Abstract

Morgan, Andrew Stacy. Design Flow Based on Sensitivity Analysis For High-speed Digital Circuits. (Under the guidance of Dr. Paul D. Franzon)

The purpose of this work is to develop a design flow for high-speed digital circuits that may be used to increase the quality of circuit performance and improve the ability of inexperienced circuit designers. This design flow meshes the use of hand and simulation analysis to increase intuitive understanding of the dominant relationships and most significant circuit parameters that determine performance. The research relies heavily on determining the sensitivity of chosen performance measures to variation in selected circuit parameters, such as transistor gate width. Four detailed examples that follow the generalized design flow are included to illustrate practical application. The examples consist of the following circuits: source-follower, gate-isolated voltage sense-amplifier, Schmidt trigger, and dual-rail domino logic gate. The examples include design specifications, topology advantages and disadvantages, a suggested design approach, and detailed sensitivity analysis including quantitative simulation results supporting drawn conclusions.

DESIGN FLOW BASED ON SENSITIVITY ANALYSIS FOR HIGH-SPEED DIGITAL CIRCUITS

By
Andrew Stacy Morgan

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the requirements
for the Degree of Master of Science

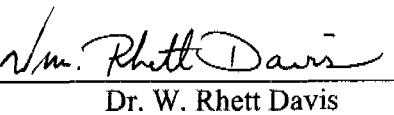
Electrical Engineering

Raleigh

2004

APPROVED BY:

Dr. Griff Bilbro



Dr. W. Rhett Davis

Dr. Paul D. Franzon
Chair of Advisory Committee

Biography

Andrew Stacy Morgan was born in Raleigh, North Carolina on January 18, 1980 and raised in Clayton, North Carolina. After completing Clayton High School in 1998, he pursued an engineering degree at North Carolina State University. He graduated Valedictorian with a Bachelor of Science Degree in Electrical Engineering in December 2002. Andy continued his education at NC State University as a Dean's Fellow to pursue a Master of Science Degree in Electrical Engineering. His primary interests lie in the area analog and mixed-signal circuit design. While working toward his M.S. degree, he held a teaching assistantship for an undergraduate analog electronics course during spring and fall semesters. He also interned at Analog Devices, Inc. during the summers, where he developed an even greater interest in transistor-level circuit design.

Acknowledgements

I wish to express my appreciation to my advisor Dr. Paul Franzon for his continuous support, guidance, and inspiration through the course of this project and through the course of my graduate study as a whole. It has been an honor and pleasure to work with and learn from him.

I would also like to thank Dr. Griff Bilbro and Dr. W. Rhett Davis for serving on my committee and also providing guidance throughout my undergraduate and graduate studies. It has been a pleasure working with Dr. Bilbro as a teaching assistant to strengthen undergraduate analog electronics and inspire young circuit designers. I would also like to recognize the valuable experience I gained while working with Dr. Davis early in my Master's program, and express my appreciation for his influence through my coursework and individual study.

Special thanks goes to my mother and father for their unconditional love and support throughout the duration of my college career. Thanks, particularly to my mother, for her gentle encouragement to pursue a graduate degree, even when my impatience began to get the better of me. Both of my parents contributed different, yet very valuable positive influence on my personality that helped to mold the man and engineer that I am proud to have become.

A very special thanks goes to my fiancé, Jayne, whose patience, understanding, and unconditional love and encouragement fueled the accomplishments of my graduate career at NCSU. My success would be hollow without her love and support.

Last but certainly not least, I would like to thank God for the many blessings that He has bestowed upon my loved ones and myself. I am truly fortunate to have been blessed with so many gifts. I only hope that I can somehow honor His generosity and return the favor through the course of my life.

Table of Contents

List of Tables

Page

List of Figures

Page

1. Introduction

The focus of this thesis is to develop a design flow for high-speed digital circuits that share common design characteristics. The design flow will include an analysis of the sensitivity of overall performance to changes in various circuit parameters and identify design "pressure points". The ultimate goal is to produce a collection of design examples that can be used as an instructional tool for students taking ECE733 – Digital Electronics at NC State University. The examples should document the thought process for circuit optimization and provide a finished product that can be used as a reference point for similar designs. The intent is that students can derive a better understanding of how the circuit works and hopefully develop a general approach to design that can be applied to other problems. The approach to be presented will utilize analysis and design techniques presented in lecture and textbooks, as well as convenient simulation techniques that take advantage of the power of SPICE. The performance data collected during the analysis is presented in both tabular and graphic form to best illustrate the circuits' sensitivity due to scaling the values of select variables.

2. Background and Motivation

In general, classic circuit design and analysis techniques using hand analysis yield very poor results when compared to simulation. Fundamental governing equations (such as those for drain current, transconductance, output impedance, etc.) reveal trends and rough relationships between circuit parameters, but lack the accuracy needed to design a circuit to precise values. In addition, describing the behavior of even simple circuits with hand analysis quickly becomes unmanageable due to the complex mathematical equations involved. As the math increases in complexity, the designer's intuitive feel for what's happening in the circuit

decreases. To reduce the complexity of the math, approximations and assumptions must be made to eliminate variables, further reducing the accuracy of the results from hand analysis.

Because these techniques demonstrate the concepts that govern circuit operation, circuits must be taught using hand analysis. One can understand how a circuit works and get a sense of what parameters are significant by examining the generic topology and governing equations. The approximations, assumptions, and "ignoring for now" reductions must be made to understand the "big picture." However, the neat and clean results that hand analysis yields describing circuit behavior can give the designer a false sense of the expected performance of his circuit. This is not to imply that hand analysis and design should not be practiced, but suggest that the presentation of such analyses be supplemented by practical application.

Fortunately, simulators exist to handle the multi-variable modeling of the circuit we hope to design, without resorting to extreme approximations. The key to utilizing the design tool is to understand the differences between the results of a simulation and the results of hand analysis. Simulators are much more accurate and efficient than hand analysis if one knows how to interpret the results. However, simulators are not perfect, nor always correct. Therefore the intuitive feel of hand analysis must be used to double check the simulator and explain the simulation results. Design through simulation and blind iteration (guess and check) alone may yield a design that works, but fails to optimize the design or improve the ability of the designer. Once the two analysis methods can be used together, a designer can tap into the power of tools, such as Hspice, to yield better designs in less time.

3. General Design Flow

Given the preceding discussion, a combination of hand and simulation analyses seems to provide the most thorough presentation of circuit design. Therefore I propose to structure the design flow as follows:

3.1. Design Goals: When first approaching a design problem, the goals for circuit performance must be clearly understood before it can be optimized. Once the most important and/or challenging design specifications are identified, a topology that most nearly meets the desired performance may be analyzed and chosen. Without a clear design goal, the tradeoffs among performance parameters may appear endless.

3.2. Hand Analysis: After a topology has been chosen, every design must still begin with some sort of hand analysis to approximate the values for independent variables that the designer can control, such as device dimensions, passive component values, bias voltages or currents, etc. This initial analysis of the circuit will characterize its general operation and identify which parameters are most significant. The equations derived for the circuit will allow the designer to develop "rules of thumb" for adjusting variables within the circuit to move towards the desired output.

3.3. Parametric Analysis: Once the designer has an intuitive understanding of how the circuit is supposed to work, he must analyze the accuracy of his equations versus models that will be used in simulation. For example, this comparison requires parametric analysis for the transistors in the process the designer plans to use. The designer should produce his own I_d curves over ranges of V_{ds} and V_{gs} to create a reference during design, as was done in Figure 3.1. This analysis will give the designer an idea about the magnitude of the discrepancies he will see between what

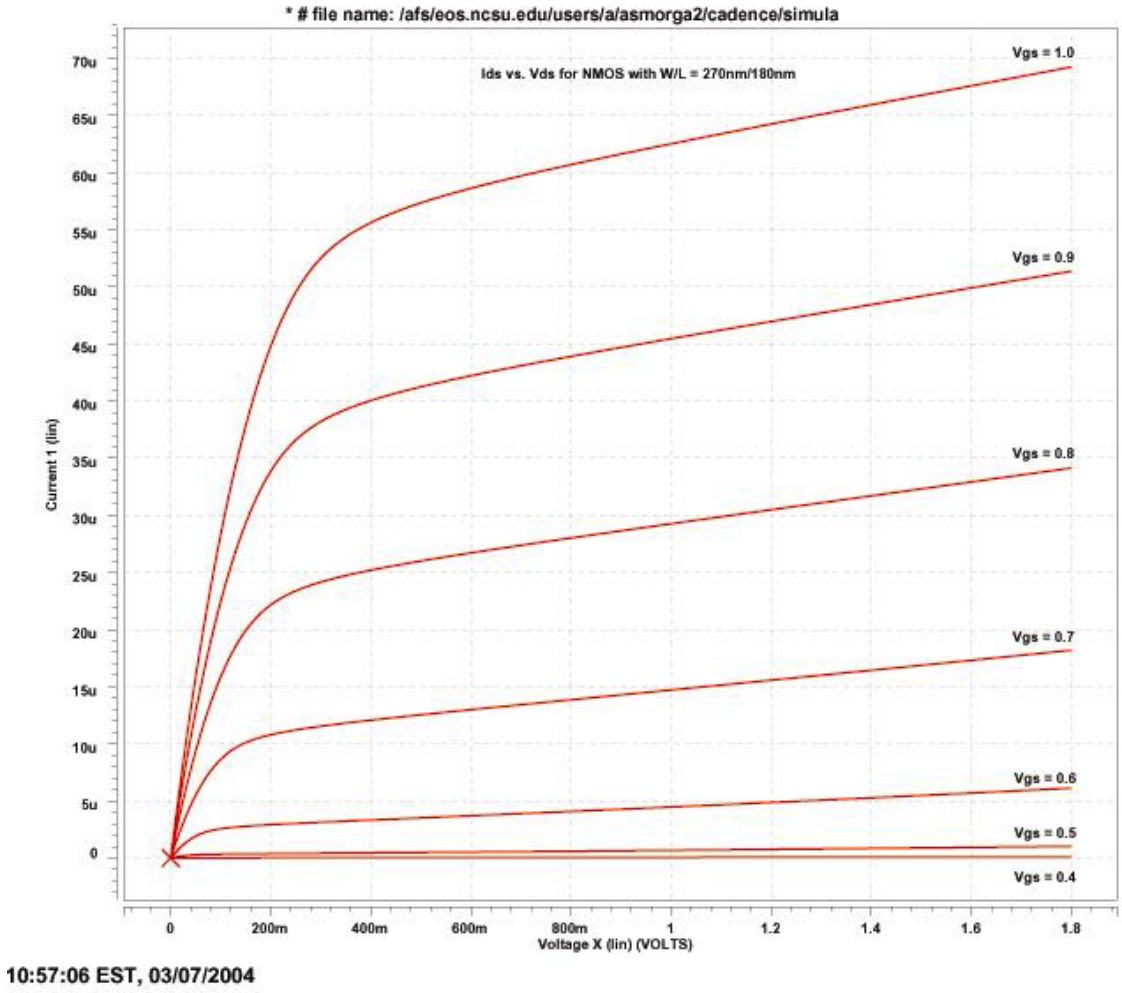


Figure 3.1 - Id Curves for NMOS with W/L=270nm/180nm in 0.18um (Deep) Process

he designs on paper and what he sees in simulation. Realizing these differences is very important when considering current drive capability, output impedance, threshold voltage, and biasing conditions. The curves in Figure 3.1 were produced using the circuit in Figure 3.2 and simulated using Hspice. Similar plots, shown in Figures A.1-A.3 of Appendix A, were created using scaled transistor widths of 2.7um, 27um, and 270um to examine how closely the drain current scales with width. To determine the effective process parameters from the Id curves, the Hspice results

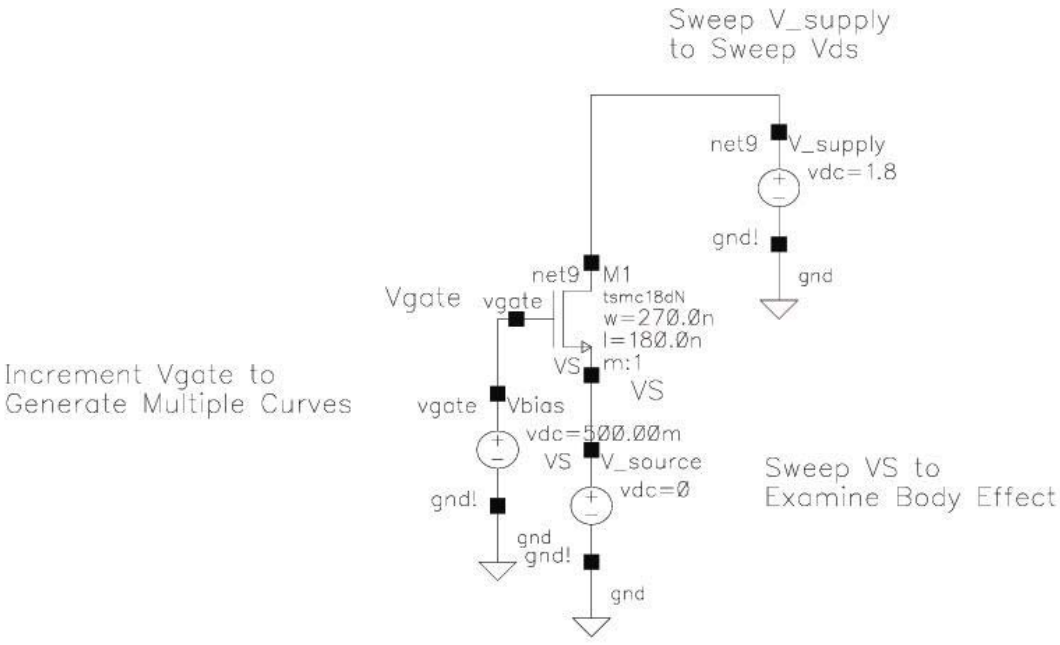


Figure 3.2 - Parametric Analysis Test Circuit

were compared to Matlab simulation results. Using the unified model equations for drain current and beginning with process parameters[1] taken from the TSMC 0.18um (Deep) model file, the parameters were adjusted until the Matlab I_d vs. V_{ds} curves closely matched those of the Hspice results.[2] This adjusted set of process parameters was used for the examples presented in the later sections. In those examples, the transistors are assumed saturated and the drain current is modeled using the first order approximation in Eqn. 3.1, where $\mu_n C_{ox} = 100 {uA}/{V^2}$, $V_{ton} = 0.43V$. [3]

[1] "process parameters" refers to the process transconductance parameter (μC_{ox}), the threshold voltage (V_{th}), and the saturation drain voltage (V_{dsat}).
[2] The Matlab Id curves, as well as the code used to generate them, can be found in Appendix A.
[3] PMOS parameters were not required for the examples presented here and are therefore not listed.

$$I_d = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{gs} - V_t)^2 \qquad \text{Eqn. 3.1}$$

Keeping modeling error in mind, the designer can estimate the rough sizes of transistors and begin the design in the right ballpark. General biasing conditions can also be determined from hand analysis. Then, using Hspice DC analysis, one can determine a more exact DC operating range and the preferred bias condition for the application. Analyzing the voltages, currents, and capacitances at critical nodes for extreme values of the transient range and during transients will reveal effects that were not included in the hand calculations and must be accounted for.

3.4. Analysis of Parasitics: The parasitic capacitors within devices will add internal loads that must be accounted for during optimization. For the purposes of hand analysis, first order approximations for gate and bulk capacitances give reasonable results for the expected impact on load capacitance due to transistor scaling. After analyzing the values for C_{gs}, C_{gd}, C_{db}, and C_{sb} over a range of gate widths, an average value for effective shunt capacitance at a node can be derived in terms of gate width, W. Tables 3.1 and 3.2 demonstrate how the average value of capacitance can be obtained and presented in the units of fF/λ, where λ is the minimum feature size. The tables include all parameters required to calculate and normalize all parasitic capacitors. Using this approximation, the total load capacitance at a node can be quickly calculated within reasonable accuracy by simply multiplying the average value of capacitance by the total λ of gate width connected to the node.[4] The

[4] This calculation must account for what device terminals are connected to that node to properly include the effective C_{gs}, C_{gd}, C_{db}, or C_{sb}.

approximation for average capacitance should be calculated separately for large variances in transistor sizes. For example the capacitance per λ of gate width for a 270nm transistor would be different than that of a 27um transistor, though the differences may be slight. Keep in mind that this rough approximation is intended to provide a quick starting point for design and not a precise value.

Table 3.1 - NMOS Parasitic Capacitance Calculations

NOMS Transistor Parameters (Model tsmc18dn)						
lamda=	9.00E-08		9.00E-08		9.00E-08	
scaleW)=	3.00E+00		3.00E+01		3.00E+02	
scaleL =	2		2		2	
W=(gate width)	2.70E-07		2.70E-06		2.70E-05	
L= (gate length)	1.80E-07		1.80E-07		1.80E-07	
Leff=	1.80E-07		1.80E-07		1.80E-07	
Wd/Ws=	2.70E-07		2.70E-06		2.70E-05	
Ld/Ls=	9.00E-07		9.00E-07		9.00E-07	
Ad/As=	2.43E-13		2.43E-12		2.43E-11	
Pd/Ps=	2.07E-06		4.50E-06		2.88E-05	
un= (	2.63E-02		2.63E-02		2.63E-02	
Er=	3.90E+00		3.90E+00		3.90E+00	
tox= (m)	4.00E-09		4.00E-09		4.00E-09	
Cox= (F/m^2)	0.00862875		0.00862875		0.00862875	
Cj= (F/m^2)	9.73E-04		9.73E-04		9.73E-04	
Cjsw= (F/m)	2.61E-10		2.61E-10		2.61E-10	
CGDO=	7.16E-10		7.16E-10		7.16E-10	
CGSO=	7.16E-10	Norm. Cap	7.16E-10	Norm. Cap	7.16E-10	Norm. Cap.
Csb=	8.23E-16	2.74E-16	4.01E-15	1.34E-16	3.59E-14	1.20E-16
Cdb=	7.76E-16	2.59E-16	3.54E-15	1.18E-16	3.11E-14	1.04E-16
Cgs=	4.73E-16	1.58E-16	4.73E-15	1.58E-16	4.73E-14	1.58E-16
Cgd=	1.93E-16	6.44E-17	1.93E-15	6.44E-17	1.93E-14	6.44E-17
Avg. Norm. Cap (F).=	1.14866E-16					

Table 3.2 - PMOS Parasitic Capacitance Calculations

PMOS Transistor Parameters (Model tsmc18dp)						
lamda=	9.00E-08		9.00E-08		9.00E-08	
scaleW)=	3.00E+00		3.00E+01		3.00E+02	
scaleL =	2		2		2	
W=(gate width)	2.70E-07		2.70E-06		2.70E-05	
L= (gate length)	1.80E-07		1.80E-07		1.80E-07	
Leff=	1.80E-07		1.80E-07		1.80E-07	
Wd/Ws=	2.70E-07		2.70E-06		2.70E-05	
Ld/Ls=	9.00E-07		9.00E-07		9.00E-07	
Ad/As=	2.43E-13		2.43E-12		2.43E-11	
Pd/Ps=	2.07E-06		4.50E-06		2.88E-05	
up= (	1.18E-02		1.18E-02		1.18E-02	
Er=	3.90E+00		3.90E+00		3.90E+00	
tox= (m)	4.00E-09		4.00E-09		4.00E-09	
Cox= (F/m^2)	0.00862875		0.00862875		0.00862875	
Cj= (F/m^2)	1.18E-03		1.18E-03		1.18E-03	
Cjsw= (F/m)	2.14E-10		2.14E-10		2.14E-10	
CGDO=	6.79E-10		6.79E-10		6.79E-10	
CGSO=	6.79E-10	Norm. Cap.	6.79E-10	Norm. Cap.	6.79E-10	Norm. Cap.
Csb=	7.85E-16	2.62E-16	4.39E-15	1.46E-16	4.04E-14	1.35E-16
Cdb=	7.28E-16	2.43E-16	3.82E-15	1.27E-16	3.47E-14	1.16E-16
Cgs=	4.63E-16	1.54E-16	4.63E-15	1.54E-16	4.63E-14	1.54E-16
Cgd=	1.83E-16	6.11E-17	1.83E-15	6.11E-17	1.83E-14	6.11E-17
Avg. Norm. Cap (F).=	1.1938E-16					

3.5. Software Utilization: Hand analysis is, of course, not restricted to performing all calculations by hand with paper and a scientific calculator. Programs such as Matlab, Mathcad, Maple, and Excel are excellent tools that reduce design time and streamline repetitive calculations. Matlab and Excel were primarily used for the examples presented in this project. Table 3.1 and 3.2 in section 3.4, generated using Excel spreadsheets, can be instantly altered to reflect the average capacitance values for various sizes. The optimization performed in section 4.6.3 of the source follower

example was done using a Matlab script that is found in Appendix A. The Schmidt trigger analysis was also streamlined with an Excel spreadsheet that generates the estimated V_{IH} and V_{IL} and is listed in Appendix A. The point is that the extra time involved for setting up software automation results in countless timesavings during design changes, verification and testing, and future designs. The data collected from the automated analysis also provides an efficient method of double-checking that Hspice simulation results are valid.

3.6. Simulation Utilization: Design tools such as Cadence, Hspice, or Awaves may be frustrating and somewhat intimidating when first approached, but are extremely useful for gathering data when used efficiently. Knowing the existence of simple functions built into the design tools is the first step in understanding how to utilize the tool. Therefore, several of the methods used to streamline the analysis of the examples presented in sections 4-7 are presented here[5].

3.6.1. The most valuable Hspice statement that can be added to a netlist to quickly gather data is the ".measure" command. Using variations of the measure statement, data such as propagation delay, rise-times, average values for voltage, current, or power, and derivatives or integrals of signals can be accurately extracted from a simulation. Recording data in this manner is efficient, repeatable, and precise such that it was the main tool used to accomplish the later analyses.

3.6.2. Another valuable Hspice command is the ".alter" statement, which is ideal for iterative analyses. The inclusion of an alter statement allows the designer to

[5] These very simple examples of tool use are meant to highlight functions that reduce design time and are not intended to represent the bulk of design tool capability or functionality.

change a portion of the netlist and automatically re-run the simulation with the new code as many times as needed with a single execution of the netlist. The data for multiple runs is saved in separate output files so that no outputs are overwritten.

3.6.3. A handy feature in Cadence Virtuoso's Analog Artist is the ability to create variables that can be used within the instance properties of a device. Editing variables within the Analog Artist window proves to be much more efficient than editing individual device instances. However, variables should not be numerically manipulated within the device properties definition (i.e. width=2*A). Errors will not be generated, but not all processes in Cadence will recognize the notation and the netlist generation will be inaccurate. The most notable erroneous calculation is that of the area and perimeter of transistor junctions when generating the netlist.

3.7. Interpreting Results: As mentioned earlier, simulations will not always provide accurate results. Simulation results are only as good as the models used to create them. In many cases, real circuit phenomena, such as the "soft region" of operation in a MOSFET when it transitions from operating in the linear mode to the saturated mode, are not modeled precisely. Therefore the circuit's behavior should be double checked against hand analysis results to ensure that the simulator has not mistreated a circuit element and/or that the designer has not erroneously set up the circuit simulation.

The following design examples encompass the design flow discussed in sections 3.1-3.7 and are meant to illustrate the ideas presented in section 2, which motivate this project.

4. Source Follower Design Example

4.1. Description: The core of this circuit is composed of a single transistor amplifier measuring the output voltage at the source node and supplying the input to the gate node. The schematic for this circuit, which also includes a Thevinin input source and external load capacitance, is shown in Figure 4.1. The name "source follower" is derived from the fact that for a given bias current, the source will "follow" the gate as the input in modulated in order to maintain a constant drain current.

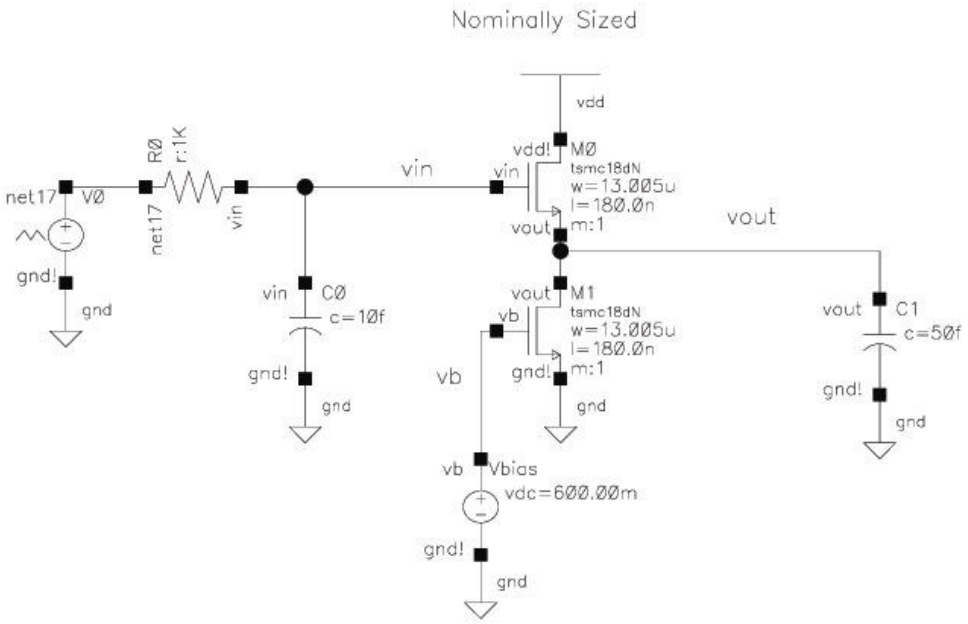


Figure 4.1 - Source Follower Schematic w/ Nominal Sizes

4.2. Common Applications

4.2.1. Level shifter – The V_{gs} drop from input to output can be utilized to level-shift the common mode of a signal. The magnitude of the shift is controlled by process parameters, device geometry, and the bias current.

4.2.2. Impedance buffer – The high input impedance with low input capacitance, coupled with the moderate output impedance and ability to drive large capacitive loads provides a mechanism to buffer a small swing signal being driven by a high output impedance.

4.3. Specific Application: For this example, we will design a source follower to be used as a buffer to take a small swing signal off-chip through a probe pad for the purpose of measurement. The following specifications will dictate design parameters:

- Input signal has 300mV peak-to-peak swing, with rise-time (T_r) equal to 200ps, and common mode voltage of 1.0V.
- Operating frequency of 500MHz.
- Output resistance of preceding stage 1kΩ with 10fF input capacitance.
- Required Gain is as close to unity as possible, probably in the range of 0.8.
- NMOS current source load used to establish bias current.
- External capacitive load of 50fF.
- The bias current chosen will determine power dissipation.
- Output waveform should maintain an eye opening of 600ps by 0.225V.

4.4. Advantages of Topology

- This topology was chosen for this application because of the circuit's low input capacitance and moderate output impedance, which allows it to drive fairly large capacitive loads.

4.5. Disadvantages of Topology

- Because the source node is not grounded, the common-drain transistor suffers from the body effect. The dependence of the threshold voltage on signal level introduces a less than unity non-linear gain. However, this non-linearity is mostly only a concern when using this circuit to buffer analog signals, rather than a digital pulse.
- The positive and negative slew rates are unequal due to the circuit's ability to charge and discharge the load capacitance. When the input is rising, I_{d0}, the current through M0, minus I_{d1}, the current through M1, charges the load capacitance. I_{d0} is initially large because M0 sees a large overdrive voltage as the input rises, and the output tries to follow. Therefore, the load capacitance is charged quickly. However, on the falling edge of the input, the load capacitance is discharged by I_{d1} alone, which is fixed by the bias voltage and relatively small. This causes the output to be discharged more slowly than it is charged. [5]

4.6. Design Approach

4.6.1. The primary concern in this application is the ability to drive the large load capacitance and maintain signal integrity. The rise-time of the input signal must be preserved as closely as possible, therefore Eqn. 4.1, the constant current

method for approximating voltage rise, is a good starting point for estimating the current needed to drive the load at the necessary speed.

$$T_r = \frac{\Delta V_{out} C}{I_d} \qquad \text{Eqn. 4.1}$$

The estimated gain is 0.85, therefore $\Delta V_{out} = 0.85 * \Delta V_{in}$. Substituting 300mV for ΔV_{in}, C_L= 50fF for C, and 200ps for T_r, a rough approximation for the required Id is found to be 63.75uA

4.6.2. The value just found for Id is used to determine which Id curves from the parametric analysis should be used. (Note that value of the parameters in the drain current equation, restated here in Eqn. 4.2, may vary slightly according to the range of gate width being used, which is discussed in section 3.3.) Working backwards from the estimated value of current and the Id equation[1], the initial gate width for M1 can be determined. But first, the internal capacitance of M1 must be added to the analysis to more accurately estimate the needed gate width.

$$I_d = \tfrac{1}{2} \mu_n C_{ox} \tfrac{W}{L} (V_{gs} - V_t)^2 \qquad \text{Eqn. 4.2}$$

Note[2]: $\mu_n C_{ox} = 100 \tfrac{uA}{V^2}, V_{ton} = 0.43V$

4.6.3. Though C_L is the estimated external capacitance, the parasitic capacitance of M1 and M0 will also be added to the output node and must be considered when scaling these transistors. This is particularly important to note when scaling large transistors, where the internal capacitance could begin to dominate the load

[1] For this approximation, body effect and channel length modulation are neglected to simplify the analysis.
[2] Process parameters extrapolated for transistors with gate width between W=2.7um and W=27um in Section 3.3.

and further scaling for current drive becomes counterproductive. To include the parasitics, we will utilize the 0.11fF per λ of gate width approximation for parasitic capacitance[3] to add C_{int} to the capacitance in Eqn. 4.1. The resulting equations are shown in Eqn. 4.3. and Eqn. 4.4, where C_{int} represents internal parasitic capacitance of M0 and M1 and W_1 represents gate width.

$$T_r = \frac{\Delta V_{out}\left(C_L + C_{INT}\right)}{I_d} \qquad \text{Eqn. 4.3}$$

$$C_{INT} = 0.11fF\left(\frac{W_1}{\lambda}\right) \qquad \text{Eqn. 4.4}$$

Id can now be represented as a function of only W_1 if the process parameters determined in section 4.6.2, minimum gate length (L), $V_{gs} = V_{bias}$, and $V_t=V_{to}$ are substituted into Eqn. 4.2. The resulting equation for Id is shown in Eqn. 4.5.

$$I_d \sim 8.023 * W_1 \qquad \text{Eqn. 4.5}$$

Combining equations 4.3, 4.4, and 4.5 into a single equation for T_r, and substituting 90nm for λ:[4]

$$T_r = \frac{\Delta V_{out}\left(C_L + 1.22_E{}^{-9} * W_1\right)}{8.028 * W_1} \qquad \text{Eqn. 4.6}$$

The rise time of the output node (T_r) can now be estimated for a given W_1 and vice versa. In this case, W_1 was solved for a T_r of 200ps and found to be approximately 13um. This will be the starting point for the gate widths of M1 and M0.

[3] This is a rough approximation for NMOS and PMOS transistors having gate widths in the expected size range. The values are derived from Table 3.1.

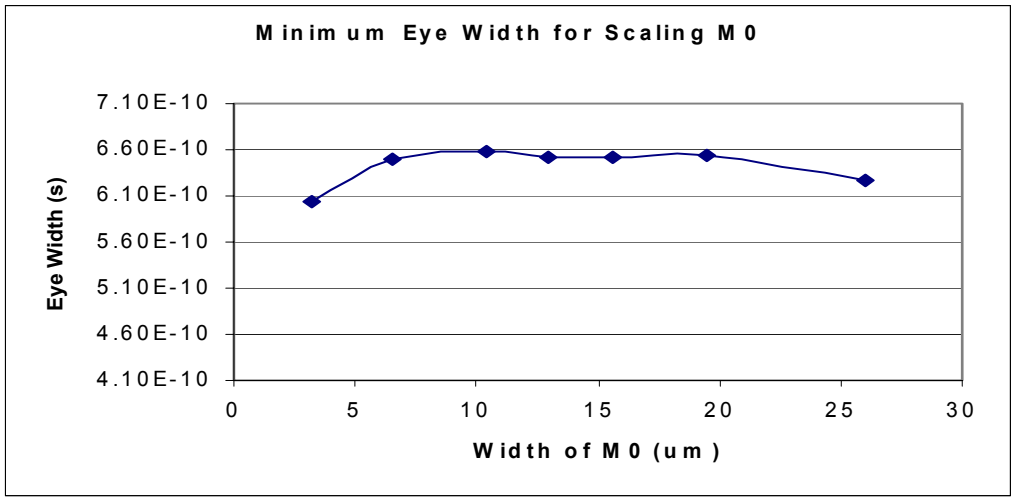
[4] Note that we are using the TSMC 0.18um Deep CMOS process.

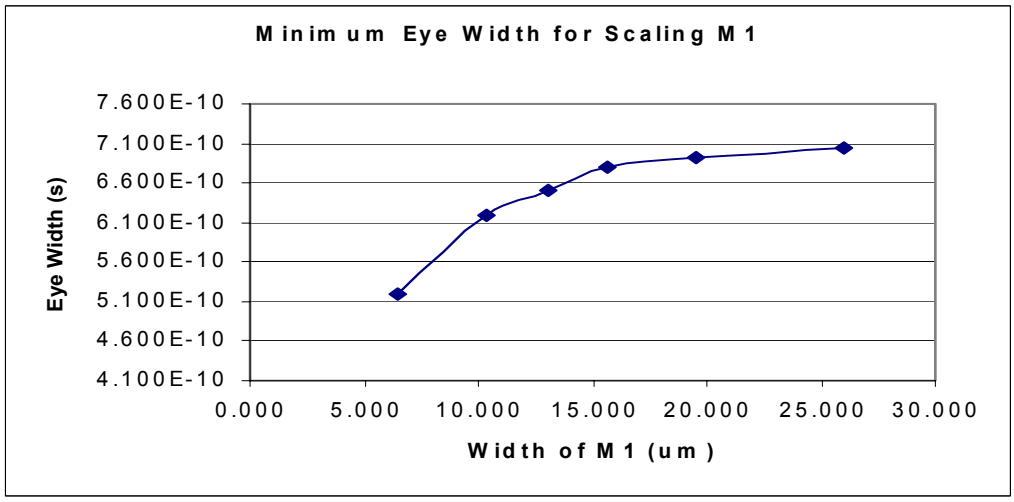
4.6.4. W_1=W_0=13um will be picked as the nominal gate widths of M0 and M1. These sizes are sufficient to meet the current specification. However, W_0 can be scaled to adjust the common-mode shift, and W_1 can be scaled to change the output edge-rate. The sensitivity of the performance to scaling W_0 and W_1 is discussed in section 4.8.

4.7. Analysis Setup and Approach: Both transistors were scaled separately and simultaneously to investigate the sensitivity of the noise margin to these changes. The dimensions of the eye diagram for an arbitrary pseudo-random bit stream are taken as the primary figure of merit reflecting the noise margin. The sensitivity of the output common-mode shift was also analyzed for possible level-shifting applications. The tests that involved scaling only one transistor at a time proved most interesting. Scaling M0 and M1 together resulted in minor performance differences other than changes in power dissipation, and were therefore not explored as heavily.

4.8. Sensitivity Analysis

4.8.1. Scaling M0 has a significant effect on the common-mode shift of the output with respect to the input, but very little effect on the width of the eye diagram. The top plot of Figure 4.2 shows that the eye width changes very little for M0 gate widths, except for the most extreme scaling cases. The top plot of Figure 4.3 reveals that $V_{out,CM}$ is most sensitive to scaling M0, showing a change of ~100mV over the scale range. The percentage changes in Table 4.1 show that M0's only real influence is on the common-mode shift.





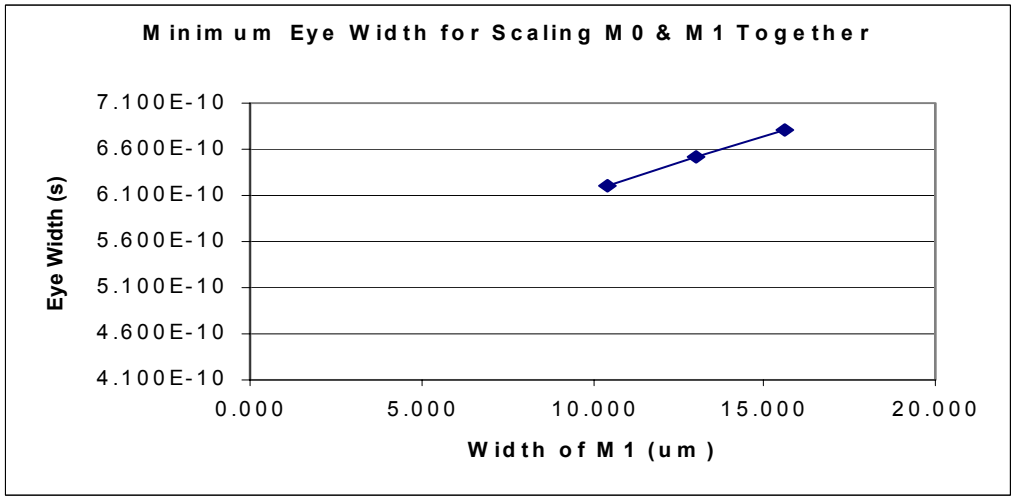
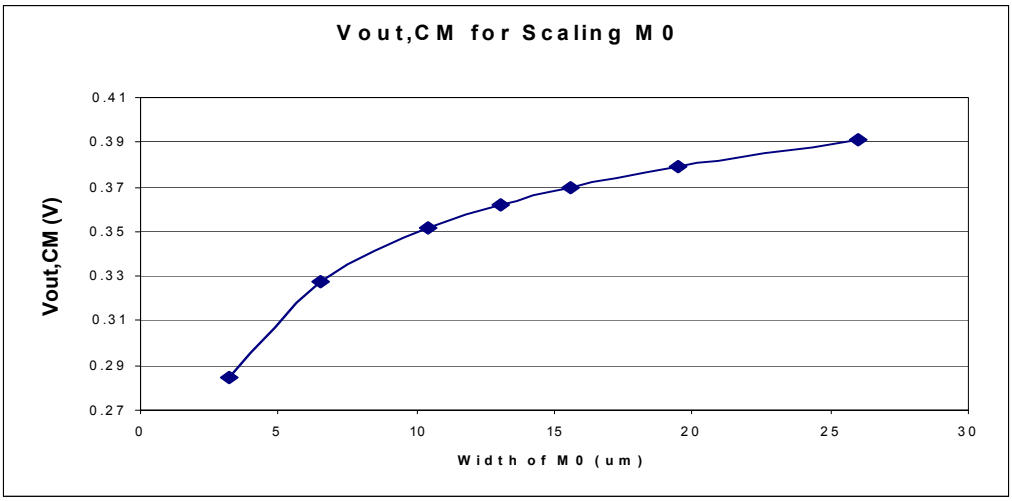
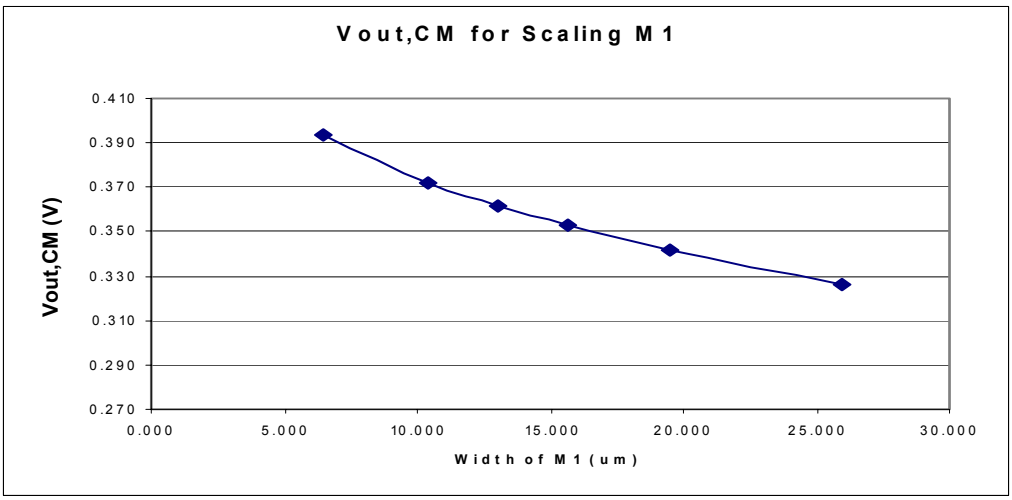


Figure 4.2 - Minimum Eye Width Sensitivity Comparison





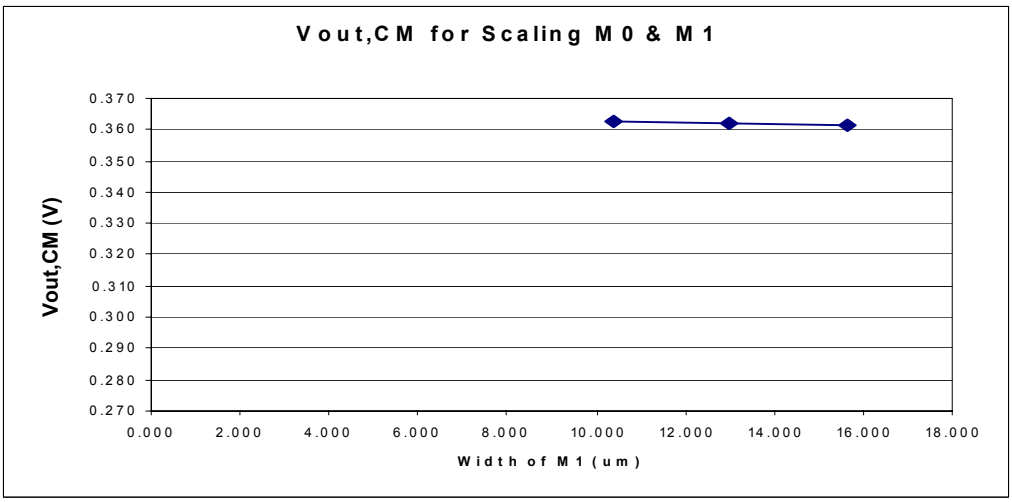


Figure 4.3 - $V_{out,CM}$ Sensitivity Comparison

Table 4.1 - Sensitivity Results for Scaling M0

M0 Scaled Alone Scaled:	25%	50%	80%	Nominal	120%	150%	200%
Gate Width M0 (um)	3.24	6.48	10.395	13.005	15.615	19.458	26.010
Gate Width M1 (um)	13.005	13.005	13.005	13.005	13.005	13.005	13.005
Min. Eye Width	6.04E-10	6.49E-10	6.580E-10	6.510E-10	6.520E-10	6.530E-10	6.270E-10
Improv. from Nom.	-4.700E-11	-2.000E-12	7.000E-12		1.000E-12	2.000E-12	-2.400E-11
% Change	-7.220%	-0.307%	1.075%		0.154%	0.307%	-3.687%
Min. Eye Height	0.227	0.232	0.234	0.235	0.235	0.236	0.238
Improv. from Nom.	-8.000E-03	-3.000E-03	-0.001		0.000	1.000E-03	3.000E-03
% Change	-3.404%	-1.277%	-0.426%		0.000%	0.426%	1.277%
Rise-time	2.44E-10	2.33E-10	2.550E-10	2.540E-10	2.540E-10	2.620E-10	2.700E-10
Improv. from Nom.	-1.000E-11	-2.100E-11	-1.000E-12		0.000E+00	8.000E-12	1.600E-11
% Change	-3.937%	-8.268%	-0.394%		0.000%	3.150%	6.299%
Fall-time	3.11E-10	2.88E-10	3.130E-10	3.110E-10	3.110E-10	3.120E-10	3.280E-10
Improv. from Nom.	0.000E+00	-2.300E-11	-2.000E-12		0.000E+00	1.000E-12	1.700E-11
% Change	0.000%	-7.395%	-0.643%		0.000%	0.322%	5.466%
Vcm	0.2847	0.3276	0.352	0.362	0.370	0.379	0.391
Change from Nom.	-7.720E-02	-3.430E-02	-0.010		0.008	1.750E-02	2.930E-02
% Change	-21.332%	-9.478%	-2.874%		2.238%	4.836%	8.096%
Avg. Power	1.88E-04	1.93E-04	1.959E-04	1.971E-04	1.980E-04	1.993E-04	2.004E-04
Improv. from Nom.	9.500E-06	4.100E-06	1.200E-06		-9.000E-07	-2.200E-06	-3.300E-06
% Change	4.820%	2.080%	0.609%		-0.457%	-1.116%	-1.674%

4.8.2. Scaling M1 dominates the rise and fall times of the output, and therefore the width of the eye diagram. With a fixed bias voltage, the width of M1 alone controls the bias current available to charge and discharge the load capacitance. The 200ps change in eye width over the scale range for M1 shown in Figure 4.2 demonstrates the noise margin's dependence on M1. The most heavily affected noise margin parameter is the fall-time, which is recorded in Table 4.2. The fall-time is more dependent than the rise-time on the size of M1 since the discharge path does not benefit from the current drive of M0, as the charge path does. Figures 4.4 and 4.5 illustrate the eye diagram variance from scaling 50% of nominal to 200% of nominal width of M1 respectively.

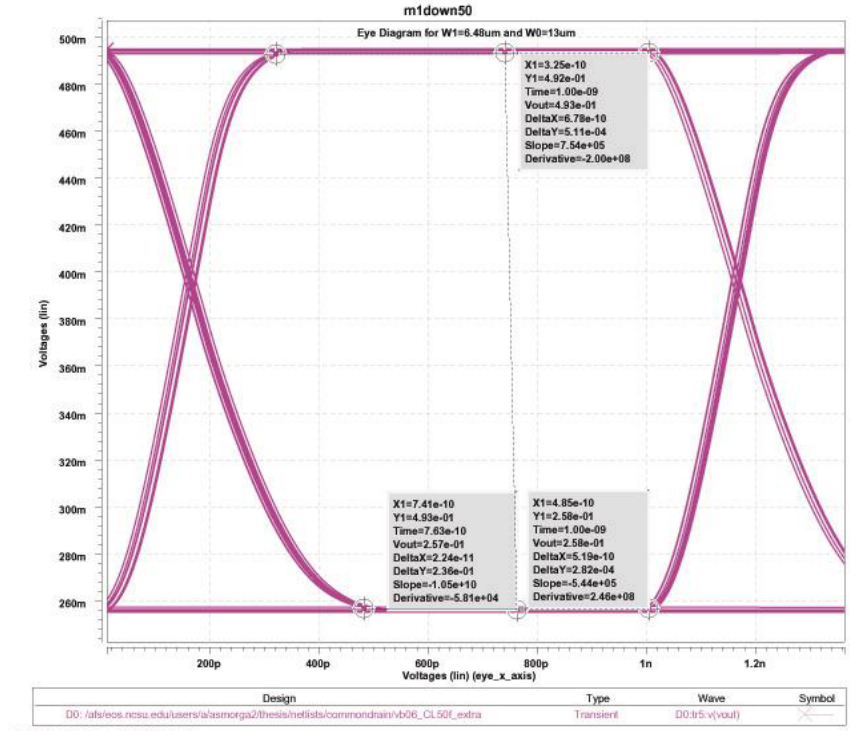


Figure 4.4 - Eye Diagram for W1 Scaled Down by 50%

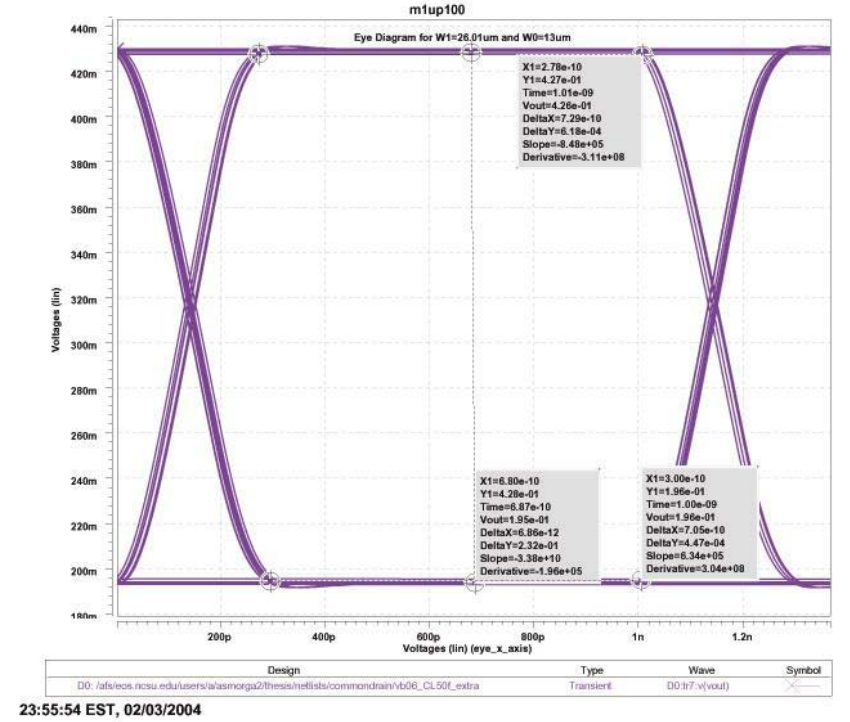


Figure 4.5 - Eye Diagram for W1 Scaled Up by 100%

Table 4.2 - Sensitivity Results for Scaling M1

M1 Scaled Alone Scaled:	50%	80%	Nominal	120%	150%	200%
Gate Width M0 (um)	13.005	13.005	13.005	13.005	13.005	13.005
Gate Width M1 (um)	6.480	10.395	13.005	15.615	19.485	26.010
Min. Eye Width	5.190E-10	6.200E-10	6.510E-10	6.800E-10	6.910E-10	7.050E-10
Improvement from Nom.	-1.320E-10	-3.100E-11		2.900E-11	4.000E-11	5.400E-11
% Change	-20.276%	-4.762%		4.455%	6.144%	8.295%
Min. Eye Height	0.236	0.236	0.235	0.234	0.233	0.232
Improvement from Nom.	1.000E-03	0.001		-0.001	-2.000E-03	-3.000E-03
% Change	0.426%	0.426%		-0.426%	-0.851%	-1.277%
Rise-time	2.490E-10	2.570E-10	2.540E-10	2.480E-10	2.350E-10	2.390E-10
Improvement from Nom.	-5.000E-12	-3.000E-12		6.000E-12	-1.900E-11	-1.500E-11
% Change	-1.969%	-1.181%		2.362%	-7.480%	-5.906%
Fall-time	4.120E-10	3.380E-10	3.110E-10	2.930E-10	2.630E-10	2.530E-10
Improvement from Nom.	-1.010E-10	-2.700E-11		1.800E-11	-4.800E-11	-5.800E-11
% Change	-32.476%	-8.682%		5.788%	-15.434%	-18.650%
Vcm	0.393	0.372	0.362	0.353	0.342	0.326
Change from Nom.	3.140E-02	0.011		-0.009	-1.990E-02	-3.570E-02
% Change	8.676%	2.901%		-2.404%	-5.499%	-9.865%
Avg. Power	1.024E-04	1.595E-04	1.971E-04	2.345E-04	2.894E-04	3.810E-04
Improvement from Nom.	9.470E-05	3.760E-05		-3.740E-05	-9.230E-05	-1.839E-04
% Change	48.047%	19.077%		-18.975%	-46.829%	-93.303%

4.8.3. Scaling M0 and M1 simultaneously produces the same noise margin results as scaling M1 alone. The only benefit to scaling M0 with M1 is the ability to maintain a constant level-shift as M1 is scaled. The increased size of M0 compensates for the increased current drawn by M1 and maintains a constant $V_{gs0,}$ as shown in Table 4.3, as long as M0 and M1 are scaled proportionally.

Table 4.3 - Sensitivity Results for Scaling M0 & M1

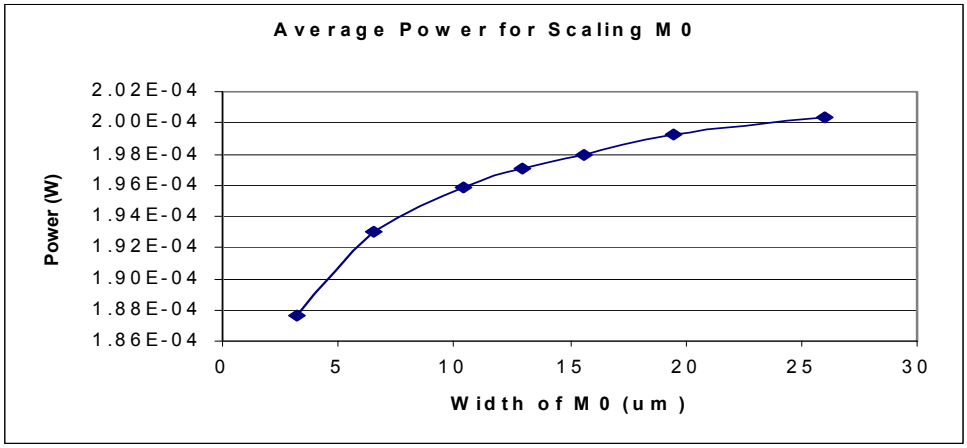
M0 & M1 Scaled Together			
Scaled:	**80%**	**Nominal**	**120%**
Gate Width M0 (um)	10.395	13.005	15.615
Gate Width M1 (um)	10.395	13.005	15.615
Min. Eye Width	6.200E-10	6.510E-10	6.800E-10
Improvement from Nom.	-3.100E-11		2.900E-11
% Change	-4.762%		4.455%
Min. Eye Height	0.237	0.235	0.234
Improvement from Nom.	0.002		-0.001
% Change	0.851%		-0.426%
Rise-time	2.570E-10	2.540E-10	2.480E-10
Improvement from Nom.	-3.000E-12		6.000E-12
% Change	-1.181%		2.362%
Fall-time	3.380E-10	3.110E-10	2.930E-10
Improvement from Nom.	-2.700E-11		1.800E-11
% Change	-8.682%		5.788%
Vcm	0.363	0.362	0.362
Change from Nom.	0.001		0.000
% Change	0.193%		-0.083%
Avg. Power	1.583E-04	1.971E-04	2.358E-04
Improvement from Nom.	3.880E-05		-3.870E-05
% Change	19.685%		-19.635%

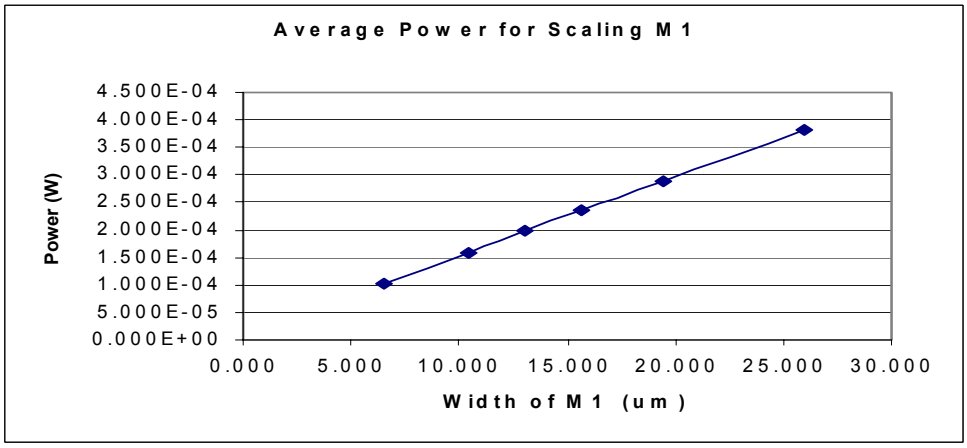
4.9. Key Considerations

4.9.1. Though increasing the gate width of M1 improves the width of the eye diagram, the resulting costs in power dissipation are huge. Figure 4.6 shows that average power dissipation changes linearly with the gate width of M1. Comparing the results for percentage change in eye width versus percentage change in average power dissipation in Table 4.2, one can see that diminishing returns are quickly reached when increasing the width of M1.

4.9.2. W_0 is not the only variable that will affect the common-mode output level. As the middle plot of Figure 4.3 shows, scaling M1 alone will also change $V_{out,CM}$ because V_{gs0} must change to compensate for the current in M1 if M0 is not

scaled proportionally. When designing for a specific level-shift, both transistors must be considered together.





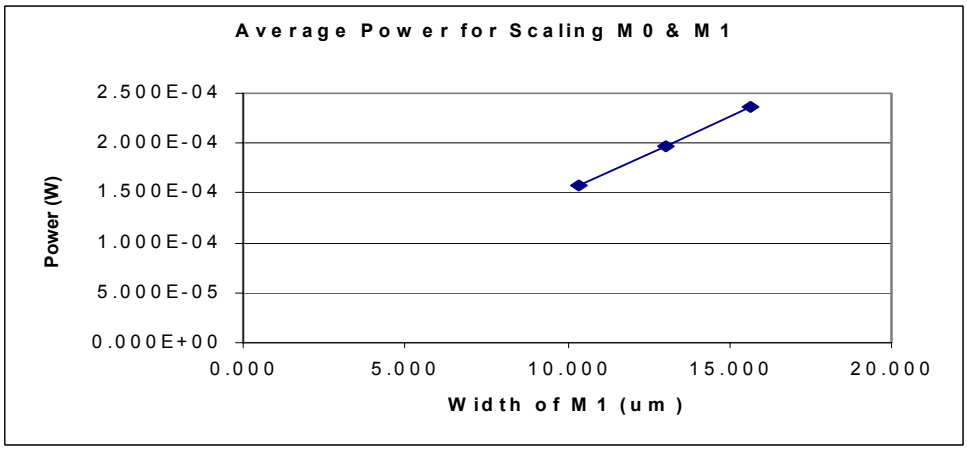
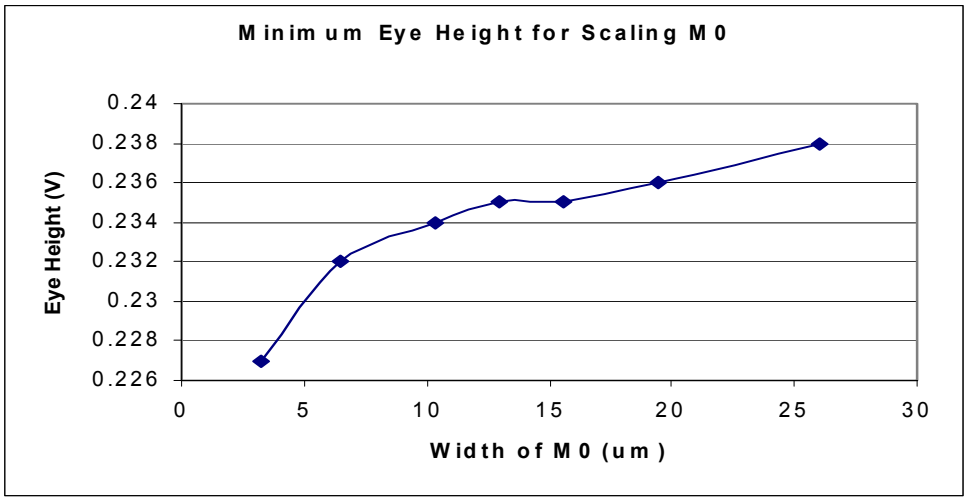
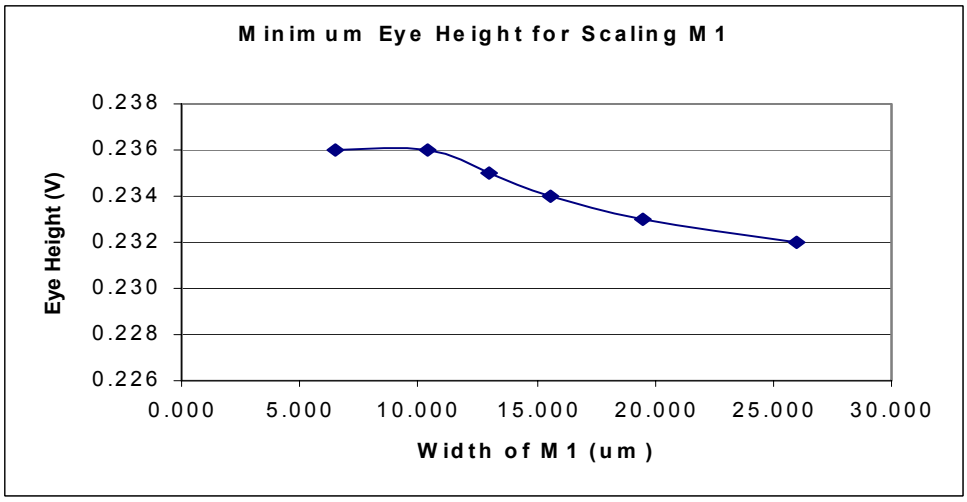


Figure 4.6 - Average Power Comparison

4.9.3. Though M1 dominates the eye width, the height of the eye diagram will be dictated by the amount of common-mode shift. Figure 4.7 shows that scaling either transistor will affect the height of the eye diagram. This is mainly due to $V_{out,CM}$ for the given data point. Comparing figures 4.3 and 4.7, one will see that the eye height tracks the trend for $V_{out,CM.}$ $V_{out,CM}$ decreases for larger values of bias current or smaller values of W_0. Either condition requires a larger V_{gs0} regardless of the signal input. Therefore, the output swing will be limited under these conditions as well, e.g. decreased eye height for decreased $V_{out,CM}$.





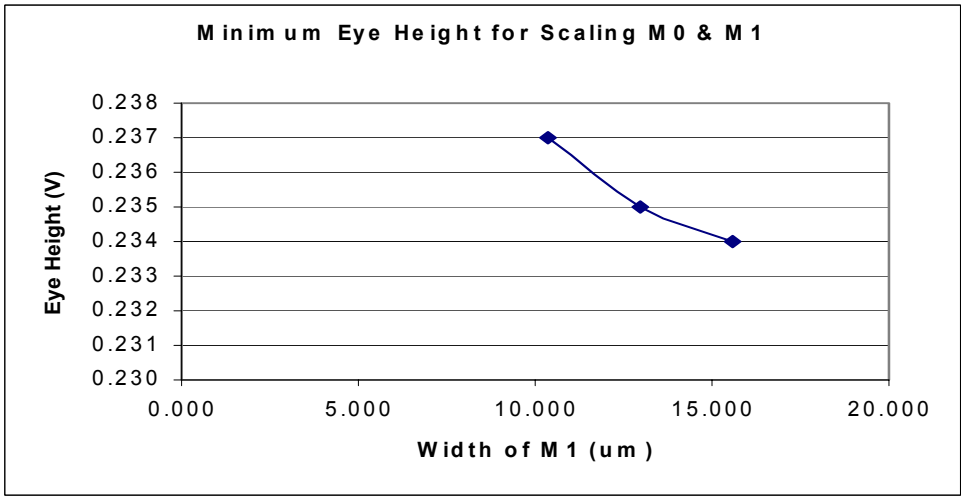


Figure 4.7 - Minimum Eye Height Sensitivity Comparison

5. Gate-Isolated Voltage Sense-Amplifier Design Example

5.1. Description: This circuit is composed of 10 transistors, 6 NMOS and 4 PMOS. A schematic with nominal transistor sizing is shown in Figure 5.1. The circuit is controlled by clocking the gate voltages of N6, P1 and P4. During the low phase of the clock, all internal nodes are pre-charged high through P1 and P4. The rising edge of the clock begins the evaluation phase, as N6 is turned on and provides a tail current. During the evaluate phase, the sense-amp detects the polarity of the input signal and latches the complement to the output. The input pair, N1 and N2, transforms a differential input voltage into a differential current. The difference in current between the two sides of the load and the resultant charge imbalance at the source nodes of N3 and N4 (which are considered the "sense nodes") causes one of the output nodes to fall faster than the other. P2, N3, P3, and N4 act as cross-coupled inverters and provide positive feedback to the output. Once the output nodes drift far enough apart, the regenerative action of the positive feedback pulls the high output to VDD and the low output to ground. Just before the regenerative action begins, N5 begins to conduct, shorting the source nodes of N3 and N4 together. This effectively isolates the regenerative nodes from the input signal. On the falling edge of the clock, all internal nodes of the amplifier are again pre-charged high. The regenerative action of this circuit allows it to exponentially amplify very small signals that have very high bandwidth. The required aperture time is mainly determined by the capacitance at the output nodes and the conduction of the evaluation path, as well as the fall time of the clock signal. [1], [3]

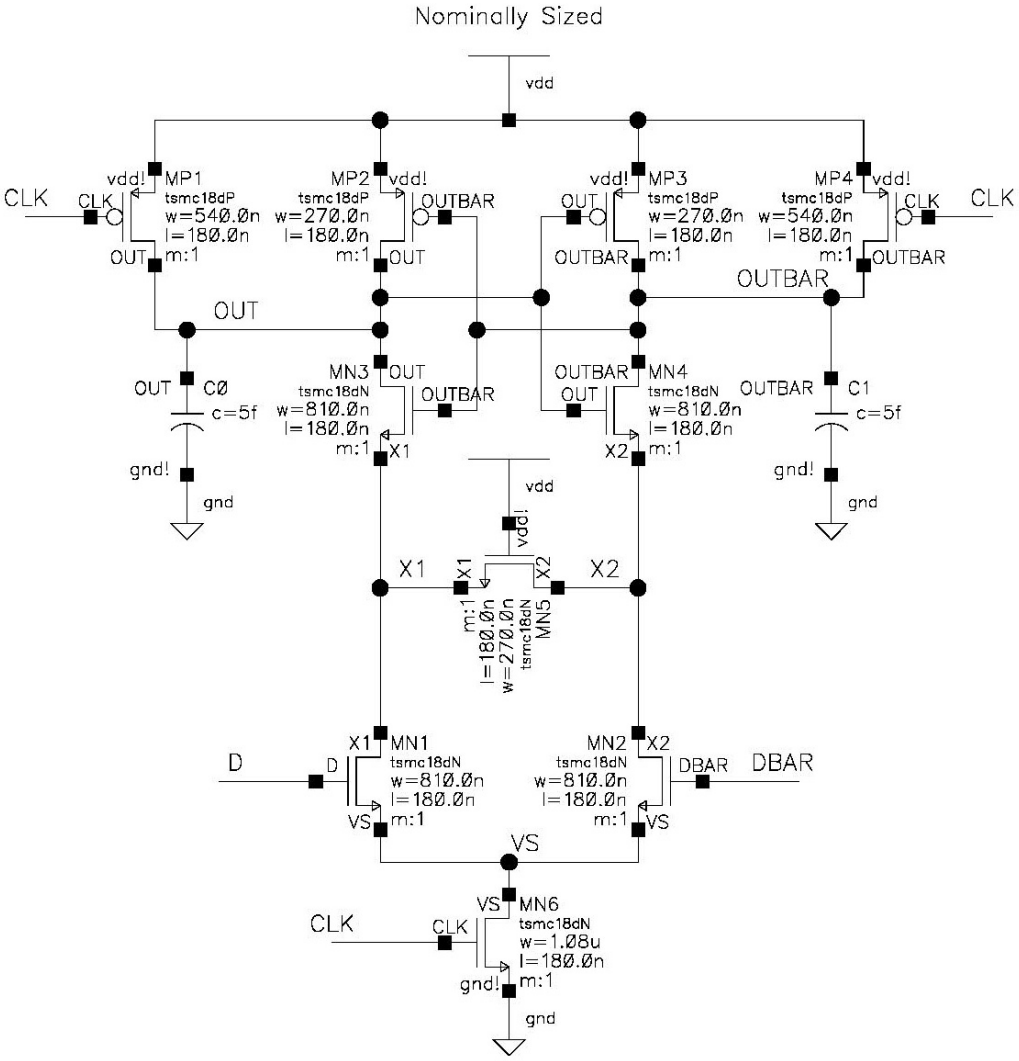


Figure 5.1 - Sense-Amplifier Schematic w/ Nominal Sizes

5.2. Specific Application: For this example, the sense-amp is used to receive a small differential pulsed signal from a lossy transmission line. The sense-amp is the front-end of a sense-amplifier-based flip-flop, therefore the amplifier's load is the input to a differential slave latch. Only the approximate input capacitance of the slave latch is known. The operating specifications are to be as follows:

- Minimum differential input voltage is 200mV peak-to-peak (100mV single-ended pulse).
- Clock frequency of 1.5GHz. (Assume clock rise time of ~60ps)
- Single-ended external capacitive load of 5fF.
- Common-mode input voltage of 0.9V.
- Average Power Dissipation < 90uW
- Maximum clock-to-output delay of 140ps.
- Nominal aperture time of 200ps

5.3. Advantages of Topology

- The input isolation provided by N5 protects against unwanted changes on the input data affecting the sense nodes once regeneration begins. [2]
- The gate-isolated topology suffers from less charge injection than other conventional clocked amplifiers since the input signal is not gated directly to the sense nodes. When the clock falls, only the tail current is turned off and charge cannot be injected from the sense nodes back into the input. [1]

5.4. Disadvantages of Topology

- This amplifier has a limited common-mode range due to the source-coupled pair. The input common-mode voltage must be high enough to saturate N1, N2, and N6 for speedy evaluation. In addition, N1 and N2 must be over-driven enough to completely discharge the output to ground before the end of the evaluation phase. Low common-mode input will drastically reduce the maximum clock speed. [1]

5.5. Design Approach:

5.5.1. The primary design goal is speed. Therefore the conductance of the evaluation path should be maximized, while the capacitance is minimized. However, the sizing must be balanced such that the evaluation path can be quickly turned off and the output pulled high during the pre-charge phase. Unlike the source follower design, the complexity of this circuit does not lend itself to hand analysis using explicit equations for individual transistors. Instead, an analysis of the function of each transistor and the load each sees and presents to the circuit is a better tool for finding approximate sizes. The "fanout-of-4" (FO-4) and 0.11fF per λ of gate width[1] approximations are suitable for the initial analysis. Since each output node drives 5fF of external load capacitance, the FO-4 approximation shows that the internal capacitance on the output node should be about 1.25fF. Using the second approximation, 1.25fF equates to approximately 11.4 λ of gate width. This rule of thumb can now be used as a guide for initial transistor scaling.

[1] This is a rough approximation for NMOS and PMOS transistors having gate widths in the expected size range. The values are derived from Table 3.1 and 3.2.

5.5.2. Initial Transistor Sizing:

5.5.2.1. P1 and P4 are responsible for pre-charging the internal nodes to a voltage near VDD in a half clock cycle. Therefore, P1 and P4 must be large enough to pull the output high before the rising edge of the clock, but small enough as to not add too much capacitance. Depending on the clock period, minimum width or 2x minimum width is a good starting point.

5.5.2.2. P2, P3, N3, and N4 make up cross-coupled inverters. N3 and N4 lie in the evaluation path; therefore their resistance should be as low as possible. Since P2 and P3 are not in the evaluation path, they mainly add capacitance to the output. The previous two points show that N3 and N4 should be large in comparison to P2 and P3. Given the preceding argument, P2 and P3 can be left minimum sized, and N3 and N4 should be 2x or 3x minimum sizing to provide strong pull-down.

5.5.2.3. N1, being in series with N3, can be initially sized equal to N3. The same is true for the N2, N4 pair. Therefore to maintain matching, N1-N4 can be sized and scaled alike.

5.5.2.4. N5 should remain minimum sized to reduce capacitance. Increasing the width of N5 would negligibly improve its ability to short the sense nodes.

5.5.2.5. N6 should be relatively large to improve current drive and increase speed. 3x or 4x minimum size is a reasonable starting point.

5.5.2.6. The circuit must remain exactly symmetric to maintain proper operation. Any imbalance in the capacitance on the sense nodes translates to an offset

voltage during amplification. The nominal values for gate width are listed below in Table 5.1.

Table 5.1 - Nominal Gate Widths for Sense-Amplifier

Ref. Designator:	N1	N2	N3	N4	N5	N6	P1	P2	P3	P4
Nominal Sizes (um)	0.810	0.810	0.810	0.810	0.270	1.080	0.540	0.270	0.270	0.540

5.6. Analysis Setup and Approach: Since the primary goal is speed, only certain transistors in the circuit were analyzed for performance sensitivity in the following tests:

5.6.1. The tail transistor, N6, was scaled alone to analyze the effect of changing the bias current.

5.6.2. N1, N2, N3, and N4 were scaled to analyze the effect of changing the transconductance of the evaluation path.

5.6.3. N1-N4 and N6 were also scaled together to see the combined effects of the first two tests.

5.6.4. Finally, P1 and P4 were scaled to analyze the trade-offs between pre-charge rise-time and added regeneration delay due to increased PMOS loading.

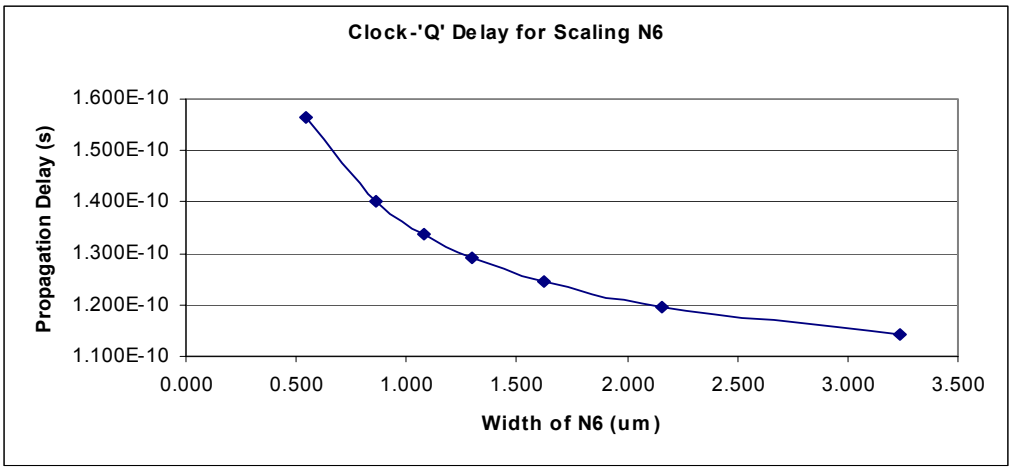
5.7. Sensitivity Analysis

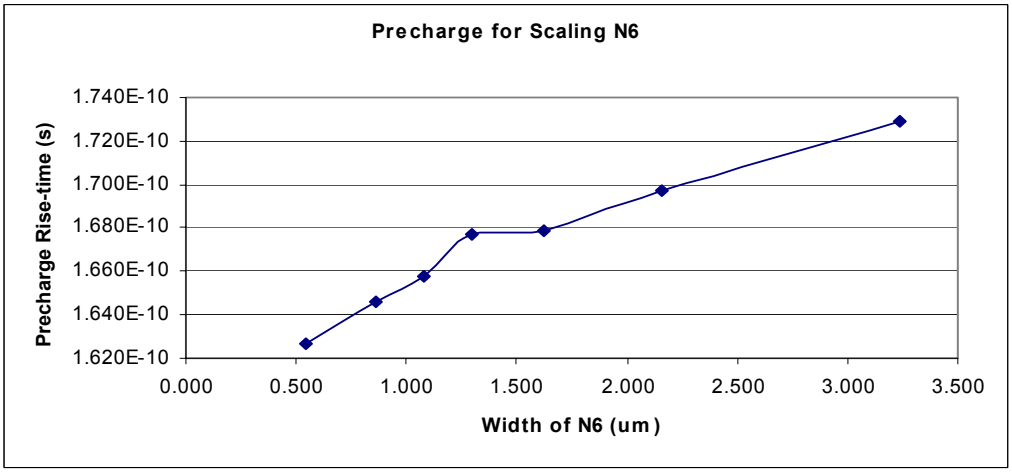
5.7.1. N6 – The tail transistor has a moderate effect on the delay when scaled alone. Table 5.2 shows that a small improvement in delay can be realized by increasing the gate width of N6 up to ~180% of its nominal value, with a moderate increase in power consumption and pre-charge rise-time compared to that seen when scaling N1-N4 alone. These results show that tweaking the size of N6 with respect to N1-N4 can produce positive results at minimal costs. Figure 5.2

indicates that scaling N6 beyond 180-200% continues to decrease delay, but with diminishing returns.

Table 5.2 - Sensitivity Results for Scaling N6

N6 Scaled:	**50%**	**80%**	**Nominal**	**120%**	**150%**	**200%**	**300%**
Gate Width (um)	0.540	0.864	1.080	1.296	1.620	2.160	3.240
Delay	1.564E-10	1.401E-10	1.338E-10	1.293E-10	1.246E-10	1.195E-10	1.141E-10
Improv. from Nom.	-2.260E-11	-6.300E-12		4.500E-12	9.200E-12	1.430E-11	1.970E-11
% Change	-16.891%	-4.709%		3.363%	6.876%	10.688%	14.723%
Prchg. Risetime	1.627E-10	1.646E-10	1.658E-10	1.677E-10	1.679E-10	1.697E-10	1.729E-10
Improv. from Nom.	3.100E-12	1.200E-12		-1.900E-12	-2.100E-12	-3.900E-12	-7.100E-12
% Change	1.870%	0.724%		-1.146%	-1.267%	-2.352%	-4.282%
Avg. Power	7.891E-05	8.304E-05	8.468E-05	8.598E-05	8.758E-05	8.975E-05	9.314E-05
Improv. from Nom.	5.770E-06	1.640E-06		-1.300E-06	-2.900E-06	-5.070E-06	-8.460E-06
% Change	6.814%	1.937%		-1.535%	-3.425%	-5.987%	-9.991%
PDP (J)	1.234E-14	1.163E-14	1.133E-14	1.112E-14	1.091E-14	1.073E-14	1.063E-14





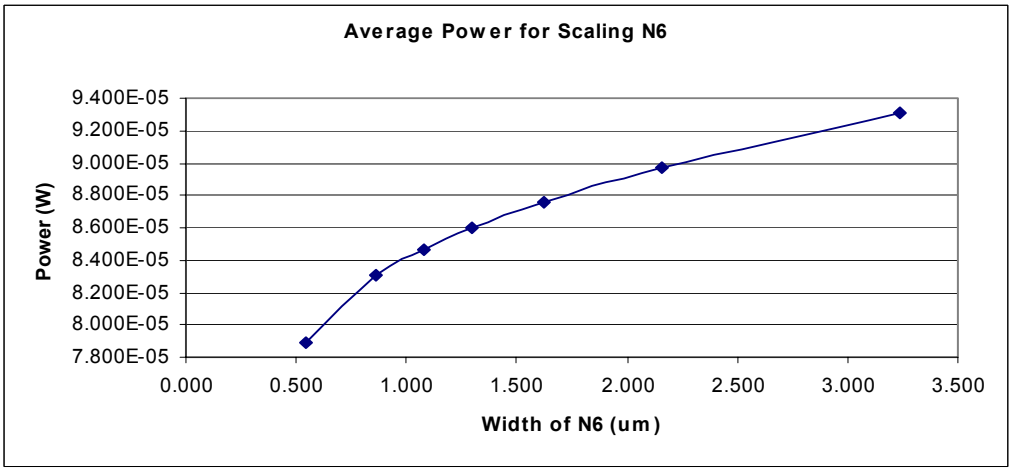


Figure 5.2 - Sensitivity Results for Scaling N6

5.7.2. N1-N4 – Since these transistors make up the majority of the evaluation path, N1-N4 have a greater effect on the delay than does N6 alone, but at a greater cost in power and particularly pre-charge time due to their direct influence on the load capacitance. Table 5.3 shows that the decrease in delay when scaling N1-N4 is almost twice that when scaling N6, but the increase in power and pre-charge rise-time is approximately 3 and 4 times greater respectively. Figure 5.3 also shows that increasing N1-N4 beyond 200% of their nominal values actually produces a negative effect on the delay. These numbers show that scaling N1-N4 alone produces few outstanding positive results.

Table 5.3 - Sensitivity Results for Scaling N1-N4

N1-N4 Scaled:	**50%**	**80%**	**Nominal**	**120%**	**150%**	**200%**	**300%**
Gate Width (um)	0.405	0.648	0.810	0.972	1.215	1.620	2.430
Delay	1.819E-10	1.456E-10	1.334E-10	1.256E-10	1.185E-10	1.138E-10	1.211E-10
Improv. from Nom.	-4.850E-11	-1.220E-11		7.800E-12	1.490E-11	1.960E-11	1.230E-11
% Change	-36.357%	-9.145%		5.847%	11.169%	14.693%	9.220%
Prchg. Risetime	1.471E-10	1.576E-10	1.654E-10	1.734E-10	1.854E-10	2.056E-10	2.422E-10
Improv. from Nom.	1.830E-11	7.800E-12		-8.000E-12	-2.000E-11	-4.020E-11	-7.680E-11
% Change	11.064%	4.716%		-4.837%	-12.092%	-24.305%	-46.433%
Avg. Power	7.380E-05	8.110E-05	8.474E-05	8.817E-05	9.310E-05	1.014E-04	1.165E-04
Improv. from Nom.	1.094E-05	3.640E-06		-3.430E-06	-8.360E-06	-1.666E-05	-3.176E-05
% Change	12.910%	4.295%		-4.048%	-9.865%	-19.660%	-37.479%
PDP (J)	1.342E-14	1.181E-14	1.130E-14	1.107E-14	1.103E-14	1.154E-14	1.411E-14

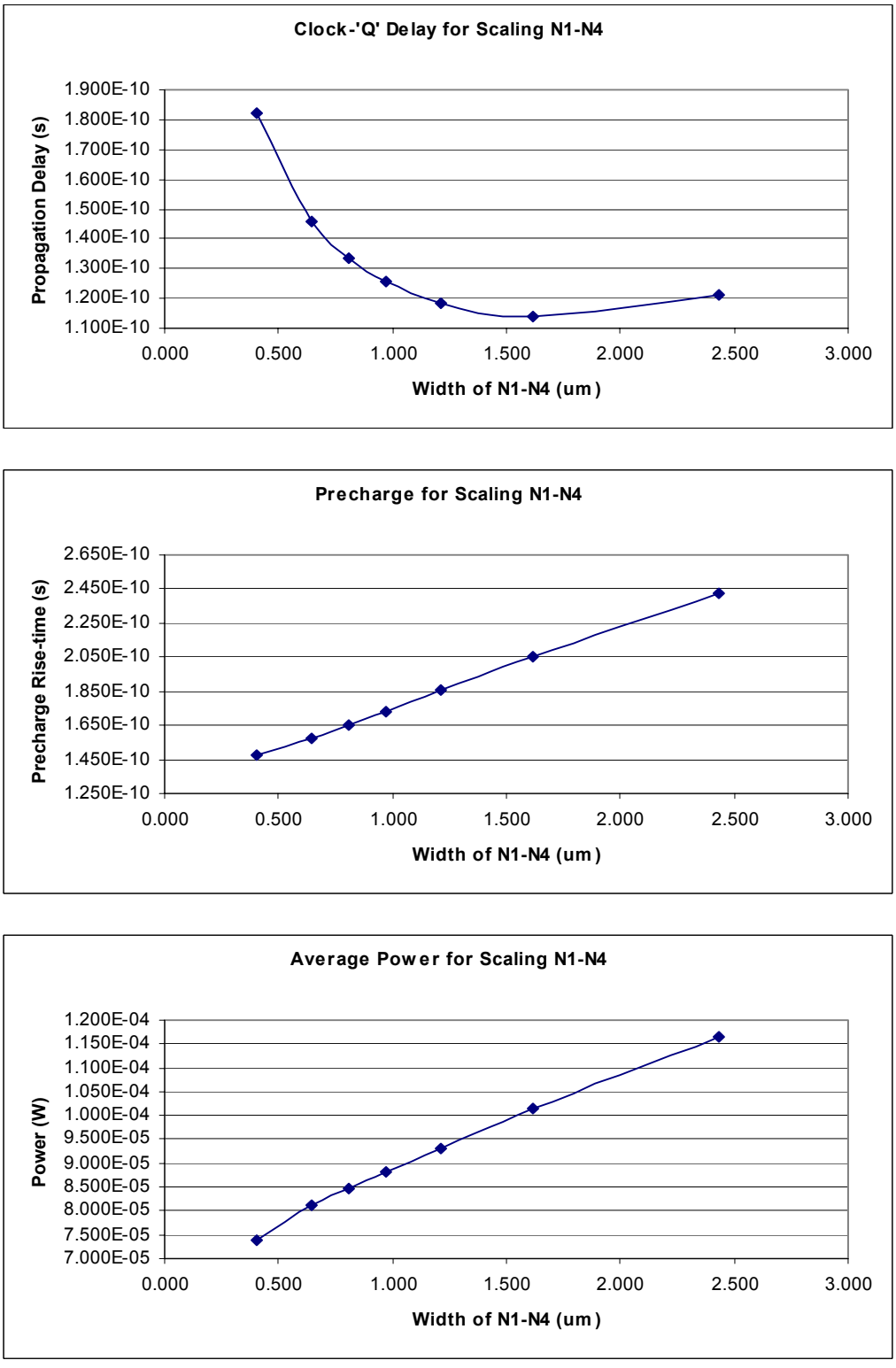


Figure 5.3 - Sensitivity Results for Scaling N1-N4

5.7.3. N1-N4 & N6 – The results from sections 5.7.1 and 5.7.2 indicate that scaling N6 or N1-N4 individually produces mediocre results. However, Table 5.4 shows that scaling these five transistors together decreased the delay by the greatest percentage for a moderate power cost if speed is the primary concern. Though the pre-charge rise-time still suffers, P1 and P4 can be scaled to balance the charge and discharge phases and account for the added capacitance of N3 and N4. Figure 5.4 shows that the lowest delay can be achieved in this case, compared to the cases discussed in the previous two sections.

5.7.4. P1 & P4 – The pre-charge rise-time is very sensitive to scaling P1 and P4, decreasing by ~14% and 28% for a 20% and 50% increase in gate width respectively (shown in Table 5.5). However, Figure 5.5 shows that increasing the width of P1 and P4 beyond 200% of the nominal value produces diminishing returns and begins to significantly increase the delay due to the added capacitance. The pre-charge rise-times are compared for all 4 scaling cases in Figure 5.6.

5.7.5. Common-mode (CM) input shift – Though the CM input level is specified at 0.9V, it is important to understand the effects of a change in the CM input voltage. Below 0.7V, the input transistors are not supplied sufficient over-drive voltage and the sense-amplifier failed to regenerate the output during the evaluation phase. If the CM level drifts as low as 0.7V with nominal sizing, Table 5.6 shows that the delay almost doubles and Figure 5.7a shows that the amplifier is on the verge of not working at all. Figure 5.7 indicates that

increasing the CM level beyond 0.9V continues to decrease the delay, but with diminishing returns.

Table 5.4 - Sensitivity Results for Scaling N1-N4 & N6

N1-N4 & N6							
Scaled:	**50%**	**80%**	**Nominal**	**120%**	**150%**	**200%**	**300%**
Gate Width N1-N4(um)	0.405	0.648	0.810	0.972	1.215	1.620	2.430
Gate Width N6 (um)	0.540	0.864	1.080	1.296	1.620	2.160	3.240
Delay	2.016E-10	1.513E-10	1.330E-10	1.204E-10	1.077E-10	9.600E-11	failed
Improv. from Nom.	-6.860E-11	-1.830E-11		1.260E-11	2.530E-11	3.700E-11	#VALUE!
% Change	-51.579%	-13.759%		9.474%	19.023%	27.820%	#VALUE!
Prchg. Risetime	failed	1.566E-10	1.653E-10	1.741E-10	1.872E-10	2.088E-10	failed
Improv. from Nom.	#VALUE!	8.700E-12		-8.800E-12	-2.190E-11	-4.350E-11	#VALUE!
% Change	#VALUE!	5.263%		-5.324%	-13.249%	-26.316%	#VALUE!
Avg. Power	6.608E-05	7.941E-05	8.475E-05	8.953E-05	9.650E-05	1.078E-04	failed
Improv. from Nom.	1.867E-05	5.340E-06		-4.780E-06	-1.175E-05	-2.305E-05	#VALUE!
% Change	22.029%	6.301%		-5.640%	-13.864%	-27.198%	#VALUE!
PDP (J)	1.332E-14	1.201E-14	1.127E-14	1.078E-14	1.039E-14	1.035E-14	

Table 5.5 - Sensitivity Results for Scaling P1 & P4

P1 & P4							
Scaled:	**50%**	**80%**	**Nominal**	**120%**	**150%**	**200%**	**300%**
Gate Width (um)	0.270	0.432	0.540	0.648	0.810	1.080	1.620
Delay	1.035E-10	1.287E-10	1.336E-10	1.341E-10	1.366E-10	1.405E-10	1.474E-10
Improv. from Nom.	3.010E-11	4.900E-12		-5.000E-13	-3.000E-12	-6.900E-12	-1.380E-11
% Change	22.530%	3.668%		-0.374%	-2.246%	-5.165%	-10.329%
Prchg. Risetime	4.118E-10	1.982E-10	1.671E-10	1.442E-10	1.200E-10	9.471E-11	6.860E-11
Improv. from Nom.	-2.447E-10	-3.110E-11		2.290E-11	4.710E-11	7.239E-11	9.850E-11
% Change	-146.439%	-18.612%		13.704%	28.187%	43.321%	58.947%
Avg. Power	7.887E-05	8.320E-05	8.471E-05	8.566E-05	8.673E-05	8.831E-05	9.132E-05
Improv. from Nom.	5.840E-06	1.510E-06		-9.500E-07	-2.020E-06	-3.600E-06	-6.610E-06
% Change	6.894%	1.783%		-1.121%	-2.385%	-4.250%	-7.803%

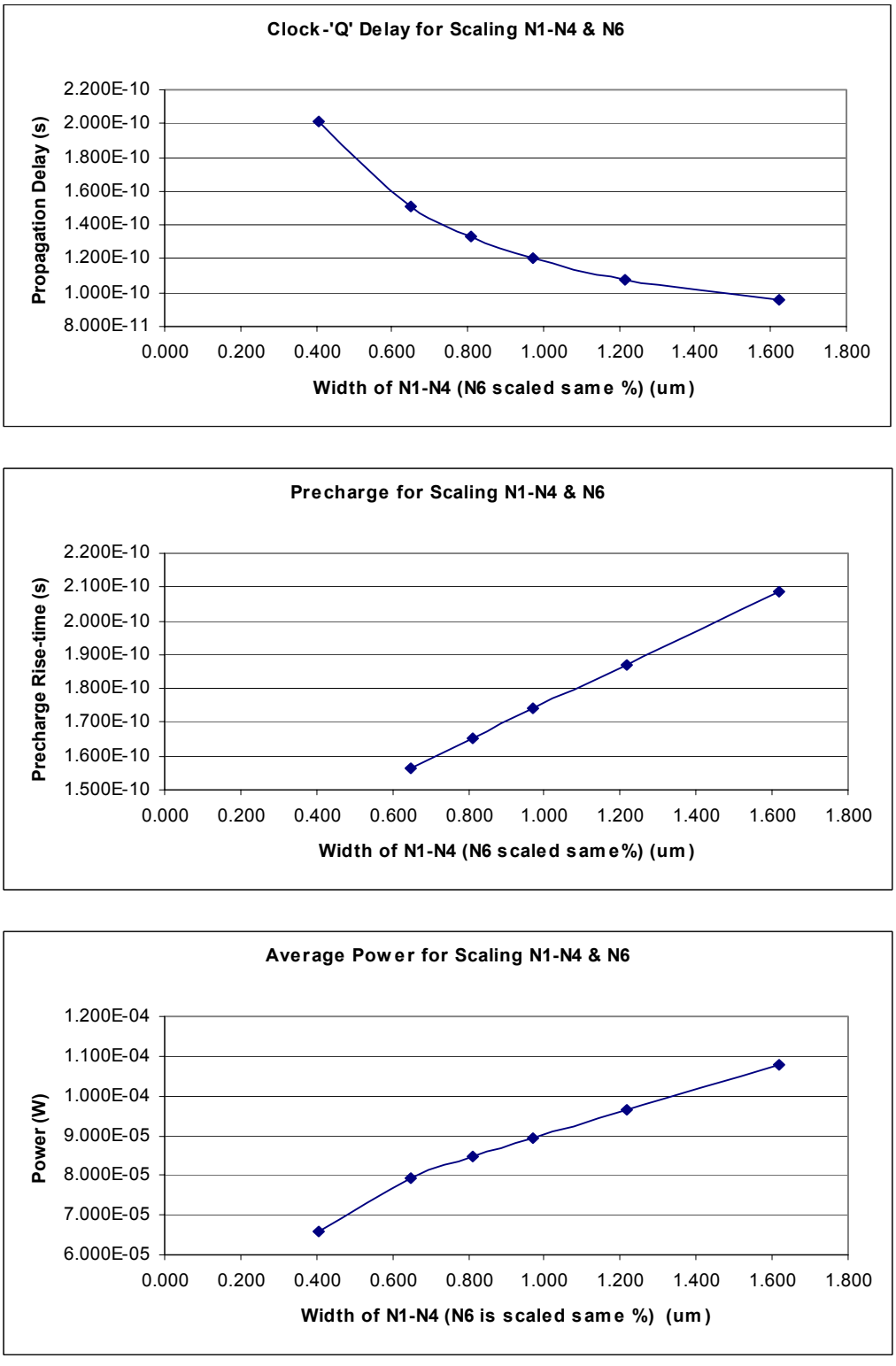
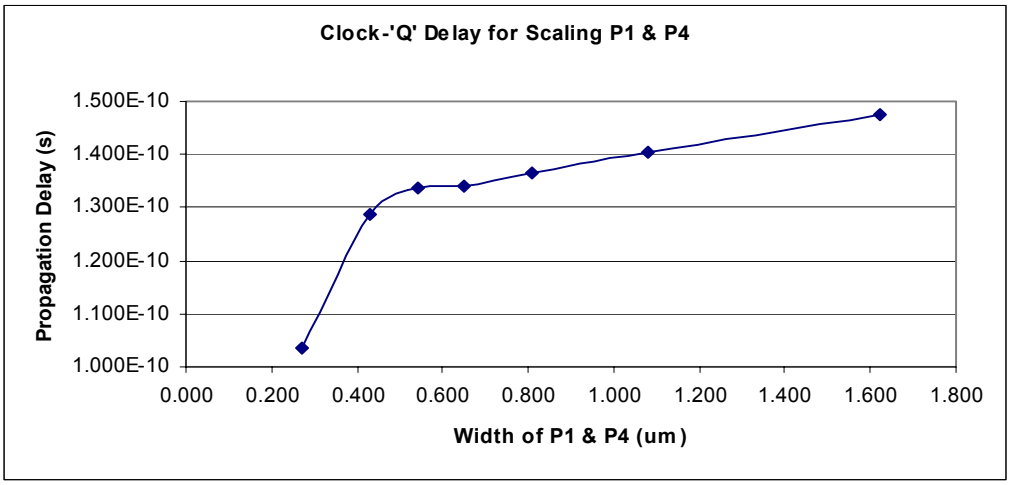
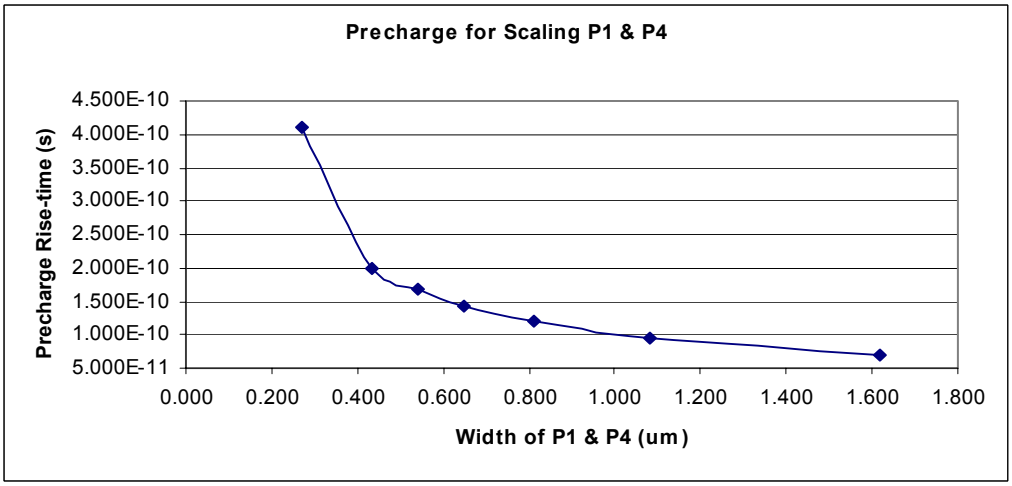


Figure 5.4 - Sensitivity Results for Scaling N1-N4 & N6





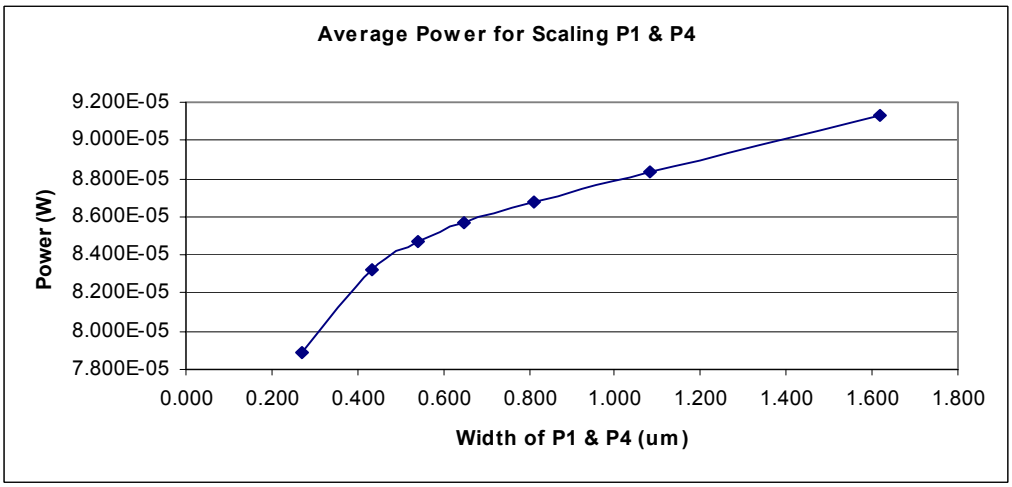


Figure 5.5 - Sensitivity Results for Scaling P1 & P4

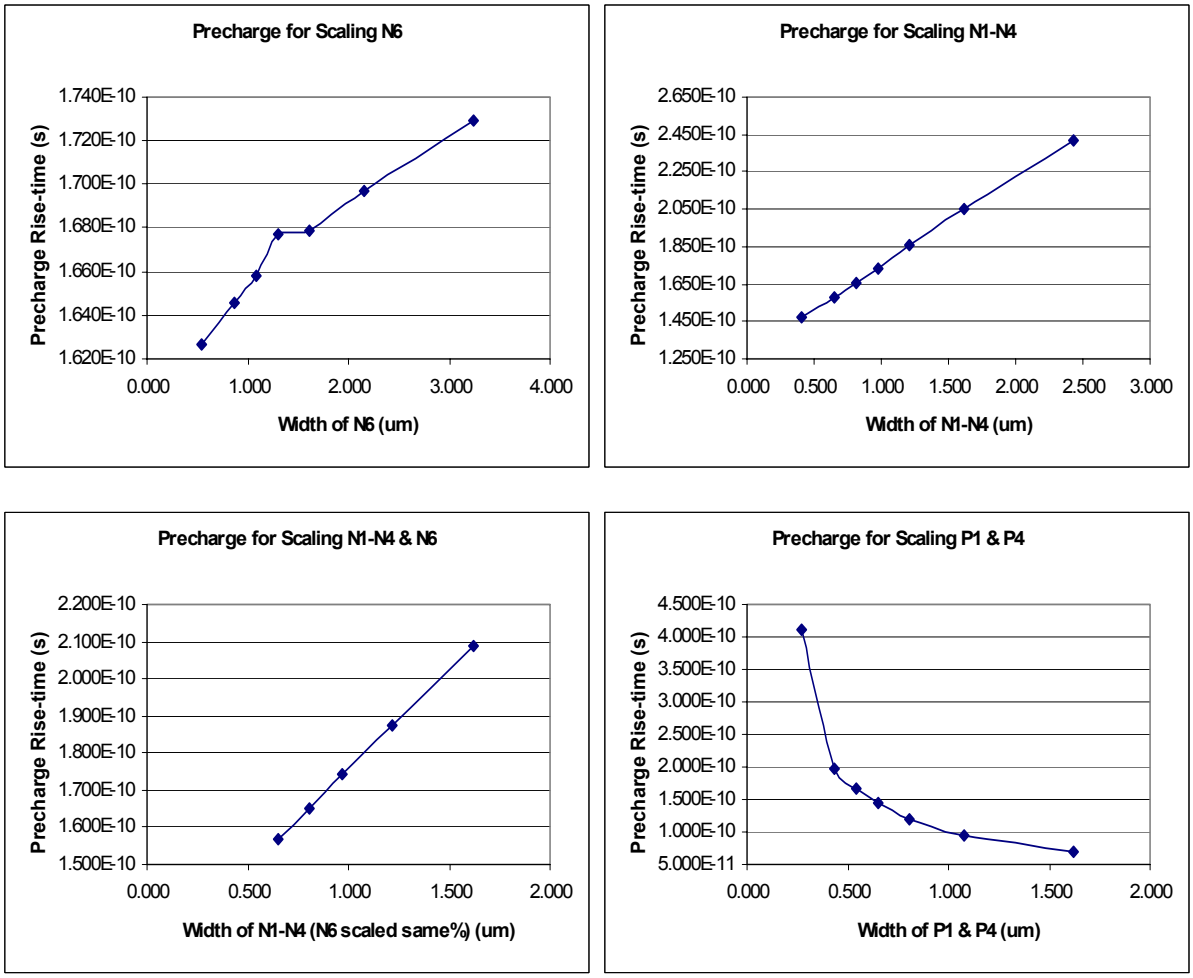


Figure 5.6 - Pre-charge Sensitivity Comparison

Table 5.6 - Delay Sensitivity to Common-mode Input Voltage Shift

Input CM Voltage (V):	0.700	0.800	0.900	1.000	1.100
Delay	2.644E-10	1.74E-10	1.369E-10	1.17E-10	1.049E-10
Improvement from Nom.	-1.275E-10	-3.720E-11		2.000E-11	3.200E-11
% Change	-93.134%	-27.173%	Nominal	14.609%	23.375%

Input CM Voltage (V):	1.200	1.300	1.400	1.500	1.600
Delay	9.782E-11	9.393E-11	9.174E-11	9.079E-11	9.077E-11
Improvement from Nom.	3.908E-11	4.297E-11	4.516E-11	4.611E-11	4.613E-11
% Change	28.546%	31.388%	32.988%	33.682%	33.696%

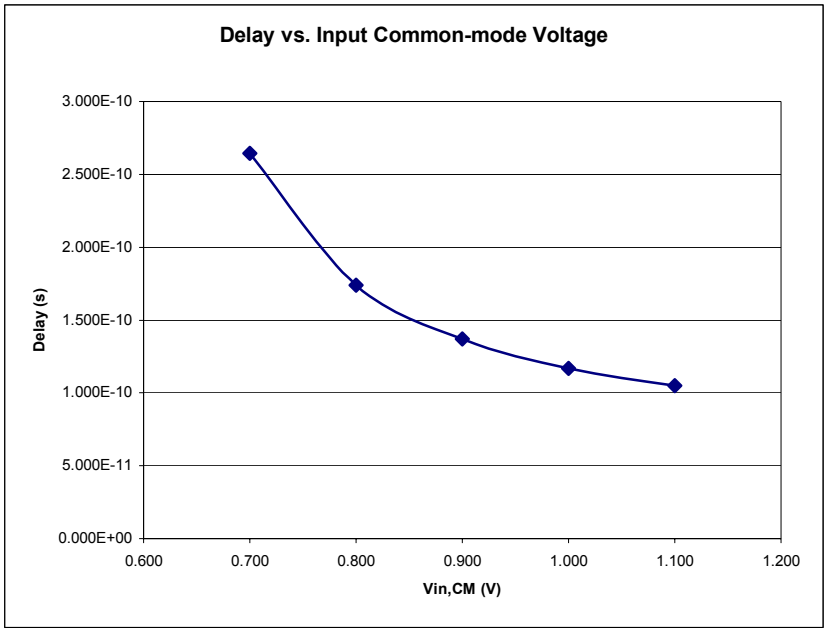


Figure 5.7 - Delay Sensitivity to Common-mode Input Voltage Shift

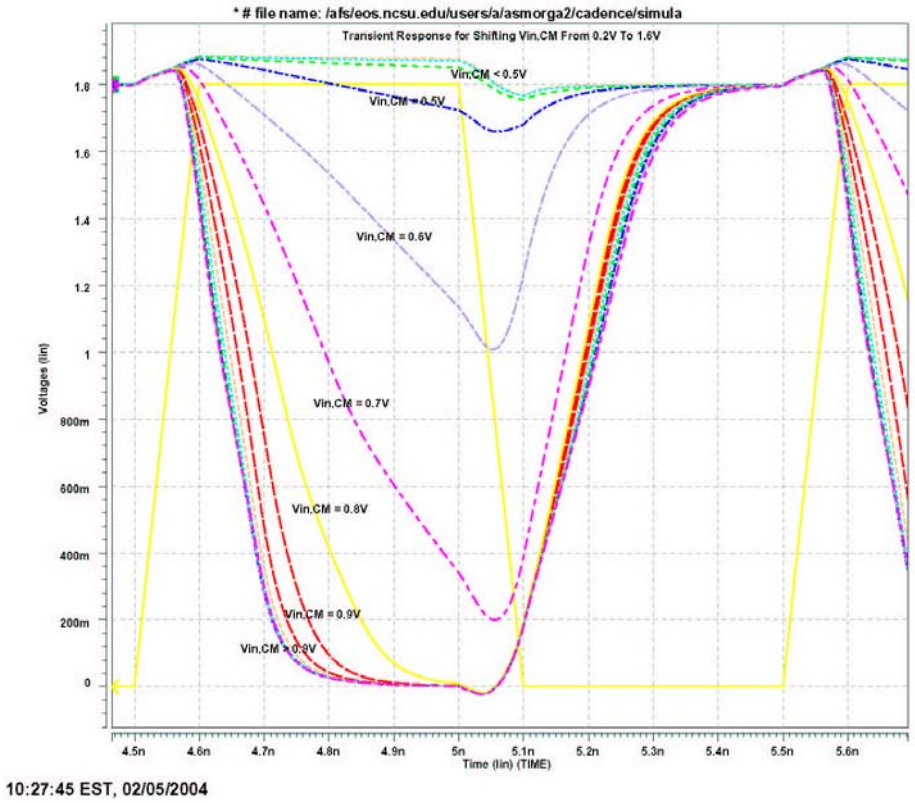


Figure 5.7a - Single-Ended Transient Response Plots Common-mode Shifting

5.7.6. Aperture Time – The nominal aperture time specification of 200ps is sufficient to swing the sense nodes into regeneration, but a decrease in the time that the data is stable at the inputs begins to increase the delay as shown in Table 5.7 and Figure 5.8. Aperture times of less than 50ps are not sufficient to regenerate the output, which can be seen in Figure 5.8a.

Table 5.7 - Delay Sensitivity to Allowed Aperture Time

Aperture Time	5.000E-11	7.500E-11	1.000E-10	1.500E-10	5.000E-10
Delay	1.714E-10	1.494E-10	1.418E-10	1.373E-10	1.369E-10
Improvement from Nom.	-3.450E-11	-1.250E-11	-4.900E-12	-4.000E-13	
% Change	-25.20%	-9.13%	-3.58%	-0.29%	Nominal
(measured from 50% point of CLK to 50% point of Data)					

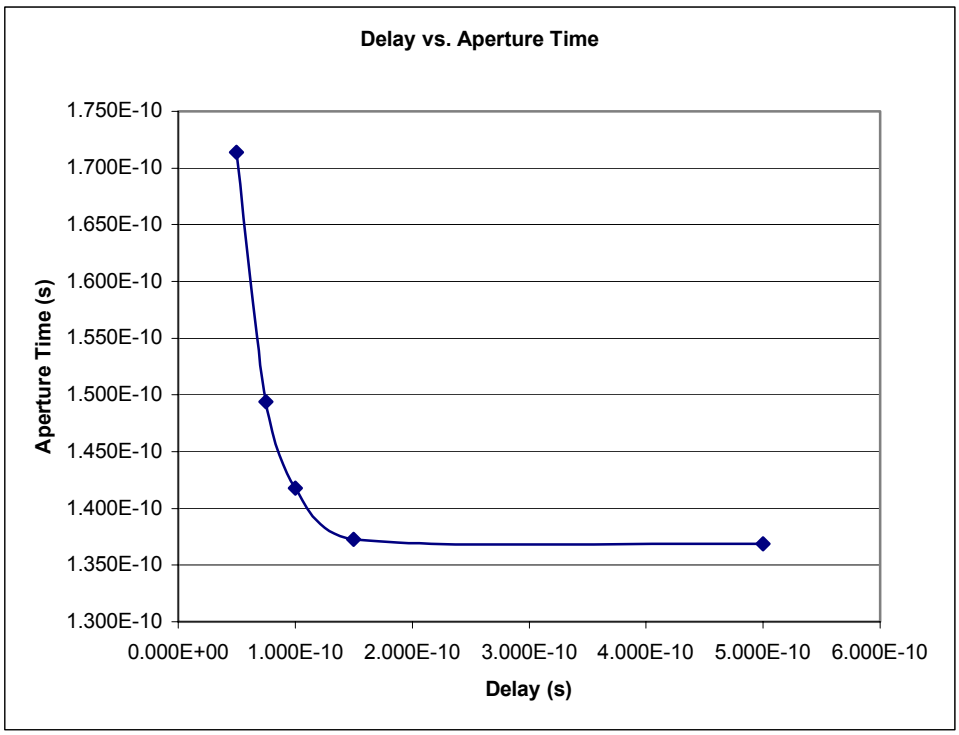


Figure 5.8 - Delay Sensitivity to Allowed Aperture Time

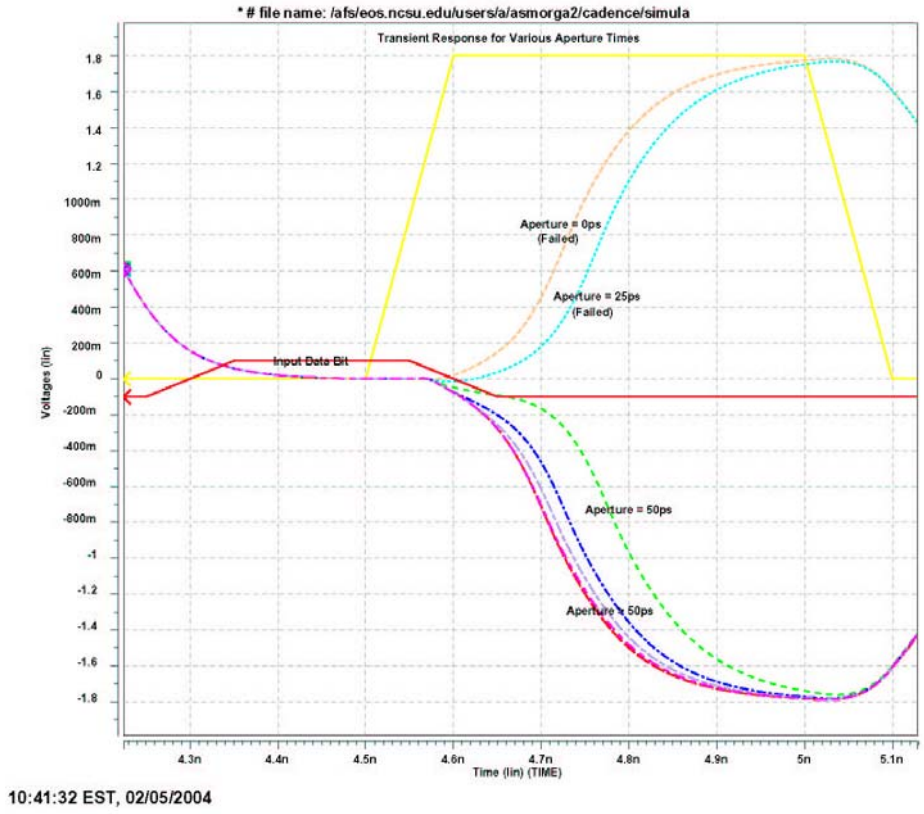


Figure 5.8a - Single-Ended Transient Response Plots for Aperture Time Sweep

5.8. Key Considerations

5.8.1. Power-Delay Product (PDP) – As with most digital circuits, there is a significant trade-off between power and delay. The delay results for all four of scaling cases are presented together for comparison in Figure 5.9. Likewise, the power results are presented together in Figure 5.10. However, Figure 5.11, which shows the power-delay product for scaling N1-N4 and & N6 together and separately, best illustrates the power-delay trade-off. The three graphs show that N6 is the key to minimizing PDP, but that scaling beyond 200% of the nominal values yields a small decrease in PDP for the area that will be consumed.

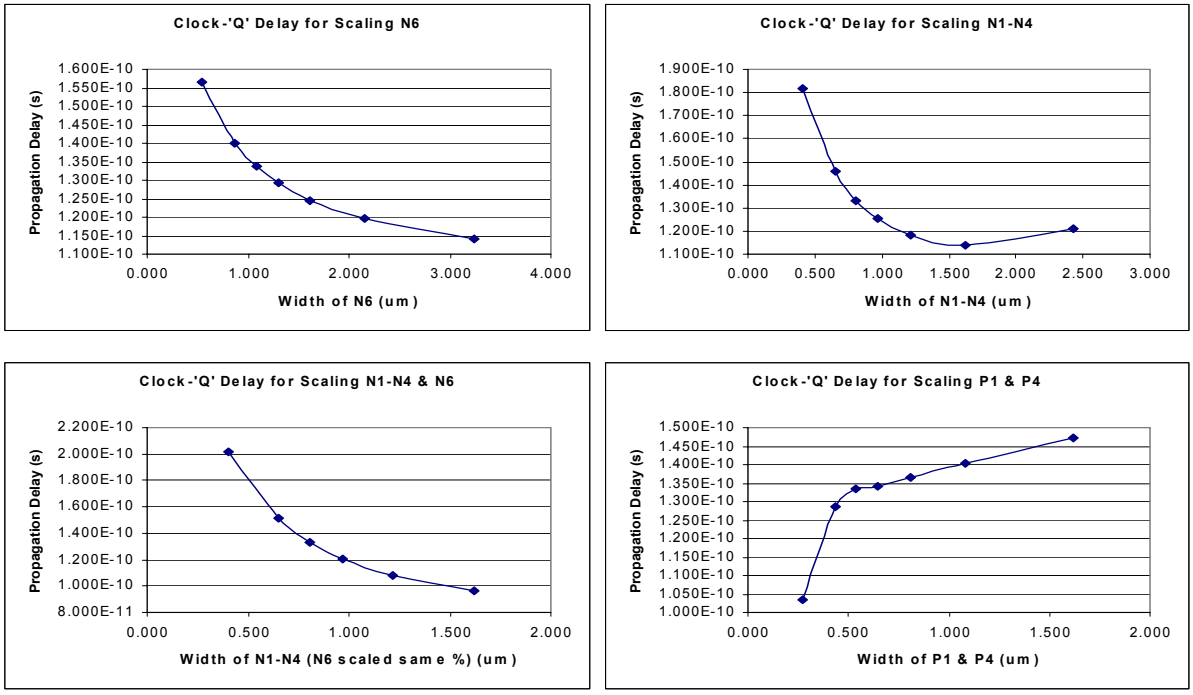


Figure 5.9 - Clock-to-Output Sensitivity Comparison

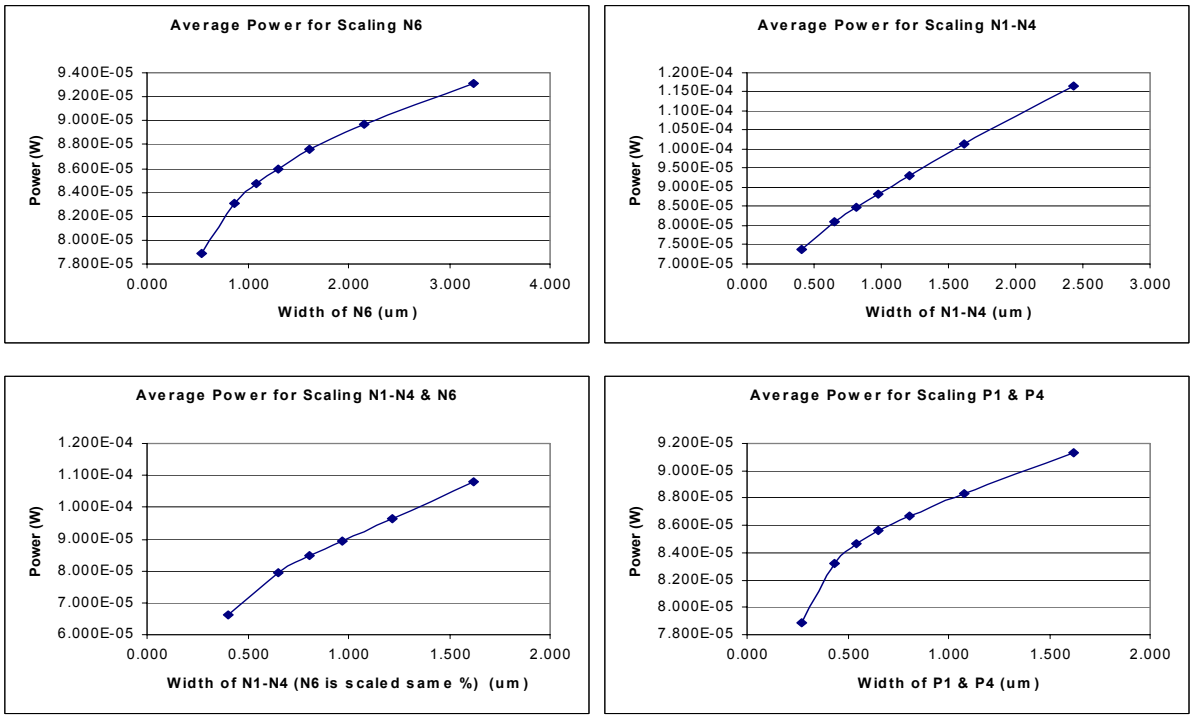
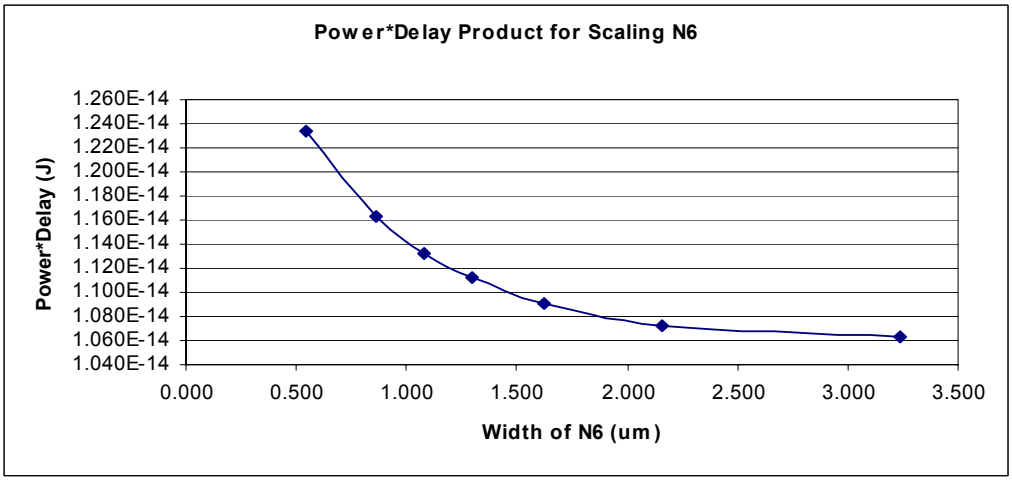
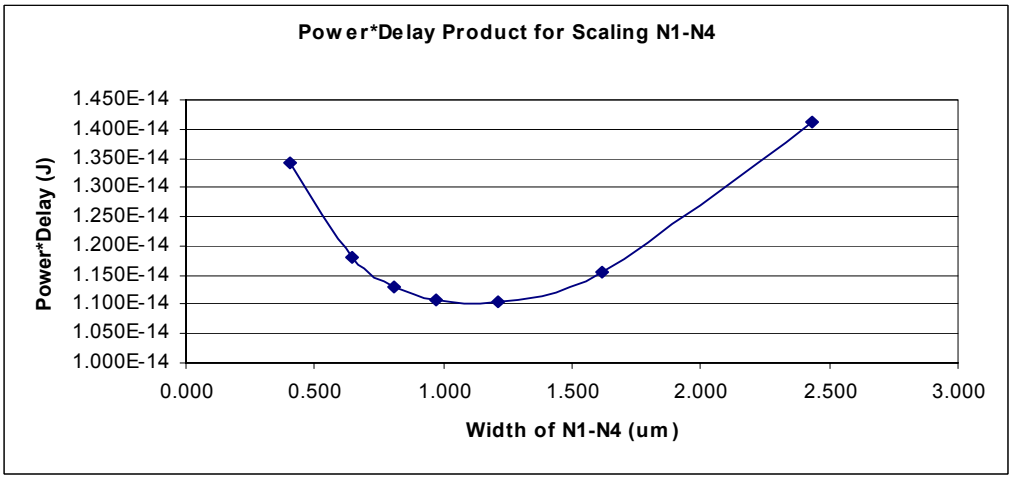


Figure 5.10 - Average Power Comparison





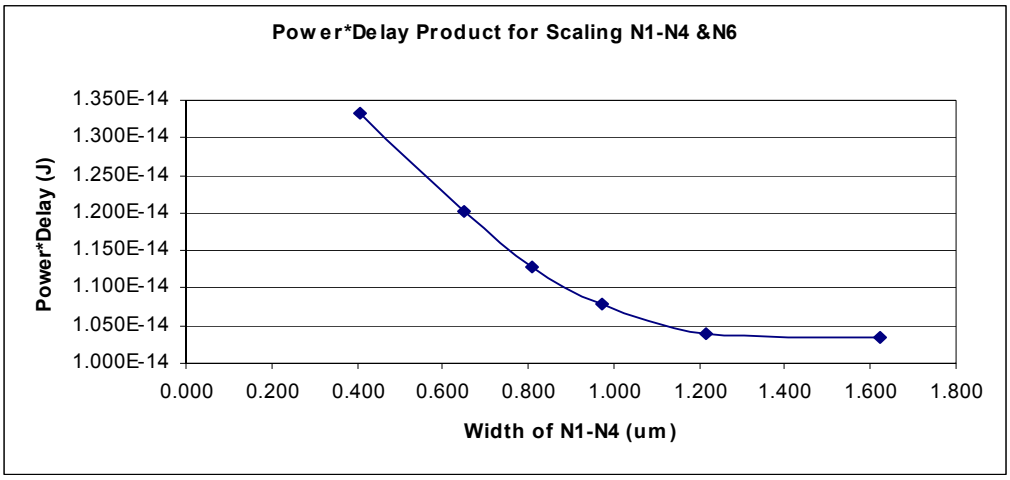


Figure 5.11 – Power-Delay Product Sensitivity Comparison

5.8.2. Elmore Delay – Though not explored in this analysis, transistors N3 and N4 could be scaled separately from N1 and N2 to examine benefits from optimizing the Elmore delay in the evaluation path. Adjusting the ratio of N3 to N1 (likewise for N4 to N2) to keep the same low series resistance, but lower the capacitance closest to the output nodes, would likely further reduce the delay.

5.8.3. Phase Balancing – Though delay minimization is key, the sizing of the pull-down transistor with respect to the pre-charge transistors is important to ensure that all internal nodes are completely pre-charged and avoid offset on the sense nodes. If the sense nodes do not pre-charge to the same voltage, the regeneration of the differential input voltage is corrupted due to the imbalanced charge remaining on the sense nodes. This may result in a bit error.

6. Schmidt Trigger Design Example

6.1. Description: This circuit is essentially an inverter with differing positive and negative switching thresholds, though it requires 3 times the number of transistors as a simple inverter. A Schmidt trigger employs positive feedback to increase the threshold at which the output switches for rising inputs and decrease the threshold at which the output switches for falling inputs. The difference between the threshold for a low-to-high transition (V_{IL}) and the threshold for a high-to-low transition (V_{IH}) is called the hysteresis voltage ($V_{HYS} = V_{IL} - V_{IH}$). A major advantage of separating the positive and negative thresholds is the ability to reduce unwanted switching due to a noisy input signal wandering at mid-level[1]. There are several Schmidt trigger topologies, but this example focuses on the circuit in Figure 6.1. Transistors NF and PF provide the positive feedback that ultimately controls the values for V_{HYS} .

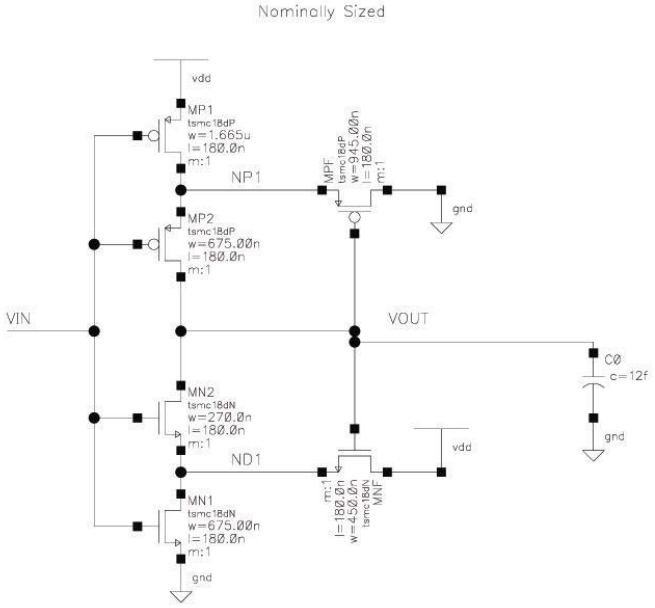


Figure 6.1 - Schmidt Trigger Schematic w/ Nominal Sizes

[1] The advantages of the Schmidt Trigger are discussed in detail in section 6.3.

6.2. Specific Application: For this example, a Schmidt trigger will be used to clean up a noisy digital signal before it is to be driven off chip. These signals have become ridden with glitches and lost edge-rate due to having traveled a considerable distance across chip. The waveform must be cleaned up before it is buffered off-chip. The estimated external capacitive load is 12fF, which is the approximate input capacitance of a 2.5x sized inverter. The operating specifications are to be as follows:

- Clock frequency of 1.5GHz. (Assume clock rise time of ~60ps)
- Full-swing (0-1.8V) digital input signal
- Single-ended external capacitive load of 12fF.
- V_{HYS} = 400mV, with V_{IH} = 700mV and V_{IL} =1.1V
- Average Power Dissipation < 300uW
- Maximum Propagation Delay[2], $T_p = 0.5*[T_{PHL} + T_{PLH}] = 150ps$

6.3. Advantages of Topology

- The hysteresis added by moving both the positive and the negative thresholds increases the noise margin of both the high and the low states. This is superior to hysteresis added in an inverter, where only one state's noise margin can be increased at the expense of the opposite state. Specific benefits include:
    - Protecting against high signals glitching low, and low signals glitching high.
    - Removing ringing in steady-state

[2] T_{PHL} = Propagation delay for a high-to-low transition on the output, and T_{PLH} = Propagation delay for a low-to-high transition on the output

- Squaring up a slowly changing edge that may even change direction in the hysteresis region
- Remove erratic switching behavior due to slow input signals

- Reduce power consumption by sharpening edges and reducing direct-path currents. [6]

6.4. Disadvantages of Topology

- This topology requires 3 times the number of transistors as an equivalent inverter. Moreover, the feedback transistors, NF and PF, are typically larger than the input transistors depending on the amount of needed hysteresis.
- The increased stack, compared to an inverter, plus the added capacitance of NF and PF, limit the speed of this circuit.
- Though the direct-path current through the input transistors is limited by the hysteresis, the paths formed by NF-N1 and PF-P1 draw some static current outside of switching events.

6.5. Design Approach:

6.5.1. There are two general relationships to examine when beginning the design of this type of Schmidt trigger. The first, and most basic, is the sizing relation among the pure "evaluation" transistors, N2 and P2, and the "hysteresis" transistors, N1, NF, P1, and PF, in order to achieve the desired speed. The second relationship is the size ratio of N1 to NF, and likewise the ratio of P1 to PF. These ratios determine the hysteresis voltage and their derivation will be discussed in the later subsections of section 6.5.

6.5.2. For now, we will begin by sizing N2 and P2 with minimum gate width, but maintaining a 2.5/1 reference inverter P/N ratio. This yields W_{P2}=675nm and W_{N2}=270nm as shown in Figure 6.1. These transistors are kept small to minimize the output capacitance, because NF and PF will already add substantial capacitance to the output and internal nodes. We can afford the increased resistance of the small N2 and P2 devices because N1 and P1 will likely need to be larger to produce the needed hysteresis, and will therefore contribute lower series resistances. This arrangement of placing the larger devices, and therefore larger capacitance, near the power and ground nodes will reduce the Elmore delay to the output node.

6.5.3. To approximate the gate widths of N1 vs. NF and P1 vs. PF, a mathematical analysis of the DC transfer characteristic must be performed. Since the sizes of NF and PF ultimately control the value of V_{IL} and V_{IH} respectively, an expression for these voltage thresholds must be derived in terms of gate widths. First an expression for V_{IL}:

- Begin by assuming that as V_{in} increases (from zero) above V_{tn1}, N1 turns on in saturation and NF is already in saturation. Therefore current flows through NF and N1 and there is a nonzero voltage on the drain of N1. Also assume that V_{out} remains at V_{dd} until N2 begins to turn on and discharge the output. N2 will not begin to turn on until V_{in} rises above $V_{dsN1} + V_{tn2}$ (ignore body effect and assume $V_{tn2}=V_{ton}$). To solve the value of V_{in} for which N2 turns on, V_{ds1} must first be found using the current equation.

Equating the saturation currents in NF and N1, noting that $V_{gsN1} = V_{in}$, and also noting that $V_{gsF} = V_{out} - V_{dsN1}$ yields Eqn. 6.1.

$$\tfrac{1}{2}\mu_n C_{ox} {W_{NF}}/{L_{NF}} \left(V_{out} - V_{dsN1} - V_{ton}\right)^2 = \tfrac{1}{2}\mu_n C_{ox} {W_{N1}}/{L_{N1}} \left(V_{in} - V_{ton}\right)^2 \quad \text{Eqn. 6.1}$$

Eqn. 6.2 shows the results after canceling like terms, moving the W/L ratio for NF to the right side, and recognizing that $V_{out} = V_{dd}$ in this state.

$$\left(V_{DD} - V_{dsN1} - V_{ton}\right)^2 = \frac{{W_{N1}}/{L_{N1}}}{{W_{NF}}/{L_{NF}}} \left(V_{in} - V_{ton}\right)^2 \quad \text{Eqn. 6.2}$$

Squaring both sides and solving for V_{ds1} yields Eqn. 6.3:

$$V_{dsN1} = V_{DD} - V_{ton} - \sqrt{\frac{{W_{N1}}/{L_{N1}}}{{W_{NF}}/{L_{NF}}}} \left(V_{in} - V_{ton}\right) \quad \text{Eqn. 6.3}$$

Recalling that V_{in} must be greater than $V_{dsN1} + V_{tn2}$ (where $V_{tn2}=V_{ton}$) for N2 to turn on, an expression for the input voltage that switches the output low is shown in Eqn. 6.4. This input voltage is referred to as V_{IL}. After some algebraic manipulation, a more useful expression for V_{IL} is shown in Eqn. 6.5.

$$V_{in} = V_{DD} - \sqrt{\frac{{W_{N1}}/{L_{N1}}}{{W_{NF}}/{L_{NF}}}} \left(V_{in} - V_{ton}\right) \quad \text{Eqn. 6.4}$$

$$V_{IL} = \frac{V_{DD} + \sqrt{\frac{{W_{N1}}/{L_{N1}}}{{W_{NF}}/{L_{NF}}}} V_{ton}}{1 + \sqrt{\frac{{W_{N1}}/{L_{N1}}}{{W_{NF}}/{L_{NF}}}}} \quad \text{Eqn. 6.5}$$

6.5.4. The analysis for the approximation for V_{IH} is similar to that of V_{IL}. Assuming that P1 turns on in saturation as the input falls below $V_{DD} - |V_{tp1}|$, and that PF is

already in saturation, the current equations for P1 and PF can be equated to find an expression for V_{sdP1}. For P2 to turn on and switch the output high, the input must fall below V_{DD} - $|V_{tp2}|$ - V_{sdP1} where $V_{tp2}=V_{top.}$ The derivation for an expression for V_{sdP1} matches the derivation for V_{dsN1} so closely that it is not shown here. However, the final expression for V_{IH}, which is the input voltage that switches the output high, is shown in Eqn. 6.6.

$$V_{IH} = \frac{\sqrt{\frac{W_{P1}/L_{P1}}{W_{PF}/L_{PF}}}\left(V_{DD} + |V_{top}|\right)}{1 + \sqrt{\frac{W_{P1}/L_{P1}}{W_{PF}/L_{PF}}}} \qquad \text{Eqn. 6.6}$$

6.5.5. Equations 6.5 and 6.6 give reasonable approximations for estimating the required ratios of W_{N1}/W_{NF} and W_{P1}/W_{PF} for given V_{IL} and V_{IH} respectively. However, the assumptions made to reduce mathematical complexity greatly reduced the accuracy of the expressions and further transistor scaling will be required to achieve the exact values of hysteresis. Specifically, the equation for V_{IH} appears to be fairly inaccurate. After starting with these equations and adjusting transistor sizes for performance, the values listed in Table 6.1 were chosen as nominal for the given specification.

Table 6.1 - Nominal Gate Widths for Schmidt Trigger

Ref. Designator:	**NF**	**PF**	**N1**	**P1**	**N2**	**P2**
Nominal Sizes (um)	0.450	0.945	0.675	1.665	0.270	0.675

6.6. Analysis Setup and Approach: To investigate the sensitivity of the two main performance criteria: speed and noise margin, the following tests were performed:

6.6.1. NF and PF were scaled above and below their nominal values, while holding all other transistors with constant nominal gate width. This test examines the sensitivity of the threshold voltages to scaling the denominators of the W_{N1}/W_{NF} and W_{P1}/W_{PF} ratios. This also measures the impact of the added gate capacitance on the output, with no added benefit of current drive.

6.6.2. N1 and P1 were also scaled above and below their nominal values, while holding all other transistors with constant gate width to examine the sensitivity of the thresholds to scaling the numerator of the W_{N1}/W_{NF} and W_{P1}/W_{PF} ratios. This test provides a comparison that might suggest which transistor in the ratio provides the most efficient hysteresis, while maintaining speed.

6.6.3. N1, NF, P1, and PF were scaled together, while maintaining constant W_{N1}/W_{NF} and W_{P1}/W_{PF} ratios and constant N2 and P2 sizes to examine the effect on speed versus scaling N2 and P2 alone to provide current drive.

6.7. Sensitivity Analysis

6.7.1. NF and PF – Scaling NF and PF provides approximately the same change in hysteresis as scaling N1 and P1 in the opposite direction, as equations 6.5 and 6.6 suggest. This relationship is shown in Figure 6.2, where curves for the threshold values versus transistor width have approximately equal, but opposite slopes for scaling the feedback transistors versus the input transistors around the nominal values. However, the data in Tables 6.2 and 6.3 shows that increasing the size of the feedback transistors, rather than reducing the size of N1 and P1,

adds hysteresis for a smaller increase in propagation delay. The added capacitance of larger NF and PF have a smaller negative effect than the reducing the current drive with smaller N1 and P1.

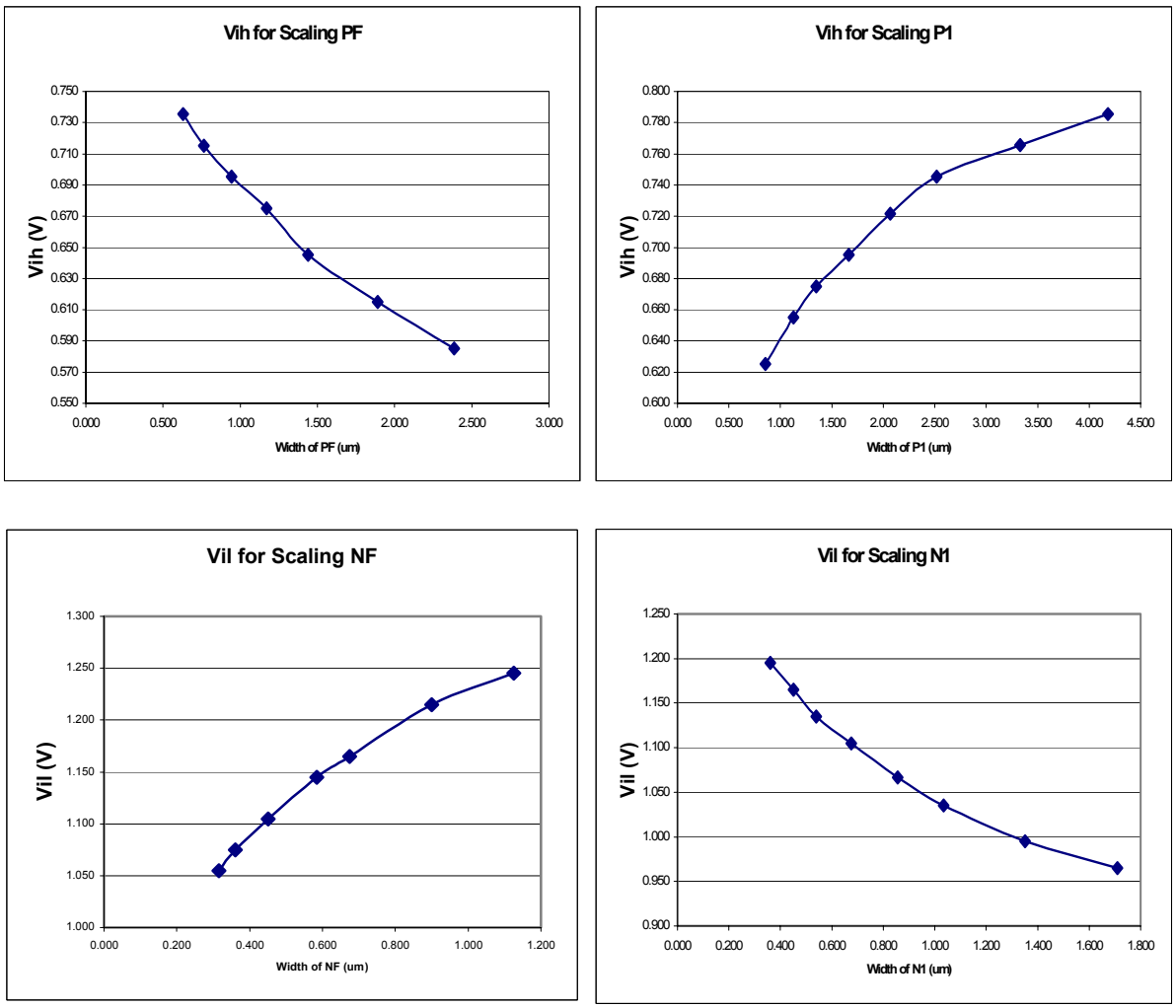


Figure 6.2 - Sensitivity of V_{IH} and V_{IL} to Scaling P1, PF, N1, and NF

6.7.2. N1 and P1 – For reducing hysteresis, increasing the size of N1 and P1 is the obvious choice since the delay decreases as well. For best results, also reducing the size of NF and PF will allow more control over V_{IH} and V_{IL} without having to oversize N1 and P1.

Table 6.2 - Sensitivity Results for Scaling NF & PF

NF and PF Scaled:	**50%**	**67%**	**80%**	**Nominal**	**125%**	**150%**	**200%**	**250%**
Gate Width NF (um)	N/A	0.315	0.360	0.450	0.585	0.675	0.900	1.125
Gate Width PF (um)	N/A	0.630	0.765	0.945	1.170	1.440	1.890	2.385
Delay (High-Low)	N/A	1.264E-10	1.295E-10	1.350E-10	1.428E-10	1.487E-10	1.610E-10	1.722E-10
Improv. from Nom.	N/A	8.600E-12	5.500E-12		-7.800E-12	-1.370E-11	-2.600E-11	-3.720E-11
% Change	N/A	6.370%	4.074%		-5.778%	-10.148%	-19.259%	-27.556%
Delay (Low-High)	N/A	1.542E-10	1.584E-10	1.644E-10	1.721E-10	1.802E-10	1.948E-10	2.104E-10
Improv. from Nom.	N/A	1.020E-11	6.000E-12		-7.700E-12	-1.580E-11	-3.040E-11	-4.600E-11
% Change	N/A	6.204%	3.650%		-4.684%	-9.611%	-18.491%	-27.981%
Avg. Prop. Delay	N/A	1.403E-10	1.440E-10	1.497E-10	1.575E-10	1.645E-10	1.779E-10	1.913E-10
Improv. from Nom.	N/A	9.400E-12	5.750E-12		-7.750E-12	-1.475E-11	-2.820E-11	-4.160E-11
% Change	N/A	6.279%	3.841%		-5.177%	-9.853%	-18.838%	-27.789%
Rise-time	N/A	2.316E-10	2.371E-10	2.449E-10	2.552E-10	2.660E-10	2.856E-10	3.067E-10
Improv. from Nom.	N/A	1.330E-11	7.800E-12		-1.030E-11	-2.110E-11	-4.070E-11	-6.180E-11
% Change	N/A	5.431%	3.185%		-4.206%	-8.616%	-16.619%	-25.235%
Fall-time	N/A	1.580E-10	1.611E-10	1.664E-10	1.738E-10	1.802E-10	1.934E-10	2.073E-10
Improv. from Nom.	N/A	8.400E-12	5.300E-12		-7.400E-12	-1.380E-11	-2.700E-11	-4.090E-11
% Change	N/A	5.048%	3.185%		-4.447%	-8.293%	-16.226%	-24.579%
Vih	N/A	0.735	0.715	0.695	0.675	0.645	0.615	0.585
Change from Nom.	N/A	0.040	0.020		-0.020	-0.050	-0.080	-0.110
% Change	N/A	5.768%	2.891%		-2.891%	-7.192%	-11.522%	-15.837%
Vil	N/A	1.055	1.075	1.105	1.145	1.165	1.215	1.245
Change from Nom.	N/A	-0.050	-0.030		0.040	0.060	0.110	0.140
% Change	N/A	-4.534%	-2.724%		3.620%	5.439%	9.956%	12.680%
Avg. Power	N/A	2.566E-04	2.652E-04	2.790E-04	2.975E-04	3.132E-04	3.437E-04	3.731E-04
Improv. from Nom.	N/A	2.240E-05	1.380E-05		-1.850E-05	-3.420E-05	-6.470E-05	-9.410E-05
% Change	N/A	8.029%	4.946%		-6.631%	-12.258%	-23.190%	-33.728%

Table 6.3 - Sensitivity Results for Scaling N1 & P1

N1 and P1 Scaled:	50%	67%	80%	Nominal	125%	150%	200%	250%
Gate Width N1 (um)	0.360	0.450	0.540	0.675	0.855	1.035	1.350	1.710
Gate Width P1 (um)	0.855	1.125	1.350	1.665	2.070	2.520	3.330	4.185
Delay (High-Low)	1.533E-10	1.487E-10	1.426E-10	1.350E-10	1.278E-10	1.229E-10	1.174E-10	1.135E-10
Improv. from Nom.	-1.830E-11	-1.370E-11	-7.600E-12		7.200E-12	1.210E-11	1.760E-11	2.150E-11
% Change	-13.556%	-10.148%	-5.630%		5.333%	8.963%	13.037%	15.926%
Delay (Low-High)	2.129E-10	1.899E-10	1.770E-10	1.644E-10	1.535E-10	1.454E-10	1.363E-10	1.308E-10
Improv. from Nom.	-4.850E-11	-2.550E-11	-1.260E-11		1.090E-11	1.900E-11	2.810E-11	3.360E-11
% Change	-29.501%	-15.511%	-7.664%		6.630%	11.557%	17.092%	20.438%
Avg. Prop. Delay	1.831E-10	1.693E-10	1.598E-10	1.497E-10	1.407E-10	1.342E-10	1.269E-10	1.222E-10
Improv. from Nom.	-3.340E-11	-1.960E-11	-1.010E-11		9.050E-12	1.555E-11	2.285E-11	2.755E-11
% Change	-22.311%	-13.093%	-6.747%		6.045%	10.387%	15.264%	18.403%
Rise-time	3.124E-10	2.801E-10	2.623E-10	2.449E-10	2.300E-10	2.187E-10	2.059E-10	1.978E-10
Improv. from Nom.	-6.750E-11	-3.520E-11	-1.740E-11		1.490E-11	2.620E-11	3.900E-11	4.710E-11
% Change	-27.562%	-14.373%	-7.105%		6.084%	10.698%	15.925%	19.232%
Fall-time	1.958E-10	1.842E-10	1.756E-10	1.664E-10	1.582E-10	1.527E-10	1.465E-10	1.422E-10
Improv. from Nom.	-2.940E-11	-1.780E-11	-9.200E-12		8.200E-12	1.370E-11	1.990E-11	2.420E-11
% Change	-17.668%	-10.697%	-5.529%		4.928%	8.233%	11.959%	14.543%
Vih	0.625	0.655	0.675	0.695	0.722	0.745	0.765	0.785
Improv. from Nom.	-0.070	-0.040	-0.020		0.026	0.050	0.070	0.090
% Change	-10.083%	-5.768%	-2.891%		3.783%	7.192%	10.098%	12.975%
Vil	1.195	1.165	1.135	1.105	1.067	1.035	0.995	0.965
Improv. from Nom.	0.090	0.060	0.030		-0.038	-0.070	-0.110	-0.140
% Change	8.155%	5.430%	2.724%		-3.475%	-6.335%	-9.965%	-12.689%
Avg. Power	2.456E-04	2.589E-04	2.680E-04	2.790E-04	2.911E-04	3.031E-04	3.222E-04	3.414E-04
Improv. from Nom.	3.340E-05	2.010E-05	1.100E-05		-1.210E-05	-2.410E-05	-4.320E-05	-6.240E-05
% Change	11.971%	7.204%	3.943%		-4.337%	-8.638%	-15.484%	-22.366%

6.7.3. N1, NF, P1, and PF – Scaling all four transistors together to reduce propagation delay, while maintaining a constant hysteresis voltage, is only slightly effective. Table 6.4 shows only a small percentage change in the delay for large changes in gate widths. The diminishing returns beyond ~150% is illustrated in the lower left plot of Figure 6.3. The increase in area and capacitance clearly negates any benefit from reduced delay.

Table 6.4 - Sensitivity Results for Scaling NF, PF, N1, & P1

NF, PF, N1, and P1 Scaled:	50%	67%	80%	Nominal	125%	150%	200%	250%
Gate Width NF (um)	N/A	0.315	0.360	0.450	0.585	0.675	0.900	1.125
Gate Width PF (um)	N/A	0.630	0.765	0.945	1.170	1.440	1.890	2.385
Gate Width N1 (um)	N/A	0.450	0.540	0.675	0.855	1.035	1.350	1.710
Gate Width P1 (um)	N/A	1.125	1.350	1.665	2.070	2.520	3.330	4.185
Delay (High-Low)	N/A	1.388E-10	1.365E-10	1.350E-10	1.347E-10	1.343E-10	1.368E-10	1.398E-10
Improv. from Nom.	N/A	-3.800E-12	-1.500E-12		3.000E-13	7.000E-13	-1.800E-12	-4.800E-12
% Change	N/A	-2.815%	-1.111%		0.222%	0.519%	-1.333%	-3.556%
Delay (Low-High)	N/A	1.767E-10	1.702E-10	1.644E-10	1.602E-10	1.575E-10	1.560E-10	1.569E-10
Improv. from Nom.	N/A	-1.230E-11	-5.800E-12		4.200E-12	6.900E-12	8.400E-12	7.500E-12
% Change	N/A	-7.482%	-3.528%		2.555%	4.197%	5.109%	4.562%
Avg. Prop. Delay	N/A	1.578E-10	1.534E-10	1.497E-10	1.475E-10	1.459E-10	1.464E-10	1.484E-10
Improv. from Nom.	N/A	-8.050E-12	-3.650E-12		2.250E-12	3.800E-12	3.300E-12	1.350E-12
% Change	N/A	-5.377%	-2.438%		1.503%	2.538%	2.204%	0.902%
Rise-time	N/A	2.650E-10	2.540E-10	2.449E-10	2.395E-10	2.375E-10	2.397E-10	2.466E-10
Improv. from Nom.	N/A	-2.010E-11	-9.100E-12		5.400E-12	7.400E-12	5.200E-12	-1.700E-12
% Change	N/A	-8.207%	-3.716%		2.205%	3.022%	2.123%	-0.694%
Fall-time	N/A	1.746E-10	1.699E-10	1.664E-10	1.650E-10	1.648E-10	1.687E-10	1.740E-10
Improv. from Nom.	N/A	-8.200E-12	-3.500E-12		1.400E-12	1.600E-12	-2.300E-12	-7.600E-12
% Change	N/A	-4.928%	-2.103%		0.841%	0.962%	-1.382%	-4.567%
Vih	N/A	0.695	0.695	0.695	0.695	0.695	0.695	0.695
Change from Nom.	N/A	0.000	0.000		0.000	0.000	0.000	0.000
% Change	N/A	0.000%	0.000%		0.000%	0.000%	0.000%	0.000%
Vil	N/A	1.115	1.105	1.105	1.105	1.095	1.105	1.105
Change from Nom.	N/A	0.010	0.000		0.000	-0.010	0.000	0.000
% Change	N/A	0.905%	0.000%		0.000%	-0.905%	0.000%	-0.009%
Avg. Power	N/A	2.395E-04	2.552E-04	2.790E-04	3.110E-04	3.420E-04	4.024E-04	4.680E-04
Improv. from Nom.	N/A	3.950E-05	2.380E-05		-3.200E-05	-6.300E-05	-1.234E-04	-1.890E-04
% Change	N/A	14.158%	8.530%		-11.470%	-22.581%	-44.229%	-67.742%

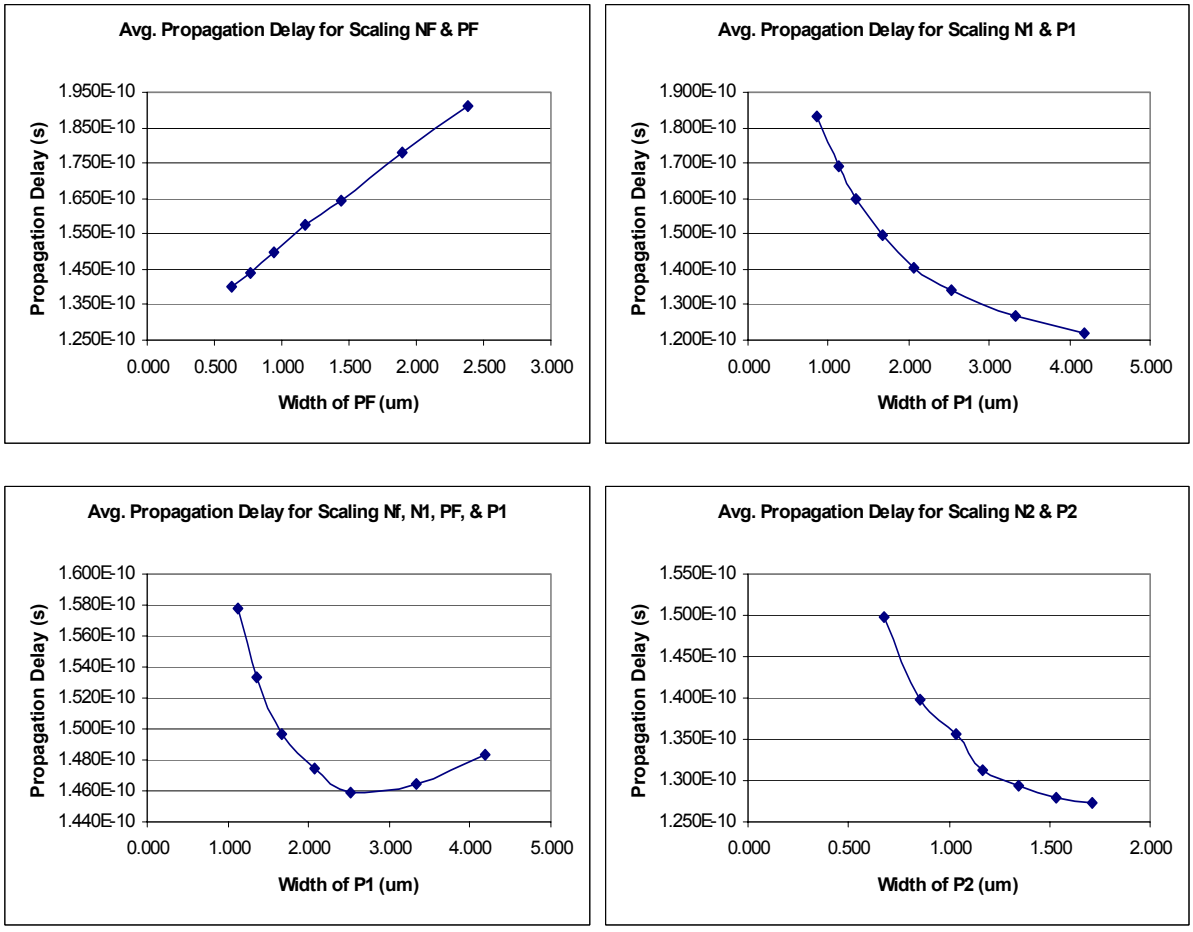


Figure 6.3 - Propagation Delay (T_P) Sensitivity Comparison

6.7.4. N2 and P2 – Increasing the size of N2 and P2 alone has a fairly significant impact on delay when scaling up from the nominal values, as shown in the steep initial slope of the lower right plot of Figure 6.3. However, the returns begin to diminish without additional scaling in N1 and P1. The most notable effect of scaling N2 and P2 is the dramatic decrease in rise and fall times as these transistors increase in gate width. Figures 6.4 and 6.5 reflect the dominance of N2 and P2 compared to scaling the other 4 transistors. Table 6.5 also shows that the considerable decrease in rise/fall times is accompanied by only a very small increase in power.

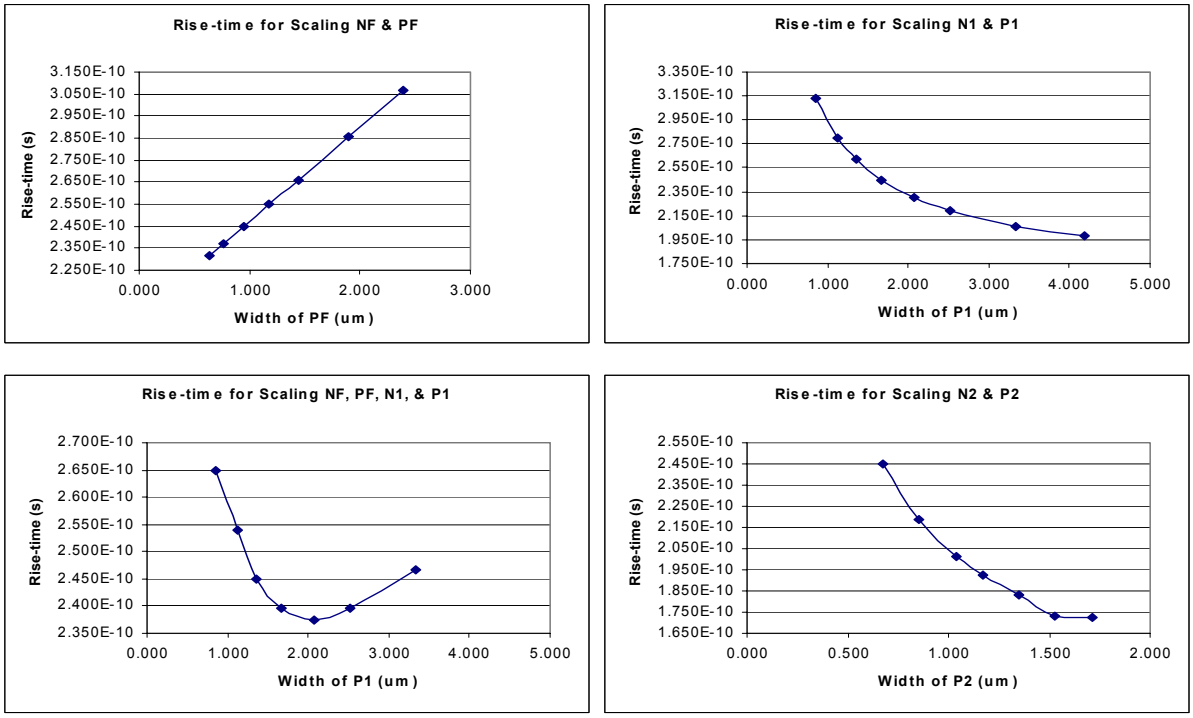


Figure 6.4 - Rise Time Sensitivity Comparison

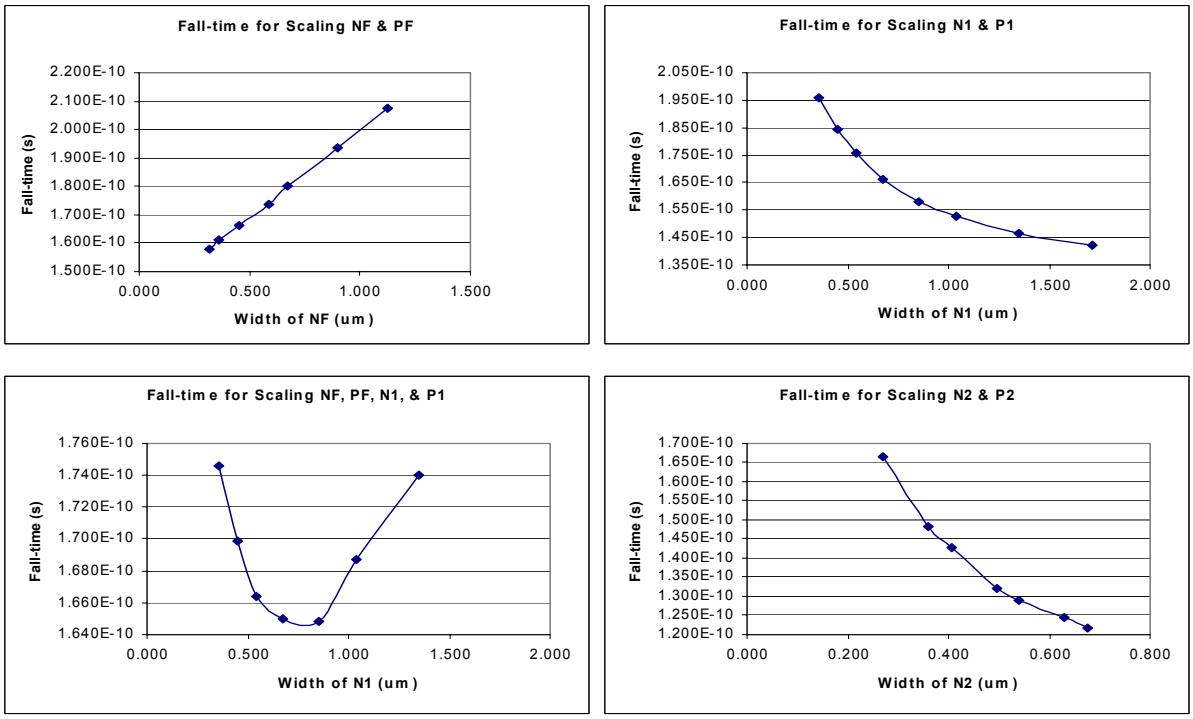


Figure 6.5 - Fall Time Sensitivity Comparison

Table 6.5 - Sensitivity Results for Scaling N2 & P2

N2 and P2 Scaled:	Nominal	125%	150%	175%	200%	225%	250%
Gate Width N2 (um)	0.270	0.360	0.405	0.495	0.540	0.630	0.675
Gate Width P2 (um)	0.675	0.855	1.035	1.170	1.350	1.530	1.710
Delay (High-Low)	1.350E-10	1.262E-10	1.256E-10	1.197E-10	1.202E-10	1.208E-10	1.191E-10
Improv. from Nom.		8.800E-12	9.400E-12	1.530E-11	1.480E-11	1.420E-11	1.590E-11
% Change		6.519%	6.963%	11.333%	10.963%	10.519%	11.778%
Delay (Low-High)	1.644E-10	1.535E-10	1.455E-10	1.427E-10	1.387E-10	1.350E-10	1.353E-10
Improv. from Nom.		1.090E-11	1.890E-11	2.170E-11	2.570E-11	2.940E-11	2.910E-11
% Change		6.630%	11.496%	13.200%	15.633%	17.883%	17.701%
Avg. Prop. Delay	1.497E-10	1.399E-10	1.356E-10	1.312E-10	1.295E-10	1.279E-10	1.272E-10
Improv. from Nom.		9.850E-12	1.415E-11	1.850E-11	2.025E-11	2.180E-11	2.250E-11
% Change		6.580%	9.452%	12.358%	13.527%	14.562%	15.030%
Rise-time	2.449E-10	2.189E-10	2.012E-10	1.927E-10	1.834E-10	1.733E-10	1.727E-10
Improv. from Nom.		2.600E-11	4.370E-11	5.220E-11	6.150E-11	7.160E-11	7.220E-11
% Change		10.617%	17.844%	21.315%	25.112%	29.236%	29.481%
Fall-time	1.664E-10	1.483E-10	1.426E-10	1.319E-10	1.290E-10	1.245E-10	1.218E-10
Improv. from Nom.		1.810E-11	2.380E-11	3.450E-11	-3.740E-11	4.190E-11	4.460E-11
% Change		10.877%	14.303%	20.733%	-22.476%	25.180%	26.803%
Vih	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Change from Nom.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
% Change	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Vil	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Change from Nom.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
% Change	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Avg. Power	2.790E-04	2.794E-04	2.831E-04	2.855E-04	2.907E-04	3.009E-04	3.025E-04
Improv. from Nom.		-4.000E-07	-4.100E-06	-6.500E-06	-1.170E-05	-2.190E-05	-2.350E-05
% Change		-0.143%	-1.470%	-2.330%	-4.194%	-7.849%	-8.423%

6.8. Key Considerations

6.8.1. Increasing the speed inevitably increases power consumption. However, significant increases in speed can be achieved with a minimal increase power by carefully choosing which transistors will give the most “bang for your buck.” Figure 6.6 shows that, although every scaling scenario results in higher average power for increased transistor width, scaling N2 and P2 results in the least power increase. Keeping in mind that Figure 6.3 shows significant speed

increases for scaling N2 and P2 up from nominal, these transistors appear to be an attractive pressure point for achieving lower delay and faster transitions for minimal cost in power.

6.8.2. Depending on the nature of the signal and the logic following the Schmidt trigger, either the rising or falling edges of the output might be more significant. For these cases, the sensitivities of high-to-low and low-to-high transitions are shown separately in Figures 6.7 and 6.8 respectively.

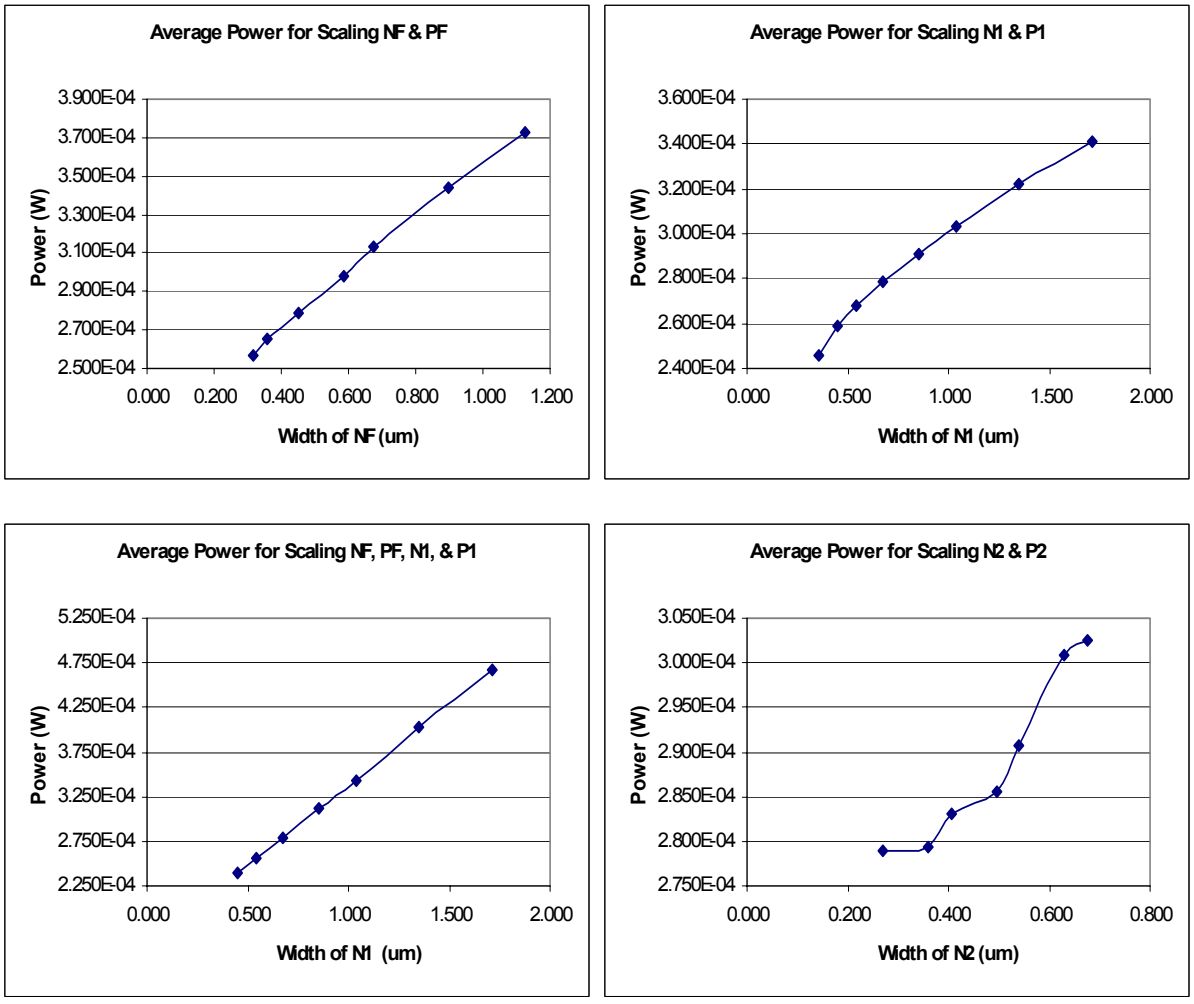


Figure 6.6 - Average Power Comparison

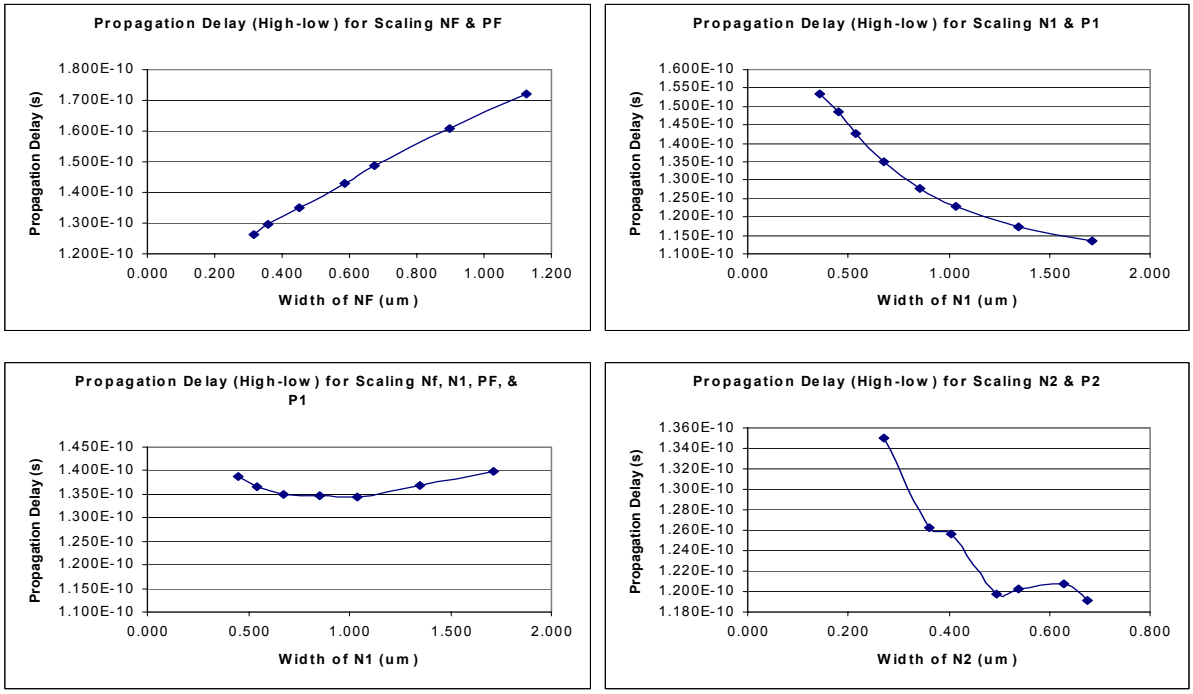


Figure 6.7 - Propagation Delay High-to-Low (T_{PHL}) Sensitivity Comparison

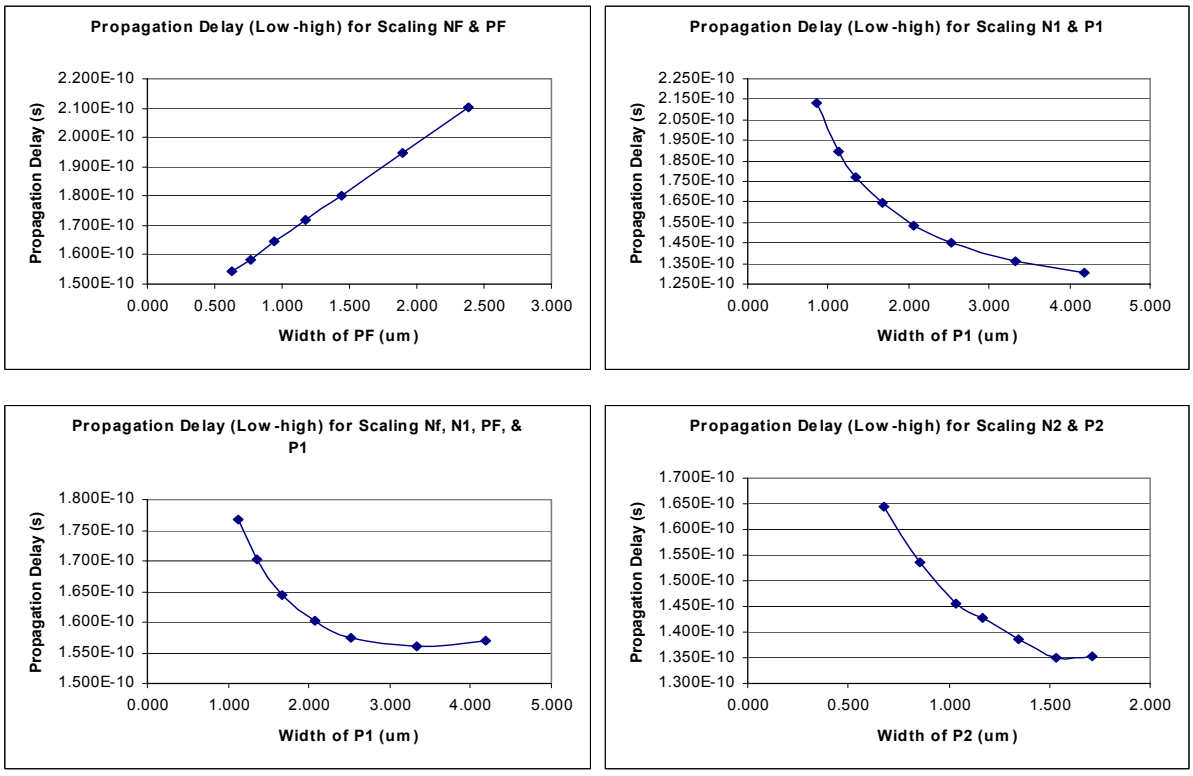


Figure 6.8 - Propagation Delay Low-to-High (T_{PLH}) Sensitivity Comparison

7. Dual-rail Domino Logic Design Example

7.1. Description: The circuit featured in this example implements a 2-input AND gate in dual-rail domino logic style. This logic style is referred to as “dual-rail” because its inputs and output are passed differentially, meaning the desired function and its complement are both computed. This topology is still classified as domino logic because all signals are monotonically rising when buffers are used on the outputs. The dual nature of the outputs allows any given function to be implemented, regardless of whether the function itself is monotonic. As shown in Figure 7.1, the logic circuitry is located between the pull-up devices and a clocked NMOS pull-down device, Me. M0-M3 implement the AND function, as well as NAND. Mp1 and Mp2 pre-charge the intermediate output nodes, ‘c’ and ‘cbar’ during the low phase of the clock. Mf1 and Mf2 provide positive feedback during transition and allow static operation by “half-latching” the outputs after transition. The buffer inverters convert the results to rising outputs, increase noise resistance, and increase the allowable fan-out. The flexibility and speed of dual-rail domino logic makes it an attractive solution for arithmetic applications. [1]

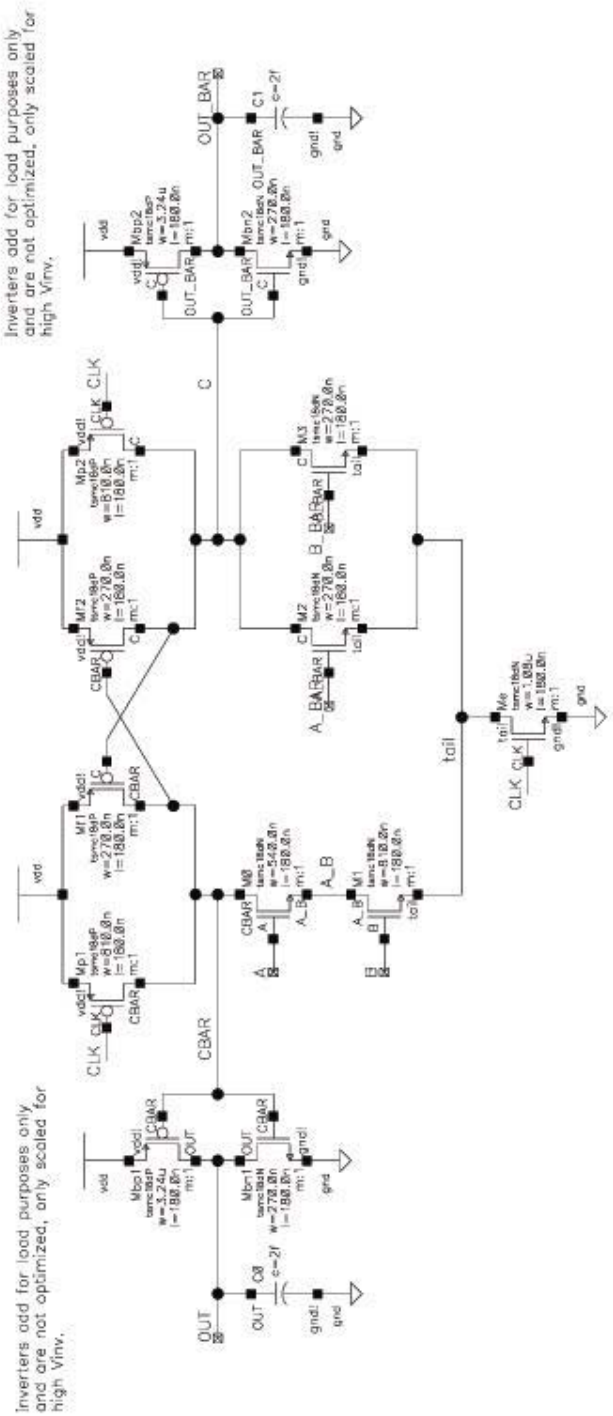


Figure 7.1 - Dual-rail Domino AND Gate Schematic w/ Nominal Sizes

7.2. Specific Application: For this example, the simple 2-input AND logic block is part of the arithmetic unit of a modern microprocessor. Only the logic circuit itself is examined in the analysis, leaving the optimization of the buffer inverters for future analysis. Therefore, the inverters included in the analysis serve as practical loads for the dual-rail AND gate. The minimum operating specifications for the logic circuit are to be as follows:

- Clock frequency of at least 2GHz. (Assume clock rise time of ~50ps)
- Full-swing (0-1.8V) differential digital input signals (A, A_BAR, B, and B_BAR required)
- Buffer inverters scaled for W_p/W_n ratio[1] equal to 12/1 provide capacitive load.
- Buffer inverters arbitrarily loaded with 2fF.
- Average Power Dissipation < 200uW
- Maximum Propagation Delay[2], T_{pHL} = 75ps

7.3. Advantages of Topology

- Any arbitrary logic function can be implemented (versus limited function domino logic).
- Very high speeds can be achieved since only a rising edge delay exists.
- The function result and its complement are automatically available, without the need for additional inversion. This removes latency between the outputs by removing an additional gate delay.

[1] The W_p/W_n ratio is chosen to provide an arbitrarily high inverter trip voltage to increase the evaluation speed.
[2] T_{PHL} = Worst case propagation delay for a high-to-low transition on the output nodes 'c' or 'cbar'. The worst case delay will be input dependent. T_{PLH} not considered during evaluation since outputs are pre-charged.

- Crowbar currents are reduced (versus DCVSL[3])
- The pull-up devices are not precisely ratioed, reducing design time (versus DCVSL). [6]

7.4. Disadvantages of Topology

- Dual-rail requires up to twice as many devices than single-ended domino logic.
- Power dissipation is increased since a transition occurs every clock cycle, regardless of the inputs.
- The timing of the clock with respect to the inputs must be carefully designed for proper operation.
- Clock noise greater than the NMOS and/or PMOS thresholds can cause the pull-up and pull-down to be on at the same time, resulting in decreased noise immunity.
- When buffer inverters are added, the large W_p/W_n ratio reduces noise margin high. [6], [7]

7.5. Design Approach

7.5.1. Since the primary goal for the design is speed for minimal power consumption, the evaluation path is the most critical set of transistors to be scaled. The evaluation path will consist of one of the three NMOS pull-down paths and the PMOS pull-up in the buffer inverter. Since the buffer is not optimized in this analysis, we will focus on the three NMOS paths and measure the output at 'c' and 'cbar'. Since the output is dual, the critical path during evaluation will either consist of M0-M1-Me, M2-Me, or M3-Me, depending on

[3] DCVSL = Differential Cascode Voltage Switch Logic

the inputs. The M0-M1-Me chain can be scaled to 3x minimum size to account for the increased path resistance for a stack of 3 and adjust the delay so that it is comparable to that of a minimum sized reference inverter. However, the delay can be optimized by progressively sizing the stack to reduce the capacitance seen closer to the load, while maintaining the same total series resistance [7]. Shifting the capacitance closer to ground by putting the largest transistors at the bottom of the stack reduces the Elmore delay. The nominal values chosen to optimize Elmore delay are listed in Table 7.1. Once the size of Me is determined to optimize the M0-M1-Me path, the sizing of M2 and M3 is relatively simple. Since Me is larger than necessary to reduce the M2-Me path resistance and M2 and M3 are in parallel, M2 and M3 can initially remain minimum sized. Additional scaling might be necessary to balance delay, depending on the input pattern.

Table 7.1 - Nominal Gate Widths for Dual-rail Domino AND Gate

Ref. Designator:	**MP1 & MP2**	**MF1 & MF2**	**M0**	**M1**	**M2**	**M3**	**Me**
Nominal Sizes (um)	0.810	0.270	0.540	0.810	0.270	0.270	1.620

7.5.2. Mf1 and Mf2 may initially remain minimum sized to reduce capacitance on the output, increase speed, and lower power. However, these keeper transistors might need to be enlarged (strengthened) to fight off incident switching due to glitchy inputs and increase noise margin.

7.5.3. Mp1 and Mp2 must be sized large enough to charge the output nodes high in a half clock-cycle. Therefore to handle the loading of the pull-down, keepers, and buffer inverters, Mp1 and Mp2 should be at least 3x minimum sized.

7.6. Analysis Setup and Approach[4]: To investigate the sensitivity of the propagation delay and pre-charge rise-time the following tests were performed:

7.6.1. Mp1 and Mp2 were scaled above and below their nominal values to analyze how fast the output can be pre-charged without greatly increasing evaluation propagation delay (T_{PHL}).

7.6.2. Mf1 and Mf2 were scaled up from their nominal values to analyze their effect on propagation delay for cases where the keepers must be scaled to increase noise margin.

7.6.3. M0-M3 and Me were simultaneously scaled up from their nominal values, while keeping the pull-up transistors with nominal gate width. This test is meant to focus on how efficiently the propagation delay decreases for the resulting power increase.

7.6.4. The sizes of the transistors in the progressively sized pull-down path were scaled relative to each other to examine alternative sizing arrangements that yield similar series resistance, but different distributed capacitance. In other words, the distribution of gate width among the stack was systematically adjusted to move parasitic capacitance closer and farther away from the output node.

7.6.5. Me was also individually scaled down, while keeping all other transistors with nominal gate width to examine any savings in power and/or area versus the increase in propagation delay.

[4] Though T_{PHL} was examined and recorded for all input scenarios, AB=01 yielded the worst case propagation delay and was therefore the only input case plotted for comparisons. Note that the % change in T_{PHL} for all of the input combinations is approximately equal, though the absolute values differ.

7.7. Sensitivity Analysis

7.7.1. Mp1 & Mp2 – As shown in Figures 7.2 and 7.3, increasing the sizes of Mp1 and Mp2 had the most dramatic effect on the pre-charge rise-times. The rise-time was reduced by up to nearly 60% for a 2.5 times increase in gate width. The propagation delay from the falling edge of the clock to the rising edge of the output (T_{pLH}) was also reduced by over 40% for equivalent scaling. The reduction in pre-charge delay is accompanied by an increase in the propagation delay of the falling edge during evaluation. However, notice in Table 7.2 that the decrease in delay for the rising edge on node[5] 'c' is almost twice the increase for the falling edge for the same scaling. Therefore, higher clock rates can be achieved by increasing the pre-charge strength, assuming the added delay during evaluation is acceptable. If the pre-charge is not strengthened, lower propagation delay for the result can be achieved, but the clock must still be slow enough to give the output time to pre-charge. Also note in Table 7.2 that the reduction in pre-charge delay will increase the power dissipation.

7.7.2. Mf1 & Mf1 – Increasing the size of the keeper transistors appears to produce negative effects on all performance parameters. Though Table 7.3 shows up to 10% deterioration in performance for most scaling, these tests do not reflect the increase in noise immunity that will result from strengthening Mf1 and Mf2. Noise immunity was not analyzed in this example, but is important and likely worth a small reduction in performance, depending on the application.

[5] This comparison is done for worst-case conditions. The change in rise-time on node 'c' is more dramatic than that of node 'cbar' because there is more parasitic capacitance loading node 'c' due to the parallel combination of M2 and M3.

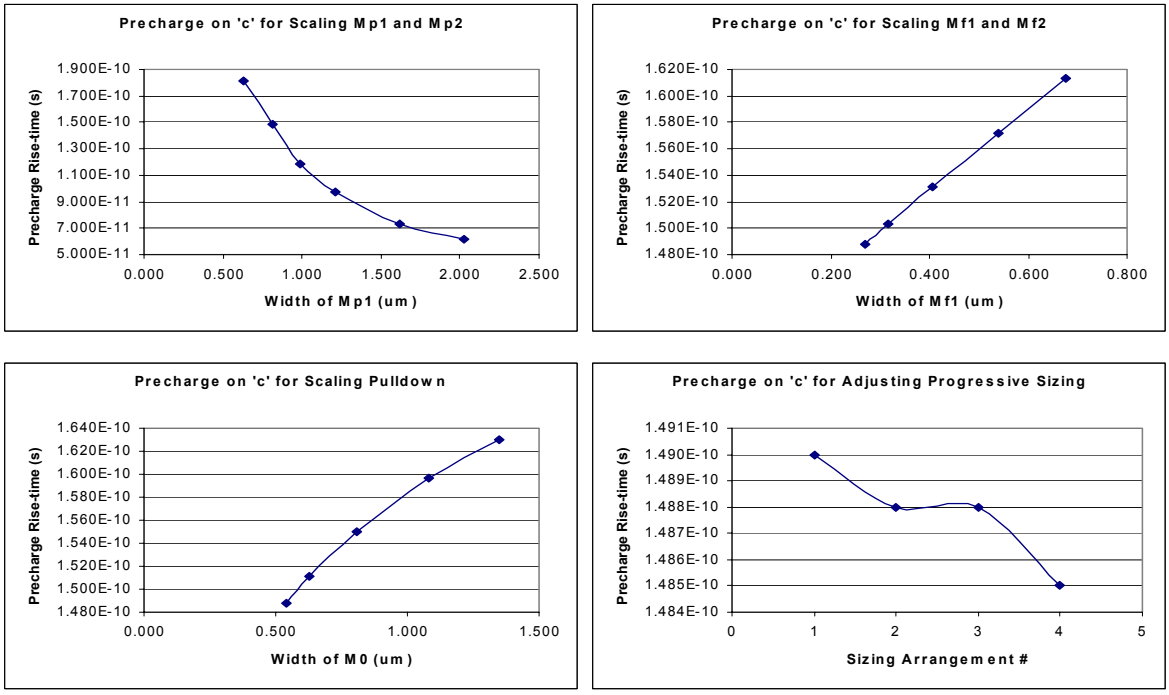


Figure 7.2 - Pre-charge Rise-time on Node ‘c’ Sensitivity Comparison

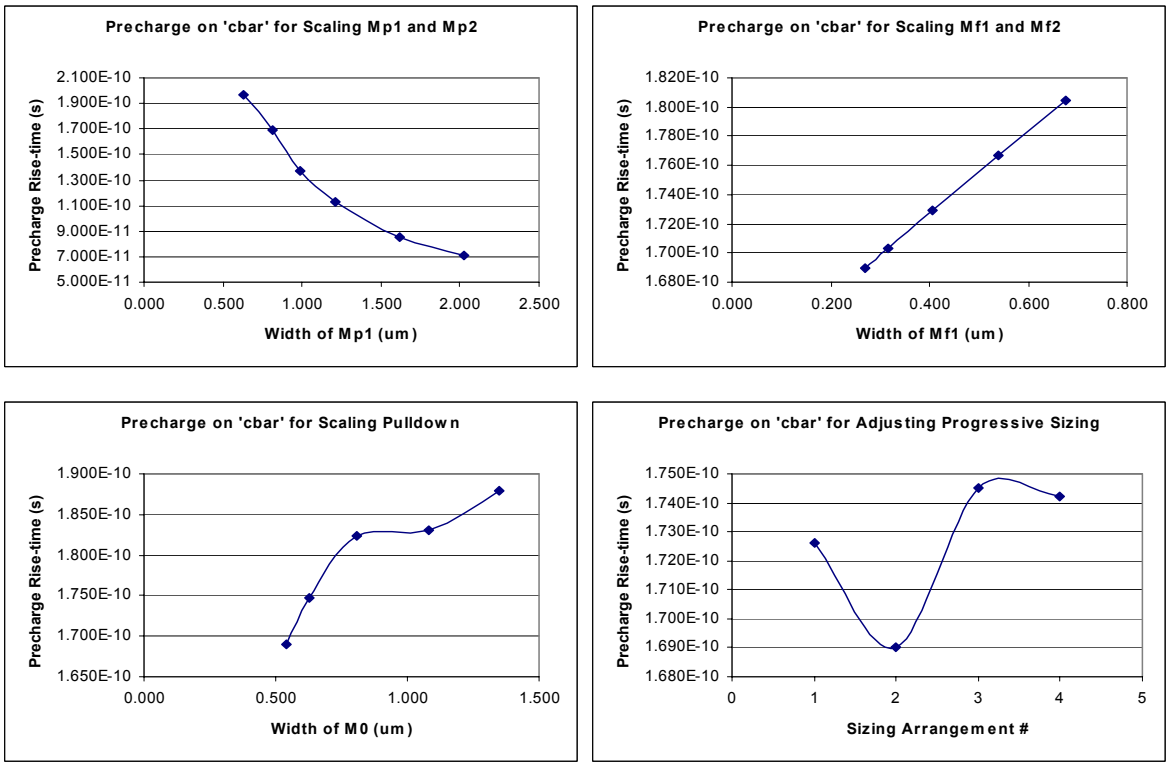


Figure 7.3 - Pre-charge Rise-time on Node ‘cbar’ Sensitivity Comparison

Table 7.2 - Sensitivity Results for Scaling Mp1 & Mp2

MP1 & MP2 Scaled:	80%	Nominal	120%	150%	200%	250%
Gate Width (um)	0.630	0.810	0.990	1.215	1.620	2.025
TpHL(AB=11, cbar)	6.248E-11	6.487E-11	6.711E-11	6.970E-11	7.417E-11	7.840E-11
Improvement from Nom.	2.390E-12		-2.240E-12	-4.830E-12	-9.300E-12	-1.353E-11
% Change	3.684%		-3.453%	-7.446%	-14.336%	-20.857%
TpHL(AB=10, c)	6.598E-11	6.895E-11	7.160E-11	7.469E-11	7.995E-11	8.500E-11
Improvement from Nom.	2.970E-12		-2.650E-12	-5.740E-12	-1.100E-11	-1.605E-11
% Change	4.307%		-3.843%	-8.325%	-15.954%	-23.278%
TpHL(AB=01, c)	6.371E-11	6.994E-11	7.371E-11	7.729E-11	8.280E-11	8.790E-11
Improvement from Nom.	6.230E-12		-3.770E-12	-7.350E-12	-1.286E-11	-1.796E-11
% Change	8.908%		-5.390%	-10.509%	-18.387%	-25.679%
TpHL(AB=00, c)	3.700E-11	4.245E-11	4.521E-11	4.747E-11	5.053E-11	5.339E-11
Improvement from Nom.	5.450E-12		-2.760E-12	-5.020E-12	-8.080E-12	-1.094E-11
% Change	12.839%		-6.502%	-11.826%	-19.034%	-25.771%
Prchg. Risetime (AB=00, c)	1.810E-10	1.488E-10	1.189E-10	9.700E-11	7.332E-11	6.122E-11
Improvement from Nom.	-3.220E-11		2.990E-11	5.180E-11	7.548E-11	8.758E-11
% Change	-21.640%		20.094%	34.812%	50.726%	58.858%
TpLH(AB=00, c)	9.538E-11	8.196E-11	7.264E-11	6.431E-11	5.442E-11	4.800E-11
Improvement from Nom.	-1.342E-11		9.320E-12	1.765E-11	2.754E-11	3.396E-11
% Change	-16.374%		11.371%	21.535%	33.602%	41.435%
Prchg. Rise. (AB=11, cbar)	1.965E-10	1.690E-10	1.375E-10	1.125E-10	8.561E-11	7.130E-11
Improvement from Nom.	-2.750E-11		3.150E-11	5.650E-11	8.339E-11	9.770E-11
% Change	-16.272%		18.639%	33.432%	49.343%	57.811%
TpLH(AB=11, cbar)	5.888E-10	5.765E-10	5.681E-10	5.604E-10	5.511E-10	5.450E-10
Improvement from Nom.	-1.230E-11		8.400E-12	1.610E-11	2.540E-11	3.150E-11
% Change	-2.134%		1.457%	2.793%	4.406%	5.464%
Avg. Power	1.787E-04	1.833E-04	1.873E-04	1.919E-04	2.000E-04	2.080E-04
Improvement from Nom.	4.600E-06		-4.000E-06	-8.600E-06	-1.670E-05	-2.470E-05
% Change	2.510%		-2.182%	-4.692%	-9.111%	-13.475%
PDP (Worst Case, AB=01)	1.138E-14	1.282E-14	1.381E-14	1.483E-14	1.656E-14	1.828E-14
Improvement from Nom.	1.435E-15		-9.859E-16	-2.012E-15	-3.740E-15	-5.463E-15
% Change	11.194%		-7.690%	-15.694%	-29.173%	-42.615%

Table 7.3 - Sensitivity Results for Scaling Mf1 & Mf2

MF1 & MF2 Scaled:	Nominal	120%	150%	200%	250%
Gate Width (um)	0.270	0.315	0.405	0.540	0.675
TpHL(AB=11, cbar)	6.487E-11	6.550E-11	6.648E-11	6.876E-11	7.065E-11
Improvement from Nom.		-6.300E-13	-1.610E-12	-3.890E-12	-5.780E-12
% Change		-0.971%	-2.482%	-5.997%	-8.910%
TpHL(AB=10, c)	6.895E-11	6.971E-11	7.128E-11	7.360E-11	7.587E-11
Improvement from Nom.		-7.600E-13	-2.330E-12	-4.650E-12	-6.920E-12
% Change		-1.102%	-3.379%	-6.744%	-10.036%
TpHL(AB=01, c)	6.994E-11	7.057E-11	7.181E-11	7.356E-11	7.518E-11
Improvement from Nom.		-6.300E-13	-1.870E-12	-3.620E-12	-5.240E-12
% Change		-0.901%	-2.674%	-5.176%	-7.492%
TpHL(AB=00, c)	4.245E-11	4.274E-11	4.329E-11	4.405E-11	4.473E-11
Improvement from Nom.		-2.900E-13	-8.400E-13	-1.600E-12	-2.280E-12
% Change		-0.683%	-1.979%	-3.769%	-5.371%
Prchg. Risetime (AB=00, c)	1.488E-10	1.503E-10	1.531E-10	1.572E-10	1.613E-10
Improvement from Nom.		-1.500E-12	-4.300E-12	-8.400E-12	-1.250E-11
% Change		-1.008%	-2.890%	-5.645%	-8.401%
TpLH(AB=00, c)	8.196E-11	8.262E-11	8.392E-11	8.586E-11	8.783E-11
Improvement from Nom.		-6.600E-13	-1.960E-12	-3.900E-12	-5.870E-12
% Change		-0.805%	-2.391%	-4.758%	-7.162%
Prchg. Rise. (AB=11, cbar)	1.690E-10	1.703E-10	1.729E-10	1.767E-10	1.805E-10
Improvement from Nom.		-1.300E-12	-3.900E-12	-7.700E-12	-1.150E-11
% Change		-0.769%	-2.308%	-4.556%	-6.805%
TpLH(AB=11, cbar)	5.765E-10	5.771E-10	5.784E-10	5.801E-10	5.819E-10
Improvement from Nom.		-6.000E-13	-1.900E-12	-3.600E-12	-5.400E-12
% Change		-0.104%	-0.330%	-0.624%	-0.937%
Avg. Power	1.833E-04	1.843E-04	1.862E-04	1.890E-04	1.919E-04
Improvement from Nom.		-1.000E-06	-2.900E-06	-5.700E-06	-8.600E-06
% Change		-0.546%	-1.582%	-3.110%	-4.692%
PDP (Worst Case, AB=01)	1.282E-14	1.301E-14	1.337E-14	1.390E-14	1.443E-14
Improvement from Nom.		-1.860E-16	-5.510E-16	-1.083E-15	-1.607E-15
% Change		-1.451%	-4.298%	-8.446%	-12.535%

7.7.3. M0-M3 & Me together – Scaling the entire pull-down path together is a very effective method of reducing the delay, as shown in Table 7.4. Observing the slope of the lower left plot of Figure 7.4, the delay continues to significantly decrease when scaling beyond 250% of the nominal gate widths. However, this is a very brute force approach that will consume a lot of area. Though the average power dissipation significantly increases with increased gate width, the power-delay product (PDP) actually continues to decrease. The improvement in PDP is illustrated in the lower left plot of Figure 7.5. The decrease in delay during evaluation comes at the cost of slower pre-charge if Mp1 and Mp2 are not scaled to compensate for the added capacitance and strength of the pull-down. Figures 7.6 and 7.7 illustrate how strengthening the pull-down increases the delay of the pre-charge.

7.7.4. Progressive Sizing Adjustments – Redistributing the parasitic capacitance in the pull-down path has a significant effect on the propagation delay. Table 7.5 shows that the effect on propagation delay due to the redistribution varies depending on the input values. However, the general trend revealed in the lower right plot of Figure 7.4 is that the delay increases as the capacitance is moved towards the output node. This trend agrees with the theory behind Elmore delay which states that the capacitance distributed farther away from ground is discharged through a higher series resistance than capacitance located closer to ground, thereby increasing the total discharge time constant. Note that the sizing arrangements in Table 7.5 change from having larger devices closer to ground in Arrangement 1 to all devices being equal in Arrangement 4 for the

Table 7.4 - Sensitivity Results for Scaling M0-M3 & Me

Pull-down Scaled Together Scaled:	Nominal	120%	150%	200%	250%
M0 Gate Width (um)	0.540	0.630	0.810	1.080	1.350
M1 Gate Width (um)	0.810	0.990	1.215	1.620	2.025
M2 & M3 Gate Width (um)	0.270	0.315	0.405	0.540	0.675
Me Gate Width (um)	1.620	1.935	2.430	3.240	4.050
TpHL(AB=11, cbar)	6.487E-11	5.894E-11	5.203E-11	4.512E-11	4.079E-11
Improvement from Nom.		5.930E-12	1.284E-11	1.975E-11	2.408E-11
% Change		9.141%	19.793%	30.446%	37.120%
TpHL(AB=10, c)	6.895E-11	6.334E-11	5.524E-11	4.709E-11	4.171E-11
Improvement from Nom.		5.610E-12	1.371E-11	2.186E-11	2.724E-11
% Change		8.136%	19.884%	31.704%	39.507%
TpHL(AB=01, c)	6.994E-11	6.430E-11	5.590E-11	4.730E-11	4.132E-11
Improvement from Nom.		5.640E-12	1.404E-11	2.264E-11	2.862E-11
% Change		8.064%	20.074%	32.371%	40.921%
TpHL(AB=00, c)	4.245E-11	3.877E-11	3.367E-11	2.791E-11	2.408E-11
Improvement from Nom.		3.680E-12	8.780E-12	1.454E-11	1.837E-11
% Change		8.669%	20.683%	34.252%	43.274%
Prchg. Risetime (AB=00, c)	1.488E-10	1.511E-10	1.550E-10	1.597E-10	1.630E-10
Improvement from Nom.		-2.300E-12	-6.200E-12	-1.090E-11	-1.420E-11
% Change		-1.546%	-4.167%	-7.325%	-9.543%
TpLH(AB=00, c)	8.196E-11	8.531E-11	9.118E-11	9.972E-11	1.076E-10
Improvement from Nom.		-3.350E-12	-9.220E-12	-1.776E-11	-2.564E-11
% Change		-4.087%	-11.249%	-21.669%	-31.284%
Prchg. Rise. (AB=11, cbar)	1.690E-10	1.748E-10	1.824E-10	1.830E-10	1.879E-10
Improvement from Nom.		-5.800E-12	-1.340E-11	-1.400E-11	-1.890E-11
% Change		-3.432%	-7.929%	-8.284%	-11.183%
TpLH(AB=11, cbar)	5.765E-10	5.791E-10	5.835E-10	5.901E-10	5.970E-10
Improvement from Nom.		-2.600E-12	-7.000E-12	-1.360E-11	-2.050E-11
% Change		-0.451%	-1.214%	-2.359%	-3.556%
Avg. Power	1.833E-04	1.903E-04	2.014E-04	2.189E-04	2.351E-04
Improvement from Nom.		-7.000E-06	-1.810E-05	-3.560E-05	-5.180E-05
% Change		-3.819%	-9.875%	-19.422%	-28.260%
PDP (Worst Case, AB=01)	1.282E-14	1.224E-14	1.126E-14	1.035E-14	9.714E-15
Improvement from Nom.		5.837E-16	1.562E-15	2.466E-15	3.106E-15
% Change		4.553%	12.182%	19.236%	24.225%

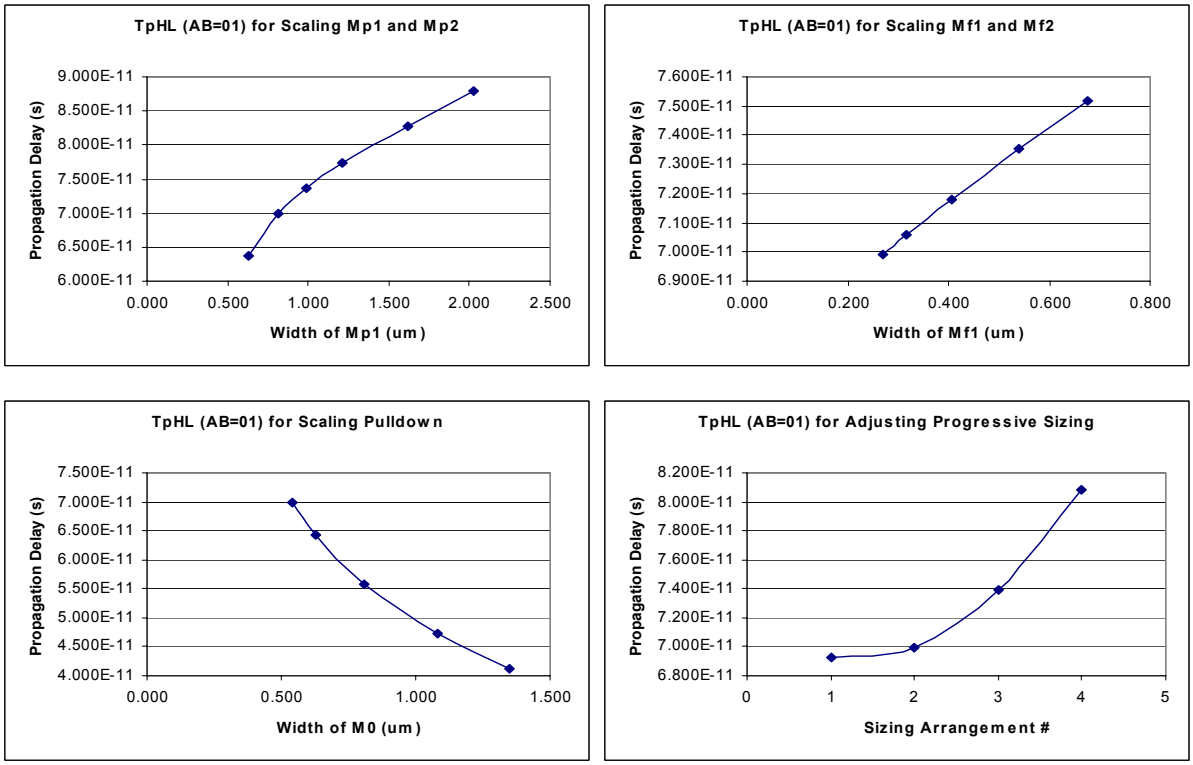


Figure 7.4 - Worst-case Propagation Delay (for inputs AB=01) Sensitivity Comparison

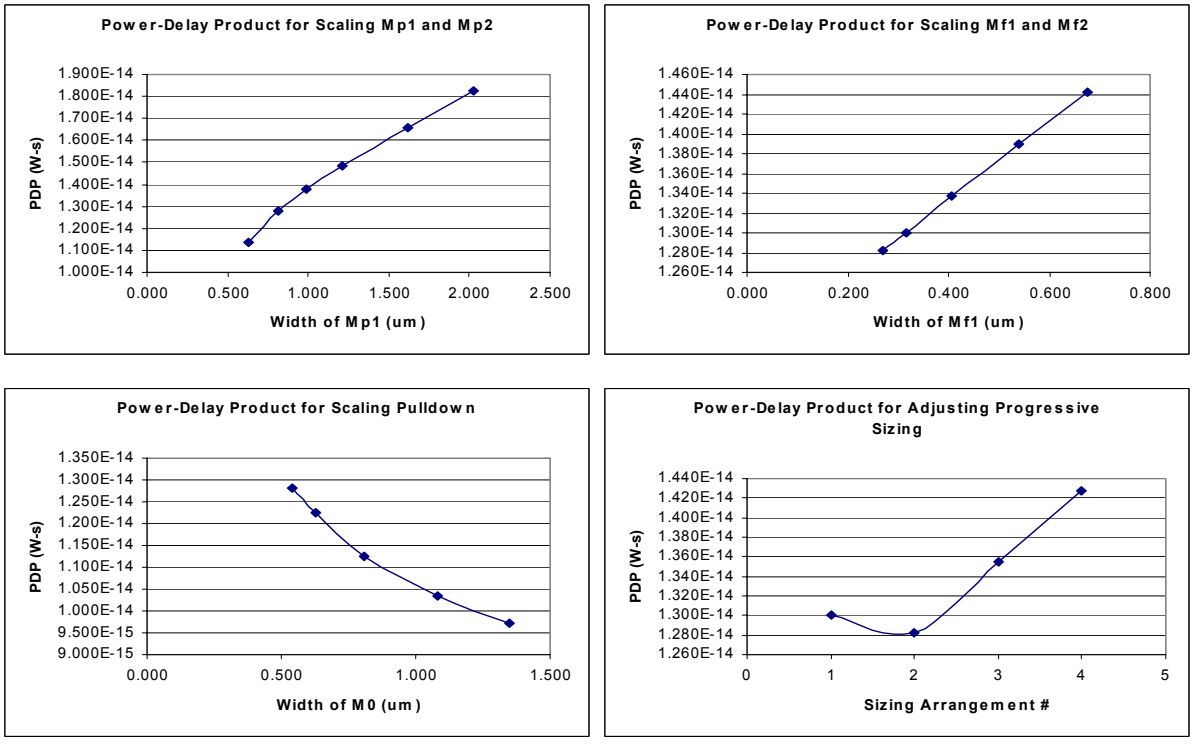


Figure 7.5 - Power-Delay Product Sensitivity Comparison

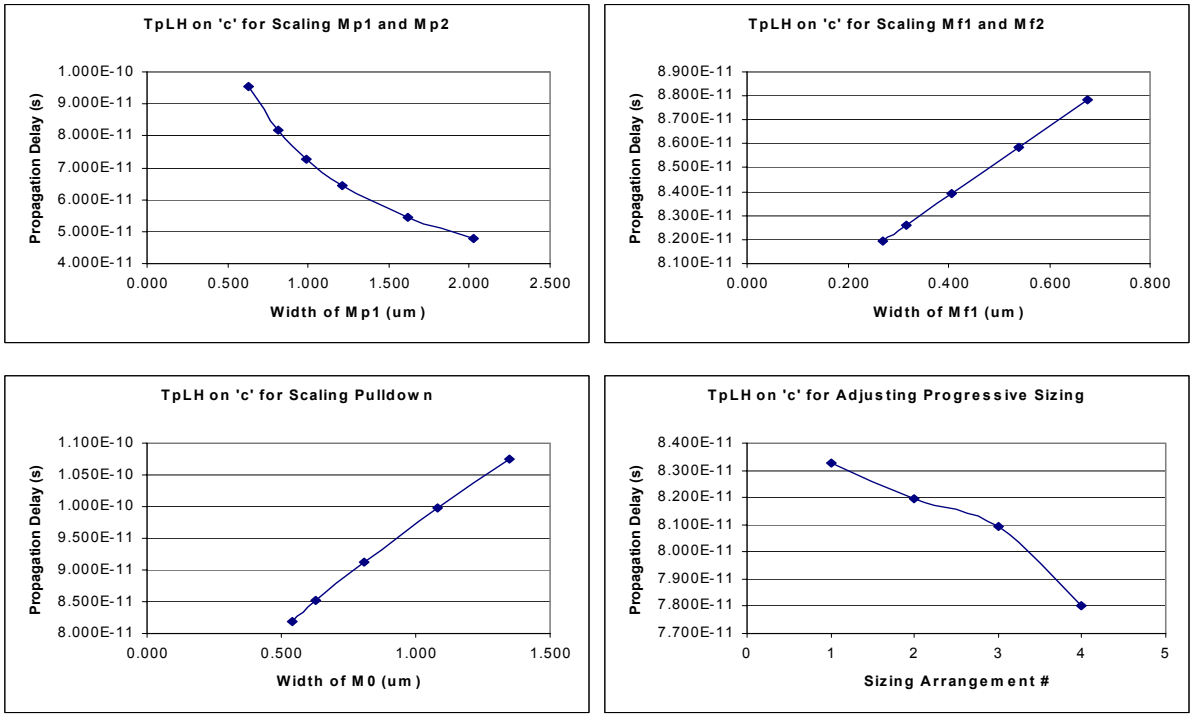


Figure 7.6 - Pre-charge Propagation Delay on Node ‘c’ Sensitivity Comparison

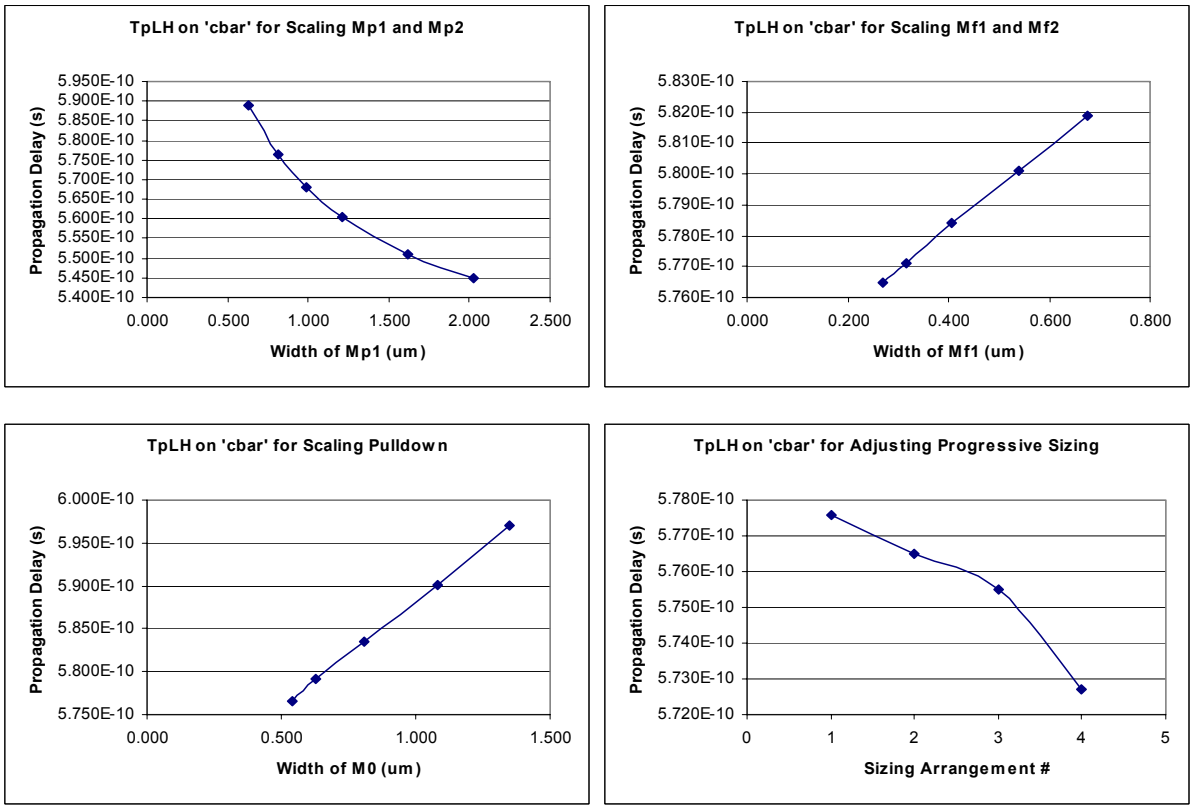


Figure 7.7 - Pre-charge Propagation Delay on Node ‘cbar’ Sensitivity Comparison

Table 7.5 - Sensitivity Results for Altering Progressive Sizing

Progressive Sizing Adjustment				
Arrangement #:	1	2	3	4
Scaled so that:	Cap. Lowered	Nominal	Cap. Raised 1	Cap. Raised 2
M0 Gate Width (um)	0.405	0.540	0.540	0.810
M1 Gate Width (um)	1.080	0.810	1.080	0.810
M2 & M3 Gate Width (um)	0.270	0.270	0.270	0.270
Me Gate Width (um)	1.890	1.620	1.350	0.810
TpHL(AB=11, cbar)	6.735E-11	6.487E-11	6.528E-11	6.940E-11
Improvement from Nom.	-2.480E-12		-4.100E-13	-4.530E-12
% Change	-3.823%		-0.632%	-6.983%
TpHL(AB=10, c)	6.776E-11	6.895E-11	7.099E-11	7.713E-11
Improvement from Nom.	1.190E-12		-2.040E-12	-8.180E-12
% Change	1.726%		-2.959%	-11.864%
TpHL(AB=01, c)	6.926E-11	6.994E-11	7.390E-11	8.086E-11
Improvement from Nom.	6.800E-13		-3.960E-12	-1.092E-11
% Change	0.972%		-5.662%	-15.613%
TpHL(AB=00, c)	4.074E-11	4.245E-11	4.433E-11	5.129E-11
Improvement from Nom.	1.710E-12		-1.880E-12	-8.840E-12
% Change	4.028%		-4.429%	-20.824%
Prchg. Risetime (AB=00, c)	1.490E-10	1.488E-10	1.488E-10	1.485E-10
Improvement from Nom.	-2.000E-13		0.000E+00	3.000E-13
% Change	-0.134%		0.000%	0.202%
TpLH(AB=00, c)	8.329E-11	8.196E-11	8.092E-11	7.801E-11
Improvement from Nom.	-1.330E-12		1.040E-12	3.950E-12
% Change	-1.623%		1.269%	4.819%
Prchg. Risetime (AB=11, cbar)	1.726E-10	1.690E-10	1.745E-10	1.742E-10
Improvement from Nom.	-3.600E-12		-5.500E-12	-5.200E-12
% Change	-2.130%		-3.254%	-3.077%
TpLH(AB=11, cbar)	5.776E-10	5.765E-10	5.755E-10	5.727E-10
Improvement from Nom.	-1.100E-12		1.000E-12	3.800E-12
% Change	-0.191%		0.173%	0.659%
Avg. Power	1.877E-04	1.833E-04	1.834E-04	1.766E-04
Improvement from Nom.	-4.400E-06		-1.000E-07	6.700E-06
% Change	-2.400%		-0.055%	3.655%
PDP (Worst Case, AB=01)	1.300E-14	1.282E-14	1.355E-14	1.428E-14
Improvement from Nom.	-1.801E-16		-7.333E-16	-1.460E-15
% Change	-1.405%		-5.720%	-11.387%

M0-M1-Me path. The paths involving M2 and M3 were not scaled to adjust the ratio of M2 (or M3) to Me because the capacitance was already minimized at the output for that logic path. However, M2 and M3 might be increased in size to increase current drive when only one input is high, regardless of the

distribution of capacitance. Note that the worst-case delay results when only one input is high and either M2 or M3 must by itself discharge the capacitance of both transistors.

7.7.5. Me – Reducing the width of Me results in a minor reduction in power dissipation; however, propagation delay increases proportionally more than the power is reduced. Table 7.6 shows that worst-case T_{pHL} increases by up to 8% for a 3% reduction in power. The important consideration is that the area required for Me was reduced by 33%. Depending on the application, this savings in area might be significant since Me is the largest transistor in the circuit.

7.8. Key Considerations

7.8.1. The size relationships among all of the transistors in the pull-down logic are very important to optimizing the delay for this circuit. The exact progressive sizing adjustments made in section 7.7.4 were done somewhat arbitrarily to illustrate the trend for delay versus the distribution of capacitance. More precise adjustments to the sizing may be necessary if this logic lie in a critical path of the system.

7.8.2. Though the buffer inverters are not closely analyzed in this example, they lie in the critical path for evaluation and should be optimized for the fan-out of this circuit. The W_p/W_n ratio should be made high to increase the switching thresholds for the inverters. A V_{inv} close to the supply voltage will speed up the evaluation when ‘c’ or ‘cbar’ begins to fall. However, the increased V_{inv} will reduce noise margin.

Table 7.6 - Sensitivity Results for Scaling Me

Me Reduction Scaled:	66.67%	83.33%	Nominal
Me Gate Width (um)	1.890	0.810	1.620
TpHL(AB=11, cbar)	7.003E-11	6.708E-11	6.487E-11
Improvement from Nom.	-5.160E-12	-2.210E-12	
% Change	-7.954%	-3.407%	
TpHL(AB=10, c)	7.323E-11	7.071E-11	6.895E-11
Improvement from Nom.	-4.280E-12	-1.760E-12	
% Change	-6.207%	-2.553%	
TpHL(AB=01, c)	7.569E-11	7.234E-11	6.994E-11
Improvement from Nom.	-5.750E-12	-2.400E-12	
% Change	-8.221%	-3.432%	
TpHL(AB=00, c)	4.713E-11	4.439E-11	4.245E-11
Improvement from Nom.	-4.680E-12	-1.940E-12	
% Change	-11.025%	-4.570%	
Prchg. Risetime (AB=00, c)	1.487E-10	1.487E-10	1.488E-10
Improvement from Nom.	1.000E-13	1.000E-13	
% Change	0.067%	0.067%	
TpLH(AB=00, c)	7.943E-11	8.072E-11	8.196E-11
Improvement from Nom.	2.530E-12	1.240E-12	
% Change	3.087%	1.513%	
Prchg. Risetime (AB=11, cbar)	1.685E-10	1.688E-10	1.690E-10
Improvement from Nom.	5.000E-13	2.000E-13	
% Change	0.296%	0.118%	
TpLH(AB=11, cbar)	5.742E-10	5.754E-10	5.765E-10
Improvement from Nom.	2.300E-12	1.100E-12	
% Change	0.399%	0.191%	
Avg. Power	1.779E-04	1.807E-04	1.833E-04
Improvement from Nom.	5.400E-06	2.600E-06	
% Change	2.946%	1.418%	
PDP (Worst Case, AB=01)	1.347E-14	1.307E-14	1.282E-14
Improvement from Nom.	-6.452E-16	-2.518E-16	
% Change	-5.033%	-1.964%	

8. Conclusions and Future Work

8.1. Insight into the Design Flow

The examples included in sections 4-7 share many characteristics of the design flow described in section 3. All of the designs begin with some sort of hand analysis, then quickly move to simulation, and finally follow through with a qualitative analysis of the results to develop an understanding of what determines the performance and quantify the dominant relationships. Though the design approaches are similar, the nature of the hand analysis varies considerably. Only two of the examples actually use equations to determine initial transistor sizes. The designer must be wary of spending *too* much time analyzing the governing equations and possibly over-designing the circuit on paper. The extra time required to calculate a design to within 5% error of the desired performance is almost completely wasted once you find that the best equations may only accurately predict the performance to within 20%. I learned this lesson the hard way near the beginning of the source-follower design. The other two examples require more of an intuitive analysis of what the circuit is supposed to do and estimate what role each component plays in accomplishing that function. This is quite often the more suitable approach in digital design, whereas the analytical approach beginning with exact equations lends itself more to analog design.

8.2. Future Work

This thesis covers only a small fraction of high-speed digital circuits that would greatly benefit from a more through treatment. I believe the sensitivity analysis, though somewhat cumbersome to complete in full, reveals valuable information about the "pressure points" of a circuit's topology that can be used to tweak an extra 5 or 10% of

performance, if not more, out of an already well-designed circuit. In addition, the analysis gives the designer invaluable experience and great insight into the behavior of circuits in general. I believe it would benefit others to pursue the same type of analysis on more circuits so that a collection of these examples can be expanded and available for use by aspiring designers. This future work would be beneficial to all who are involved.

9. References:

[1] Digital Systems Engineering, William J. Dally and John W. Poulton, Cambridge University Press, 1998

[2] Borivoje Nikolic, et al., "Improved Sense-Amplifier-Based Flip-flop: Design and Measurements," *IEEE J. Solid-State Circuits*, vol.35, pp.876-877, June 2000.

[3] P. Wijetunga and A. F. J. Levi, "3.3 GHz Sense-amplifier in 0.18um CMOS Technology," Advanced Interconnect and Network Technology, USC, IEEE pp.764-765, 2002.

[4] Design of Analog CMOS Integrated Circuits, Behzad Razavi, McGraw-Hill 2001

[5] J. G. Henney, et al.,"Enhanced Slew Rate Source Follower," *IEEE J. Solid-State Circuits*, vol.30, p.144, February 1995.

[6] Digital Integrated Circuits: A Design Perspective, Rabaey, Chandrakasan, and Nikolic, Prentice Hall, 2003.

[7] Paul D. Franzon, ECE 733 Lecture: "Putting It Together, Sample Logic Gates", NC State University, 2004.

Appendix A

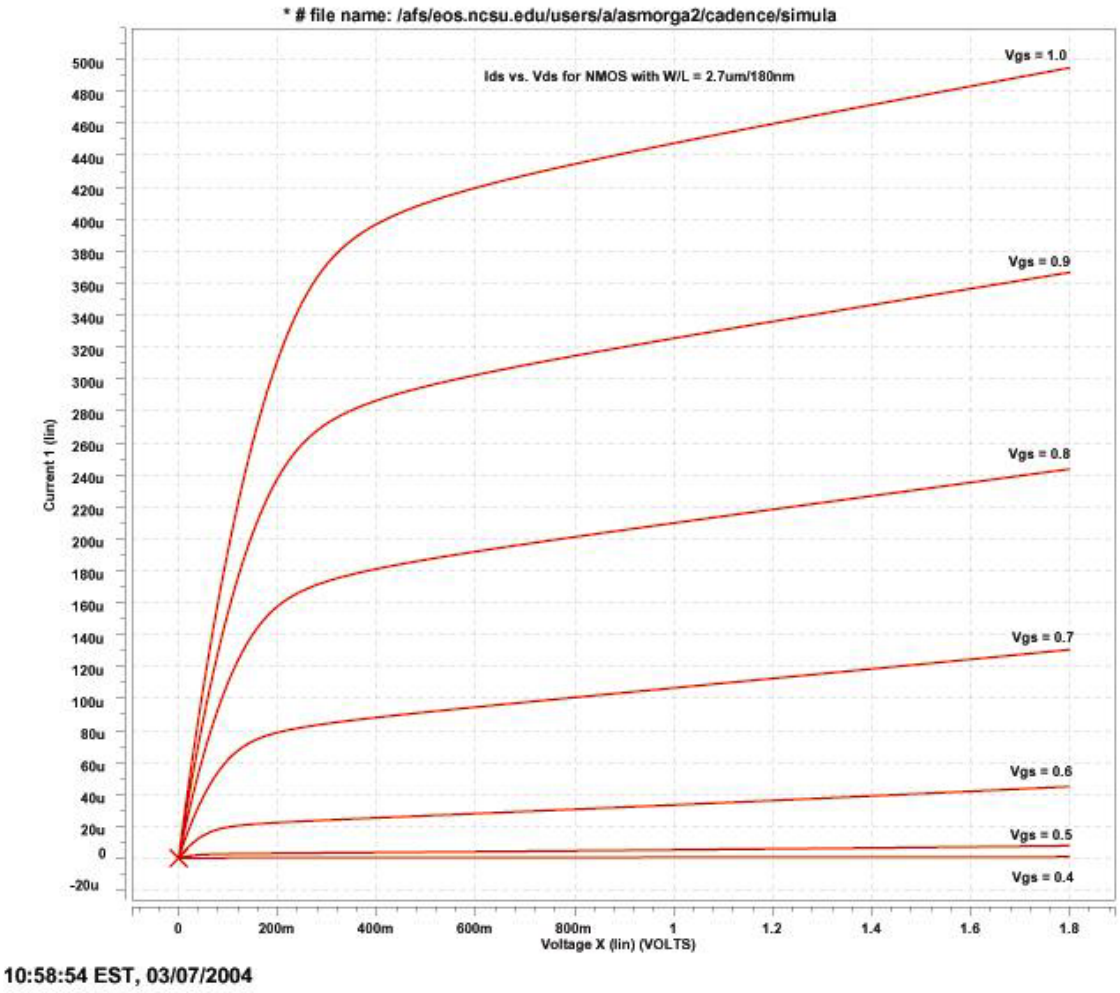


Figure A.1 - Id Curves for NMOS with W/L=2.7um/180nm in 0.18um (Deep) Process

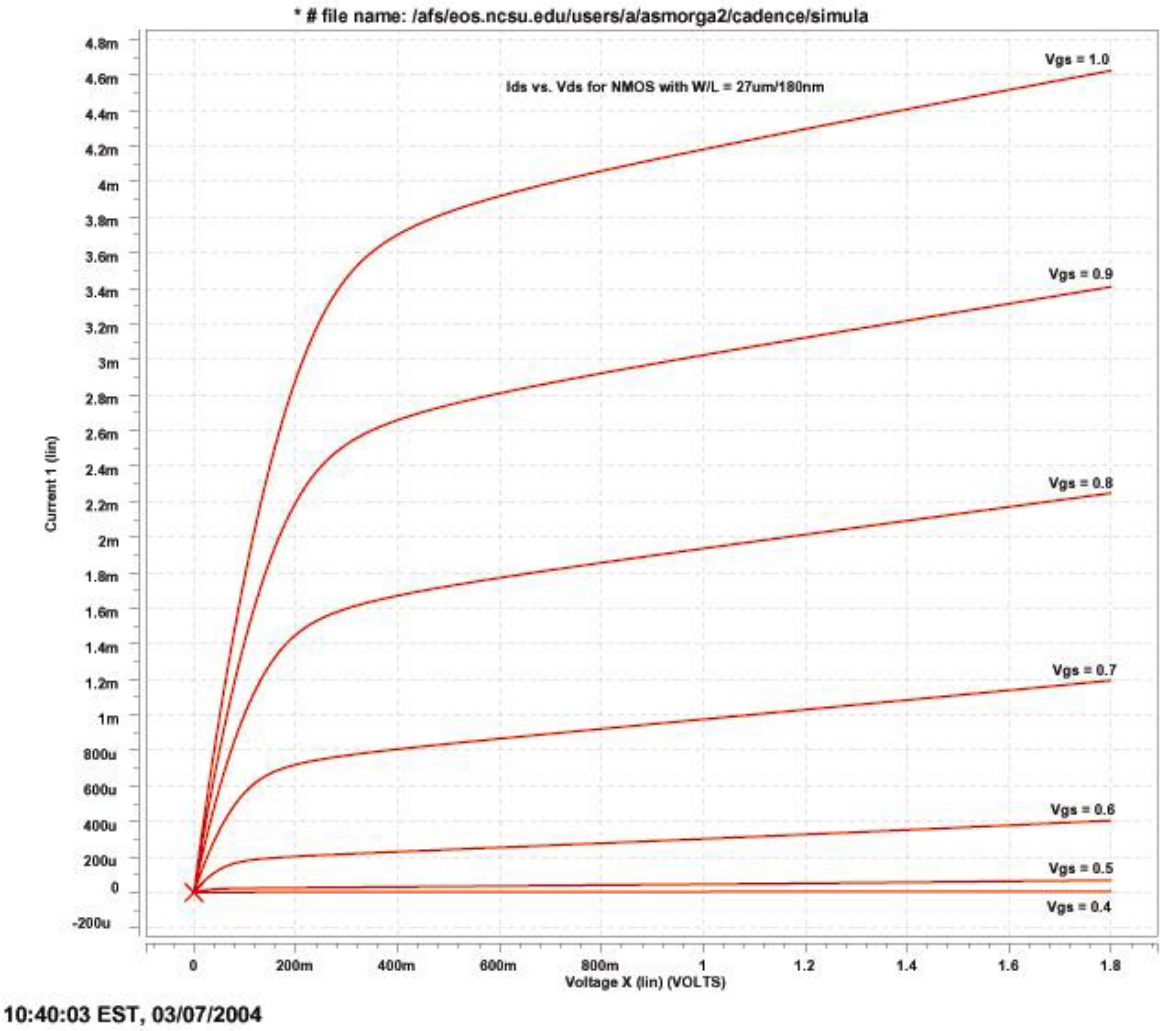


Figure A.2 - Id Curves for NMOS with W/L=27um/180nm in 0.18um (Deep) Process

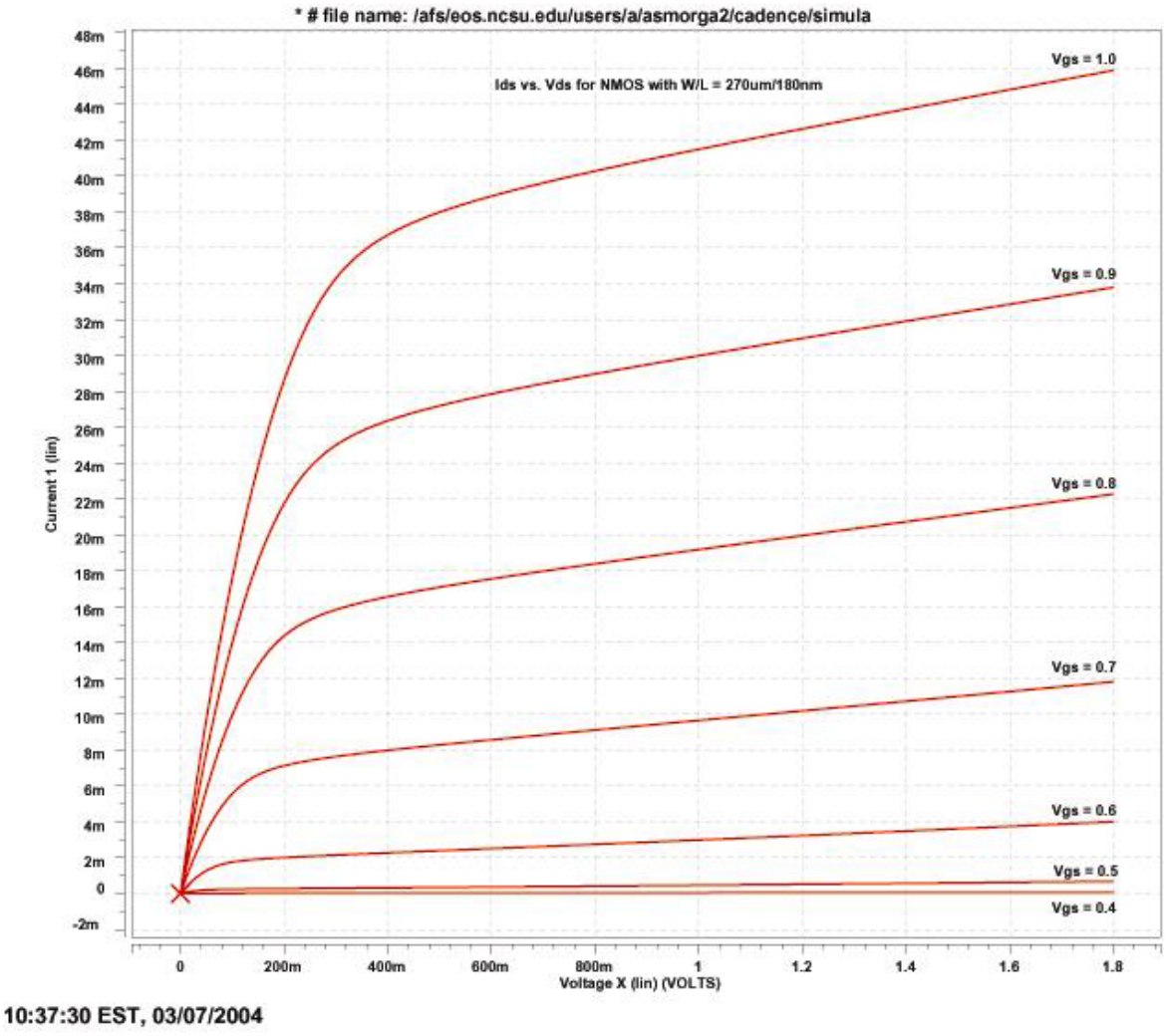


Figure A.3 - Id Curves for NMOS with W/L=270um/180nm in 0.18um (Deep) Process

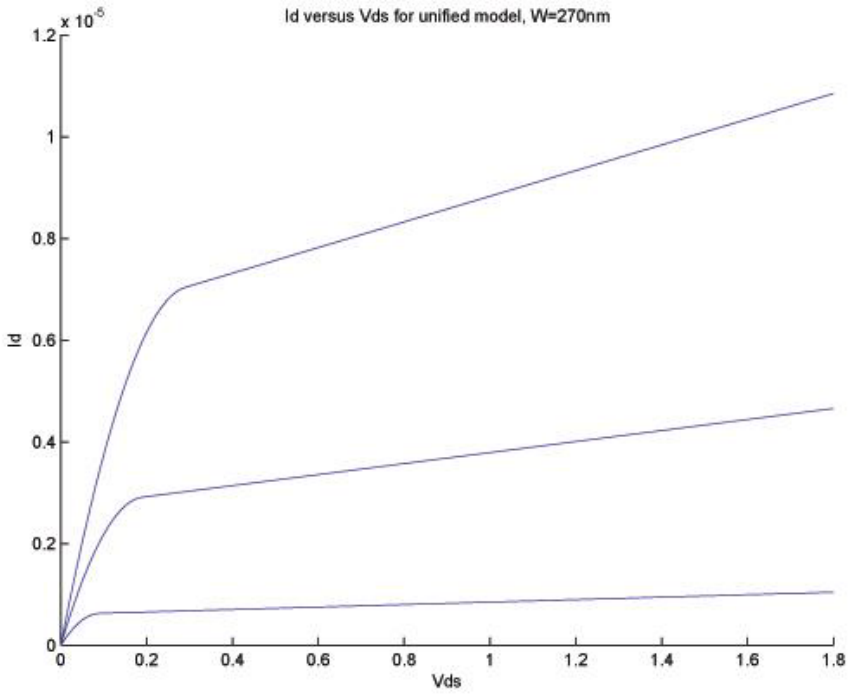


Figure A.4 - Id Curves for NMOS with W/L=270nm/180nm Generated in Matlab

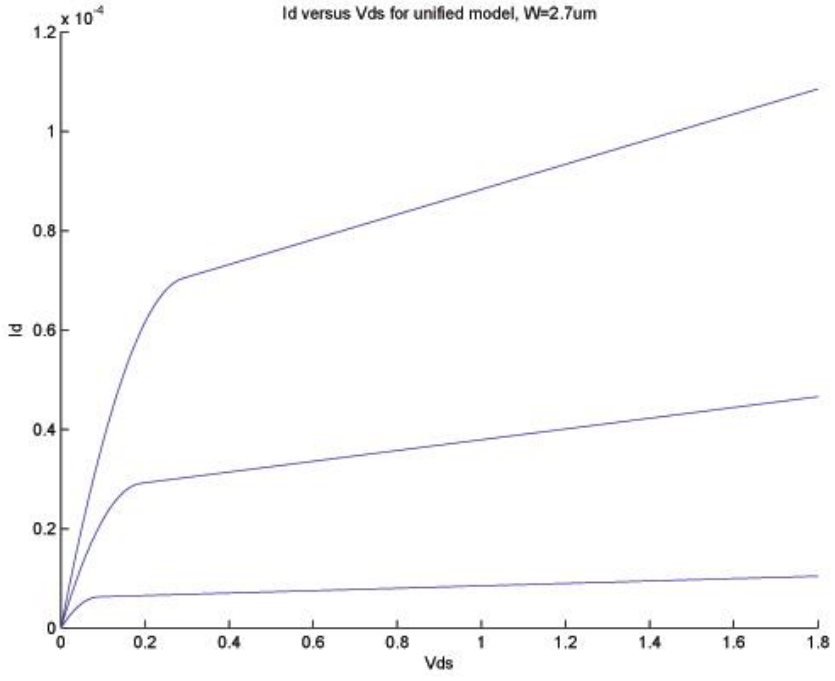


Figure A.5 - Id Curves for NMOS with W/L=2.7um/180nm Generated in Matlab

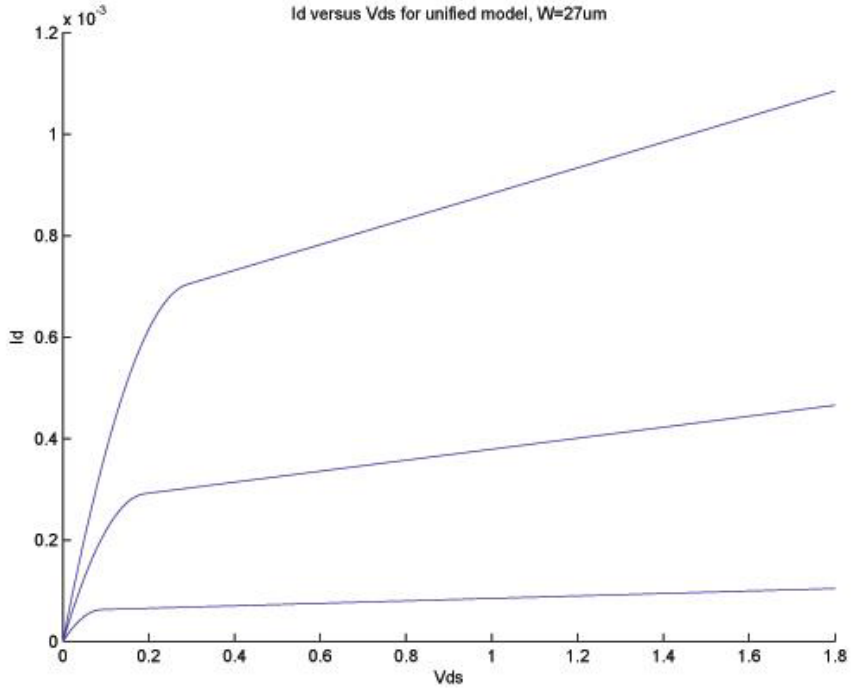


Figure A.6 - Id Curves for NMOS with W/L=27um/180nm Generated in Matlab

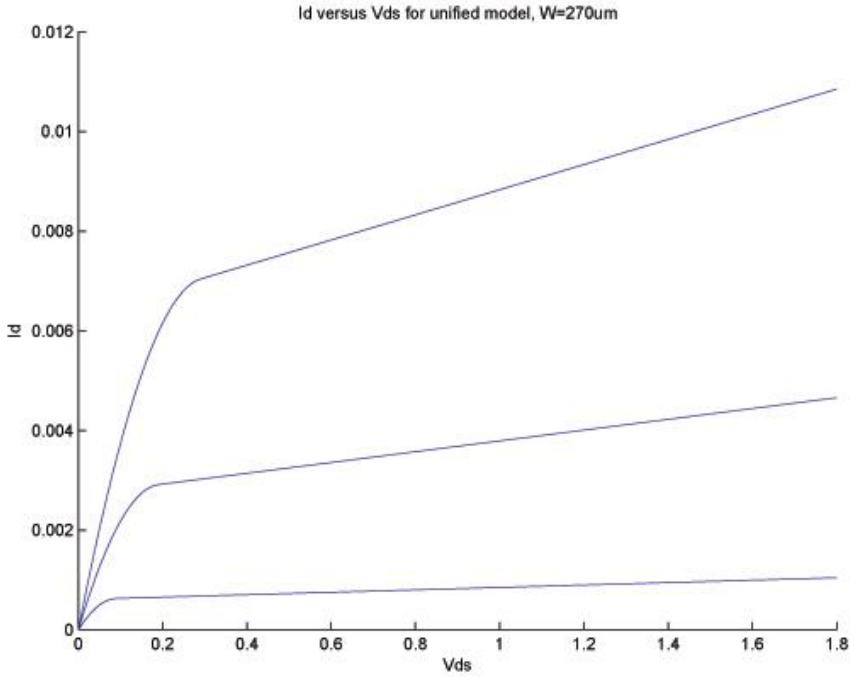


Figure A.7 - Id Curves for NMOS with W/L=270um/180nm Generated in Matlab

Matlab Script for Calculating Adjusted Process Transconductance:


```
%The following script plots the Id vs. Vds curves for
%   270nm, 2.7um, 27um, and 270um NMOS transistors with
%   the given process parameters.  "kp" and "kn" represent
%   the process transconductance of the TSMC 0.18um modeled
%   PMOS and NMOS transistors respectively.  "kn_low"
%   represents the adjusted process transconductance of the
%   NMOS to attempt to match Hspice simulation results.
%   The PMOS was not yet adjusted, nor plotted since it is
%   not used in the present examples.

clear;

%Device Parameters
%PMOS
Coxp=8.7E-15;
Cop=0.68E-15;
Cjp=1.2E-15;
mjp=0.42;
Phibp=0.86;
Cjswp=0.21E-15;
mjswp=0.27;
Phibswp=0.64;
Vtp=-0.34;
Vdsatp=-1;
kp=-51E-6;
lambdap=-0.1;  %Channel length modulation 'lambda'

%NMOS
Coxn=8.7E-15;
Con=0.72E-15;
Cjn=0.97E-15;
mjn=0.37;
Phibn=0.73;
Cjswn=0.26E-15;
mjswn=0.1;
Phibswn=0.4;
Vtn=0.41;
Vdsatn=0.48;
kn=230E-6;
kn_low=100E-6;  %REDUCED FROM DAVIS'S 150e-6 TO MATCH SPICE
lambdan=0.4;  %Channel length modulation 'lambda'  CHANGED LAMBDA FROM 0.09 TO MATCH SPICE


W=270E-9;
L=180E-9;

figure(1)
xlabel('Vds');
ylabel('Id');
title('Id versus Vds for unified model, W=270nm');

for i=0:2,
    Vgs=0.500 + i*0.100;
    for j=1:181,
        Vds=0 + (j-1)*0.01;
        [Id(j), ro(j)] = ids_unified(kn_low,W,L,lambdan,Vgs,Vtn,Vds,Vdsatn);
        Vds_axis(j) = Vds;
```

```
    end
    hold on;
    plot(Vds_axis,Id);
    hold off;
end

W=2.7E-6;
L=180E-9;

figure(2)
xlabel('Vds');
ylabel('Id');
title('Id versus Vds for unified model, W=2.7um');

for i=0:2,
    Vgs=0.500 + i*0.100;
    for j=1:181,
        Vds=0 + (j-1)*0.01;
        [Id(j), ro(j)] = ids_unified(kn_low,W,L,lambdan,Vgs,Vtn,Vds,Vdsatn);
        Vds_axis(j) = Vds;
    end
    hold on;
    plot(Vds_axis,Id);
    hold off;
end

W=27E-6;
L=180E-9;

figure(3)
xlabel('Vds');
ylabel('Id');
title('Id versus Vds for unified model, W=27um');

for i=0:2,
    Vgs=0.500 + i*0.100;
    for j=1:181,
        Vds=0 + (j-1)*0.01;
        [Id(j), ro(j)] = ids_unified(kn_low,W,L,lambdan,Vgs,Vtn,Vds,Vdsatn);
        Vds_axis(j) = Vds;
    end
    hold on;
    plot(Vds_axis,Id);
    hold off;
end

W=270E-6;
L=180E-9;

figure(4)
xlabel('Vds');
ylabel('Id');
title('Id versus Vds for unified model, W=270um');

for i=0:2,
    Vgs=0.500 + i*0.100;
    for j=1:181,
```

```
K:\thesis\netlists\trx_performance_tests\Matlab\Ids_vs_Vds.m          Page 3
March 7, 2004                                                  12:30:49 PM

        Vds=0 + (j-1)*0.01;
        [Id(j), ro(j)] = ids_unified(kn_low,W,L,lambdan,Vgs,Vtn,Vds,Vdsatn);
        Vds_axis(j) = Vds;
    end
    hold on;
    plot(Vds_axis,Id);
    hold off;
end
```

Matlab Function ‘ids_unified.m’ called in “Ids_vs_Vds.m’ script :

```
K:\thesis\netlists\trx_performance_tests\Matlab\ids_unified.m         Page 1
March 7, 2004                                                  12:34:02 PM

%This function calculates the expected Ids over discrete
%   values of Vgs for a sweep of Vds.  Ids is calculated
%   using the unified model for drain current.


function [Id, Ron] = ids_unified(k,W,L,lambda,Vgs,Vt,Vds,Vdsat)


x = [abs(Vgs-Vt) abs(Vds) abs(Vdsat)];
Vmin=min(x);

%Adjust Vmin for negative signs in PMOS
if (Vgs-Vt)<0 | Vds<0 | Vdsat<0,
    Vmin = -1*Vmin;
end

Id= k*W/L*((Vgs-Vt)*Vmin-(Vmin^2)/2)*(1+lambda*Vds);
Ron= Vds;
```

Matlab Script for Calculating Optimum Gate Width for a Given Rise-time:

```
K:\thesis\netlists\trx_performance_te...\capscaling_for_delay.m       Page 1
March 7, 2004                                                  12:48:34 PM

%  The following  script also calculates
%   the optimum gate width (Wopt) to achieve a particular rise-time
%   using the constant-current method for approximating voltage rise
%   The optimization accounts for internal parasitic capacitance
%   by using the approximation for parasistic capacitance per unit gate
%   width derived in an earlier Excel spreadsheet.


%Set parameters for specific problem to be solved
Ln=180E-9;
Vgs=0.6;
Vt=0.43;
kn=100E-6;
CL=50E-15;
Cnom=0.11E-15;
tf=200E-12;
vin=0.3;
Av=0.85;
vout = Av*vin;
beta=0.5*kn*(Vgs-Vt)^2/Ln;

[Wopt] = solve('tf-(vout*(CL+2*(Cnom/lambda)*Wopt))/(beta*Wopt)','Wopt=Wopt')
```

Table A.1 - Schmidt Trigger Hysteresis Analysis

NMOS parameters:			
un= (electrom mobility)	2.63E-02	2.63E-02	2.63E-02
Er= (relative dielectric)	3.90E+00	3.90E+00	3.90E+00
tox= (oxide thickness, m)	4.00E-09	4.00E-09	4.00E-09
Cox= (F/m^2)	8.63E-03	8.63E-03	8.63E-03
Vtn=	4.10E-01	4.10E-01	4.10E-01
Device Dimensions:			
W_N1=	2.70E-07	5.40E-07	6.75E-07
L_N1=	1.80E-07	1.80E-07	1.80E-07
W_NF=	2.70E-07	4.50E-07	4.50E-07
L_NF=	1.80E-07	1.80E-07	1.80E-07
BetaN1=	3.41E-04	6.82E-04	8.52E-04
BetaNF=	3.41E-04	5.68E-04	5.68E-04
Vil=	1.11E+00	1.07E+00	1.03E+00
Input/Feedback Ratio:			
N1/NF=	1.00E+00	1.20E+00	1.50E+00
PMOS parameters:			
up= (electrom mobility)	1.18E-02	1.18E-02	1.18E-02
Er= (relative dielectric)	3.90E+00	3.90E+00	3.90E+00
tox= (oxide thickness, m)	4.00E-09	4.00E-09	4.00E-09
Cox= (F/m^2)	0.008629	0.008629	0.008629
Vtp=	0.34	0.34	0.34
Device Dimensions:			
W_P1=	2.70E-07	1.35E-06	1.67E-06
L_P1=	1.80E-07	1.80E-07	1.80E-07
W_PF=	8.10E-07	6.75E-07	9.45E-07
L_PF=	1.80E-07	1.80E-07	1.80E-07
BetaP1=	1.53E-04	7.63E-04	9.42E-04
BetaPF=	4.58E-04	3.82E-04	5.34E-04
Vih=	7.83E-01	1.25E+00	1.22E+00
Input/Feedback Ratio:			
P1/PF=	3.33E-01	2.00E+00	1.76E+00