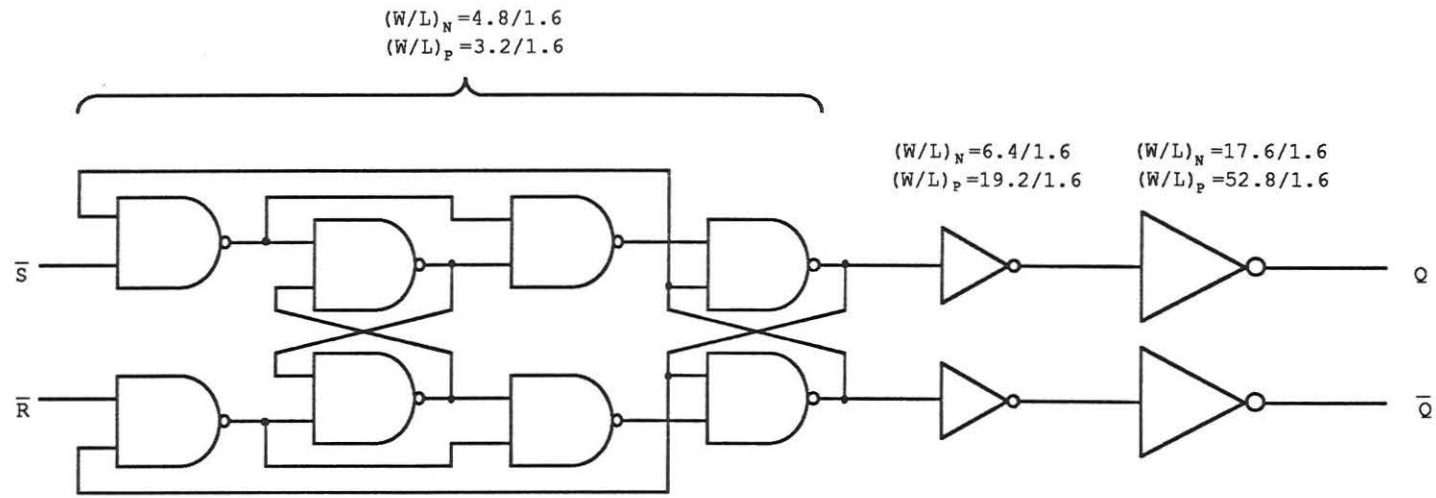


Figure 3.4. By adding inverter chains to the outputs, the flip-flop can compensate for the unequal loading.

capacitive loading at the outputs of the two inverter chains. This was accomplished by running a dummy wire next to the interconnection wire for the 8 channels. The dummy wire was connected to $\overline{Q}$ in order to simulate the load seen on Q. A picture of the entire system is shown in Figure 3.5.

3.4.2 8-Bit Shift Register A shift register was utilized to select between the 8 TVC channels. This provided the most space efficient design and was also conceptually simple. The shift register was externally reset to 00000001 and had its own clock line in order to change the channels. Subsequent clock edges moved the shift register output to 00000010, 00000100, etc. The register was wrapped around on itself so that 10000000 would shift back to 00000001. The eight bits of the shift register were connected to the eight NAND gates at the inputs of $M1$–$M8$.

The logic diagram of the register is shown in Figure 3.6. A simple one-bit register cell consists of two inverters and two transmission gates [19]. The cell is clocked with non-overlapping clocks ϕ_1 and ϕ_2 in order to shift a bit through. The non-overlapping clocks were generated as shown in Figure 3.7. This circuit takes a single external clock and generates the complementary clocks necessary to operate the shift register. Timing is shown in Figure 3.7b. The clocks were designed with a minimum of two gate delays between each non-overlapping component; this ensured a safety margin in case of any unexpected loading on the clock wiring and fanout.

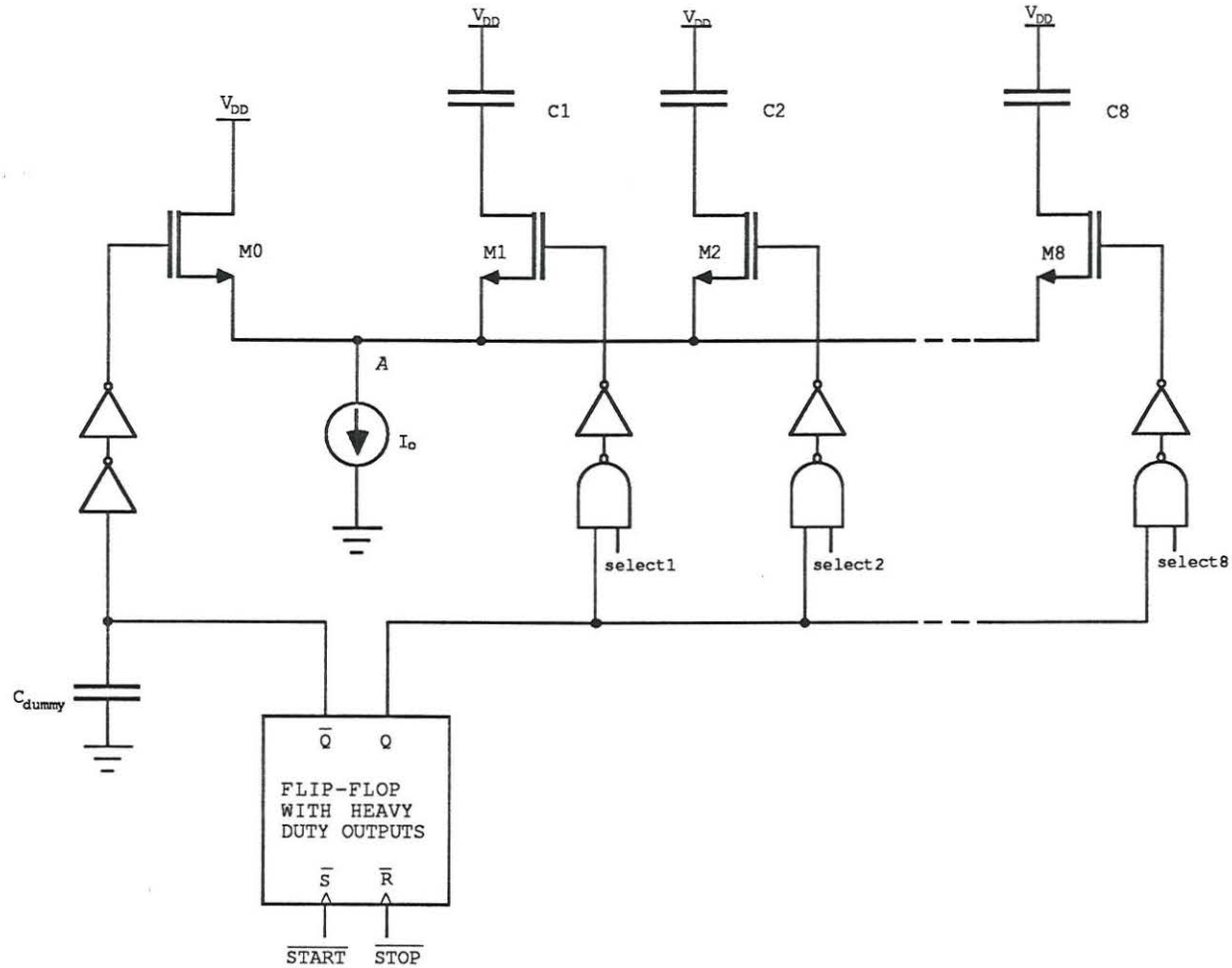
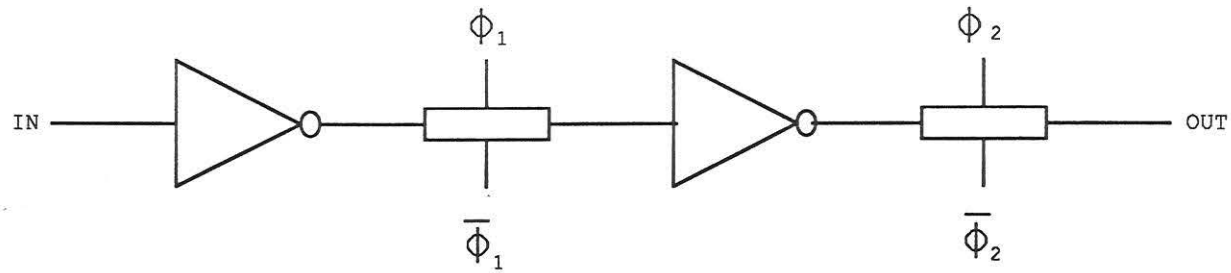
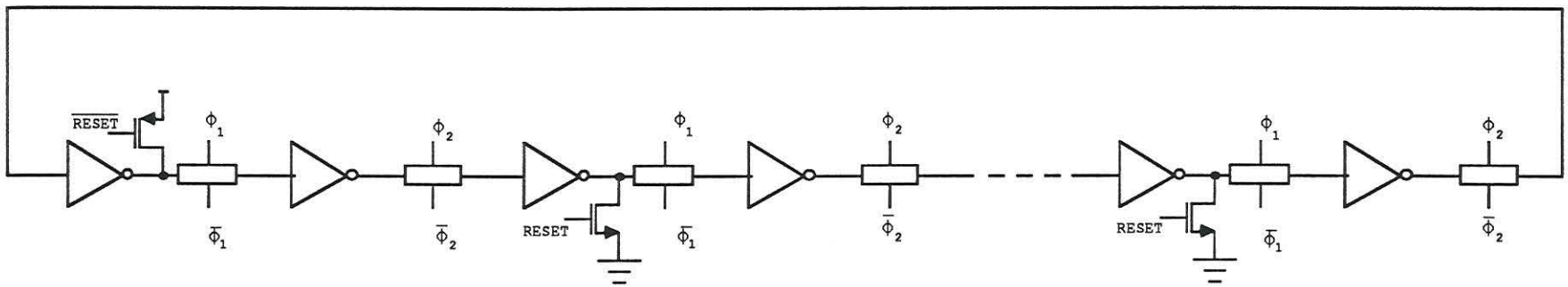


Figure 3.5. 8-channel TVC/Analog Memory system. An 8-bit shift register (not shown) was used to generate the select signals.

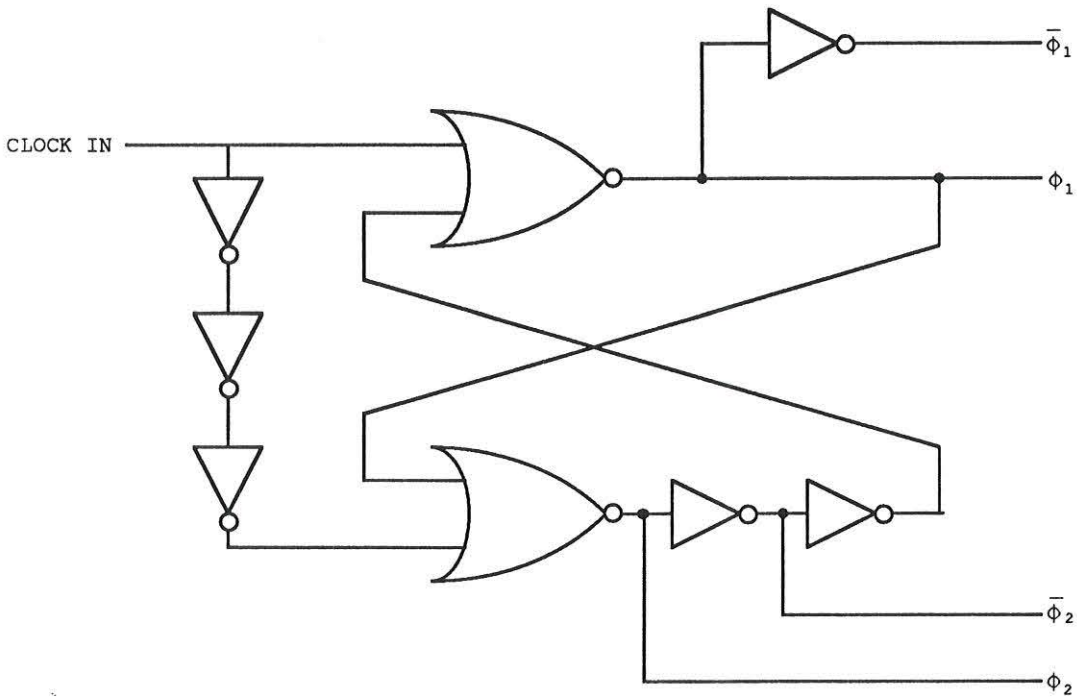


(a)

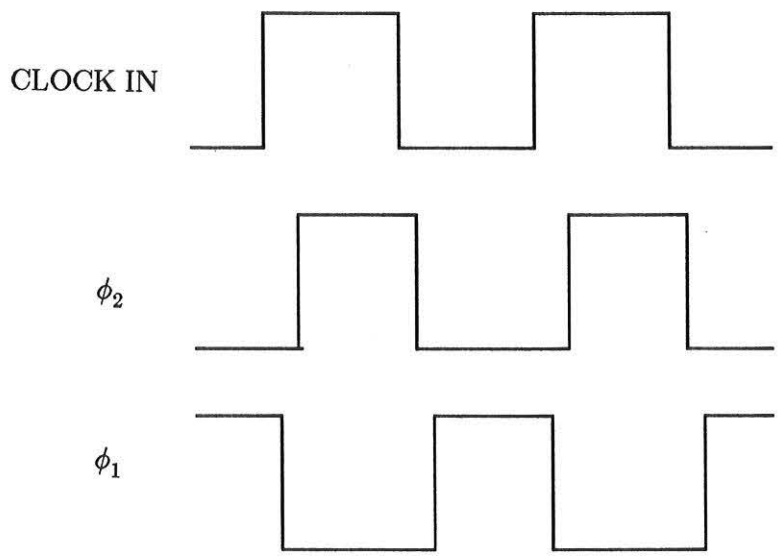


(b)

Figure 3.6. CMOS shift register. (a) 1-bit shift register cell. (b) 8-bit shift register as used in the Analog Memory.



(a)



(b)

Figure 3.7. Shift register clocking scheme. ϕ_1, $\overline{\phi}_1$, ϕ_2, and $\overline{\phi}_2$ were designed to have a minimum of two gate delays between each non-overlapping signal. (a) Clock generation circuit. (b) Clock timing.

4. Simulations and Layout

4.1 Notes on Simulation

Extensive computer simulation was done on the design in order to get a good estimate of performance before the expensive fabrication process. The modified SPICE2G6 simulator with the BSIM model [20-22] was used for all analog and most digital circuitry. BSIM is a relatively new MOS transistor model which uses a set of special laboratory-derived parameters in order to accurately model short-channel transistors. The ESIM logic simulator [23] was used to verify most of the digital logic such as the flip-flop and shift register. Simulations were performed on a VAX 8650 under the Ultrix operating system.

In order to get accurate analog simulations, several options needed to be specified in SPICE. For example, in order to get numerically stable transient solutions, it was necessary to use the Gear integration method. In order to achieve charge and current conservation, it was necessary to adjust RELTOL, CHGTOL, ABSTOL, and TRTOL. A detailed description of the SPICE techniques used in the simulations is found in Appendix A.

4.2 Simulation of the TVC

4.2.1 Clocking Schemes A great deal of information can be obtained from the simulation of the simple circuit in Figure 4.1. This circuit contains ideal voltage and

current sources, two ideal capacitors, and two short-channel MOS transistors. For the cases of single-ended and differential clocking, the relevant circuit voltages are shown in Figures 4.2 and 4.3.

For single ended clocking, it can be seen that circuit operation is excellent for small values of C_P, but is highly degraded as C_P approaches the realistic value of 0.1 pF. This is because of the slow swing at node A due to C_P. For differential clocking, the results are much more reasonable (Figure 4.3), making it clear why this clocking method was chosen. As can be seen in the figure, V_{OUT} is independent of the magnitude of C_P. The error voltage due to the overlap capacitances c_{gs} and c_{gd} can also be clearly seen in this plot. A small ramp-up is visible on V_{OUT} while V_{RIGHT} is rising *before* $M2$ turns on. A equal ramp-down can be seen while V_{RIGHT} falls after turn-off, making the net error equal to zero.

Another effect which can be observed is the feedthrough current due to the overlap capacitances (Figure 4.4). Because the gate voltages are strictly piece-wise-linear models, it is easy to see the feedthrough currents in I_{D1} and I_{S1} as pedestal offsets. This is because the gate rises and falls have constant slopes, i.e. dv/dt = constant. Recalling Equation 2.2, resulting feedthroughs will be constants. It is also interesting to note the effect of the parasitic C_P on the transistor currents. Referring to the plot of $-I_{S1}$, the first pedestal will be 10 μA feeding through c_{gs1} due to the linear rise on V_{RIGHT}. This will cause V_S to rise in order to decrease the current in $M0$ by 10 μA to compensate for the extra 10 μA coming from $M1$. The subsequent fall on V_{LEFT} will pull current *out* of node A, causing V_S to fall. However, depending on the size of C_P, V_S may or may not be able to adjust fast enough, and $-I_{S1}$ will vary as a result. The opposite effect can be seen when $M1$ turns off. Initially, $-I_{S1}$ will decrease to compensate for the feedthrough when

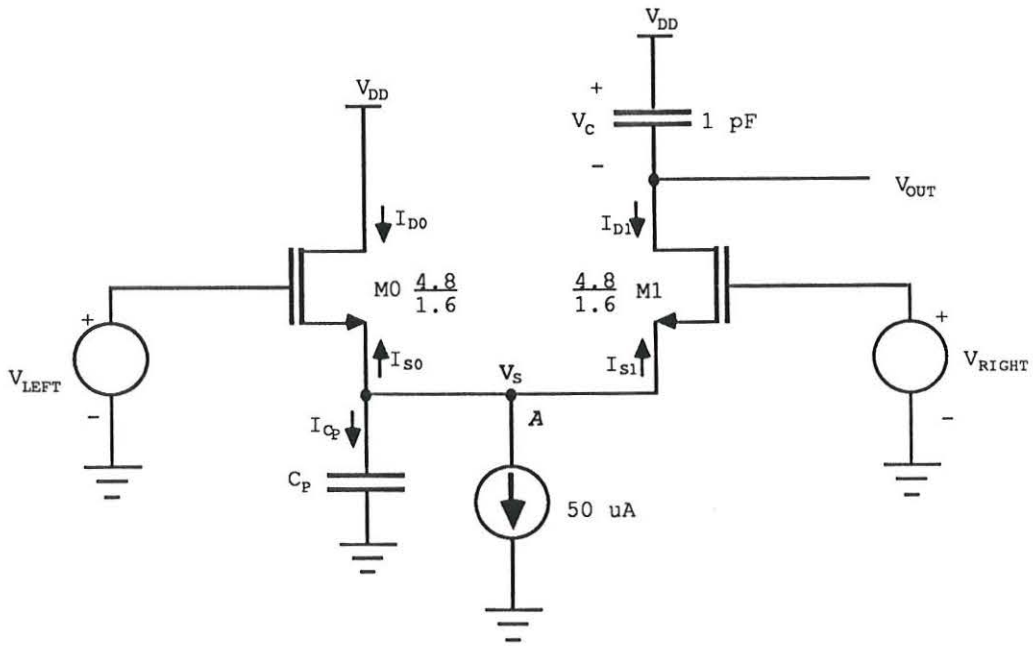


Figure 4.1. Simplified circuit used for initial SPICE simulations. CMOS realizations of the ideal parts were added later.

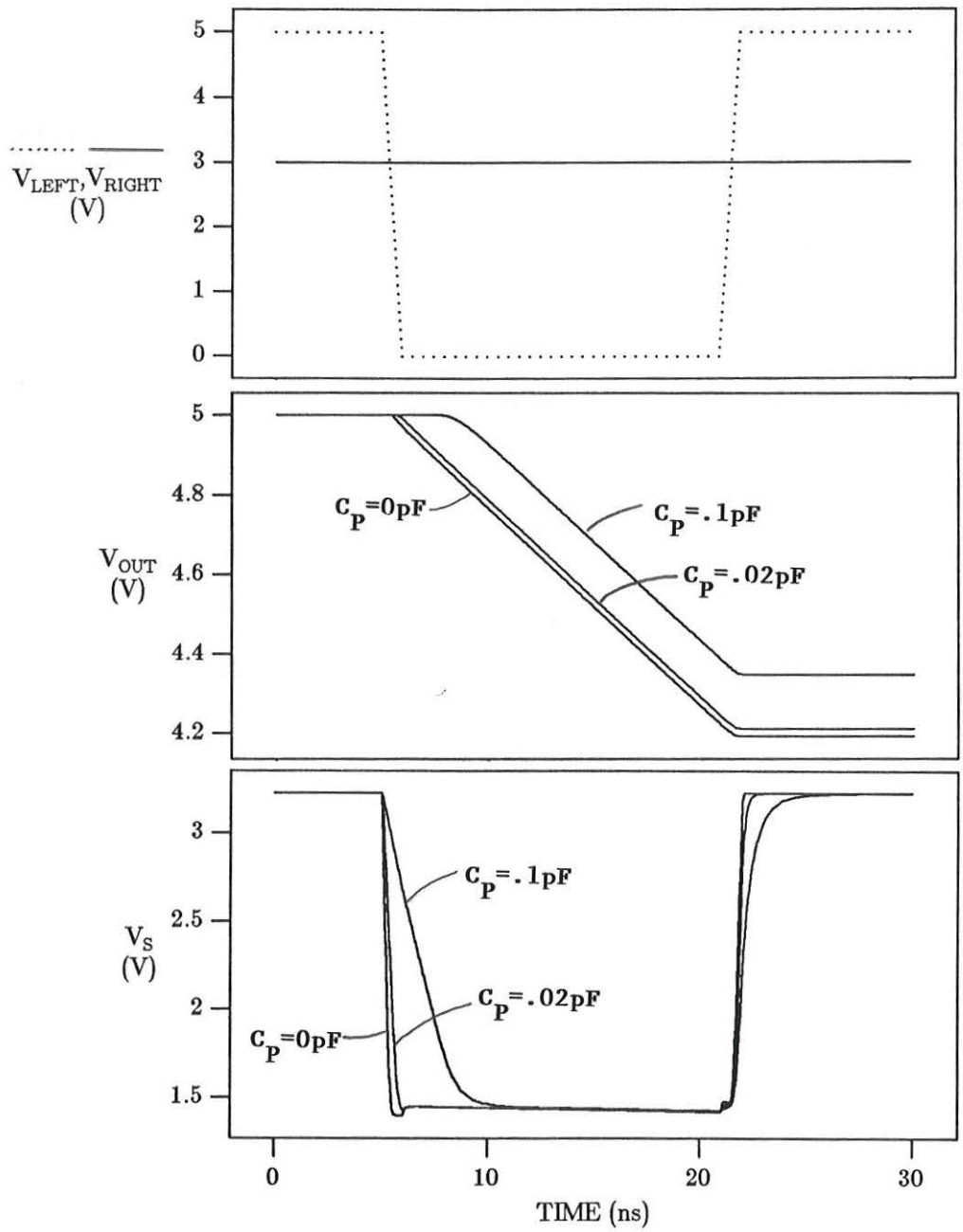


Figure 4.2. Important circuit voltages for single-ended clocking included the transistor gate voltages, the source voltage, and the TVC output voltage.

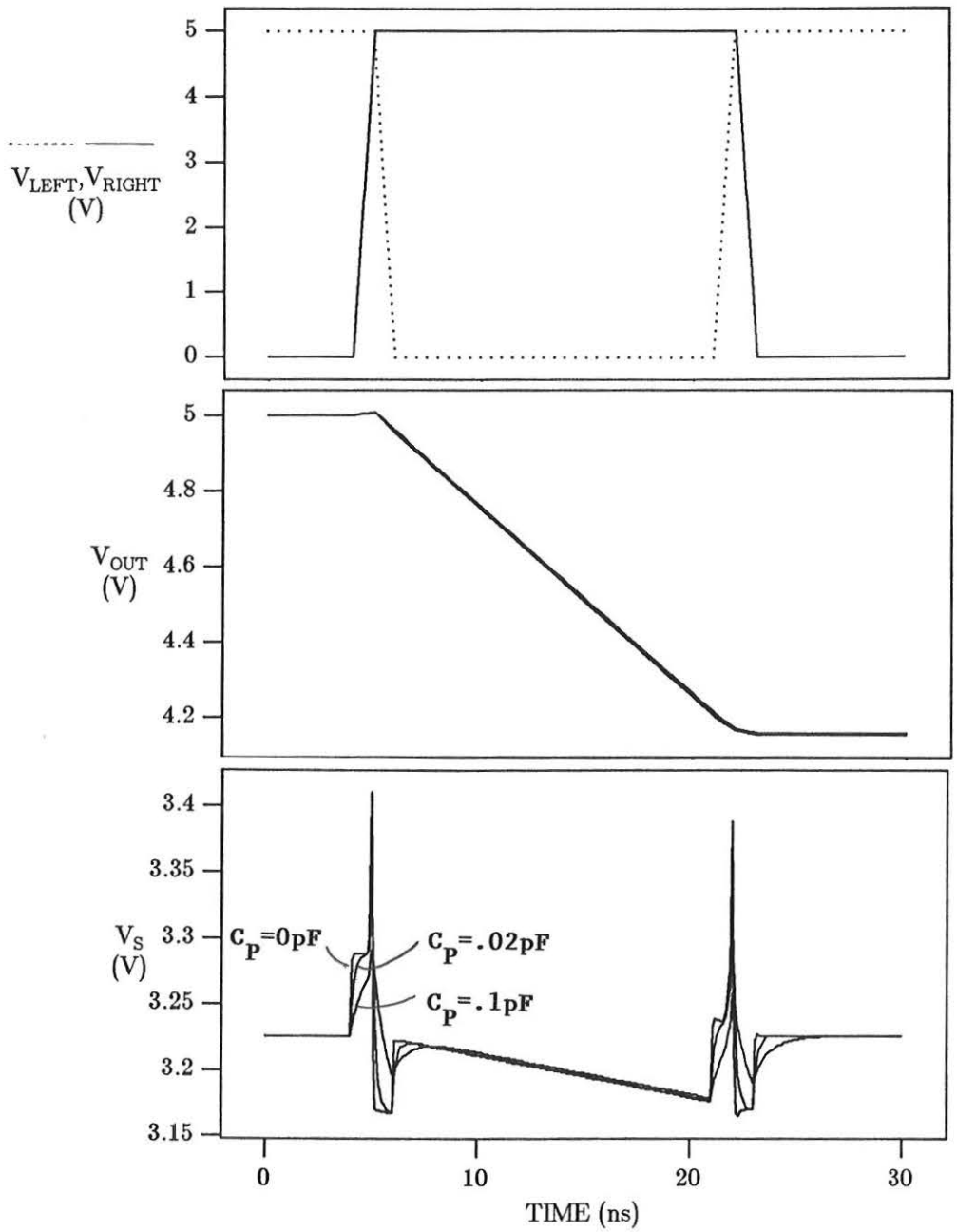


Figure 4.3. Important circuit voltages for differential clocking. Note that the output voltage is independent of the magnitude of C_P.

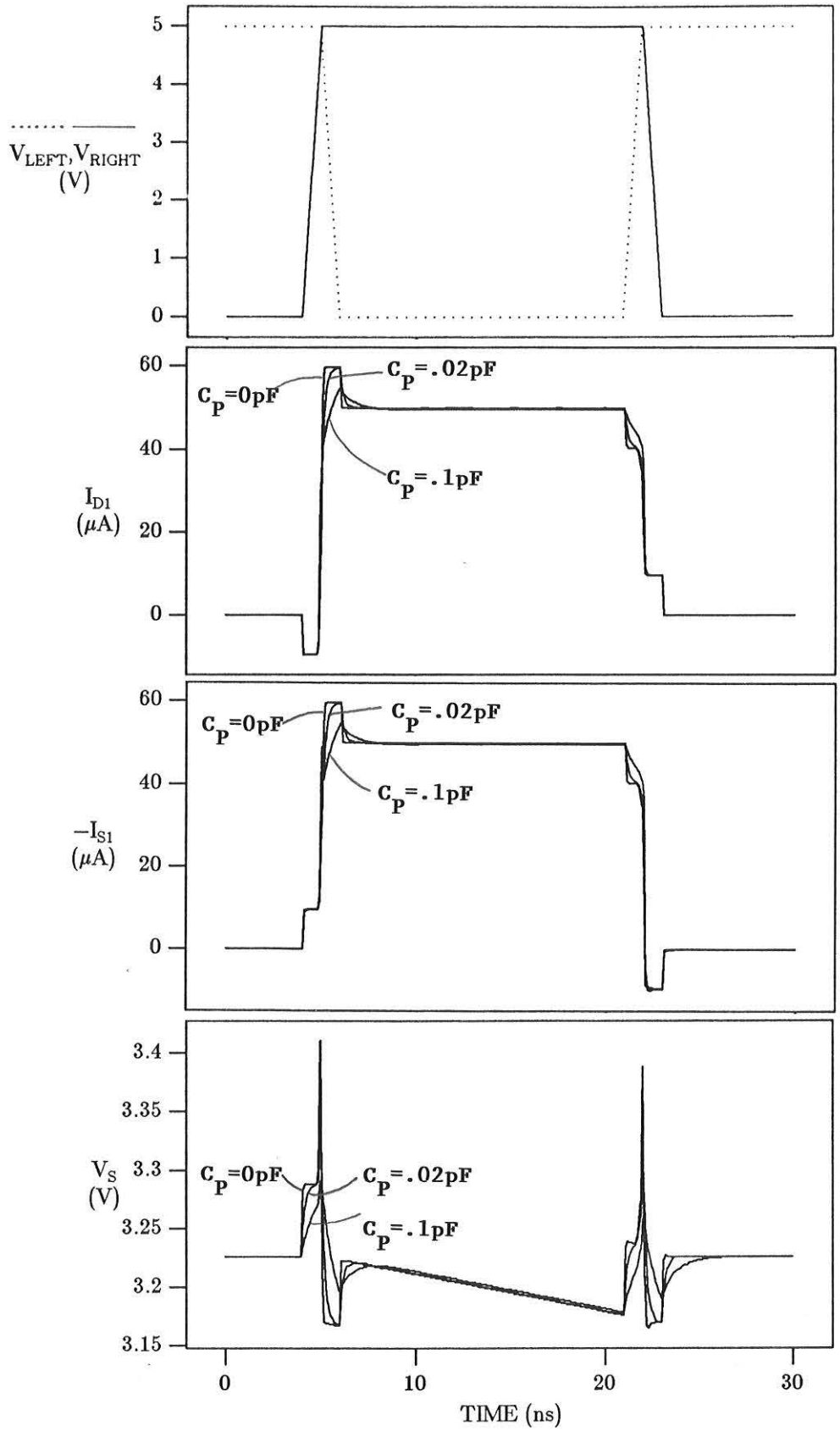


Figure 4.4. Transistor currents for differential clocking. Note the pedestals on the currents due to the overlap capacitances.

V_{LEFT} goes high; then, $-I_{S1}$ will pull charge out of node A as V_{RIGHT} goes low, causing the final pedestal.

From the simulations, it is obvious that the capacitor current (I_{D1}) is not constant during the entire time that $M1$ is on. However, as long as I_{D1} is constant over a portion of the interval, we can use this portion for the actual time measurement, and treat the non-linear part as a constant offset. This idea is borne out in Figure 4.5, where the capacitor current for two different time intervals is shown. It is clear that the non-linear charging portion is identical for both waveforms, and that the difference between the final capacitor voltages will be $\frac{I_o}{C}\Delta t$.

More information about the transistor currents may be seen when the overlap capacitances are removed from the transistor model. This was accomplished by making CGSO=CGDO=0.0 in the BSIM parameters. The resulting transistor currents are shown in Figure 4.6 (Note: subsequent simulations in this chapter will assume that $C_P = 0.1$ pF). From this simulation, it is clear that the transistor currents almost precisely follow the gate voltages, except right around turn-on and turn-off, as previously explained in section 2.3.1.

An expansion of the response right around turn-on is shown in Figure 4.7. This figure shows three currents: I_{D1}, I_{D0}, and the current through the parasitic capacitor I_{C_P}. From this simulation, we can draw a detailed picture of what is going on when the two transistors are switched. Initially, $I_{D0} = 50$ μA and $I_{D1} = 0$. When V_{RIGHT} goes high, I_{D1} will immediately jump to 50 μA because the large capacitance C_P will keep V_S and therefore V_{GS1} constant. Likewise, I_{D0} will remain at 50 μA because V_{LEFT} is still high. Thus, for a fraction of a nanosecond, both transistors will have currents of 50 μA, meaning that the two transistors will feed a total of 100 μA to node A. Since only 50 μA can exit node A

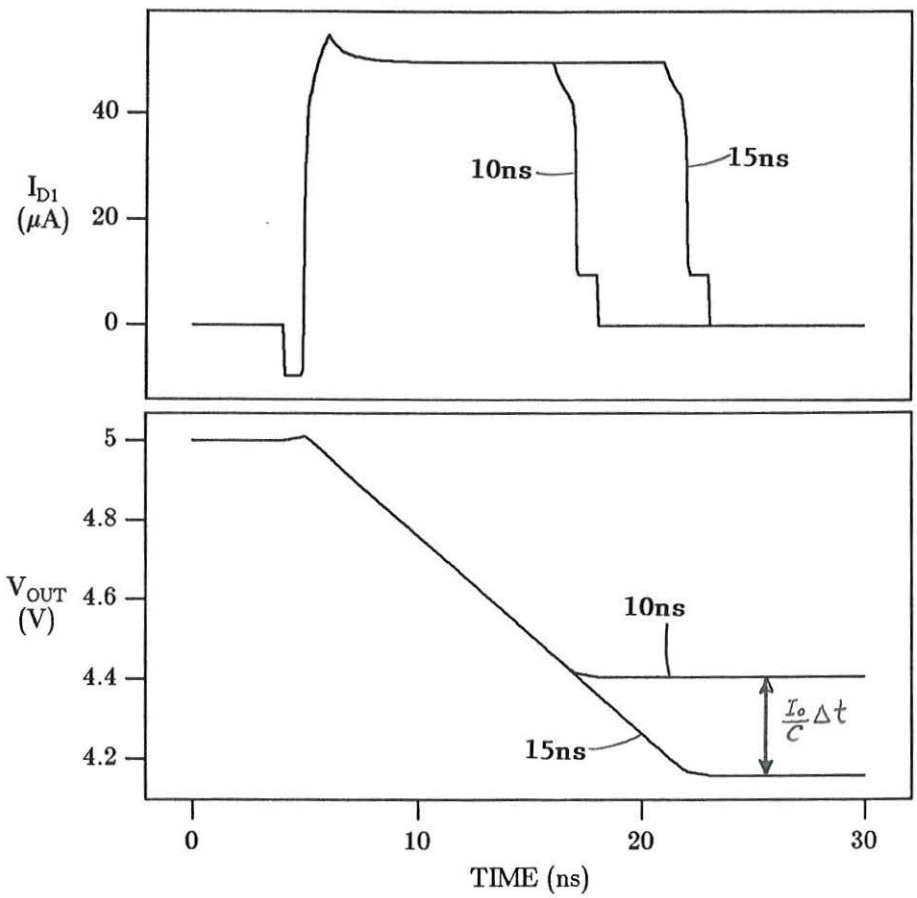


Figure 4.5. Drain current and output voltage for time intervals of 10 ns and 15 ns. The non-linear portions of the capacitor current will be identical for both measurements.

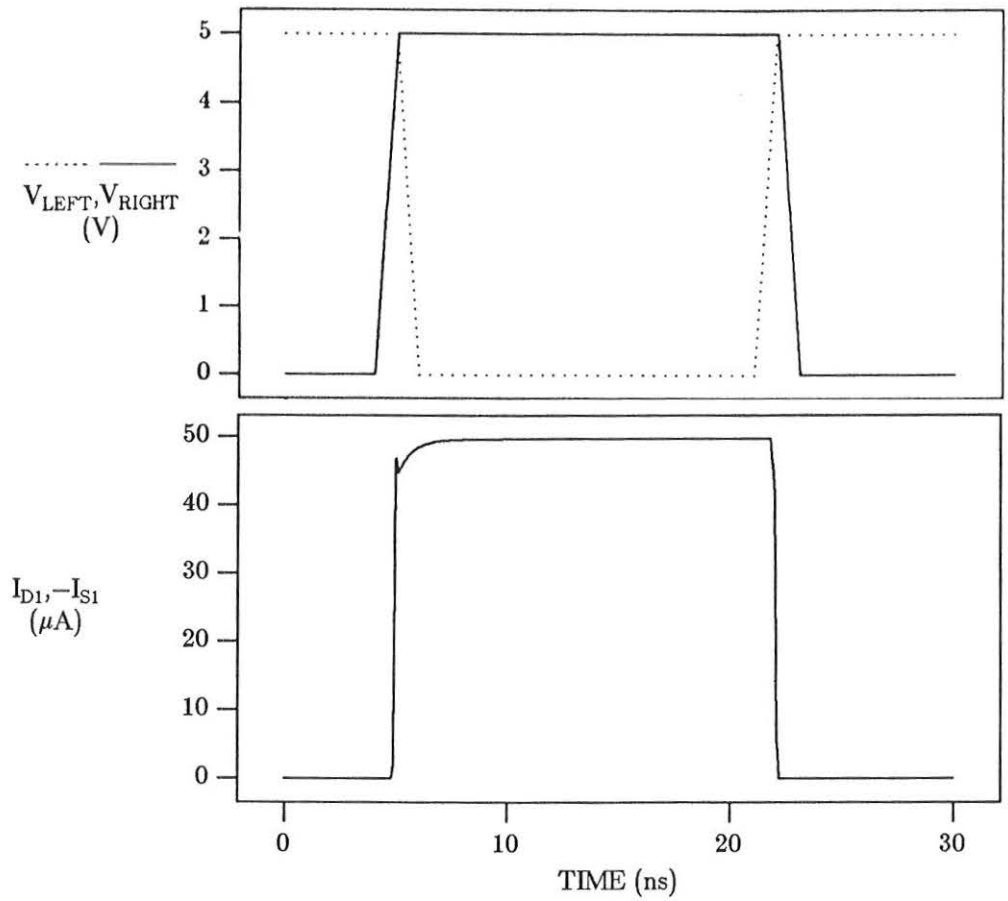


Figure 4.6. Transistor current with the overlap capacitances removed from the transistor model. Note the absence of the feedthrough currents and the small peak in the current right at turn-on.

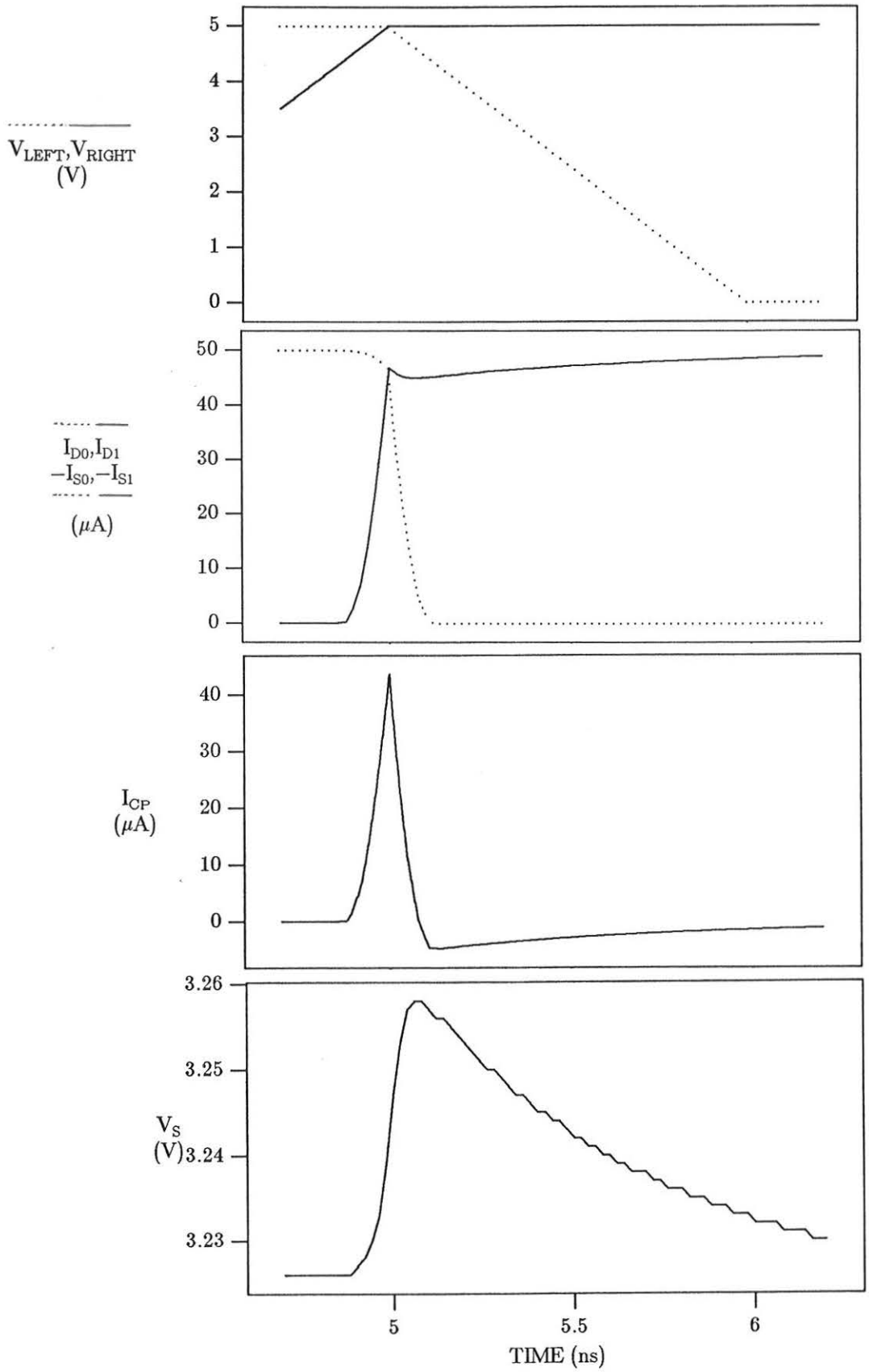


Figure 4.7. Circuit currents and voltages at the instant when $M1$ turns on. The parasitic capacitance C_P will briefly cause both transistors to be on with a current of 50 μA each. (Note: overlap capacitances have been omitted.)

via the current source, the remaining current will be used to charge C_P. C_P will continue to get charged until V_{LEFT} goes low, turning off $M0$. In the meantime, the voltage V_S at node A will have charged up to a higher value with

$$\Delta V_S = \frac{1}{C_P}\Delta Q \tag{4.1}$$

This increase in V_S will cause a decrease in V_{GS1}, and therefore force the current through $M1$ to be lower until the extra charge can be bled off C_P. This decrease in I_{D1} will be of magnitude

$$\Delta I_{DS1} = g_m \Delta V_S \tag{4.2}$$

The length of time for this small current to decay away will be approximately

$$t_{bleed} \approx \frac{\Delta Q}{\Delta I_{DS1}} = \frac{C_P \Delta V_S}{g_m \Delta V_S} = \frac{C_P}{g_m} \tag{4.3}$$

Substituting $C_P = 0.1$ pF and $g_m = 0.1$ mmho, $t_{bleed} \approx 1$ ns. This result is verified by the simulation, where it takes approximately 1 ns for the current to return to 50 μA. Equation 4.3 is also interesting because it states that the time for the current through $M1$ to reach its final value is independent of the overlap time between V_{LEFT} and V_{RIGHT}. This is not really true because the magnitude of g_m will depend on V_{GS1}, making (4.3) invalid for a large ΔV_S. However, even accepting (4.3) as an approximation considerably reduces the constraint on the gate clocks to have a specified overlap time; this is desirable because a specific overlap time is difficult to design into a digital circuit. Clearly, we would like to minimize the overlap in order to minimize ΔV_S, but the overlap will ultimately be decided by the digital characteristics of the flip-flop.

4.2.2 Voltage at the Common Source Node As shown in Figure 4.8, the voltage V_S (at node A) does not remain constant during capacitor charging (Note: subsequent simulations in this chapter *will* have the overlap capacitances included.). This effect is due to the low output resistance of short-channel $M1$. Figure 4.9 shows the I-V

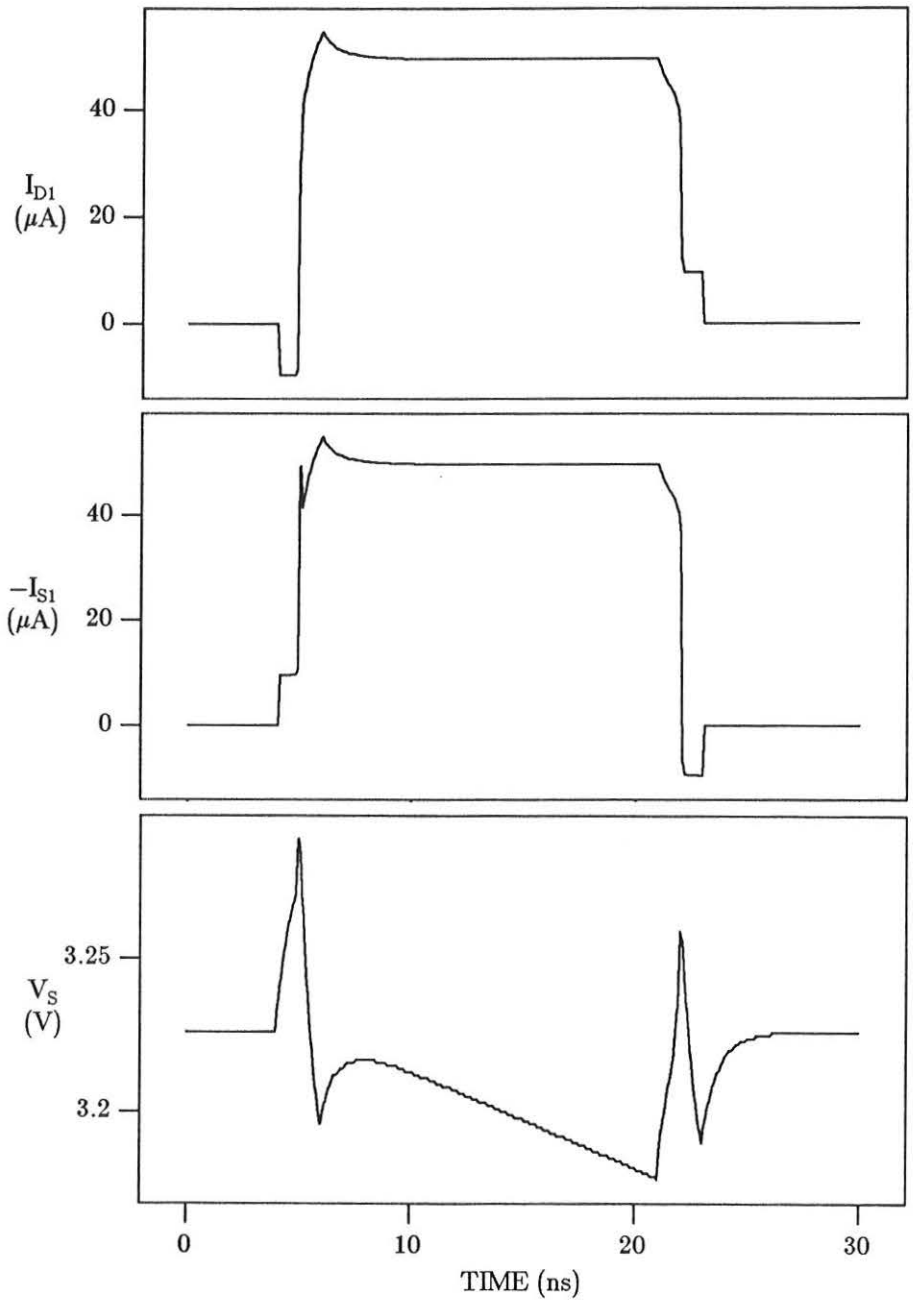


Figure 4.8. Transistor terminal currents and source voltage. The decrease in V_S during charging is due to the low output resistance of $M1$. $C_P = 0.1$ pF.

curves of $M1$ and the path of the operating point. Initially, when $M1$ is turned fully on, $V_{GS1} = V_{DS1} \approx 2$ V. Then, as the capacitor starts to charge, V_{D1} will fall, causing a drop in V_{DS1}. However, in order for I_{D1} to remain constant, the operating point needs to switch to a "higher" curve, causing a resulting increase in V_{GS1}. Since $V_{GS1} = V_{DD} - V_S$, V_S must fall as a result. For example, suppose V_{DS1} falls by some ΔV_{DS1}; a corresponding fall in I_{D1} would occur, of magnitude

$$\Delta I_{D1} = \frac{\Delta V_{DS1}}{r_o}. \tag{4.4}$$

where r_o is the output resistance of $M1$. However, we are driving $M1$ with a current source; thus, I_{D1} must remain constant, and the change in I_{D1} will be compensated for by a change in V_{GS1}. The magnitude of the change will be

$$\Delta V_{GS1} = \frac{\Delta I_{D1}}{g_m}. \tag{4.5}$$

Thus, combining 4.4 and 4.5,

$$\frac{\Delta V_{GS1}}{\Delta V_{DS1}} = \frac{\Delta V_{G1} - \Delta V_S}{\Delta V_{D1} - \Delta V_S} = \frac{1}{g_m r_o} \tag{4.6}$$

We know that $V_{G1} = V_{DD}$, so $\Delta V_{G1} = 0$. We can also make the approximation that $\Delta V_{D1} >> \Delta V_S$, since V_D changes with the capacitor voltage, and V_S will have a much smaller change. Thus,

$$\frac{\Delta V_S}{\Delta V_{D1}} = \frac{1}{g_m r_o} \tag{4.7}$$

This result simply reaffirms the usual goals of any analog circuit design, i.e. high transconductance and high output resistance are desirable. In the case of the TVC, by maximizing these parameters, we minimize the change at node A; this in turn would make the parasitic C_P less of a performance-limiting factor, and also relax the constraints on the current source design.

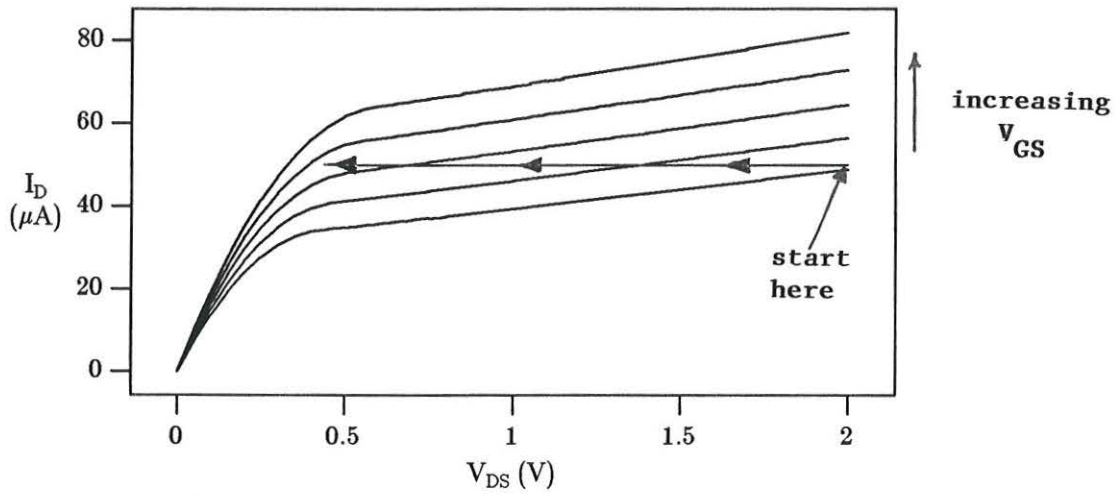


Figure 4.9. Path of the operating point of $M1$ as the capacitor is charged. Due to the low output resistance, the operating point will move to a higher curve as the capacitor is charged. V_{GS1} will increase as a result.

4.2.3 Current Source The next simulation step was to start replacing the ideal voltage and current sources in Figure 4.1 with realizable circuit components. As discussed in section 2.5, a simple current mirror was used for the 50 μA current source. The design constraints were somewhat opposite: make the transistors large in order to have a high output resistance, but small in order to have a high frequency response. The final compromise was to shoot for a minimum current source output resistance of 500 kΩ. The resulting DC characteristic and frequency response are shown in Figure 4.10.

4.2.4 Integration/Hold Capacitor Up until now, the actual implementation of the hold capacitor has not been discussed. In the HP 1.6 μm process, the following capacitors exist: metal-metal, metal-polysilicon, and MOS. Because the nominal value and matching of the capacitors was an issue, the MOS capacitor was chosen due to the fact that t_{OX} is one of the most carefully controlled process parameters. Several problems arise, however, due to the properties of the MOS capacitor.

A C-V curve of an MOS capacitor is shown in Figure 4.11. The basic MOS capacitor is both frequency and voltage dependent, so we must find a way to circumvent these dependencies. Frequency dependence becomes a non-issue due to the self-aligning technology; the only method of manufacturing a MOS capacitor is to make a transistor and short the drain and source to the bulk, thereby making the capacitor frequency-independent [19].

Regardless of the technology, the voltage dependence of the MOS capacitor will still exist. However, as seen in the C-V plot, at gate voltages above threshold, the capacitor becomes linear. Thus, one method would be to initially reset the capacitor to some voltage above threshold, say 1 V, and then commence charging from there (Figure 4.12a). There are several obvious reasons for abandoning this method, the least of which being the nuisance of generating the 1 V reset voltage. A much better approach would be to

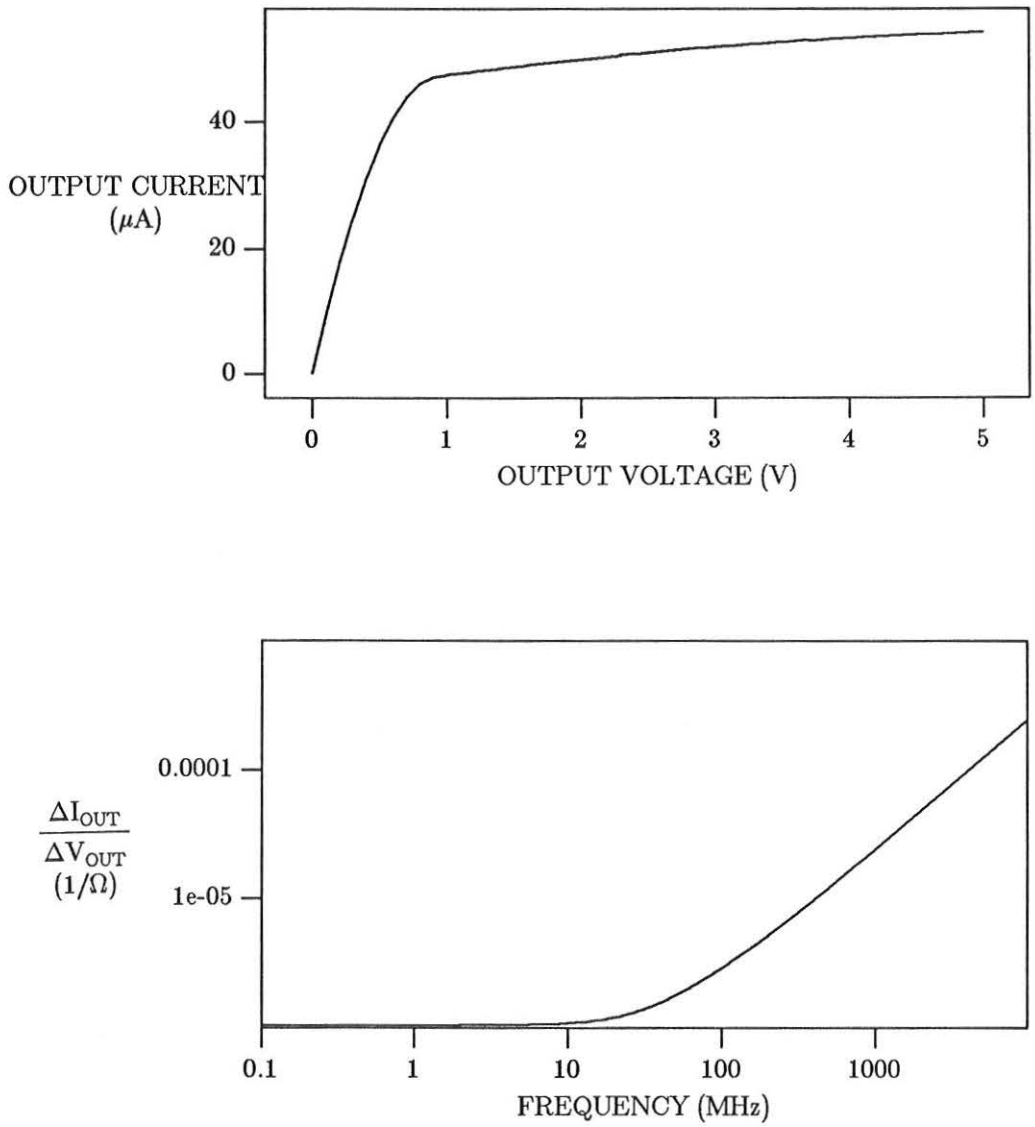
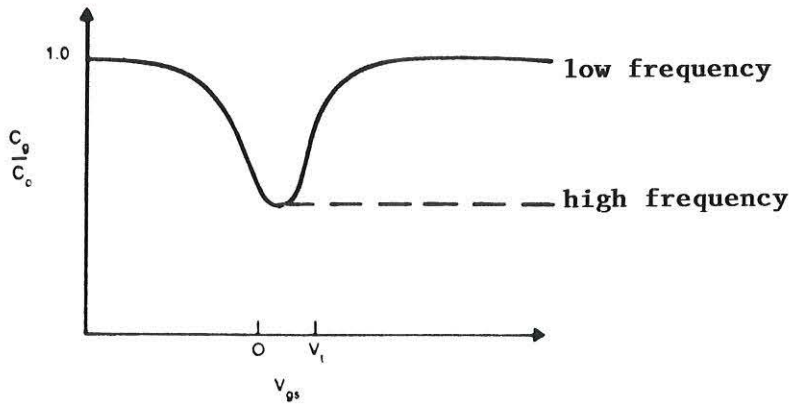
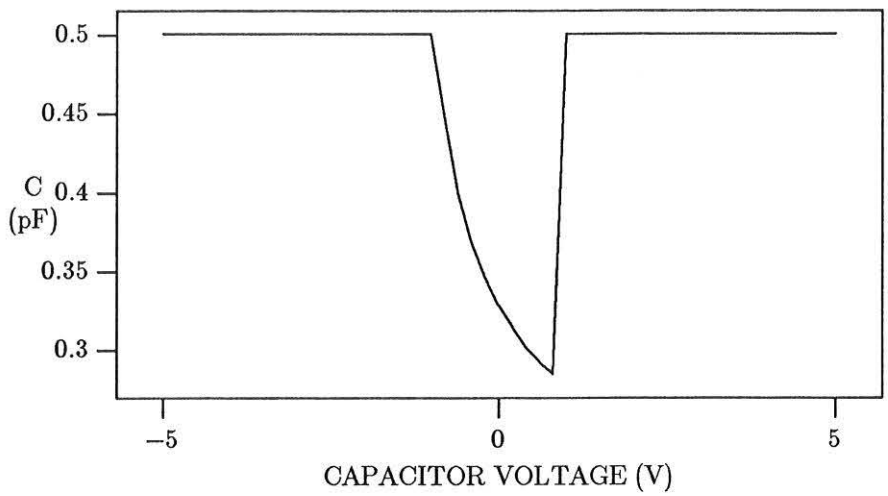


Figure 4.10. Simulated DC characteristic and frequency response of the 50 μA current source.

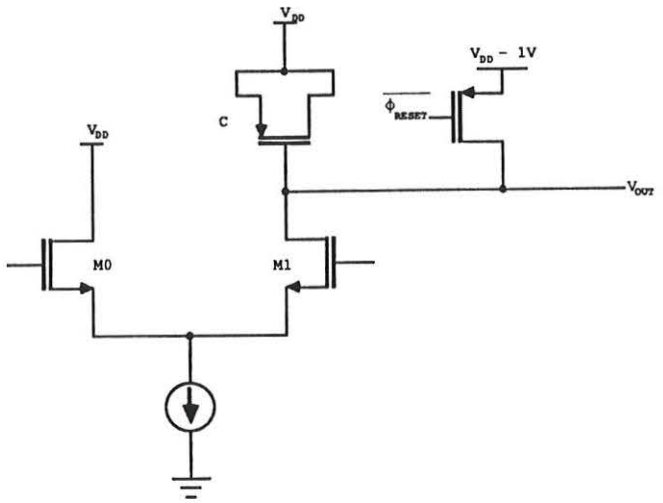


(a)

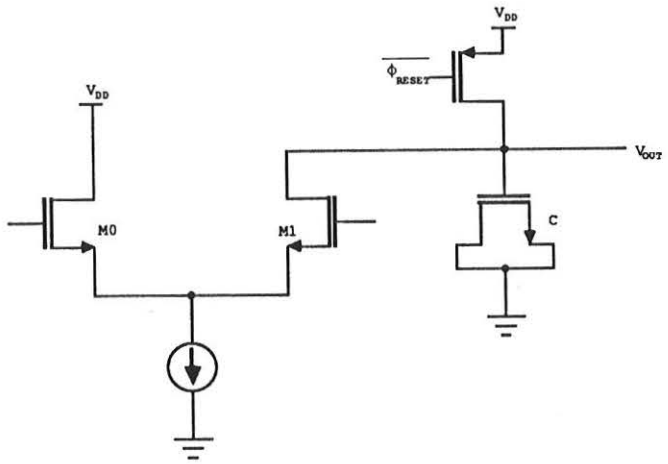


(b)

Figure 4.11. (a) C-V curve of a MOS capacitor [26]. The capacitance will be both voltage and frequency dependent. (b) SPICE simulation of the MOS capacitance characteristic for the capacitor used in the TVC. W = 3.2 μm, L = 115.2μm



(a)



(b)

Figure 4.12. Capacitor implementations for the TVC. (a) For a p-channel capacitor, the capacitor voltage must be precharged to $>V_T$ for linear charging. (b) By precharging an n-channel capacitor to V_{DD}, we avoid the uncertain non-linear region and escape the added nuisance of generating a reset voltage on the chip.

use the opposite sex of MOS capacitor (NMOS) and precharge the capacitor to V_{DD}, then discharge it during time measurement. This modification is shown in Figure 4.12b. Circuit operation will not change whether we physically charge or discharge the capacitor, and the MOS capacitor provides the best possible accuracy short of using a double-poly process.

4.2.5 Edge-triggered RS Flip-Flop Simulation of the flip-flop provided the following information: first, it characterized the behavior of the flip-flop outputs with different loads applied (remember the overlapping requirement); second, it gave the minimum resolvable time interval between the two input signals; and third, it gave a functional check of the layout via the layout extraction program.

The logic diagram of the flip-flop was shown in Figure 3.4. As described in section 3.4, it consists of eight NAND gates and two inverter chains on the outputs to boost drive capability. The approximate load on the outputs will be

$$C_L = N \times C_{inNAND} + A_{interconnect} \times C_{metal-substrate}$$

where N is the fanout (N=1 for $\overline{Q}$ and N=8 for Q). Since the interconnect areas were matched, the difference between the loads was 7 logic gates. An estimate of C_{inNAND} is 10 fF, so the flip-flop outputs needed to be able to withstand loading differences of about 0.07 pF and still overlap. This simulation is shown in Figure 4.13, where the nominal load due to the interconnect has been estimated at 0.4 pF, and the load on Q has been ranged by $\pm$0.1 pf. From the voltage waveform, it is apparent that proper overlapping will still exist even with unequal matching.

The minimum resolvable time interval for the flip-flop was simulated via trial-and-error to be 3 ns. Any smaller interval would cause the second falling edge (the $\overline{R}$ input) the be ignored. The simulated total propagation delay of the flip-flop was about 3 ns;

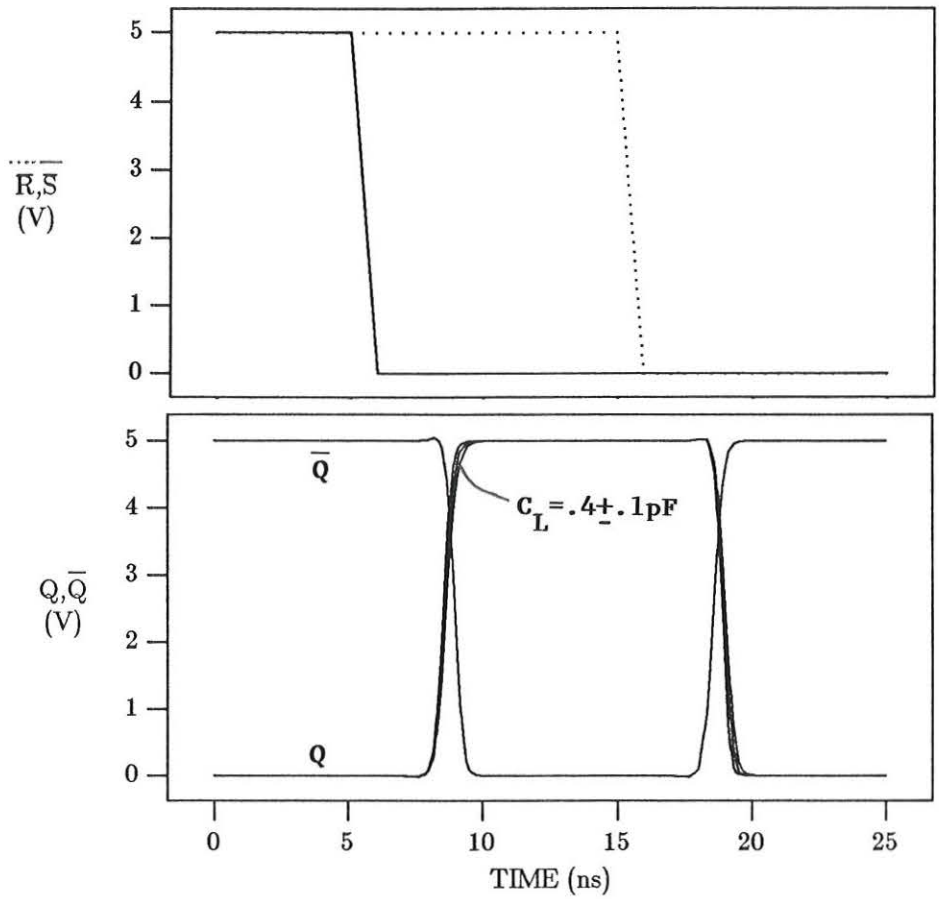


Figure 4.13. SPICE simulation results for the flip-flop. The total propagation delay and minimum resolvable time interval were 3 ns. Note the insensitivity of the flip-flop to the load capacitance. This was accomplished by using wide output devices.

this translates to a gate delay of about 0.6 ns, which is typical for the HP process.

4.2.6 Source Follower The source follower design (referring back to Figure 2.15) proceeded as follows: first, a good 1 mA current source was designed, and then $M1$ was sized to get a reasonable DC offset. The 1 mA source was fairly straightforward. A gain of 4 was inserted between $M6$–$M7$ and the biasing transistors $M3$–$M5$ in order to minimize areas. It was important to size $M2$ such that $\left(\frac{W}{L}\right)_2$ was *less* than $\frac{1}{4}\left(\frac{W}{L}\right)_{3,4,5}$. This was necessary due to the increased threshold of $M2$ and $M4$ because of the body effect. A current source output resistance of 1.1 MΩ was simulated.

The subsequent sizing of $M1$ led to a DC offset of about 1.3 V, i.e. for $V_{IN} = 5$ V, $V_{OUT} = 3.7$ V. The simulated DC characteristic and frequency response are shown in Figure 4.14.

4.2.7 Shift Register The shift register logic (inverter/transmission-gate chain) was functionally simulated with ESIM in order to check the layout. A three-bit version of the register was also verified using SPICE (SPICE simulation of the entire 8-bit register would not have been particularly enlightening and was also computationally expensive.). It *was* important, however, to thoroughly simulate the shift register clocking circuit. This was because the clock needed to be non-overlapping (somewhat opposite to our flip-flop requirement). Simulations are shown in Figure 4.15. With an estimated load of 0.15 pF on each output, the clock circuit performed well.

4.2.8 Putting It All Together It was possible to simulate the entire system, but the simulations were very computationally demanding. The most complicated circuit simulated is shown in Figure 4.16. This circuit contains all of the important pieces of the TVC/Analog Memory, including the flip-flop, differential pair, current source, source follower, MOS capacitor, and reset transistor. In order to approximate the actual capacitive

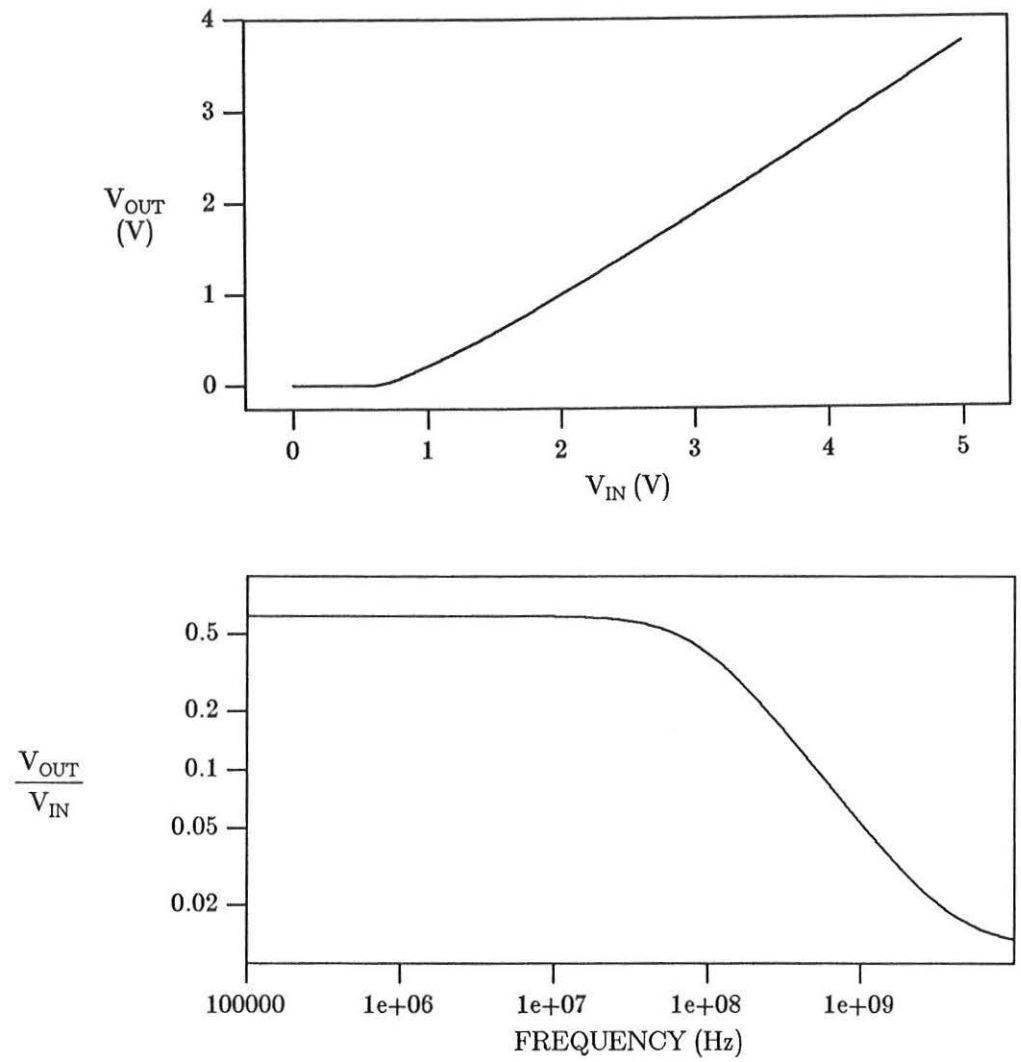


Figure 4.14. Simulated DC characteristic and frequency response of the source follower.

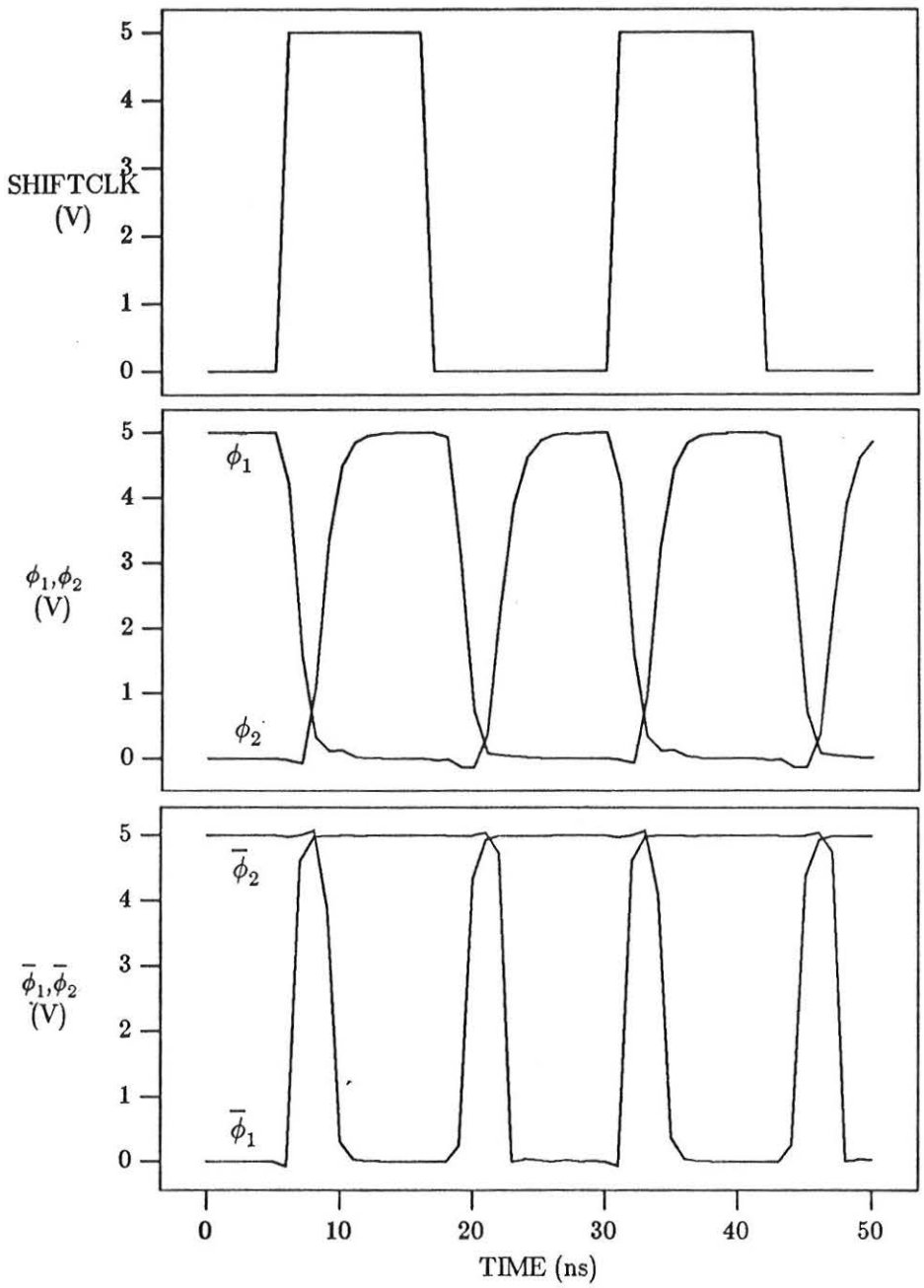


Figure 4.15. Simulation of the clock generator for the shift register. Note the non-overlapping behavior.

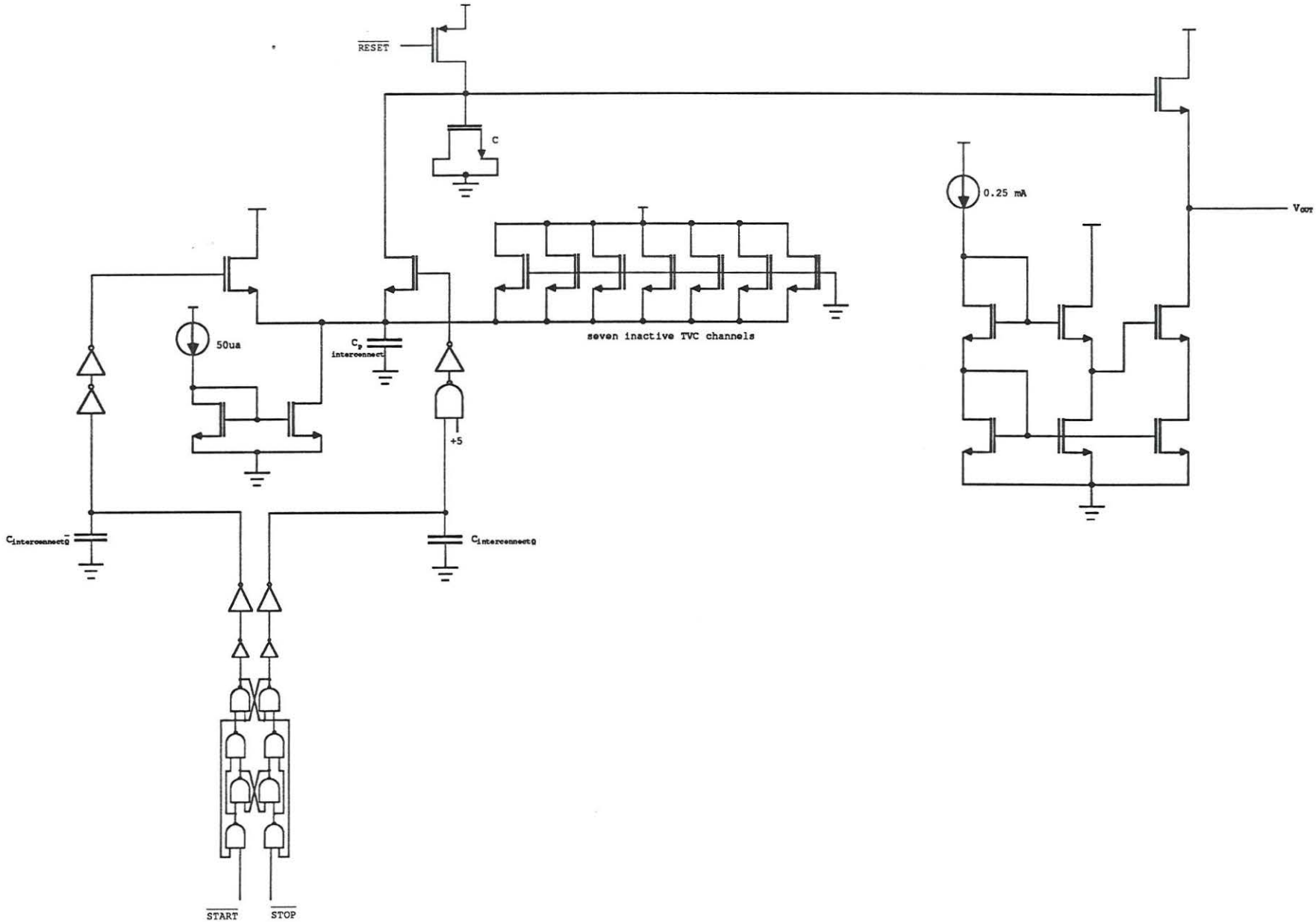


Figure 4.16. Largest simulated circuit for the 8-channel TVC/Analog Memory.

loads, $M2$–$M8$ were added to simulate the other seven inactive channels. The only major circuit component not included in this simulation was the shift register.

The SPICE input file is listed in Table 4.1, and simulation results are shown in Figures 4.17 and 4.18. The large spikes in I_{D1} are due to the overlap-capacitance feedthroughs from the sub-nanosecond rise times on V_{LEFT} and V_{RIGHT}. However, I_{D1} settles down rather quickly to 50 μA, and linear charging takes place over most of the time interval.

4.3 Layout Issues

Layout was done using the MAGIC VLSI tool [23] on an IBM RT. An IBM AT with a Professional Graphics Display was used as a terminal. Design rules were for the MOSIS scalable CMOS processes.

4.3.1 Capacitor Matching The key issue in the layout was the matching of the integration/hold capacitors between the 8 channels. This capacitance actually consisted of two components, the MOS capacitor and input capacitance of the source follower. In order to achieve optimum matching, the MOS capacitors and the source follower input transistors were laid out in a common-centroid geometry. This technique has often been used in charge-redistribution ADC's [24]. In order to implement this technique, each transistor and capacitor was split into four identical smaller components which were situated around a central point. As a result, all first-order process variations across the chip were canceled; for example, a linear oxide thickness variation in any direction would get averaged by the four components.

An example of this scheme is shown in Figure 4.19, where four channels have been laid out. Another feature of this particular method is that the total area of the interconnection wiring will be identical for each channel; although the wiring will not be common

```
DIFFERENTIAL TIME-TO-VOLTAGE 8/12/87 N CHANNEL

.SUBCKT NAND 2 4 3
VDD 1 0 5V
S1 3 2 1 1 PC1_PM1_DU2 L=1.6U W=3.2U AS=10P AD=10P PS=10U PD=10U
S2 3 4 1 1 PC1_PM1_DU2 L=1.6U W=3.2U AS=10P AD=10P PS=10U PD=10U
S3 3 2 5 0 PC1_NM1_DU1 L=1.6U W=4.8U AS=15P AD=15P PS=11U PD=11U
S4 5 4 0 0 PC1_NM1_DU1 L=1.6U W=4.8U AS=15P AD=15P PS=11U PD=11U
.ENDS NAND

.SUBCKT INV 2 3
VDD 1 0 5V
S1 3 2 1 1 PC1_PM1_DU2 L=1.6U W=7.2U AS=22P AD=22P PS=13U PD=13U
S2 3 2 0 0 PC1_NM1_DU1 L=1.6U W=2.4U AS=10P AD=10P PS=10U PD=10U
.ENDS INV

.SUBCKT BIGINV 2 3
VDD 1 0 5V
S1 3 2 1 1 PC1_PM1_DU2 L=1.6U W=19.2U AS=60P AD=35P PS=31U PD=8U
S2 3 2 0 0 PC1_NM1_DU1 L=1.6U W=6.4U AS=20P AD=20P PS=13U PD=13U
.ENDS BIGINV

.SUBCKT BIGGERINV 2 3
VDD 1 0 5V
S1 3 2 1 1 PC1_PM1_DU2 L=1.6U W=52.8U AS=162P AD=97P PS=60U PD=8U
S2 3 2 0 0 PC1_NM1_DU1 L=1.6U W=17.6U AS=54P AD=54P PS=22U PD=22U
.ENDS BIGGERINV

* EDGE-TRIGGERED JK FLOP
XE1 301 310 303 NAND
XE2 302 309 304 NAND
XE3 303 306 305 NAND
XE4 304 305 306 NAND
XE5 303 305 307 NAND
XE6 304 306 308 NAND
XE7 307 310 309 NAND
XE8 308 309 310 NAND
XB1 309 809 BIGINV
XB1A 809 819 BIGGERINV
XB2 310 810 BIGINV
XB2A 810 820 BIGGERINV
C819 819 0 .4PF
C820 820 0 .4PF
R309 309 0 1G
R310 310 0 1G
R302 302 0 1G
R301 301 0 1G

VSTOP 301 0 DC 5V PWL(0NS 5V 20NS 5V 21NS 0V)
VSTART 302 0 DC 5V PWL(0NS 5V 3.5NS 5V 4.5NS 0V)

VDD 1 0 5V

* CURRENT SOURCE
SC1 2 2 0 0 PC1_NM1_DU1 L=8U W=8U AD=14.7P AS=24.5P PD=8.4U PS=21U
SC2 4 2 0 0 PC1_NM1_DU1 L=8U W=8U AD=14.7P AS=24.5P PD=8.4U PS=21U
IBIAS 1 2 50UA
VAMMCURR 3 4 0V
CPAR 3 0 .1PF

* LEFT SIDE
S1 8 9 10 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
RBIG9 9 0 1G
VAMM1 10 3 0V
VAMM2 1 8 0V
XLEFT1 500 501 INV
XLEFT2 501 9 INV
VLEFT 500 819 0V

* RIGHT 1
S11 18 19 20 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
VAMM11 20 3 0V
VAMM12 21 18 0V
VRIGHT 502 820 0V
XRIGHT1 502 1 503 NAND
XRIGHT2 503 19 INV
S13 0 21 0 0 PC1_NM1_DU1 L=3.2U W=115.2U
RBIG19 19 0 1G
RBIG 21 0 1G
SRESET 21 800 1 1 PC1_PM1_DU2 L=1.6U W=7.2U AS=22P AD=22P PS=13U PD=13U
VRESET 800 0 DC 5V PULSE(5V 0V 40NS 1NS 1NS 25NS 80NS)

* OTHER RIGHT LEGS
S21 1 0 3 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
S31 1 0 3 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
S41 1 0 3 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
S51 1 0 3 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
S61 1 0 3 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
S71 1 0 3 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U
S81 1 0 3 0 PC1_NM1_DU1 L=1.6U W=4.8U AD=15P AS=9P PD=18U PS=8.4U

* SOURCE FOLLOWER
S14 1 21 22 0 PC1_NM1_DU1 L=1.6U W=576U AD=1750P AS=1050P PD=514U PD=8.4U
CLOAD 22 0 5PF
* CASCODE CURRENT SOURCE
SI1 102 102 103 0 PC1_NM1_DU1 L=7.2U W=9.6U AD=17P AS=29P PD=8.4U PS=22U
SI2 103 103 0 0 PC1_NM1_DU1 L=7.2U W=51.2U AD=98P AS=157P PD=8.4U PS=59U
SI3 1 102 104 0 PC1_NM1_DU1 L=7.2U W=51.2U AD=98P AS=157P PD=8.4U PS=59U
SI4 104 103 0 0 PC1_NM1_DU1 L=7.2U W=51.2U AD=98P AS=157P PD=8.4U PS=59U
SI5 105 104 106 0 PC1_NM1_DU1 L=7.2U W=204.8U AD=376P AS=626P PD=8.4U PS=390U
SI6 106 103 0 0 PC1_NM1_DU1 L=7.2U W=204.8U AD=376P AS=626P PD=8.4U PS=390U
ISFBIAS 200 102 .25M
V200 200 0 5V
VSF 22 105 0V

.NODESET V(102)=5.8215 V(103)=1.8565 V(104)=3.1715 V(106)=.9994 V(22)=2.8
+ V(9)=5V V(19)=0V V(21)=5.047 V(309)=5 V(310)=0
.IC V(21)=5.047 V(102)=5.8215 V(309)=5 V(310)=0

.TRAN .2NS 40NS
.PRINT TRAN I(VAMM1) I(VAMM11)
.PRINT TRAN V(9) V(19)
.PRINT TRAN V(3) V(21) I(VAMMCURR) V(22) I(VSF)
.PRINT TRAN I(VAMM2) I(VAMM12)
.PRINT TRAN V(309) V(310)
.OP
.OPTIONS NOMOD
.OPTIONS RELTOL=.0005
.OPTIONS CHGTOL=1E-15
.OPTIONS ABSTOL=1E-13
.OPTIONS METHOD=GEAR
.OPTIONS TRTOL=20
.PROCESS PC1 FILENAME=HP2
.WIDTH OUT=80
.OPTION ITL5=0
.END
```

TABLE 4.1. SPICE input file for the circuit of Figure 4.16.

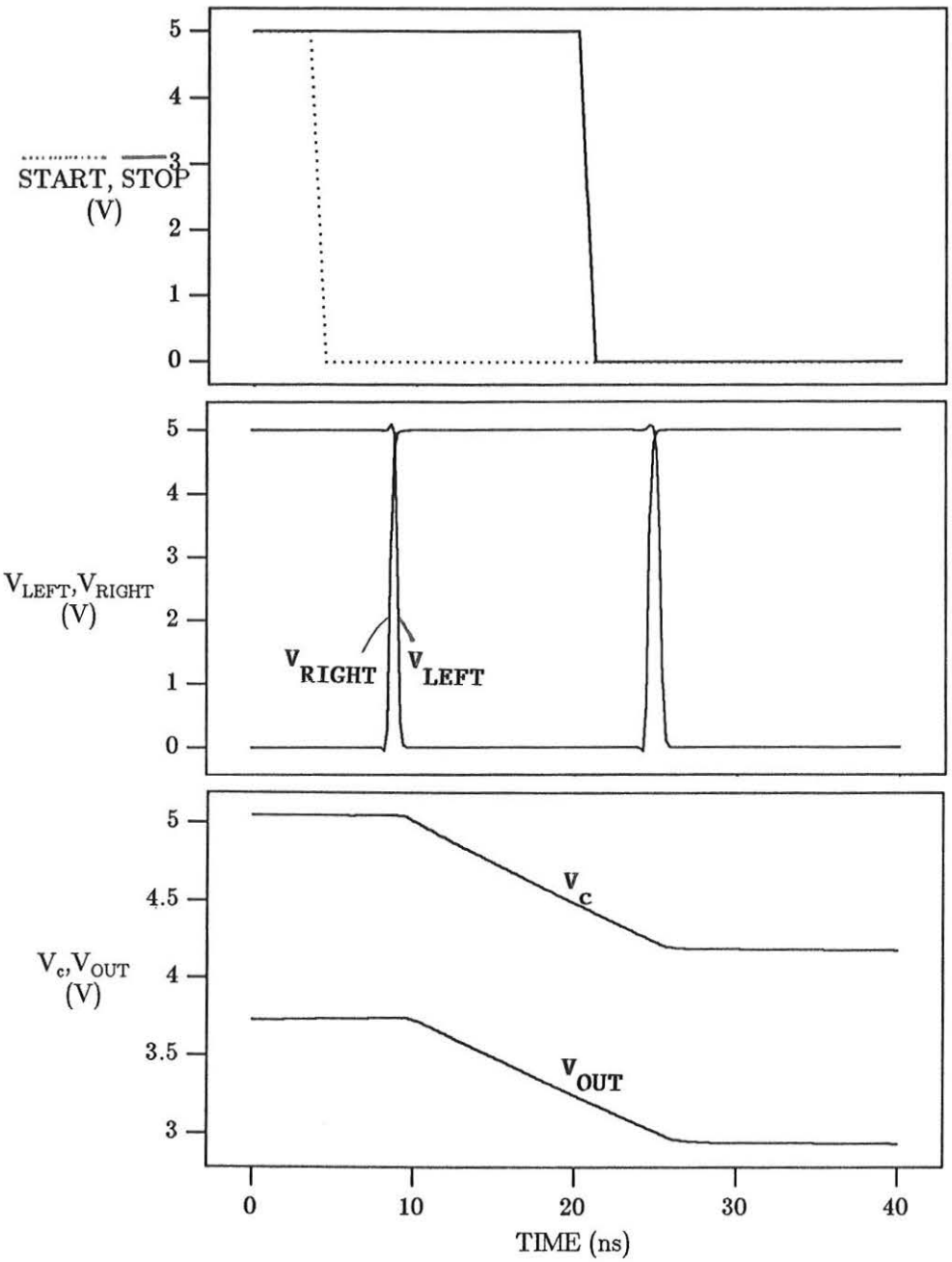


Figure 4.17. Simulation results for the circuit of Figure 4.16. Note that V_{OUT} now refers to the output of the source follower, and V_c refers to the capacitor voltage.

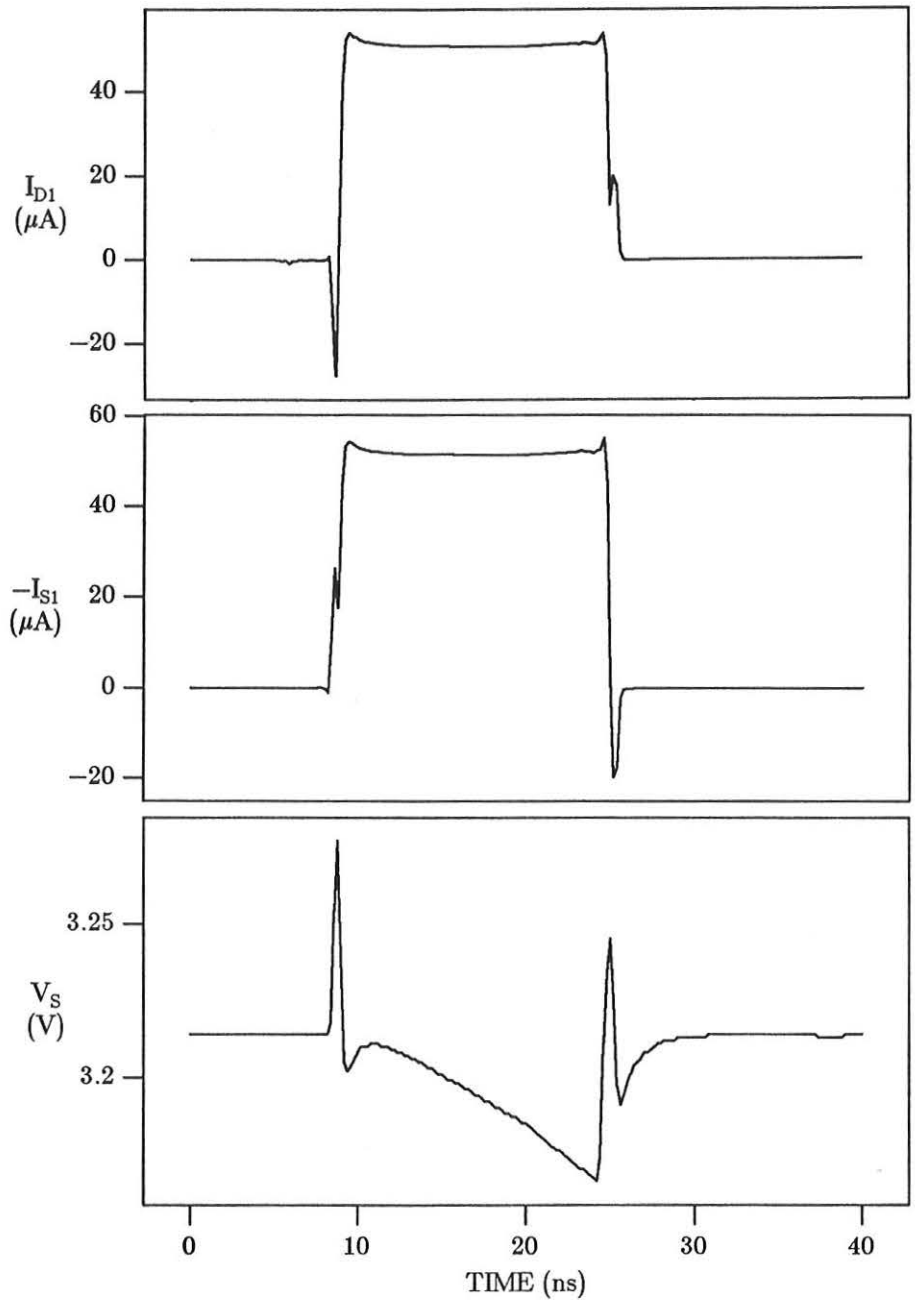


Figure 4.18. Simulation results for the circuit of Figure 4.16. The large current spikes will be due to the feedthroughs from the fast rise times at the transistor gates.

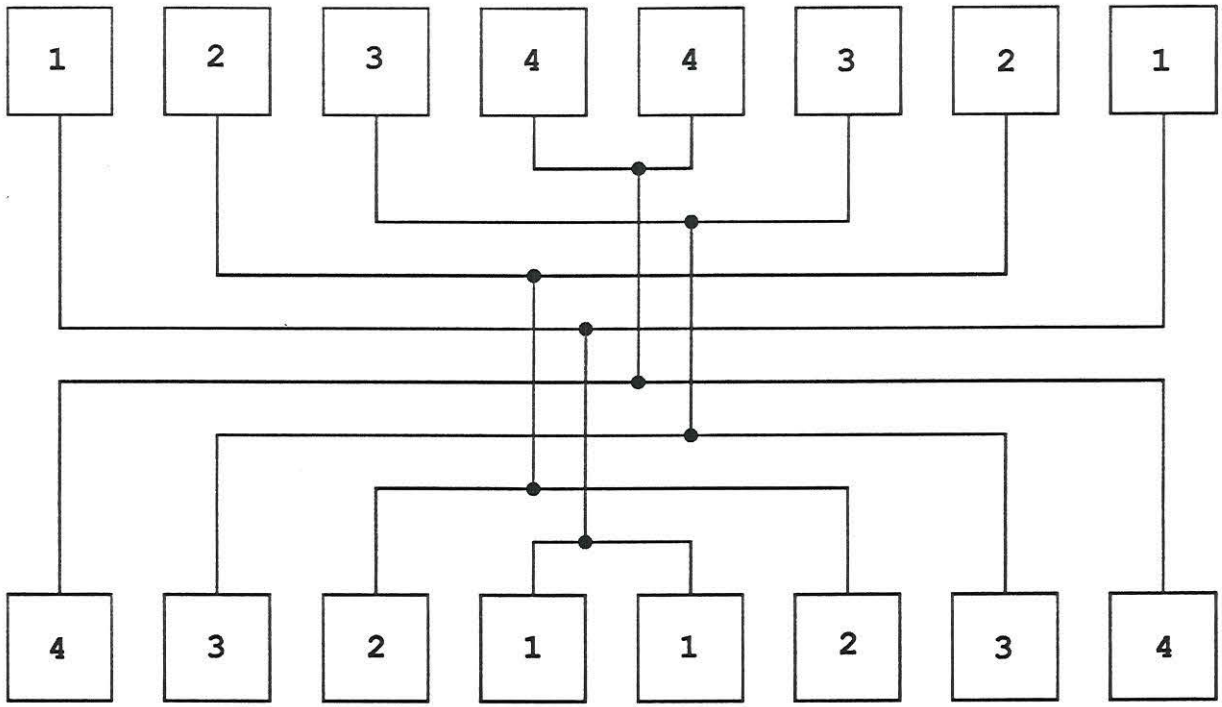


Figure 4.19. The common-centroid layout scheme used for the capacitors. In this case, four capacitors have been laid out by splitting each capacitor into four smaller capacitors, and then laying out each set around the same centroid. Note that the total interconnection wire lengths are identical for each channel.

centroid, it will be matched as closely as possible. This becomes important as more capacitors are added to the array and the interconnect capacitance begins to add a non-negligible component to the total hold capacitance.

It should be noted that the input capacitance of the source follower added a sizable component to the total hold capacitor. In fact, the contribution was about 50%, or 0.5 pF. Thus, the MOS capacitor needed to be sized for a total capacitance of 0.5 pF. The total capacitance of a capacitor-connected transistor will be

$$C_G = (W - \Delta W)(L - \Delta L)C_{OX} + W \times C_{gdo} + W \times C_{gso} + L \times C_{gbo}$$

where ΔW and ΔL are the variations in the transistor dimensions due to the lateral diffusion and undercut, and C_{gdo}, C_{gso}, and C_{gbo} are overlap capacitances. The capacitor was laid out in four pieces each with $W/L = 28.8/3.2$.

A plot of the actual layout of the capacitors and source follower input transistors is shown in Figure 4.20.

4.3.2 Complete Layout The complete layout for the 8-channel TVC/Analog Memory is shown in Figure 4.21, with parts identification in Figure 4.22. The total area was 1100 μm $\times$ 1500 μm. General layout techniques were as follows:

1. Guard rings were used generously to separate digital and analog circuit components. The guard rings consisted of p+ on p substrate (tied to ground), and n+ on n-well (tied to V_{DD}).

2. Separate digital and analog power and ground lines were used throughout.

3. Power and ground lines were made wide to minimize ohmic drops.

4. Substrate contacts were used generously to minimize substrate currents and thus eliminate the possibility of latchup.

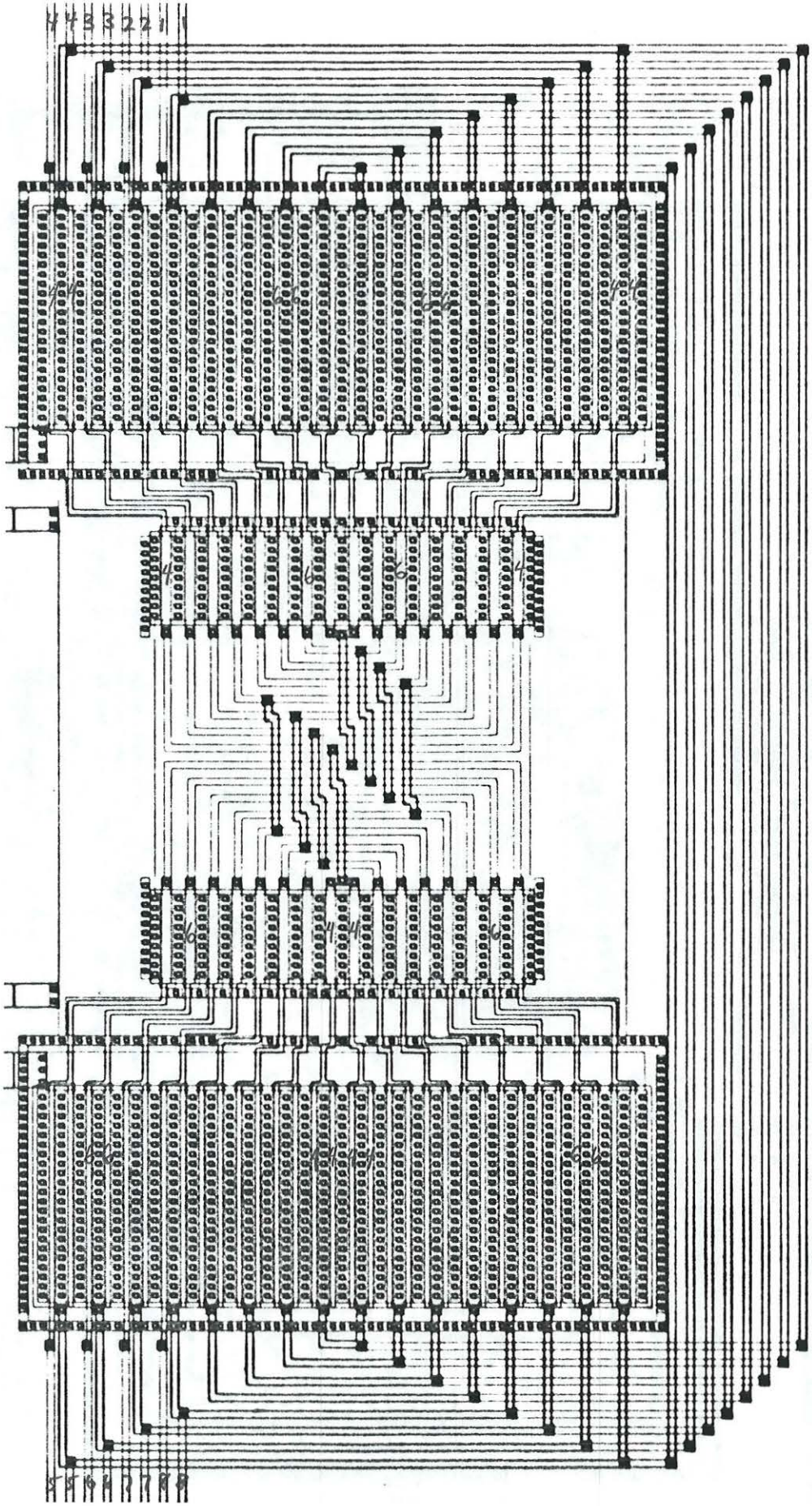


Figure 4.20. Layout plot of the capacitors and source follower input transistors.

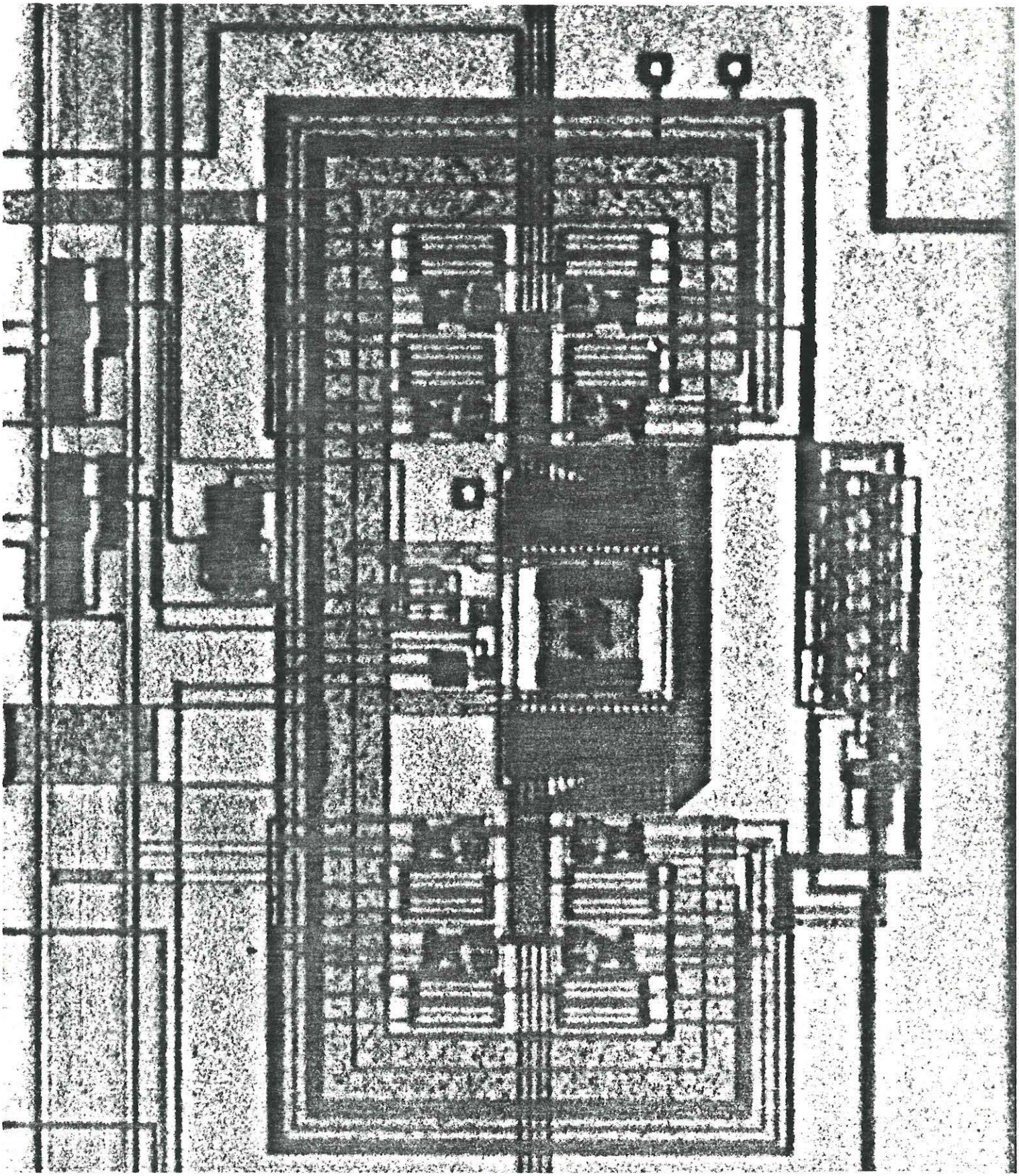


Figure 4.21. Photomicrograph of the 8-channel TVC/Analog Memory. Total area 1100 μm $\times$ 1500 μm.

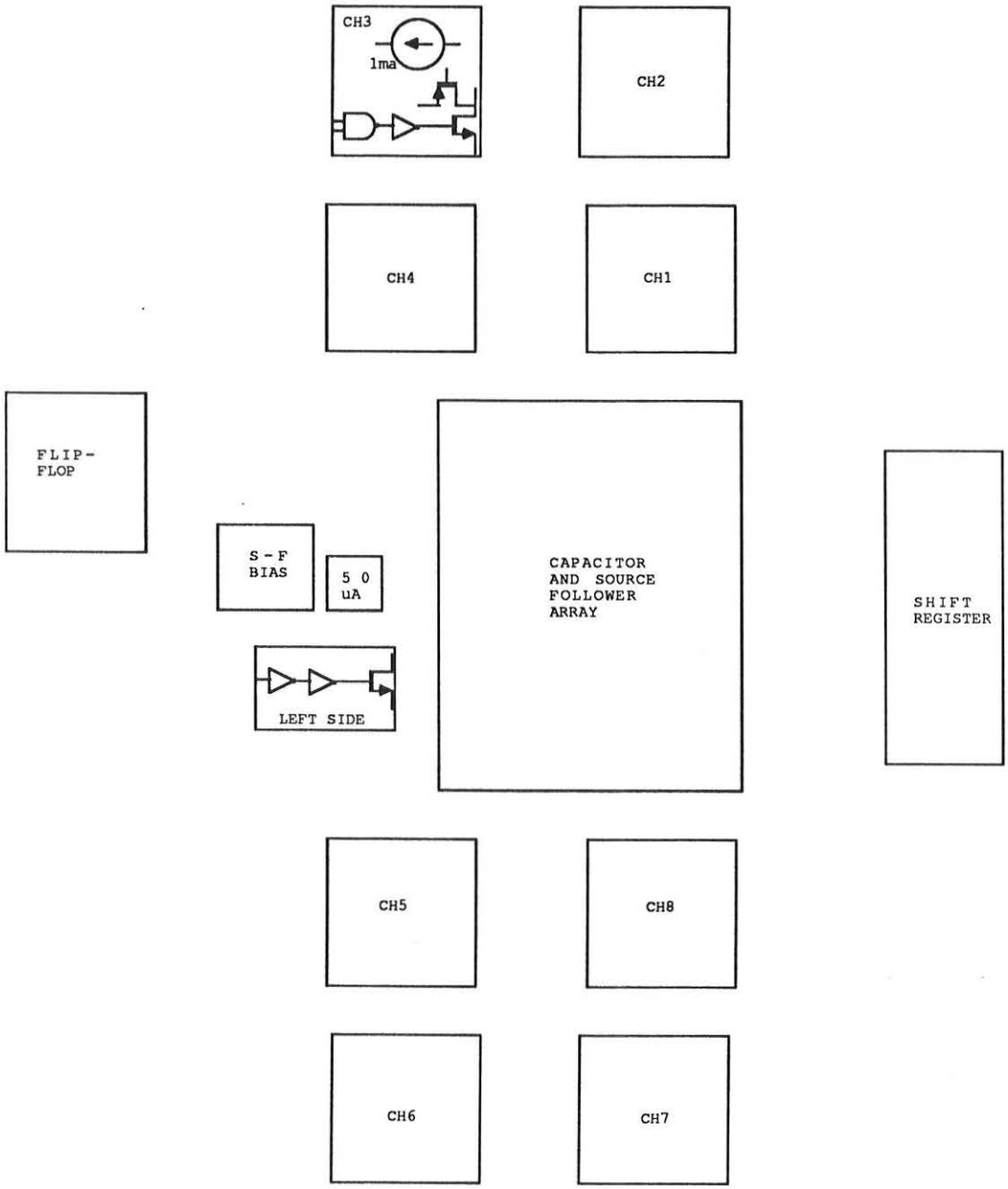


Figure 4.22. Parts identification for the 8-channel TVC/Analog Memory layout.

4.3.3 Complete Die A large number of test structures were laid out on the same die as the 8-channel TVC/Analog Memory. These structures included the various bits and pieces which comprised the TVC, such as the flip-flop, current source, source follower, etc. Thus, a complete characterization of each system part was available in case of a layout error in the main circuitry. Fortunately, however, the layout was error free. A photomicrograph of the die is in Figure 4.23. A complete list of the test structures on the chip is found in Appendix C.

4.3.4 Input/Output Pads Special pad circuitry was designed for both the analog and digital inputs and outputs to the chip. This circuitry provided two functions: it protected the packaged device from electrostatic discharge (ESD), and it provided buffering for the digital signals. The ESD protection circuit is shown in Figure 4.24, and consists of a resistor and two diodes. These components provide a safe path for any unexpected voltage spikes. The resistor was fabricated from a strip of polysilicon, and the diodes were made from diffusion-substrate junctions.

For the digital inputs to the chip, large inverter-buffers were placed between the pad and the actual circuitry (Figure 4.25a) in order to minimize external noise and transients. For outputs, an inverter chain allowed the low-power internal signals to drive the high off-chip capacitances (Figure 4.25b). Note the lack of protection diodes in this circuit; this function is automatically provided by the drain-substrate junctions of the final inverter.

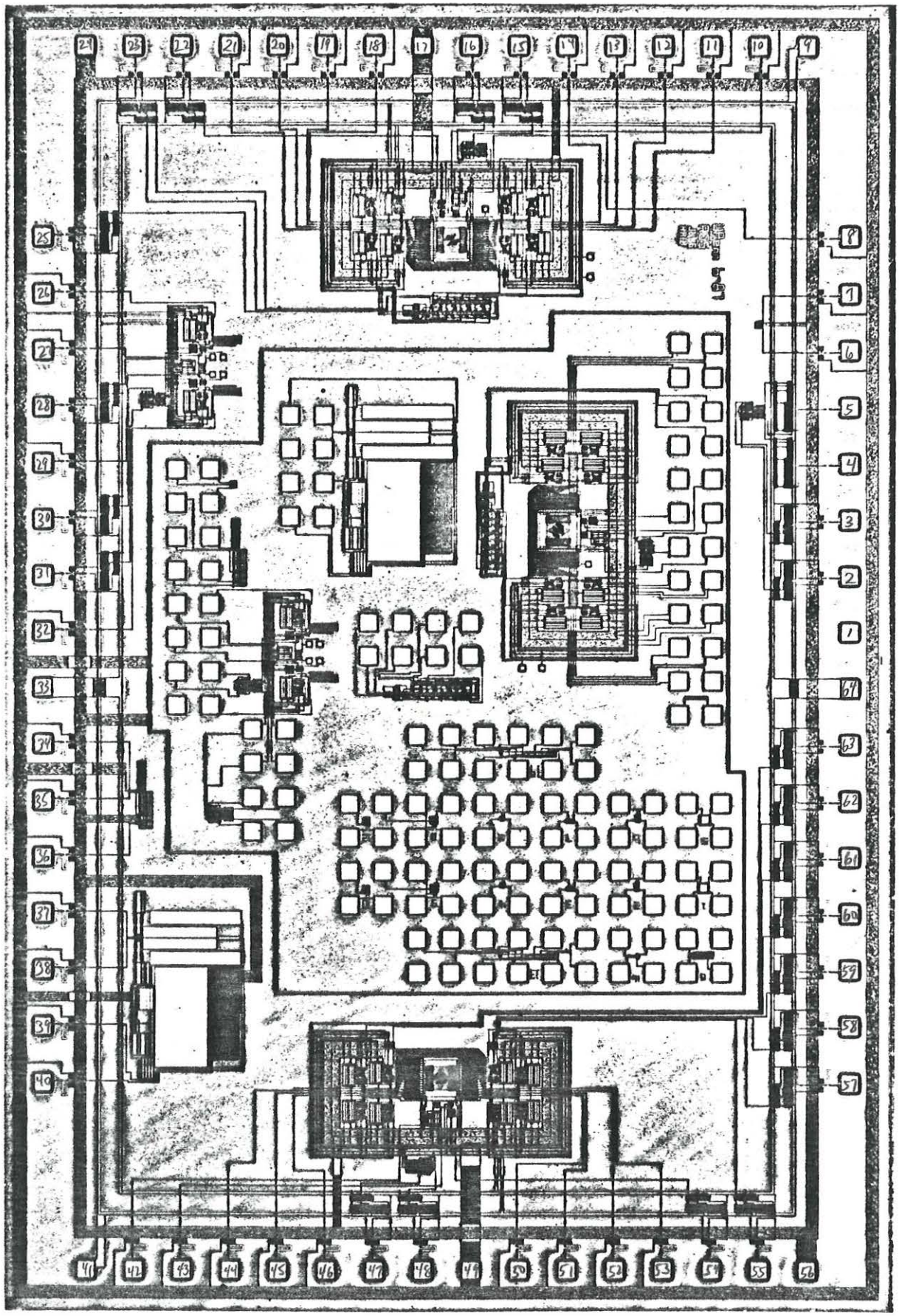


Figure 4.23. Photomicrograph of the entire die. Note that for test purposes, the entire 8-channel structure was repeated three times on the die. Total area 4600 μm $\times$ 6800 μm.

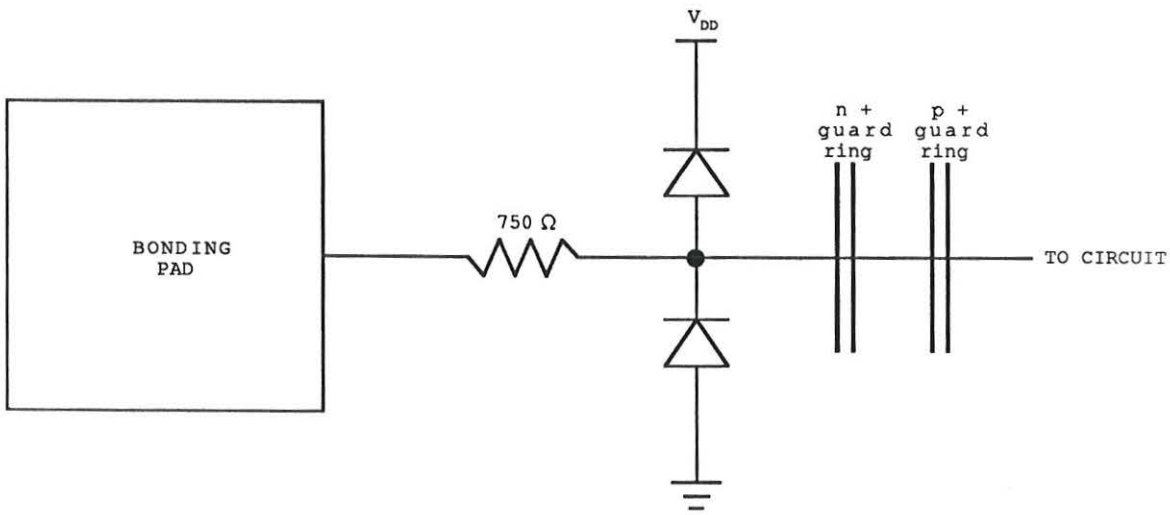
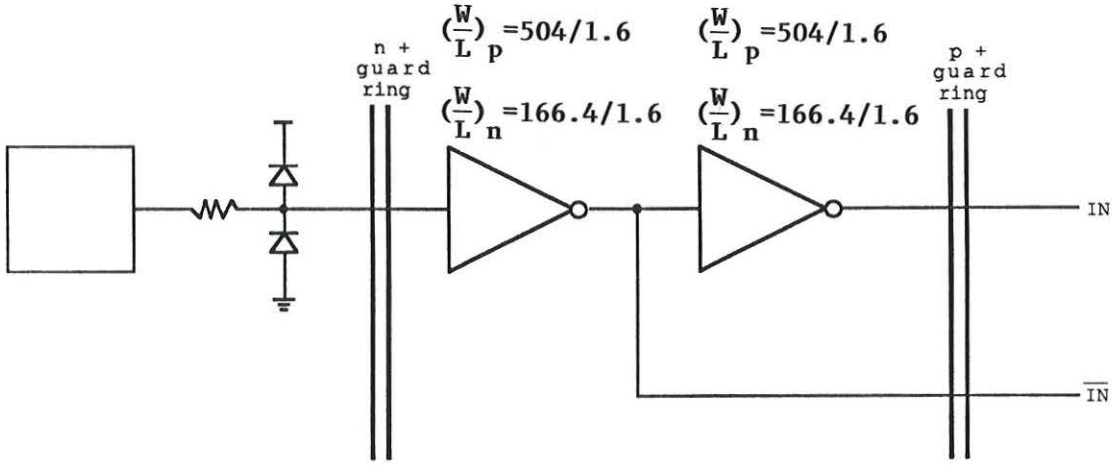
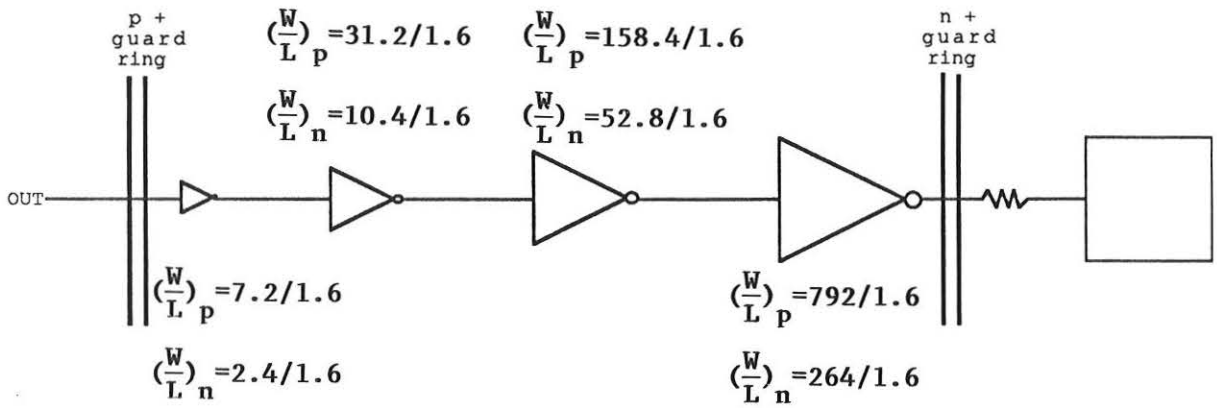


Figure 4.24. Pad protection circuitry to guard against electrostatic discharge and latchup.



(a)



(b)

Figure 4.25. (a) Digital input pad buffer. (b) Digital output pad buffer. Note the lack of diodes; these are automatically provided by the drain-bulk diffusions of the last inverter in the chain.

5. Testing

5.1 Test Set-up

An automated system was designed to test the TVC/Analog Memory device. The system consisted of two fast pulse generators, a programmable delay line, a digitizing oscilloscope, and a personal computer to control the instruments via IEEE-488 bus. A schematic of the test system is shown in Figure 5.1.

The most difficult task of the system was to generate the required START and STOP pulses for time measurement. These signals needed to have fast rise times in order to provide the required resolution of 0.5 ns. This was accomplished by using Tektronix PG502 pulse generators, which have output rise times of <1 ns. However, these pulse generators are simple, non-programmable devices. In order to generate a variable time delay between the two fast pulsers, an HP 8160A pulse generator was used as a delay element. In normal mode, this equipment runs a reference output at a specified frequency and provides two programmable signals which may be delayed from the reference by 0 to 99 ns in 100 ps steps. This was ideal for our application. Thus, the reference output (actually called the "trigger" output on the front of the machine) and one programmable output were used to trigger the two PG502's. The second programmable output was used as the reset and shift register clock. The relative timing of these signals is shown in Figure 5.2.

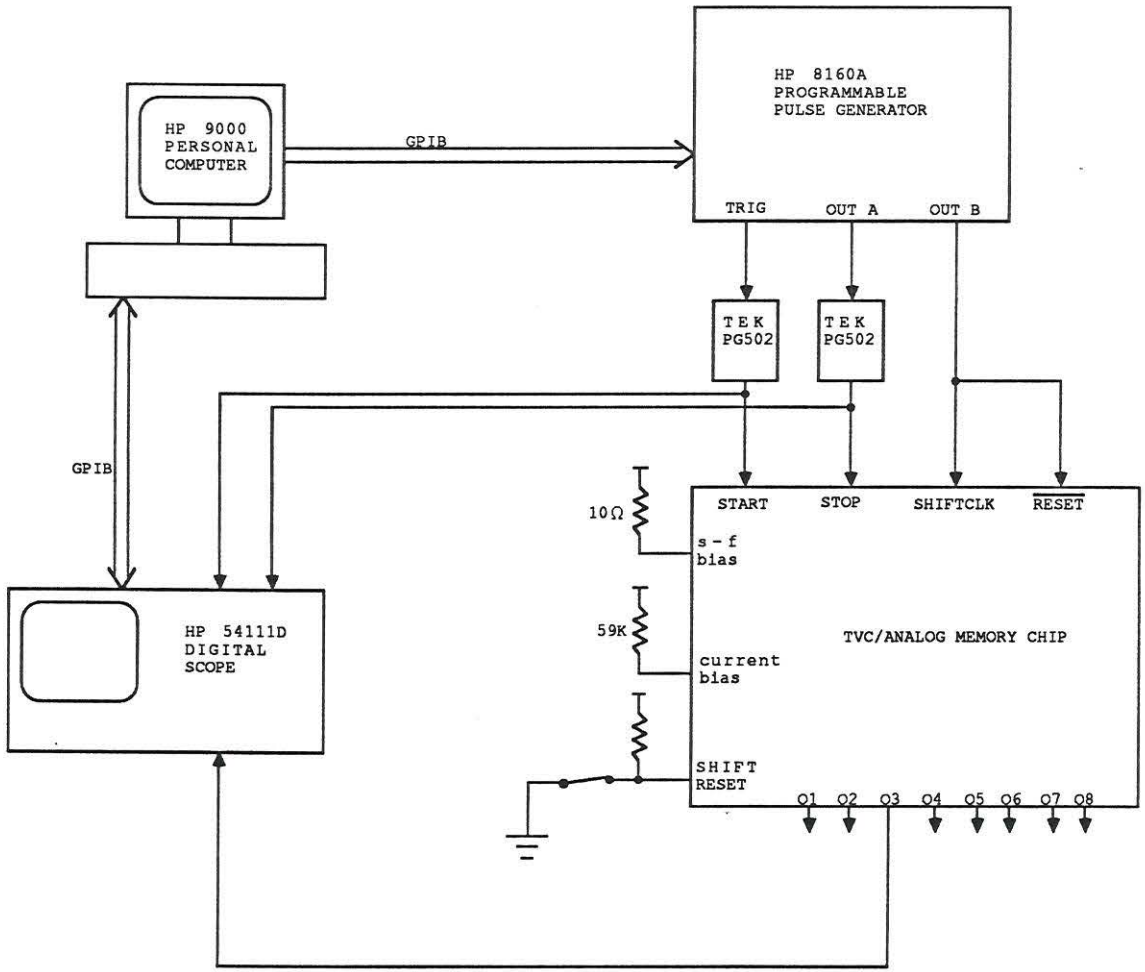


Figure 5.1. TVC/Analog Memory test system.

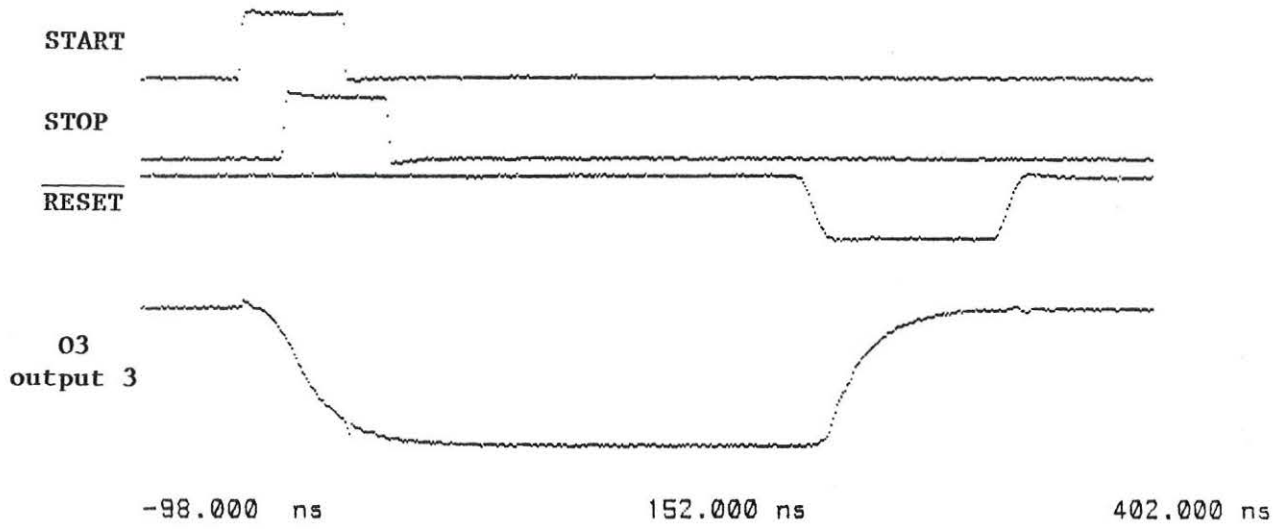


Figure 5.2. Test system signal timing.

Generation of the proper pulse timing produces a periodic output signal which is observable on an oscilloscope. The HP 54111D scope was chosen because of its high sampling rate of 1 GSample/sec. This allowed direct observation of the output waveform without the averaging which is inherent in scopes with lower sampling rates. Also, because of the high single-shot bandwidth of the scope (250 MHz), noise frequency and amplitude were easily readable.

A Hewlett-Packard Series 9000 personal computer was used to control the 8160A pulse generator and the 54111D scope via IEEE-488 bus. This allowed automated testing on individual channels of the analog memory by varying the time delay and then reading the resulting voltage from the scope. However, tests involving differences between the eight analog memory channels had to be done manually due to the necessity of physically moving the scope probe to the different channel outputs.

5.2 Description of Tests Performed

Several series of performance measurements were done on the TVC/Analog Memory device. Conceptually, these measurements can be thought of as separate tests on the TVC and Analog Memory functions of the chip. Additionally, tests were performed on the various test structures which were in the package, including a current source and a source-follower.

5.2.1 TVC Measurements Linearity of the TVC was measured by running a set of known delays through the circuit and then measuring the output voltage for one of the eight channels. The programmed input range was 5 to 35 ns in 0.5 ns steps. The output voltage was measured by taking the difference between the initial reset voltage and the final output voltage; this allowed any constant offsets between channels to be canceled out. The final data is presented as a plot of ΔV vs. time.

The second major measurement on the TVC was the resolution, i.e. how small of a variation in the time interval could be measured. This effect was observed by using the persistence capability of the storage scope. An arbitrary delay was programmed, and a range of output voltages was recorded on the scope screen. By measuring the variation in the output voltage for a given interval, the minimum interval spacing could be measured.

Measurements were made for the nominal current source value of 50 μA as well as 20 μA and 5 μA.

5.2.2 Analog Memory Measurements Tests of the Analog Memory were performed to check the matching between the eight channels and to examine possible unwanted interactions between channels. Capacitor leakage currents were also examined. Testing of channel matching involved setting an arbitrary input delay and then measuring the initial and final voltages at each of the eight outputs. This test was repeated on many samples of the chip in order to observe constant and random effects within the run.

Testing of interactions between channels required slightly modifying the test set-up (Figure 5.3). In this case, the HP 8160A was operated in "burst" mode. This allowed generation of eight START and STOP pulses before reset was performed, permitting observation of possible changes on a "held" voltage by the other channels (Figure 5.4). The modified set-up also required an additional pulser to generate a single reset signal after the eight time measurements.

5.2.3 Current Source Measurements Tests on the 50 μA current source involved finding DC characteristic and and subsequently the output resistance of the source.

5.2.4 Source-Follower Measurements Because the n-well process prevents connection of the substrate to the source on the source-follower transistor, the backgate effect

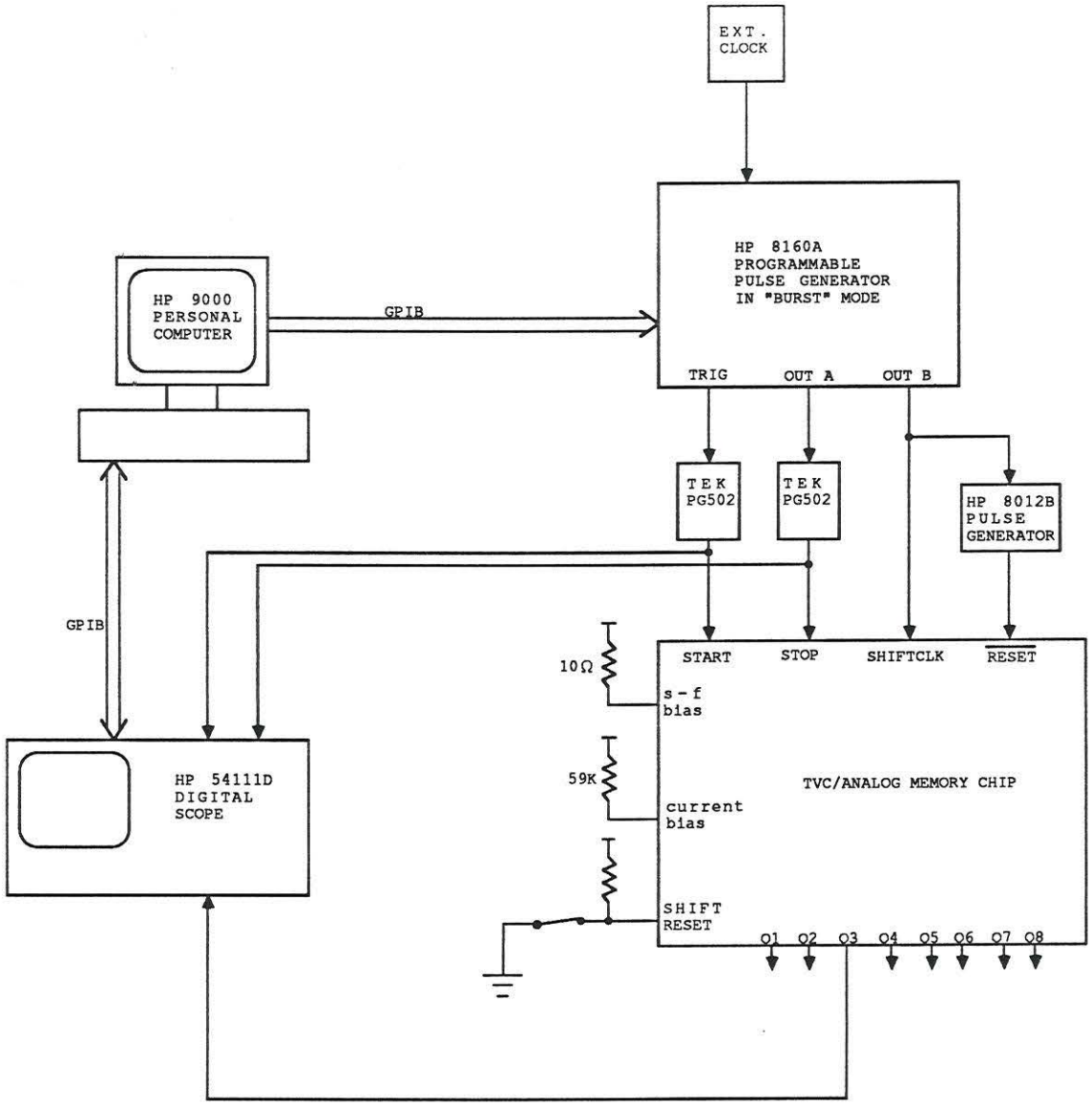


Figure 5.3. Modified test system to observe inter-channel interactions.

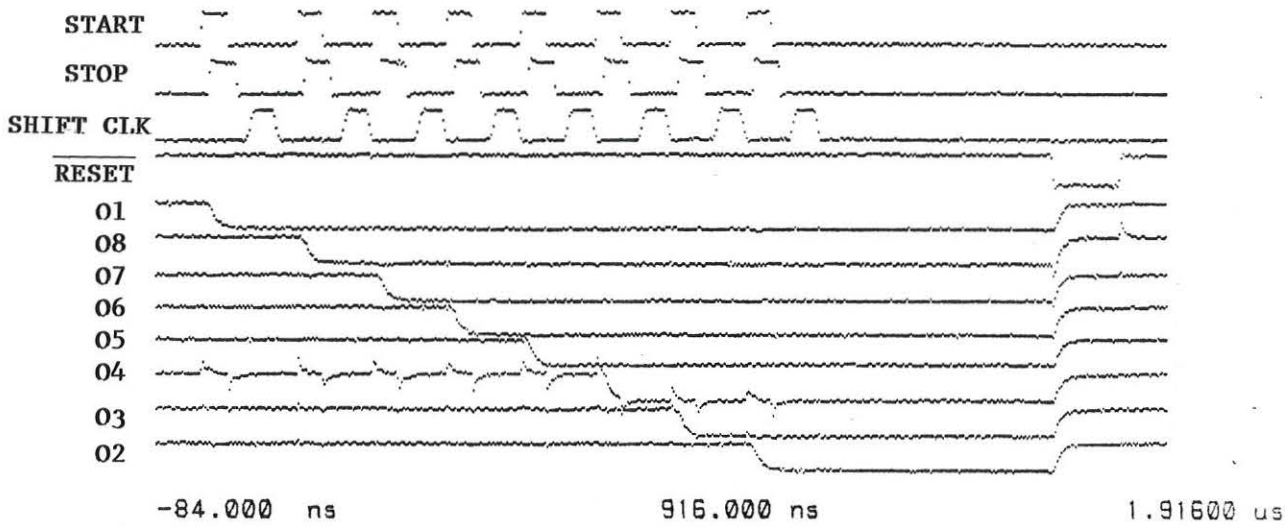


Figure 5.4. Signal timing for modified test system. All eight memory channels were charged before resetting.

will cause the source-follower to be slightly non-linear. Since the source-follower is used as the output buffer in the TVC, its non-linearity must be measured so that its effect can be canceled when calculating TVC linearity. Thus, a DC analysis was done on the source-follower. Additionally, frequency response and slew rate analyses were performed.

5.3 Test Results

5.3.1 TVC Linearity of the time-to-voltage converter was measured to be highly accurate over an input range of 7 to 25 ns. A plot of time vs. voltage is shown in Figure 5.5. It should be noted that the plotted output voltages have been averaged over many samples in order to cancel noise effects. Several observations can be made from this plot concerning circuit operation. At small time intervals, i.e. around 5 ns, linearity is poor. This corresponds to the domination of the non-linear behavior right around switching; as the time interval increases, all of the transients will have enough time to die away, and linear measurement will take place.

As the time interval approaches 25 ns, the slope of the transfer characteristic changes. This corresponds to the region of operation where the final capacitor voltage is below $V_{DD}-V_T$. This means that the right transistor is in the linear region during a portion of the time interval. Since the voltage across the current source will vary noticeably during linear transistor operation, the charging current will be affected and a non-constant integration will take place on the capacitor. This in turn causes a change in slope on the transfer curve. We can also use the 25 ns limit to approximate a transistor threshold voltage of $V_T = 1.25$ V. At first glance, this value may seem high, but remember that the body effect will cause all circuit thresholds to be high.

Linear regression was performed on the linearity plots with results shown in Figure 5.6. Regressions are computed for the entire test range of 5—35 ns, and also for the

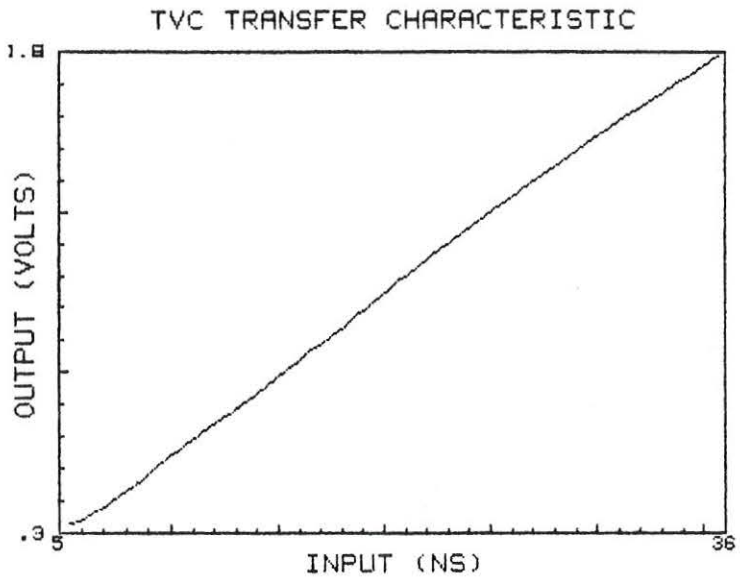
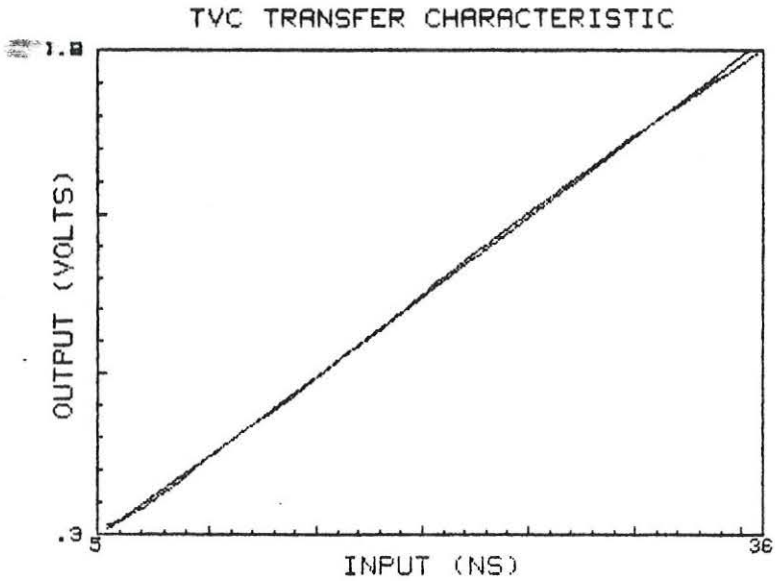
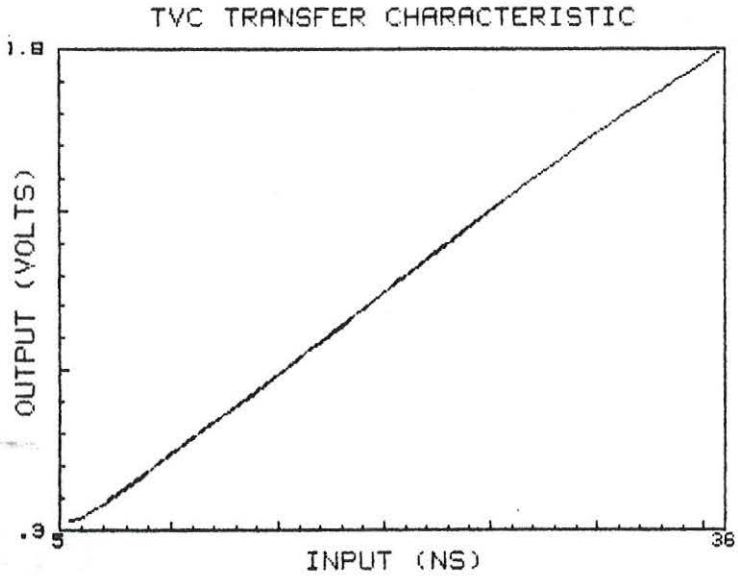


Figure 5.5. TVC transfer characteristic. The TVC was highly linear over the input range of 7—25 ns.



SLOPE= .0495887100559
INT= .0466937287144
CORR= .999719913373
AVERAGE RMS ERROR= .00133404233785

(a)



SLOPE= .0510164312233
INT= .024720791276
CORR= .999933505315
AVERAGE RMS ERROR= .000531650293544

(b)

Figure 5.6. Linear regression performed on the TVC transfer characteristic. (a) Regression performed over 5—35 ns. (b) Regression performed over 7—25 ns.

highly linear range of 7—25 ns. For the second case, the plotted regression line is almost indiscernible from the data line, and the average RMS error is about 0.5 mV.

The resolution of the TVC was directly related to the external noise from the laboratory environment. A picture of TVC noise is shown in Figure 5.7. In this photo, the scope persistence is infinite, meaning that all triggers are stored and displayed on the screen simultaneously. From the picture, the width of the waveforms can be read to be about 10 mV. This corresponds to a resolution of about 0.2 ns, well with the design target of 0.5 ns. The main source of the noise was found to be a 25 kHz component coming from the switching power supply inside the HP 8160A pulse generator (Figure 5.8). When the pulser was removed from the test system and replaced by a passive delay consisting of a piece of transmission line, the noise was reduced by about 30% (Figure 5.9). The lesson here was that the power supply rejection of the TVC was almost nil. This was because of resetting of the capacitor voltage to V_{DD}; any noise on V_{DD} was propagated through the system as a first order effect. A better method in a future design might be to add a bandgap reference on the chip in order to provide a solid reset voltage.

When the charging current was varied, the usable input time range could be adjusted to any reasonable value. For example, linearity plots for currents of 5 μA and 20 μA are shown in Figure 5.10. Clearly, the TVC is linear over the extended range, but noise levels will remain the same as for the 50 μA case. Thus, resolution will be lower; a minimum of 2 ns for 5 μA, and 0.5 ns for 20 μA.

Notes on lab testing: It was important to take an oscilloscope reading of the actual time delay fed to the circuit, rather than trust the value programmed into the HP 8160A. The variation between the programmed and measured interval was about $\pm 2\%$, which was enough to affect linearity measurements. There was also a problem with unintentional feedback between adjacent pins on the DIP chip carrier. For example, the output 1

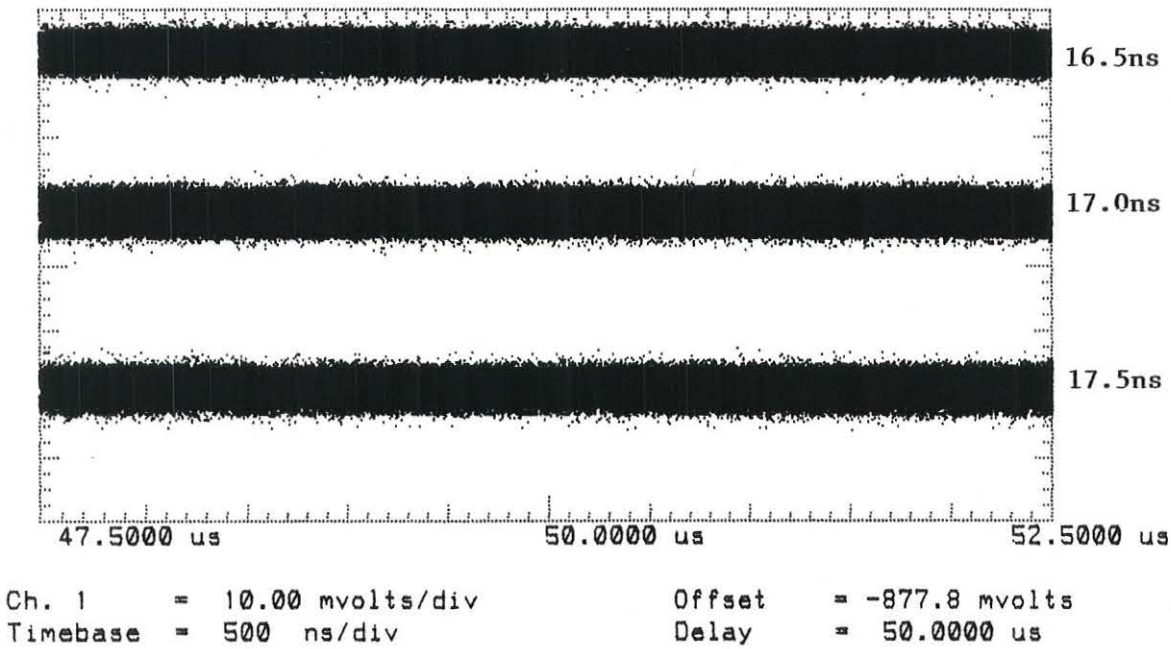


Figure 5.7. TVC resolution. These traces are approximately 1000 TVC samples for each of three different time intervals, with the scope set on infinite persistence.

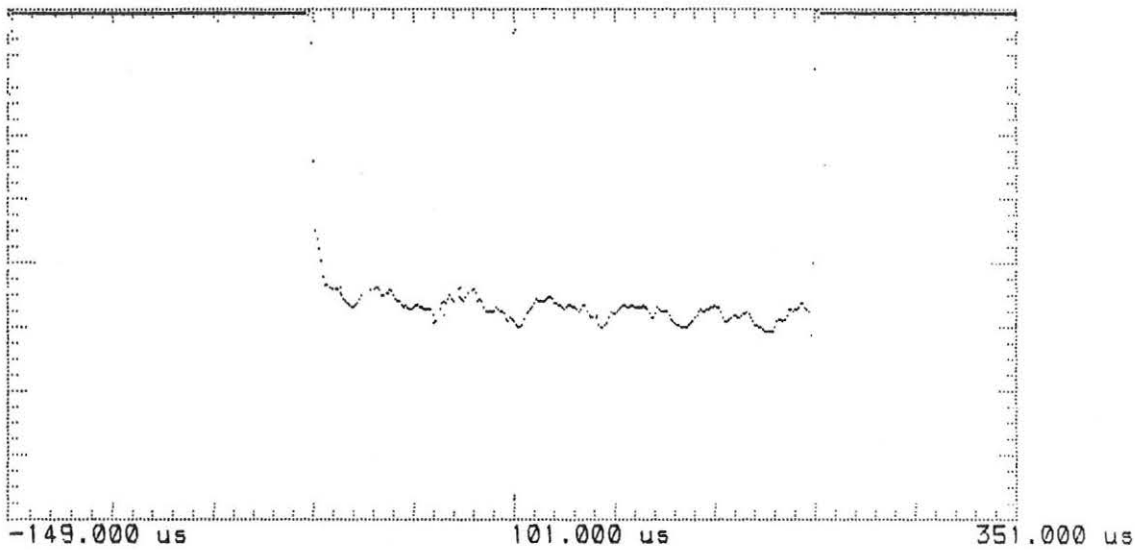


Figure 5.8. Single-shot scope measurement of a TVC output. 25 KHz noise from faulty test equipment is clearly visible.

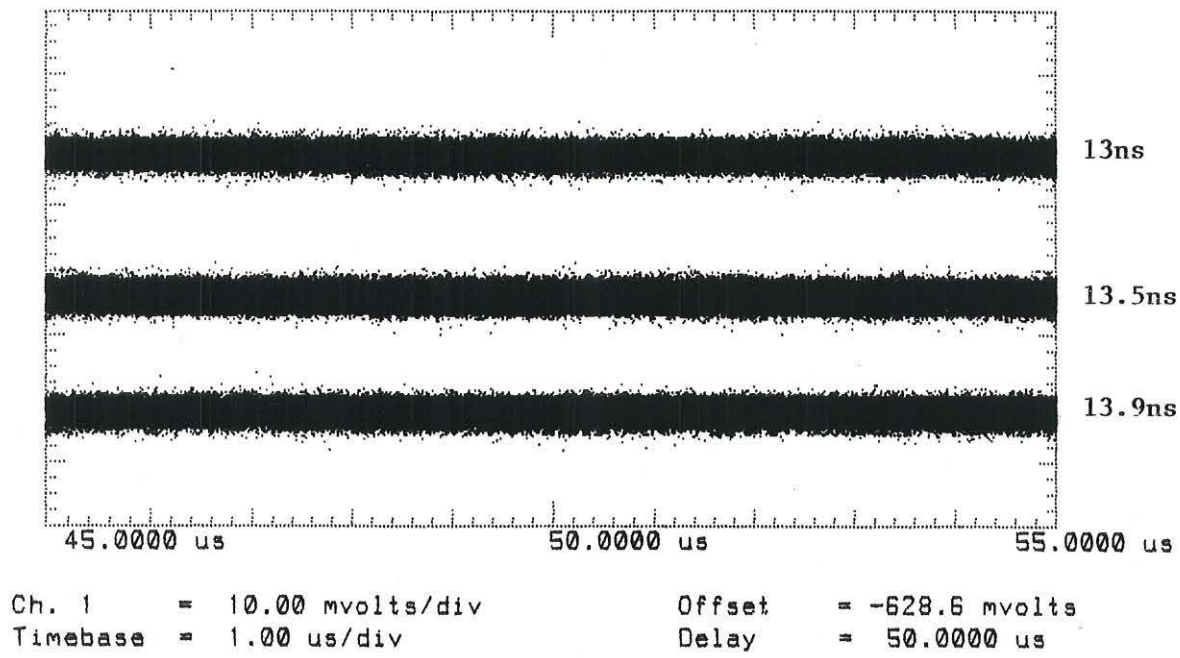
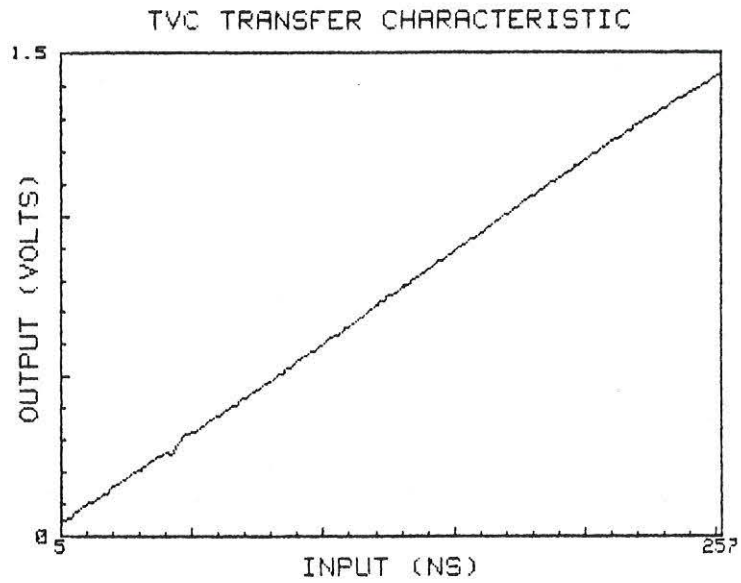
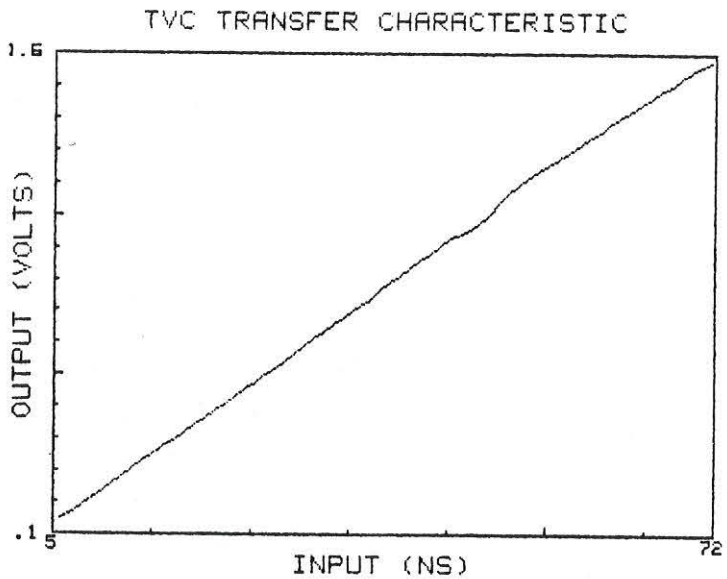


Figure 5.9. TVC resolution with the faulty test equipment removed from the set-up. Results are approximately 30% better.



(a)



(b)

Figure 5.10. TVC transfer characteristics for different currents. (a) $I_o = 5\ \mu$A. (b) $I_o = 20\ \mu$A.

pin of the analog memory was situated next to the biasing pin for the 50 μA current source. This caused a feedback loop such that the linearity on TVC channel 1 was radically different than the linearity on the other seven channels. The problem was solved by adding a bypass capacitor to ground on the bias pin.

5.3.2 Analog Memory Matching of linearity between the eight analog memory channels was quite excellent. This is reflected in the plot in Figure 5.11, where it is evident that the eight slopes are identical and that the total variation from channel to channel does not exceed 20 mV. This is well within the desired specification for time resolution of 0.5 ns. Be reminded, however, that these results are for the *change* in output voltage, i.e. the difference between the initial and final capacitor voltages as seen through a source follower. Thus, any offset variations between channels will not be reflected in the linearity data. For example, a scope picture of the initial output voltages of the eight channels before any time measurement is shown in Figure 5.12. We can see that all of the channels start off at slightly different voltages, and two of the channels (4 and 5) are noticeably off from the rest. This two-channel discrepancy was repeatable on all of the devices from the foundry, and although no satisfactory explanation has yet been formulated, it is assumed that layout was to blame. It should be noted that the offset is strictly a DC measurement, and therefore load and inter-channel capacitances were not to blame. Similar behavior is found on the output voltages after a time interval is measured. For example, Figure 5.13 shows the eight channels after an interval of 15 ns. The same offset behavior is observed, i.e. channels 4 and 5 are still off by the same amount.

Capacitor hold time is shown in Figure 5.14, where the capacitor voltage is shown over a range of 4 ms. The decay can be read as approximately 2 mV/ms, translating to a leakage current of 2 pA. These numbers are typical for CMOS.

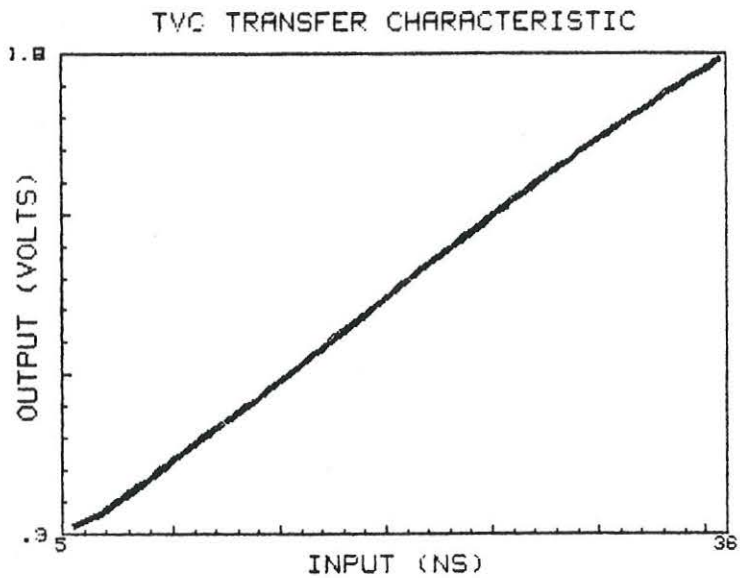


Figure 5.11. Transfer characteristic for all eight channels of the TVC/Analog Memory. Note the maximum spread of about 20 mV between the channels.

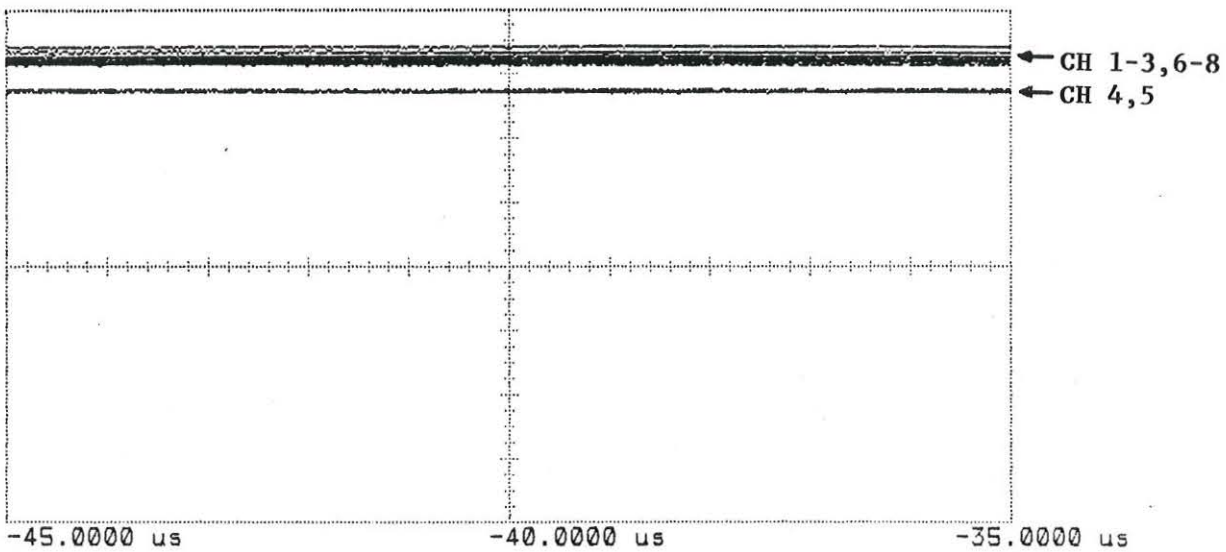


Figure 5.12. Initial voltages on the eight channels of the Analog Memory. Two of the channels (4 and 5) were consistently off from the rest.

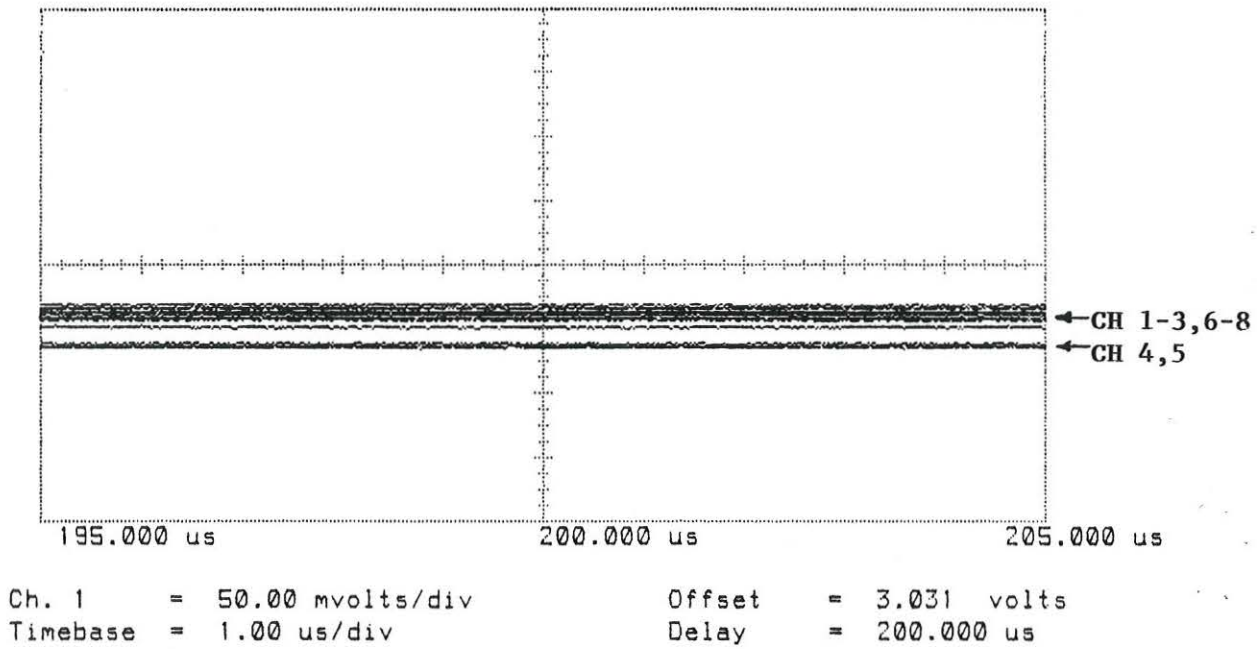


Figure 5.13. Final voltages on the eight channels of the Analog Memory after eight time measurements of 15 ns. Note that channels 4 and 5 are still off by the same amount.

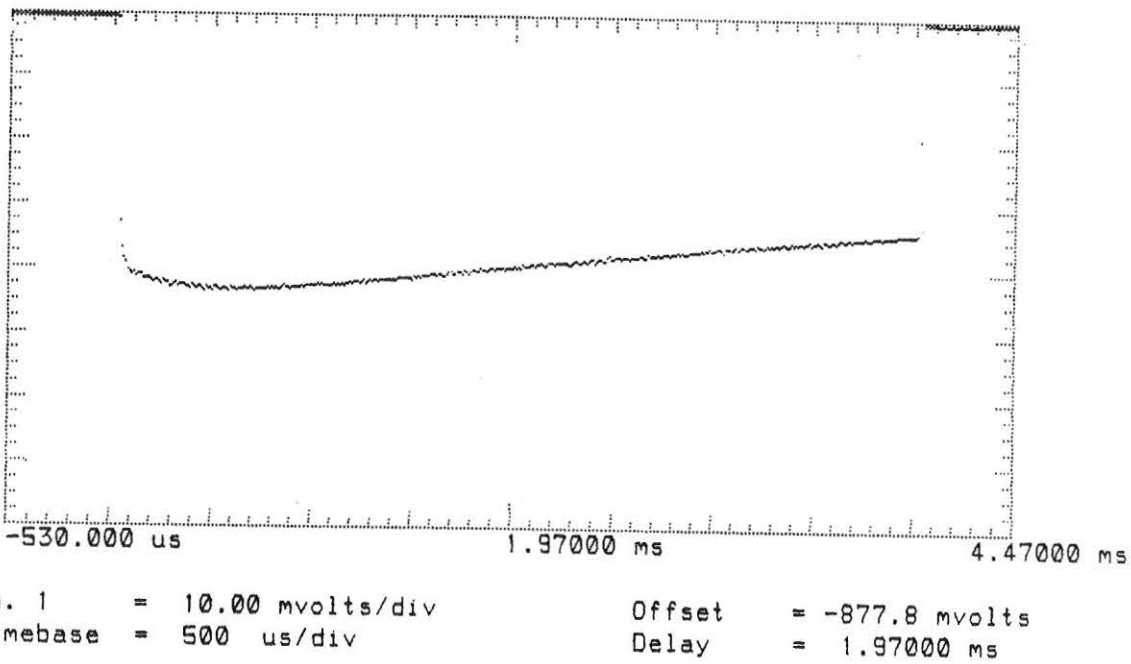


Figure 5.14. A look at the long-term hold capability of the capacitors. Over a period of 4 ms, the voltage decays by about 8 mV at room temperature.

Inter-channel effects appear as an error voltage on one channel due to a changing voltage on another channel. For example, suppose that a voltage is held on channel 3, and a subsequent time measurement is made on channel 4. The output voltage on channel 4 will drop by some appropriate amount, but a side effect of this change will be an error feedthrough onto channel 3. This comes about due to coupling from inter-wire capacitances, especially at the output pads and bonding wires. A simplified equivalent circuit for this effect is shown in Figure 5.15. From the figure, it is clear that charge leaks to the channel 3 hold capacitor via the inter-pad capacitance on the adjacent channels. Scope photos of this effect are shown in Figure 5.16. It should be noted that if channel 4 were to subsequently return to its initial voltage, an opposite feedthrough effect would appear on channel 3, and the total error voltage would be zero. However, this situation could not occur on this chip because the capacitor reset is common for all channels.

5.3.3 Current Source The DC characteristic of the current source is shown in Figure 5.17. The actual measurement is more favorable than the simulation; this simply reaffirms past observations that SPICE is very poor at predicting output resistances of transistors [25]. The measured r_o was 1.5 MΩ.

5.3.4 Source Follower The DC characteristic of the source follower is shown in Figure 5.18. The follower is very linear over the input range of 3—5 V with a gain of 0.93, as seen by the regression. This indicates that the body effect is very small; a simple back-calculation using Equation 2.6 gives a measured γ of 0.32 $\sqrt{\text{V}}$.

A comparison of the simulated and measured characteristics of the follower is shown in Figure 5.19. The relatively low response of ~12 MHz is an indication of the impossibility of directly mirroring the capacitor voltage due to its high-frequency components at the corners (the areas around turn-on and turn-off).

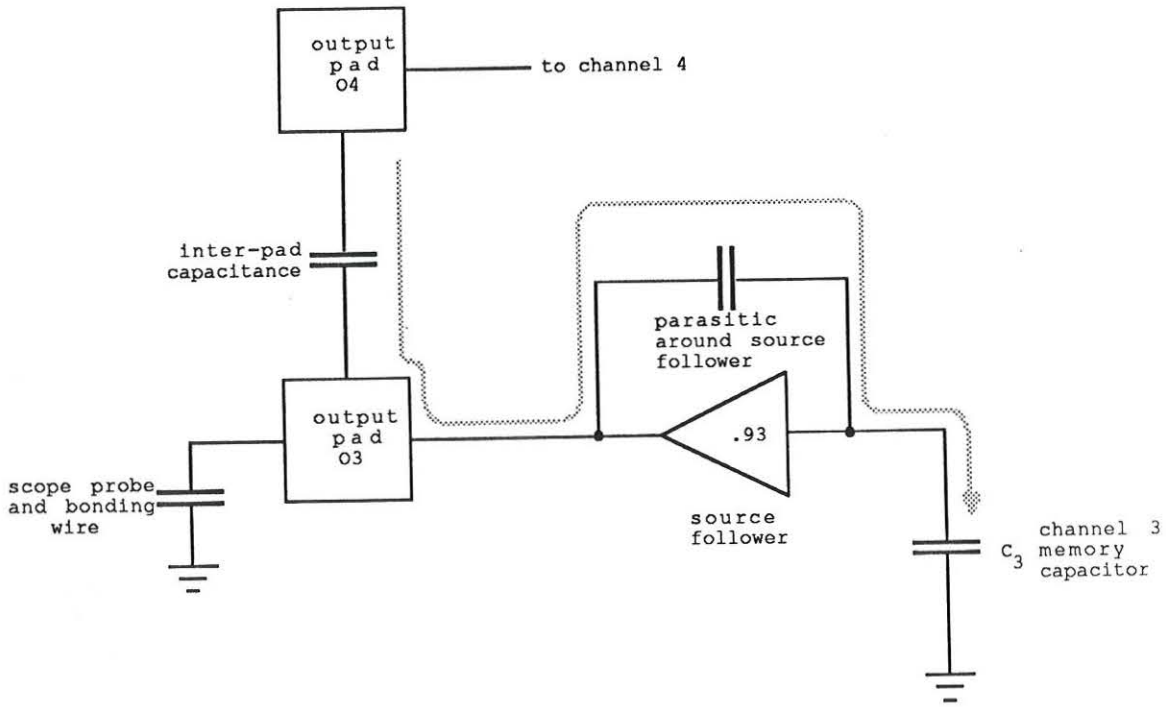
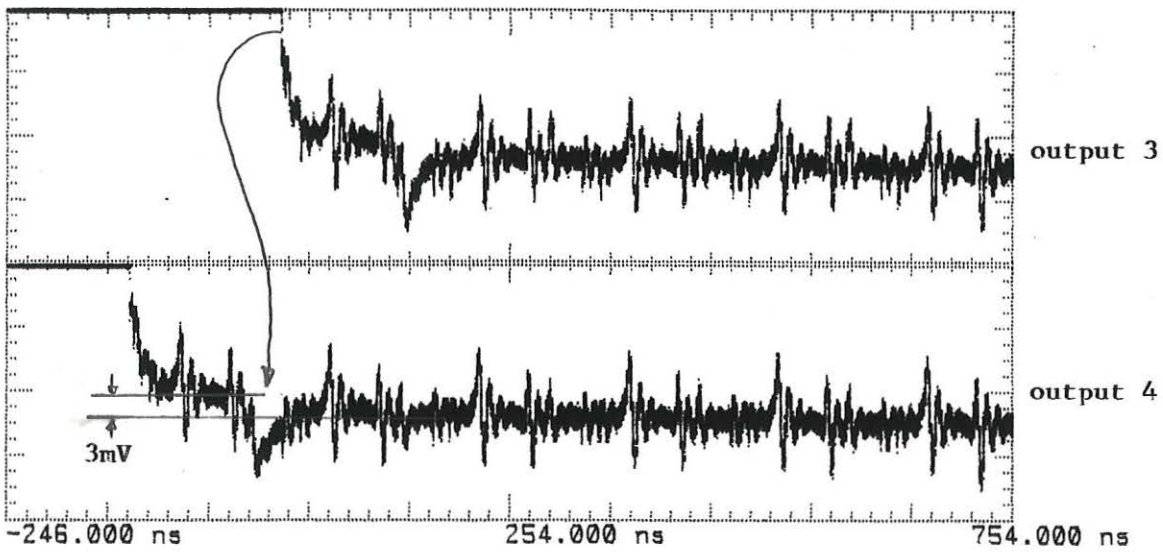
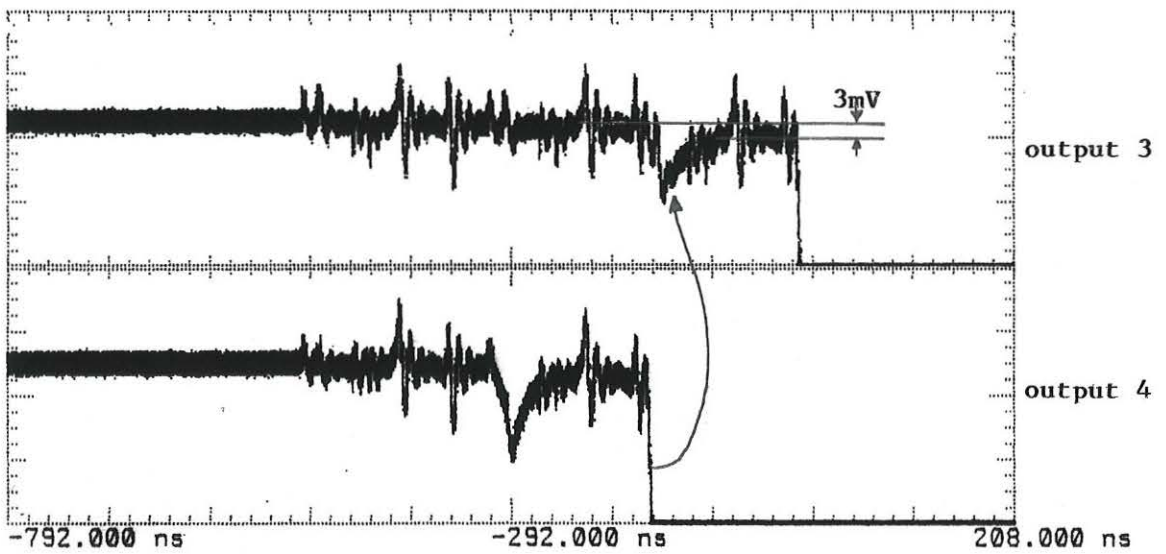


Figure 5.15. A model of the feedthrough effects between adjacent channels. The dotted line represents error charge which affects C_3 due to a voltage change at $O4$.



Ch. 1	=	40.00 mvolts/div	Offset	=	-810.4 mvolts
Ch. 2	=	40.00 mvolts/div	Offset	=	-815.0 mvolts
Timebase	=	100 ns/div	Delay	=	254.000 ns



Ch. 1	=	40.00 mvolts/div	Offset	=	49.40 mvolts
Ch. 2	=	40.00 mvolts/div	Offset	=	36.80 mvolts
Timebase	=	100 ns/div	Delay	=	-292.000 ns

Figure 5.16. Inter-channel effects. (a) A fall on channel 3 will cause a 3 mV change on the final voltage on channel 4. (b) A fall on channel 4 will cause a 3 mV change on the initial voltage on channel 3.

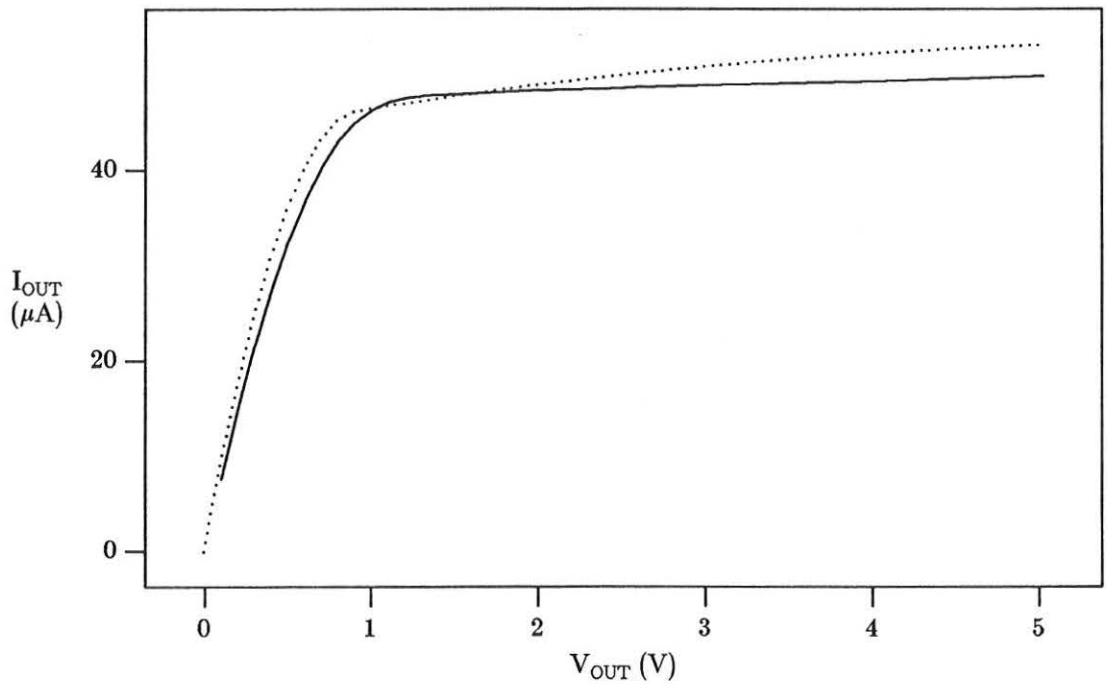
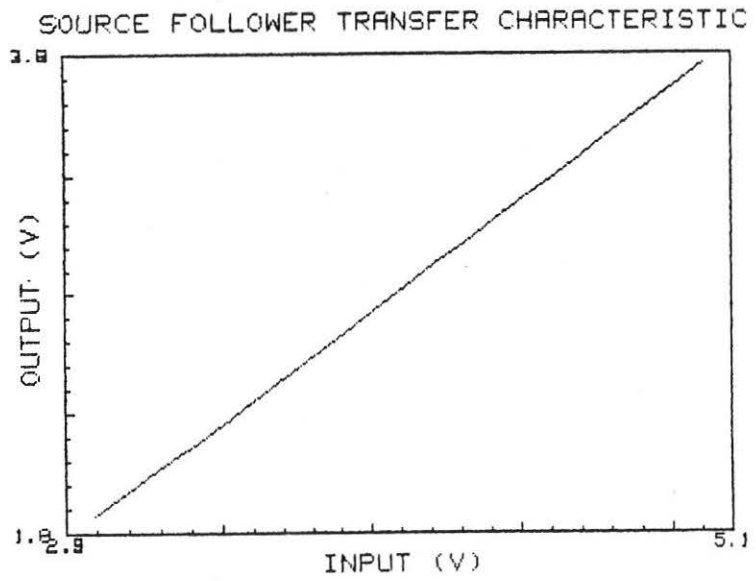
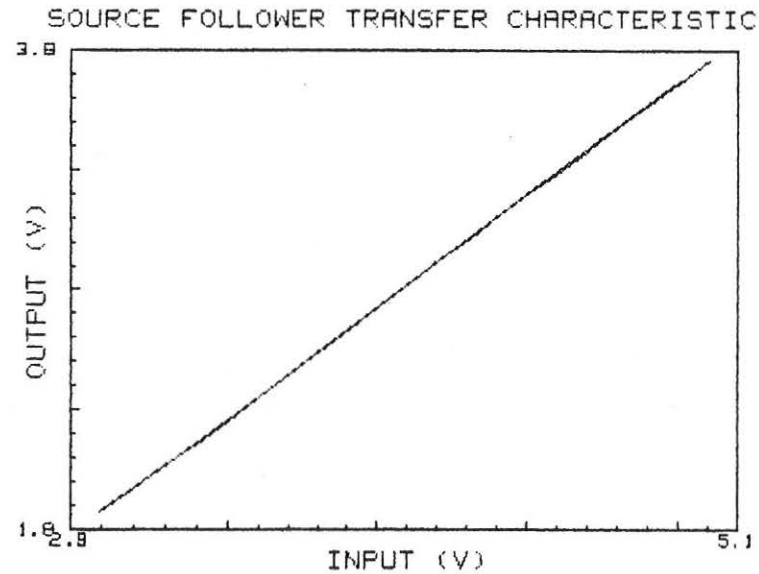


Figure 5.17. Laboratory measurement (*solid line*) and SPICE simulation (*dotted line*) of the 50 μA current source DC characteristic. SPICE is consistently poor at predicting transistor output resistances.



(a)




```
SLOPE=       .936693164617
INT=        -.928700489797
CORR=        .999986168772
AVERAGE RMS ERROR=    .000640768470685
```

(b)

Figure 5.18. (a) Measured DC characteristic of the source follower over its useful input range of 3–5 V. (b) Linear regression.

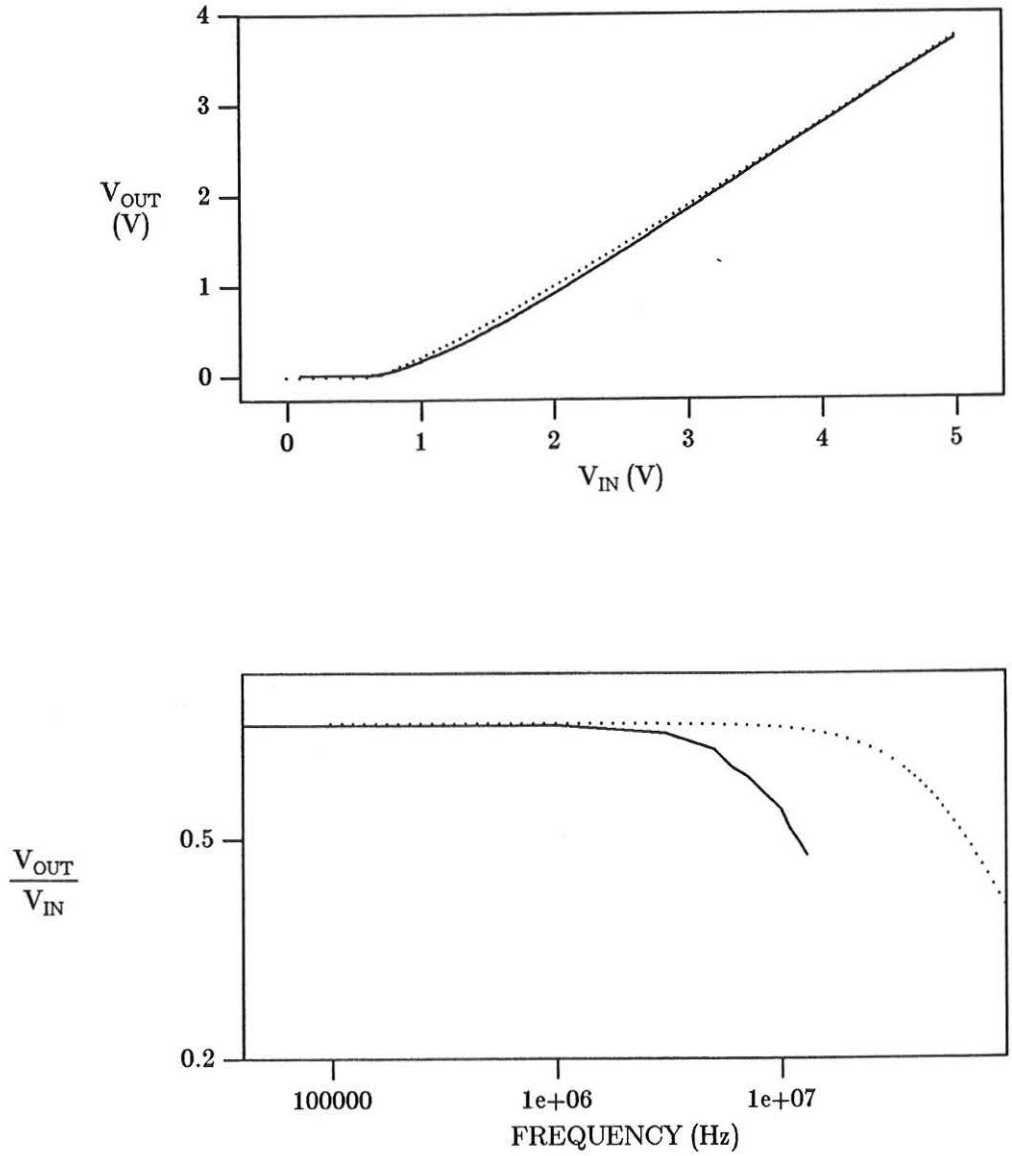


Figure 5.19. DC characteristic and frequency response of the source follower showing laboratory measurements (*solid line*) and SPICE simulations (*dotted line*).

The slew rate of the source follower was measured by overdriving the input with a fast square wave. Negative slew was limited by the biasing current and also gave an approximation of the capacitive load at the output. Results are shown in Figure 5.20. The slew rate of 17 mV/ns with a current of 0.4 mA indicates a load capacitance of 24 pF. This total includes contributions of 8 pF from the scope probe and 16 pF from the pad, bonding wire, and surrounding environment.

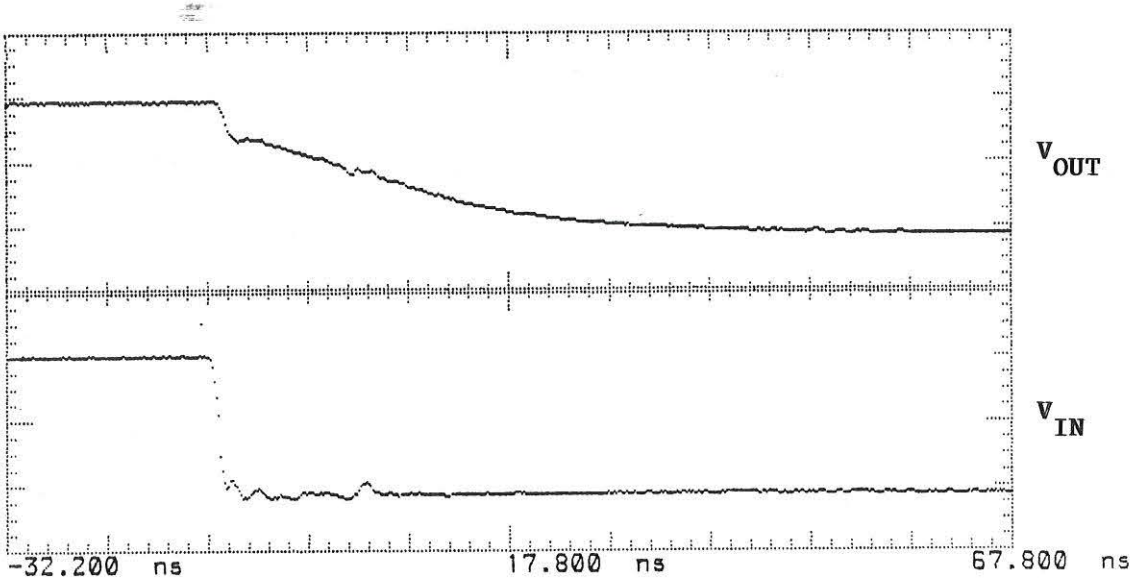
5.3.5 Chip-to-Chip Variations and Yield Twenty samples of the TVC/Analog Memory chip were laboratory tested. The following measurements were performed on each sample:

1. Measurement of initial reset voltage on each memory channel.
2. Measurement of final voltage on each channel for input intervals of 15 and 20 ns.
3. Measurement of output voltage of the source follower test structure for input voltages of 4 and 5 V.
4. Measurement of output current of the current source test structure for an imposed output voltage of 3 V.

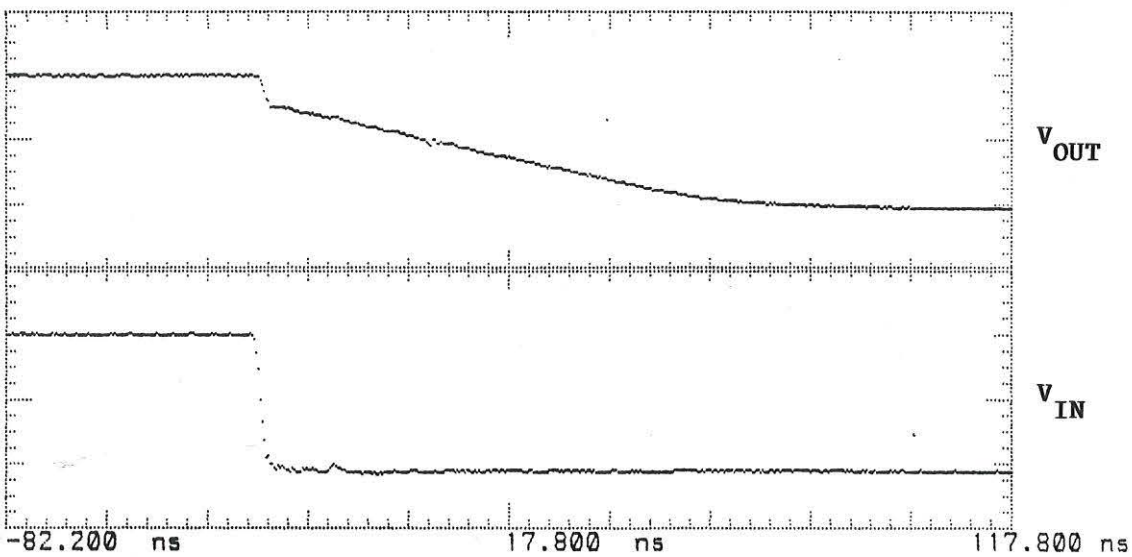
This data is presented in the form of histograms for the following quantities:

1. TVC initial voltage before any time measurement on two of the eight channels (Figure 5.21)
2. TVC transfer ratio in mV/ns for one of the eight channels (Figure 5.22)
3. Gain of the source follower test structure (Figure 5.23)
4. Output current of the current source test structure (Figure 5.24)

The TVC initial voltage plots are for one of the six "normal" channels (1–3,6–8), and one of the two "abnormal" channels (4–5). These two histograms demonstrate two

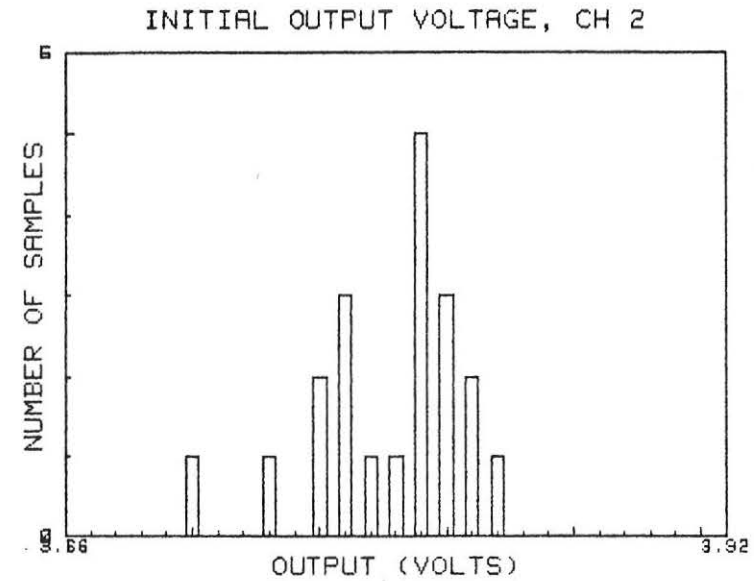


Ch. 1 = 1.000 volts/div Offset = 2.680 volts
Ch. 2 = 1.000 volts/div Offset = 3.810 volts
Timebase = 10.0 ns/div Delay = 17.800 ns

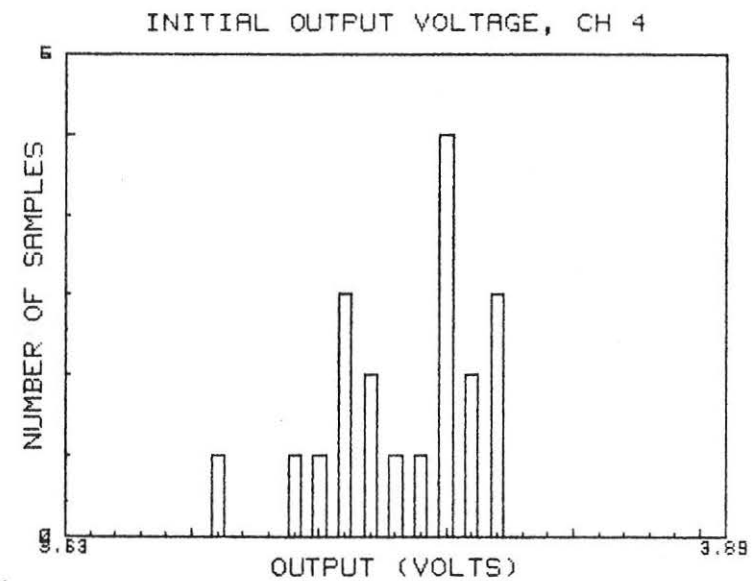


Ch. 1 = 1.000 volts/div Offset = 2.680 volts
Ch. 2 = 1.000 volts/div Offset = 3.810 volts
Timebase = 20.0 ns/div Delay = 17.800 ns

Figure 5.20. Source follower slew rate. (a) $I_D = 1$ mA. (b) $I_D = .4$ mA. By measuring the slope of the fall, the load capacitance can be estimated at about 24 pF.



MEAN= 3.788
STD DEV= .02883



MEAN= 3.763
STD DEV= .03018

Figure 5.21. Histograms of the initial output voltages before time measurement on channels 2 and 4. 20 chips were measured.

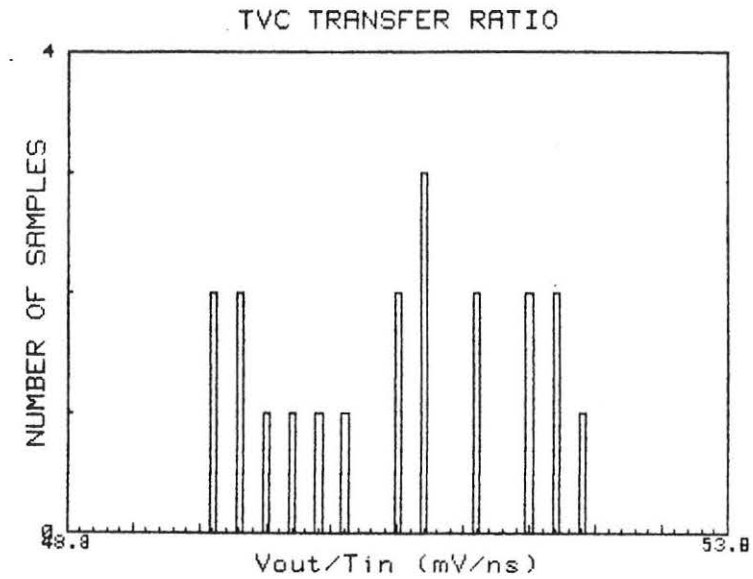


Figure 5.22. Histogram of the TVC transfer ratio for channel 6 of an 8-channel TVC/Analog Memory. 20 chips were measured.

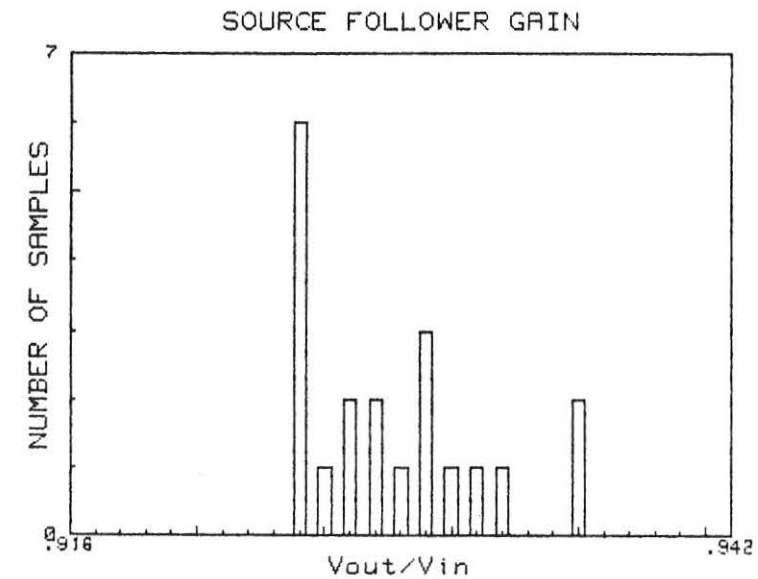


Figure 5.23. Histogram of the gain of the source follower test structure for 20 chips.

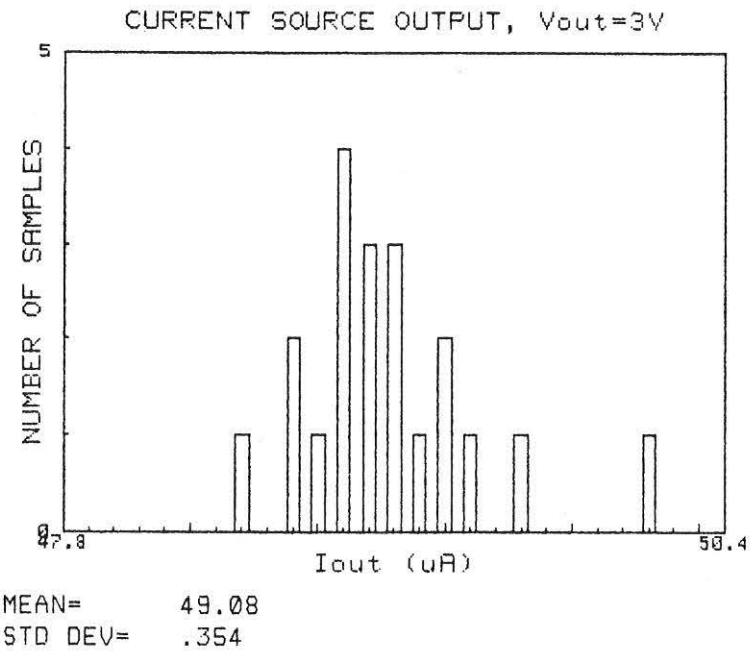


Figure 5.24. Histogram of the output current of the current source test structure for 20 chips.

things: (1) that the variation among the two different channels was approximately the same (note similar standard deviations); and (2) that the offset difference between channels 4 and 5 and the other six channels was approximately equal for all the chips (note similar histogram shapes).

We can get a nominal value for the magnitude of the hold capacitor by using the mean values for the TVC transfer ratio, the source follower gain, and the current source. Applying Equation 1.6,

$$C = \frac{I_o}{(dV_{OUT}/dt)} A_o \tag{5.1}$$

Assuming that $dV_{OUT}/dt = 51.32$ mV/ns, $I_o = 49.08$ μA, and $A_o = 0.929$, $C = 0.888$ pF. We can also expand 5.1 in order to get an estimation for the deviation of C:

$$\Delta C = \frac{\partial C}{\partial I_o}\Delta I_o + \frac{\partial C}{\partial (dV_{OUT}/dt)}\Delta(dV_{OUT}/dt) + \frac{\partial C}{\partial A_o}\Delta A_o \tag{5.2}$$

By substituting standard deviations for the deltas ($\Delta A_o = 0.0034$, $\Delta(dV_{OUT}/dt) = 0.913$, $\Delta I_o = 0.354$), the resulting variation of C will be $\sigma_C = \Delta C = 0.00615$. By examining the ratios of the standard deviation over the mean ($\frac{\sigma_x}{\bar{x}}$), we can get a better feel for these numbers. Thus, errors of 0.37% on the source follower gain, 0.72% on the current source magnitude, and 1.78% on the TVC transfer ratio implied a 0.69% error on the values of the hold capacitors. Remember that this result is a measure of the matching from chip-to-chip. It is assumed that the matching between adjacent channels on the same chip will be even better.

The yield on the run was excellent. Of the twenty devices tested in the lab, only one exhibited any problems, corresponding to a yield of 95%.

6. Conclusions

A successful design and implementation of a Time-to-Voltage Converter/Analog Memory has been presented. A new CMOS time-integration circuit based on saturation-region current switching performed very well, demonstrating that CMOS is well suited for this type of application. The TVC architecture was also easily expandable to implement the analog memory. Future versions of the chip could easily be expanded to 16 or more channels, limited by the added parasitic capacitance at one critical circuit node.

An explanation was introduced for the movement of the channel charge for a saturated transistor at turn-off. It was found that all of the excess channel charge will exit via the drain terminal of the device when it is switched off. Circuit simulators such as SPICE do not accurately model this effect.

Time resolution of the TVC was better than expected at 0.2 ns. This limit was mostly due to external noise levels of about 10 mV. The fundamental limit for single-channel resolution will be set by the intrinsic noise level of the MOS devices; this quantity was not measurable with available test equipment. Matching of the linearity between the eight TVC channels was good, showing that the common-centroid layout scheme was worthwhile. Noise levels between channels was due mainly to coupling between adjacent bond wires; this problem could be avoided in the future by using more care in the design of the pinout of the chip.

On the other hand, the source-follower readout scheme did not perform well due to DC offset variations and mismatches. However, it was known from the outset that the source followers would be eliminated from subsequent designs; they were only used on this chip to provide a simple method of reading out the hold capacitors. Clearly, some sort of multiplexed readout system would be better (and in fact is already being designed), but this was deemed too complicated for this chip.

References

[1] V. Radeka, "Signal Processing for Radiation Detectors," CERN Lecture Series, Switzerland, March 1986.

[2] J.E. Dodd, *The Ideas of Particle Physics,* Cambridge: Cambridge University Press, 1984.

[3] H.H. Williams, "Design Principles of Detectors at Colliding Beams," *Annual Review of Nuclear Particle Science,* Vol. 36, 1986, pp. 361-417.

[4] T.J. Devlin, A. Lankford, H.H. Williams, "Electronics, Triggering, and Data Acquisition for the SSC," *Proceedings of the 1986 Summer Study on the Physics of the Superconducting Super Collider,* Snowmass CO, 1986.

[5] G.M. Haller, D.R. Freytag, J.T. Walker, S.I. Chae, "Performance Report for Stanford/SLAC Multi-Channel Sample-and-Hold Device," Poster paper presented at the Nuclear Science Symposium, San Francisco CA, Oct. 1985, SLAC-PUB-3778.

[6] D.R. Freytag, J.T. Walker, "Performance Report for Stanford/SLAC Microstore Analog Memory Unit," *IEEE Transactions on Nuclear Science,* Vol. NS-32, No. 1, Feb. 1985.

[7] J.T. Walker, S.-I. Chae, S. Shapiro, R.S. Larson, "Microstore—The Stanford Analog Memory Unit," *IEEE Transactions on Nuclear Science,* Vol. NS-32, No. 1, Feb. 1985.

[8] S. Kleinfelder, presentation at International Conference on Fast Analog Integrated Circuits for Particle Physics, Philadelphia PA, Mar. 1987.

[9] G. Smarandoiu, K. Fukahori, P.R. Gray, D.A. Hodges, "An All-MOS Analog-to-Digital Converter Using a Constant Slope Approach," *IEEE Journal of Solid-State Circuits,* Vol. SC-11, June 1976, pp. 408-410.

[10] F.H. Musa, R.C. Huntington, "CMOS Monolithic 3-1/2-Digit A/D Converters," *1976 ISSCC Digest of Technical Papers,* pp. 144-145.

[11] A.G.F. Dingwall, B.D. Rosenthal, "Monolithic CMOS Dual-slope 11-bit A/D Converter," *1976 ISSCC Digest of Technical Papers,* pp. 146-147.

[12] Larry Singer, Analog Devices, Andover MA, private communication, Jan. 1987.

[13] W.B. Wilson, H.Z. Massoud, E.J. Swanson, R.T. George, Jr., R.B. Fair, "Measurement and Modeling of Charge Feedthrough in n-Channel MOS Analog Switches," *IEEE Journal of Solid-State Circuits,* Vol. SC-20, No. 6, Dec. 1985, pp. 1206-1213.

[14] B.J. Sheu, C. Hu, "Switch-Induced Error Voltage on a Switched Capacitor," *IEEE Journal of Solid-State Circuits,* Vol. SC-19, No. 4, Aug. 1984, pp. 519-525.

[15] J.B. Kuo, R.W. Dutton, Bruce A. Wooley, "MOS Pass Transistor Turn-Off Transient Analysis," *IEEE Transaction on Electron Devices,* Vol. ED-33, No. 10, Oct. 1986.

[16] R.C. Yen, P.R. Gray, "A MOS Switched-Capacitor Instrumentation Amplifier,"

IEEE Journal of Solid-State Circuits, Vol. SC-17, No. 6, Dec. 1982, pp. 1008-1013.

[17] R.E. Best, *Phase-Locked Loops,* New York: McGraw-Hill, 1984.

[18] P.R. Gray, R.G. Meyer, *Analysis and Design of Analog Integrated Circuits, 2nd ed.,* New York: John Wiley and Sons, 1984.

[19] L.A. Glasser, and D.W. Dobberpuhl, *The Design and Analysis of VLSI Circuits,* Reading, MA: Addison-Wesley, 1985.

[20] L.W. Nagel, "SPICE2: A Computer Program to Simulate Semiconductor Circuits," *Electronics Research Laboratory Memo ERL-M520,* University of California, Berkeley, May 1975.

[21] B.J. Sheu, "MOS Transistor Modeling and Characterization for Circuit Simulation," *Electronics Research Laboratory Memo ERL-M85/85,* University of California, Berkeley, Oct. 1985.

[22] B.J. Sheu, D.L. Scharfetter, P.-K. Ko, M.-C. Jeng, "BSIM: Berkeley Short-Channel IGFET Model for MOS Transistors," *IEEE Journal of Solid-State Circuits,* Vol. SC-22, No. 4, Aug. 1987, pp. 558-566.

[23] W.S. Scott, R.N. Mayo, G. Hamachi, J.K. Ousterhout, "1986 VLSI Tools: Still More Works by the Original Artists," *Computer Science Division Report No. UCB/CSD 86/272,* University of California, Berkeley, Dec. 1985.

[24] J.L. McCreary, P.R. Gray, "All-MOS Charge Redistribution Analog-to-Digital Conversion Techniques—Part I," *IEEE Journal of Solid-State Circuits,* Vol. SC-10, Dec. 1975, pp. 371-378.

[25] A.E. Stevens, "Characterization of Errors in Circuit Simulation Programs," Bachelor's Thesis, M.I.T., Cambridge, MA, May 1986.

[26] N.H.E. Weste, K. Eshraghian, *Principles of CMOS VLSI Design,*, Reading MA: Addison-Wesley, 1985.

7. Appendix A: Using SPICE

7.1 Overview

This document contains useful information for both first-time and experienced users of the SPICE2G.6 circuit simulation program. If you already know how to run SPICE, you might want to skip directly to the sections on SPICE convergence.

7.2 Introduction

SPICE is a popular circuit simulator developed at University of California, Berkeley [A.1]. SPICE is a device-level simulator, meaning that it uses mathematical equations to model the I-V characteristics of circuit elements. Since most interesting circuit elements are non-linear (BJTs, MOSFETs), SPICE must solve a large number of transcendental equations in each simulation.

When a circuit gets large ($>$ 10 transistors), the numerical algorithms used in SPICE will often fail. These problems will produce errors in the SPICE output file such as "No DC Convergence" or "Internal Timestep Too Small." There are a number of tricks you can use in order to get SPICE to converge without affecting the accuracy of the simulation. These tricks will be outlined here.

7.3 SPICE at Penn

SPICE is available on several computers at the Moore School, including *eniac* (VAX 8650), *pender* (VAX 785), and *dsl* (VAX 750). All of these machines run Ultrix (the DEC version of UNIX[1]). If you are simulating a large circuit, you will definitely want to run on the largest machine that you can access. If you are working primarily on a small computer, you can run your spice jobs on a larger machine by using the UNIX commands *rsh* or *ftp* (if you don't know how to do this, ask a system administrator).

To run SPICE in UNIX, first write a SPICE input file using a text editor such as *emacs* or *vi* (read the SPICE manual [A.2] to learn how to make SPICE input files). Then, execute your SPICE run by typing:

spice $<$*infile* $>$*outfile*

7.4 Successfully Simulating Large Circuits

When simulating a complicated circuit, it is important to start with the most simple version of your circuit possible. The bells and whistles can always be added later. This means the following:

1. For analog circuits such as op amps, start out with just the input stage. Use ideal current sources and voltage sources for biasing. Then simulate each additional stage and bias circuit separately before putting the whole thing together.

2. For digital circuits, start with simple gates. When you have successfully simulated all of the gates that you will need, make each gate into a subcircuit. Then, assemble your logic circuit by attaching subcircuits together. Start small; for example,

1. UNIX is a trademark of AT&T Bell Laboratories, Ultrix is a trademark of Digital Equipment Corporation.

make sure that you can get a 1-bit counter to work before trying to design a 4-bit counter.

For your own sanity, it is important that you start simple! The numerical algorithms in SPICE are pretty fragile; if you start out with a complicated circuit, there will be many places where the algorithms can fail, making it difficult to ever get your circuit to converge. By getting the smaller parts to converge separately, you have a much better chance of getting the whole circuit to converge.

Always do the first simulation with only a .OP analysis card,[2] i.e. leave out .DC, .AC, .TRAN, etc. This will cause SPICE to simply find the DC operating point of your circuit. This simulation should take only a few seconds, and can save you a lot of trouble later on. From the results of this simulation, you should check the following things:

1. proper hookup of circuit elements, i.e. no unconnected nodes
2. DC voltage at important nodes
3. quiescent current for important transistors
4. quiescent power consumption

Don't forget to connect your circuit inputs to their quiescent values. This means setting DC values on your independent current and voltage source cards. If you are unable get your circuit to converge to a DC operating point, look at the section entitled *.OP Convergence.*

Once you have successfully simulated your circuit at DC, you should have a much easier time doing AC and transient analyses.

---

2. Historical note: Each line in a SPICE input file is called a card, going back to the days when all computer input was done on punched cards. A "deck" refers to a SPICE input file, as in a deck of cards.

7.5 Running SPICE on MAGIC Extractions

MAGIC is an MOS integrated circuit layout program which is used extensively at Penn. MAGIC can take a layout and generate a corresponding SPICE input deck. This input deck is extracted directly from the layout to include interconnect capacitances, transistor lengths and widths, etc. Theoretically, you could then send this input deck directly to SPICE. In reality, you must modify the MAGIC extraction a great deal; these modifications will be explained here.

1. An extracted SPICE deck will contain arbitrary node numbers. If you have labeled the nodes in your MAGIC layout, you can find the corresponding SPICE node numbers in the file entitled *magicfilename*.names. This is important if your circuit has more than approximately four transistors.

2. The MAGIC extractor does not add power supply voltages such as V_{DD} or V_{SS}; you must add these yourself. Also, don't forget to add input voltages and bias currents.

3. Even if you have many substrate contacts to power or ground in your layout, the MAGIC extractor will not make these connections in the extraction. The bulk terminals of n-channel and p-channel transistors will have separate node numbers which must be connected to the appropriate voltages.

4. The extractor will often reverse the source and drain terminals on transistors. In theory, since the MOS transistor is symmetric, this reversal should not matter. In reality, SPICE will choke on this reversal. You must go through the extracted SPICE deck and unflip any reversed transistors.

5. The extractor will model parasitic source-bulk and drain-bulk capacitances as linear capacitors. In reality, these are non-linear junction capacitances. Thus, the extractor will cause these capacitances to be modeled incorrectly.

6. Don't forget to add the appropriate .MODEL cards to the SPICE deck.

It is apparent that the MAGIC extractor is less than perfect. In fact, once your circuit gets bigger than about 20 transistors, the extracted SPICE file is almost impossible to run. For this reason, I suggest the following design methodology: (a) Simulate the circuit with SPICE before doing the layout. Estimate source and drain areas and perimeters. Specify these explicitly on each transistor card so that these capacitances will be calculated correctly. (b) Lay out the circuit. (c) Extract the circuit to get an idea of the interconnect capacitances. Add these capacitances to your own working SPICE deck. Remember: These capacitances will be overestimated due to the drain and source areas.

For more information on the MAGIC extractor, see the MAGIC manual and look in the directory ~cad/src/magic/tech.

7.6 Understanding and Fixing Convergence Problems

The major reason for SPICE not to converge is because you have your nodes numbered incorrectly. This problem is also the most difficult to find. It often helps to give another person your SPICE input deck and have him derive the circuit by hand. For larger circuits, however, you are usually on your own; the best you can do is use a .OPTIONS NODE card in your SPICE input deck. This will print a node table in the output file which lists all of the elements connected to each node in your circuit.

In addition, read Appendix A from Glasser & Dobberpuhl [A.3] on SPICE techniques.

7.6.1 .OP Convergence Operating Point (.OP) Convergence problems will manifest themselves as:

- NO CONVERGENCE IN DC ANALYSIS
- PIVOT CHANGE ON FLY...

Note: SPICE will perform a .OP analysis for every simulation, regardless of whether you specify the .OP card or not. Thus, you might encounter these problems even for a .TRAN or .AC analysis. This is why you should make sure your .OP will work before running the other analyses.

SPICE finds the operating point by iterating until it converges on a solution (it uses the Newton-Raphson algorithm). It takes an educated guess at the initial voltages at each node. Sometimes, these initial voltages will cause the iteration algorithm to diverge. This is the main reason for DC convergence problems. By using a .NODESET card, you can explicitly set the initial voltages on any nodes in your circuit. Note that these initial voltages will have no effect on the final node voltages; they only help SPICE converge on the correct solution.

Another trick to help convergence is to add large resistors from floating nodes to ground. Remember that in a .OP analysis, capacitors are treated as open-circuits; these are the nodes that usually require the resistors. A resistance of 1 gigaohm should suffice. The added resistors will help SPICE in its initial guesses for the floating node voltages. You can also try adding the resistors across the output terminals of transistors, i.e. collector-emitter or drain-source. However, you should avoid adding too many resistors, because they can affect output resistances and gain calculations, especially in amplifiers with active loads.

Also, you can try to make SPICE converge by switching iteration algorithms. By specifying a .OPTIONS ITL6=30 card, SPICE will use the source-stepping method rather than Newton-Raphson. See Appendix D of the SPICE manual.

The following are several common .OP convergence problems and their solutions:

1. A long chain of transistors. For example, many analog biasing circuits have a chain

of three or four transistors connected from supply to ground in order to generate bias voltages. Cascode amplifier stages and CMOS NAND gates are also suspect. SPICE will need help to converge on the voltages in the middle of the chain. Use a .NODESET card to set initial voltage guesses on the middle nodes. Any voltage within the supply range will usually do the trick.

2. A bistable circuit. This is common in analog biasing circuits. A circuit might have an alternate solution in which all of the transistors are off. Use a .NODESET card to imbalance the circuit and get it started. Don't forget: this may be a real-life problem with your circuit, and not only a numerical problem! You might have to add start-up circuitry.

3. Don't forget to set circuit inputs to their quiescent values!

If you have difficulty, your best source of information is the table of last node voltages in the SPICE output file. For .OP convergence problems, several nodes will usually have voltages of several hundred volts. These are the nodes which need .NODESETs or resistors.

7.6.2 .DC Convergence Numerical problems in .DC analysis will cause the following error:

- NO CONVERGENCE IN DC TRANSFER CURVES AT...

A .DC analysis is identical to a series of .OP analyses. Thus, the tricks are similar. However, a problem arises in choosing the values for the .NODESET voltages. A set of values that works in one part of the .DC range may be invalid over another range. There are several possible fixes:

1. Run two separate simulations over different .DC ranges, using different .NODESET voltages.

2. Try making your step size an odd number; for example, use .95 V instead of 1 V. Also, try starting the sweep range at .001 V rather than 0 V.

3. Try using a smaller step size, or a smaller sweep range.

7.6.3 .TRAN Convergence Errors in transient analysis are most aggravating because .TRAN analyses are very computationally demanding. Your circuit might run for twenty minutes before returning the errors:

- INTERNAL TIMESTEP TOO SMALL...
- MAXIMUM ENTRY IN THIS STEP IS LESS THAN PIVTOL

These errors occur because the voltages are changing too rapidly in your circuit. Normally, on the .TRAN card, you specify a time interval and a timestep. For a slow circuit, SPICE will find the solution at each timestep in the interval. However, suppose you specify a timestep of 1 microsecond, and your circuit switches at a speed of 1 nanosecond. In order to complete the analysis, SPICE must lower its timestep internally.

Timestep problems are especially common in clocked digital circuits. You might also encounter them when simulating the step response of an analog circuit. When a signal changes quickly from rail to rail, SPICE will lower its timestep in order to simulate this transition accurately. For an abrupt change, however, SPICE will have trouble "turning the corner," i.e. making the timestep small enough to catch the transition. .TRAN also has problems with positive feedback circuits which switch fast, e.g. Schmitt triggers or VCO's. A metastable or unstable circuit will also confuse SPICE; circuits with an unstable frequency response (i.e. zero or negative phase margin) may also be unstable in numerical analysis.

When you encounter internal timestep problems, look at the SPICE output file to find the timestep where the numerical analysis breaks down. Usually, the simulation will run

for a few timesteps before crashing. This breakdown point is where you will want to examine your input PULSE's and PWL's for fast-changing signals. For example, make your PULSE rise and fall times slower, maybe by a factor of ten. Then, when you are satisfied that your circuit works, crank up the speed. Or, make the transitions more curvy by using a PWL approximation of a real waveform. Also, don't forget that the transient SIN waveform turns on abruptly after the specified time delay. You may want to lower your frequency to make this turn-on less abrupt.

Another possible source of problems for the .TRAN analysis are the initial conditions of your circuit. This is especially important for digital circuits which have state, such as flip-flops and latches. You must set the initial conditions of both the Q and $\overline{Q}$ outputs. There are two ways to set initial conditions with SPICE. One method is to set the initial voltages and currents directly on the circuit element cards. The other method is to use a .IC card to set the node voltages. The latter method is preferred; it is easier to keep track of one .IC card than dozens of element cards. Note that you normally should *not* use the UIC option on the .TRAN card.

Sometimes, it is difficult to determine the correct initial conditions for a large circuit. A method to figure them out is to run a .TRAN analysis without varying any inputs to the circuit. Theoretically, this steady-state analysis should be the same as a .OP analysis. However, .OP assumes that all capacitors are open circuits, which is not quite true in a steady-state .TRAN analysis with initial conditions (this is especially the case with MOS dynamic logic circuits). To run the analysis, first add a few important initial conditions, and then simulate the circuit for a long time interval. This will allow the circuit to settle to its quiescent state. You can then get the appropriate value for any node voltage in the circuit by specifying the node on a .PRINT card. Add the additional initial conditions to the .IC card for subsequent simulations. This will aid convergence for simulations with

complicated inputs to the circuit.

7.6.4 Transient Simulation Accuracy Even when SPICE does converge, sometimes the results can be inaccurate! As VLSI circuits get smaller and smaller, the associated currents, voltages, and charges also get smaller. By default, SPICE is accurate to 1 picoamp, 1 microvolt, and 10 femtocoulombs. Low-power analog designers especially should be aware of these numbers. Incorrect tolerances can introduce numerical noise into the simulation in the form of oscillations and bumps. These tolerances can be modified by adjusting the ABSTOL, VNTOL, and CHGTOL parameters on a .OPTIONS card. How do you know if you need to adjust these numbers? Try running a few transient simulations with tighter tolerances (make ABSTOL, VNTOL, and CHGTOL *smaller* numbers). If your simulation results are the same, then you needn't worry. Note that tightening the tolerances can cause convergence problems. Make sure that the circuit will run with default tolerances before fiddling with the tolerance values.

Another important parameter is RELTOL. This is the error tolerance of the program, i.e. the convergence accuracy. The default value is 0.1%. This means that when SPICE iterates towards a solution, it finishes when the iterations get within 0.1% of each other. Sometimes, this value is not tight enough, e.g. simulating settling times of op amps or charge feedthroughs in MOS switches. Try setting RELTOL to .01% to check your simulation accuracy. Also, adjusting RELTOL may require that TRTOL be adjusted. TRTOL controls the internal timestep size used in transient simulations. Increasing TRTOL increases the step size and therefore decreases the simulation accuracy. However, increasing TRTOL can aid convergence in cases where you have already increased the accuracy by adjusting RELTOL or using Gear integration. TRTOL defaults to 7; try simulating with a TRTOL of 10 or 15.

The numerical integration method used by SPICE can also affect simulation results. Sometimes you can get numerical overshoot, oscillation, or ringing in your transient analysis. This is especially common in very small and fast MOS circuits. The default integration method is trapezoidal. However, by specifying a .OPTIONS METHOD=GEAR card, you can use the Gear method. The Gear method is more mathematically "stiff." This means that numerical oscillation will get damped out. Be careful, though; it has been reported that the Gear method can remove *wanted* oscillations as well.

7.6.5 Floating Point Errors SPICE can cause computer faults such as:

- Floating point overflow
- Divide by zero
- Segmentation fault

Sometimes, these errors will cause the simulation to abort in the middle of the run. In this case, there will be no error indication in the SPICE output file. These types of errors can be very difficult to track down. Here are some possible reasons for floating point errors, and their fixes:

1. You have upside-down MOS transistors. Swap the drain and source.
2. You have a long chain of transistors, which will cause the DC iteration algorithm to crash the process. Use a .NODESET card.
3. You have left out LAMBDA from the MOS Level 2 model. Theoretically, if you omit this parameter, SPICE should calculate it for you. However, it has been been demonstrated that this can cause machine faults.
4. You have omitted or misplaced a parameter or node number on a SPICE card.

Check the manual for the correct syntax.

Note that your SPICE run might cause floating point errors and still run to completion. In this case, you can usually ignore the errors.

7.7 Other Random Hints

The following cards are useful when running SPICE:

- .OPTIONS NOMOD eliminates the model information from the output file. This information gets repetitive after the first few simulations.
- .WIDTH OUT=80 will make the output file only 80 columns wide. This is useful when using an 80-column terminal.
- .OPTIONS ITL5=0 will eliminate the transient analysis iteration limit. SPICE has several iteration limits which will cause the job to terminate if it uses too much CPU time. ITL5 is the most bothersome of the bunch.

7.8 Some Notes on BSIM

The Berkeley Short-channel IGFET Model (BSIM) [A.4,A.5] is a new MOS model which uses a curve-fitting approach to modeling I-V characteristics of transistors. In order to use this model, special laboratory measurements must be taken from test wafers. These measurements are sometimes available from the silicon foundry, e.g. the MOSIS 1.25 micron process.

The BSIM model is more accurate than any of the other SPICE MOS models. It also consumes less CPU time. The BSIM version of SPICE2 is available on Penn computers by typing

bsim $<$*infile* $>$*outfile*

See the BSIM supplement to the SPICE manual. Some important features of BSIM:

- BSIM does not use .MODEL cards. You must use a .PROCESS card to indicate the filename which contains the model data. This model file must be same directory as your SPICE input deck.
- BSIM MOSFET element cards begin with S, not M.
- You can specify interconnect resistances and capacitances in BSIM by just specifying the length, width, and layer of your resistors and capacitors.

7.9 Some Notes on SPLOT

SPLOT is a SPICE post-processor which was developed at Penn [A.6]. SPLOT takes a SPICE output file and displays the data nicely on a high-quality graphics screen. This makes it simple to measure rise times, frequencies, etc. You can also generate hardcopy plots of your data. Thus, .PLOT cards are unnecessary.

SPLOT runs on the IBM ATs in the Digital Systems Lab (Moore 100), or the Vaxstations in Moore 207. Ask around in Moore 100 to learn how to run this program.

7.10 Some Notes on SPICE3

SPICE3 is the new version of SPICE which is currently under development at Berkeley. In its completed form, SPICE3 will run faster, converge better, give nice output on a graphics display, generate hardcopy plots, etc. SPICE3 also contains models for gallium arsenide MESFETs, and is written in C rather than Fortran. In theory, SPICE3 should blow the doors off its predecessor. In reality, the current version of SPICE3 (SPICE3A7) is slow, causes floating point errors, and has many unimplemented features. For the time being, stick with SPICE2 and SPLOT. However, be on the lookout for a new and improved version of SPICE3.

Note: The BSIM model is implemented in SPICE3 as the MOS4 model. To convert a

SPICE2-compatible BSIM process file into SPICE3-compatible MOS4 .MODEL cards, use the program *Proc2Mod* (This program is supplied with SPICE3).

7.11 For More Information

Several sources are available for further information on SPICE. Dr. Ping Yang edits a column entitled "Circuit Simulation and Modeling" in *IEEE Circuits and Devices Magazine.* There is also a USENET newsgroup named *comp.lsi* (USENET news is available on *eniac* and *dsl*). These two sources provide a forum for the exchange of ideas and tricks concerning SPICE and other circuit simulators. They are definitely worth a look.

7.12 Acknowledgements

Thanks to Dr. Jan Van der Spiegel of University of Pennsylvania and Kurt Ware of MIT for help in preparing this document.

7.13 References

[A.1] Laurence W. Nagel, "SPICE2: A Computer Program to Simulate Semiconductor Circuits," *Electronics Research Laboratory Memo ERL-M520,* University of California, Berkeley, May 1975.

[A.2] A. Vladimirescu, Kaihe Zhang, et. al., *SPICE Version 2G User's Guide,* University of California, Berkeley, Aug. 1981

[A.3] Lance A. Glasser and Daniel W. Dobberpuhl, *The Design and Analysis of VLSI Circuits,* Reading, MA: Addison-Wesley, 1985, pp. 451-452.

[A.4] Bing J. Sheu, "MOS Transistor Modeling and Characterization for Circuit Simulation," *Electronics Research Laboratory Memo ERL-M85/85,* University of California, Berkeley, Oct. 1985.

[A.5] Bing J. Sheu, Donald L. Scharfetter, Ping-Keung Ko, Min-Chie Jeng, "BSIM: Berkeley Short-Channel IGFET Model for MOS Transistors," *IEEE Journal of Solid State Circuits,* Vol. SC-22, No. 4, Aug. 1987, pp. 558-566.

[A.6] John Bradley, "Splot: An Interactive SPICE Plotting Program," University of Pennsylvania Department of Computer and Information Science, Feb. 1987.

8. Appendix B: Pinout of TVC/Analog Memory Chip

The chip was packaged in a standard 64-pin DIP.

1. blank (substrate contact)
2. **flip-flop** S
3. **flip-flop** R
4. **flip-flop** Q
5. **flip-flop** $\overline{Q}$
6. **current source** bias
7. **current source** output
8. **8-channel TVC/Analog Memory** source-follower bias
9. digital V_{DD}
10. **8-channel TVC/Analog Memory** current source bias
11. **8-channel TVC/Analog Memory** output 1
12. **8-channel TVC/Analog Memory** output 2

13. **8-channel TVC/Analog Memory** output 3

14. **8-channel TVC/Analog Memory** output 4

15. **8-channel TVC/Analog Memory** START

16. **8-channel TVC/Analog Memory** STOP

17. analog V_{DD}

18. **8-channel TVC/Analog Memory** output 5

19. **8-channel TVC/Analog Memory** output 6

20. **8-channel TVC/Analog Memory** output 7

21. **8-channel TVC/Analog Memory** output 8

22. **8-channel TVC/Analog Memory** reset

23. **8-channel TVC/Analog Memory** shift register reset

24. analog GND

25. **8-channel TVC/Analog Memory** shift register clock

26. **1-channel TVC/Analog Memory with channel-charge compensation** output

27. **1-channel TVC/Analog Memory** source-follower bias

28. **1-channel TVC/Analog Memory** reset

29. **1-channel TVC/Analog Memory** current source bias

30. **1-channel TVC/Analog Memory** START

31. **1-channel TVC/Analog Memory** STOP

32. **1-channel TVC/Analog Memory** output

33. digital GND

34. **source-follower** output

35. **source-follower** input

36. **source-follower** bias

37. **op amp** V^+

38. **op amp** V^-

39. **op amp** bias

40. **op amp** V_{OUT}

41. digital V_{DD}

42. **8-channel TVC/Analog Memory w/o shift register** source-follower bias

43. **8-channel TVC/Analog Memory w/o shift register** current source bias

44. **8-channel TVC/Analog Memory w/o shift register** output 1

45. **8-channel TVC/Analog Memory w/o shift register** output 2

46. **8-channel TVC/Analog Memory w/o shift register** output 3

47. **8-channel TVC/Analog Memory w/o shift register** START

48. **8-channel TVC/Analog Memory w/o shift register** STOP

49. analog V_{DD}

50. **8-channel TVC/Analog Memory w/o shift register** output 4

51. **8-channel TVC/Analog Memory w/o shift register** output 5

52. **8-channel TVC/Analog Memory w/o shift register** output 6

53. **8-channel TVC/Analog Memory w/o shift register** output 7

54. **8-channel TVC/Analog Memory w/o shift register** reset

55. **8-channel TVC/Analog Memory w/o shift register** select 1

56. analog GND

57. **8-channel TVC/Analog Memory w/o shift register** select 2

58. **8-channel TVC/Analog Memory w/o shift register** select 3

59. **8-channel TVC/Analog Memory w/o shift register** select 4

60. **8-channel TVC/Analog Memory w/o shift register** select 5

61. **8-channel TVC/Analog Memory w/o shift register** select 6

62. **8-channel TVC/Analog Memory w/o shift register** select 7

63. **8-channel TVC/Analog Memory w/o shift register** select 8

64. digital GND

9. Appendix C: Test Structures on the TVC/Analog Memory Chip

9.1 Bonded Test Structures

These structures were accessible via the input/output pins of the 64-pin DIP.

1. flip-flop
2. current source
3. source follower
4. op amp
5. 8-channel TVC/Analog Memory
6. 8-channel TVC/Analog Memory w/o shift register
7. 1-channel TVC/Analog Memory
8. 1-channel charge-compensated TVC/Analog Memory

9.2 Drop-in Test Structures

These test structures were only accessible with a probe station.

1. n-channel MOS transistor W=40 μm L=40 μm
2. n-channel MOS transistor W=24 μm L=2.4 μm

3. n-channel MOS transistor W=2.4 μm L=24 μm

4. n-channel MOS transistor W=2.4 μm L=2.4 μm

5. p-channel MOS transistor W=40 μm L=40 μm

6. p-channel MOS transistor W=24 μm L=2.4 μm

7. p-channel MOS transistor W=2.4 μm L=24 μm

8. p-channel MOS transistor W=2.4 μm L=2.4 μm

9. lateral bipolar npn transistor

10. lateral bipolar npn transistor

11. lateral bipolar pnp transistor

12. vertical bipolar pnp transistor

13. n-channel MOS transistor array W=32 μm L=1.6, 2.4, 3.2, 4.0, 4.8, 8.0, 11.2, 16.0 μm

14. p-channel MOS transistor array W=32 μm L=1.6, 2.4, 3.2, 4.0, 4.8, 8.0, 11.2, 16.0 μm

15. ring oscillator

16. CMOS inverter

17. op amp

18. 8-channel TVC/Analog Memory

19. 1-channel TVC/Analog Memory

20. 1-channel TVC/Analog Memory with channel-charge compensation

21. MOS capacitor W=115.2 μm L=3.2 μm

22. source follower

23. current source

24. flip-flop

25. 8-bit shift register