1.0 Introduction

This lecture concludes the study of dynamic gates by computing their logical effort and exploring how the use of dynamic gates impacts the optimal gain per stage of paths. It then catalogs various circuit pitfalls which must be avoided in all circuit families. It concludes by describing a plethora of circuit families, some of which are useful in special applications, and many of which are best avoided.

2.0 Logical Effort of Dynamic Gates

The logical effort of a dynamic gate is the ratio of its input capacitance to a static inverter with the same pulldown resistance. Since precharge is assumed to be less critical, the size of the precharge transistor does not impact logical effort. Generally, sizing the precharge device to be half the strength of the pulldown path gives sufficiently fast precharge while keeping the clock loading and diffusion parasitics of the precharge device reasonably low. A few examples of logical effort of dynamic gates is shown in Figure 1.
Designing with dynamic gates has an interesting effect on the optimal gain per stage. Remember that the calculation of optimal gain per stage driving heavy loads were based on optimal fanout per stage of about 4. Since the buffer element in static circuits is an inverter, the logical effort is 1 and the optimal gain per stage equals the optimal fanout per stage. Dynamic circuits can buffer heavy loads with a domino buffer built from a dynamic inverter followed by a high skew static inverter. These gates have lower logical effort. Therefore, paths built from dynamic circuits should be expected to have a lower optimal gain per stage for quickly driving heavy loads.

Consider a buffer chain built from dynamic inverters alternating with high-skewed inverters. To provide a fanout of 4 per stage, the domino gate (dynamic + static gate) must have
Lecture 8: Circuits & Pitfalls

a fanout of 16. The static inverter has a logical effort of 5/6. The dynamic gate has a logical effort of 2/3 with a clocked pulldown or 1/3 without. Thus, the total gain of the two stages is $16 \times 5/6 \times (2/3 \text{ or } 1/3) = 8.9$ or 4.5, corresponding to an optimal gain per stage of about 3 when dynamic gates have clocked pulldowns or 2.1 with no clocked pulldown.

The conclusion of this analysis is that paths with dynamic gates should be sized with a gain of about 3 per stage instead of 4 per stage used for static paths.

3.0 Circuit Pitfalls

We have seen numerous circuit pitfalls in the past lectures. In this section we’ll collect them into a single list. We’ll then apply the list to understand weaknesses of many other circuit families.

3.1 Ratioed Circuits

Ratioed circuits are any circuits that depend on the relative strength of PMOS and NMOS transistors. Process variation causes this ratio to change. Thus, the ratio must be selected so that even under worst case process variation, one transistor is sufficiently stronger than the other to produce the correct result. Pseudo-NMOS circuits are a classic example of ratioed logic.

3.2 Charge Sharing

Charge sharing occurs anytime charge may be transferred onto a floating node. It can lead to data-dependent failures when particular input patterns deposit the wrong charges which get shared with the floating node. Dynamic gates exhibit the classic example of charge sharing.

3.3 Voltage Drops

Certain circuits exhibit voltage drops which prevent their output from swinging rail to rail. Classic examples are NMOS pass transistors (which only swing up to $V_{DD} - V_t$) and bipolar transistors (which only swing up to $V_{DD} - V_{BE}$). If no restoring devices are included to pull the output to the rail, the PMOS transistors of the next stage may turn slightly on and dissipate power. If the drop is too large a fraction of the power supply, the next stage may actually sense an incorrect value. This is especially severe for low-voltage circuits. $V_{BE}$ is constant at 0.6-0.7 volts, so it becomes a larger fraction of the supply. Similarly, if a circuit designed in a high-voltage process with large values of $V_t$ is then used on a low-voltage part, $V_t$ drops become more important.
3.4 Power Supply Noise

The power supply VDD and GND voltages vary across the chip and in time. Operating frequency is set by the average value of the supply over the entire cycle; if VDD droops at the beginning of the cycle but charges back up later, the first gates will run slower than average and the last gates faster, for a net average delay that sets cycle time. Unfortunately, noise margins are impacted by instantaneous power supply variation.

Since the inductance of power pins is too large to supply significant current at extremely high frequency, most gates get their instantaneous switching current from nearby on-chip bypass capacitance. For example, suppose at the beginning of a cycle, a bank of flip-flops toggle and send new values to large bus drivers. The bus drivers pull their current from the nearby power network. This causes an IR drop through the nearby power lines. The current prefers to come from nearby sources with lower series resistance to minimize this drop. Thus, only nearby capacitors can supply much of the current. These nearby capacitors include both explicit bypass capacitance and implicit capacitance provided by all the nearby gates which are not switching. Even if the overall power supply resistance is low for the chip, a high supply resistance caused by a minimum width supply line driving many simultaneously switching gates will lead to large instantaneous IR drop and high noise. This must come out of the noise margin of the next gate.

Over a longer time, such as an entire cycle, average currents flow through the power pins to recharge the on-chip capacitance to an average value.

3.5 Coupling Noise

As we have seen in detail, coupling between adjacent signals is a major source of noise.

3.6 Leakage

Transistors leak even when off due to subthreshold conduction. As long as circuits are actively driven sufficiently often, leakage is not a problem. However, if the clock is stopped indefinitely, nodes which are not driven may leak away to the wrong value. The rate of leakage increases exponentially with reduced threshold voltages.

3.7 Alpha Particle Hits

Alpha particles are helium nuclei (two protons and two neutrons) which hurtle through space at high speed. They occasionally crash into chips and dump their energy onto the node they hit. If they hit floating dynamic nodes, such as RAM cells or dynamic gate outputs, they can cause the voltage to change. To minimize this change, the energy held on the capacitance of the floating node must be much greater than the energy in the alpha particle. Typical design rules call for a minimum of several fF of capacitance on any dynamic node.
3.8 Contention

Whenever two drivers connect to a single node, contention can occur. This contention could cause glitches on sensitive inputs attached to the node. It can also cause failure in circuits like latches where feedback allows the latch to be backdriven.

3.9 Races

Whenever the ordering of two signals matters, it is important to guarantee the correct order will occur under any processing or environmental variation. A classic race problem arises constructing matched delays for dynamic logic, as was discussed in a previous lecture.

Extra margin is added to the delay which is supposed to be longer to make sure the signal always arrives last. This margin may be 50% of the nominal delay if layout information is not available. With extracted layout information, the margin can be relaxed to around 30%. These margins are very large and significantly impact chip performance, so often it is best to avoid races entirely.

Races are crucial to operation of a chip. Since races are based on relative delays instead of on a clock, a failed race will force the chip to be scrapped, rather than just run more slowly.

3.10 Noise Margins

Many of these effects described in this section contribute noise to a system. The combined noise of all the effects may cause circuits to fail even when any single source is too small. Thus, noise budgets are used to divide the total noise margin of a circuit among the various sources of noise. The chance of failure can never be totally eliminated, but it can be reduced to acceptable levels.

Noise margins are largest for static gates with switching thresholds near VDD/2. Skewed static gates have smaller, but still hefty margins. Dynamic gates have input noise margins of only $V_t$. Sense amplifiers have even smaller noise margins.

Since the noise margins of dynamic gates are so low, being overly conservative about noise budgets results in undesignable circuits. In particular, coupling noise is a function of layout, so decreased coupling noise can be traded for increased layout area. By taking advantage of specific information like the fact that dynamic NOR gates never experience charge sharing, some dynamic gates can allow higher coupling noise and thus easier layout.

3.11 Tool limitations

Microprocessors have become so complex that no human can verify a design. Instead, automatic tools must be used to analyze circuits and check design rules. Hence, special
circuits which cannot be handled by available tools must not be used on a wide-spread basis.

4.0 More Circuit Families

Having discussed a variety of circuit pitfalls, let us move on to examine more circuit families. Keep the pitfalls in mind while analyzing the circuit families.

4.1 NORA

Domino logic is fast because it reduces the input capacitance of dynamic gates by eliminating PMOS transistors and because it responds to smaller input swings. Unfortunately, there is still a static inverter between each dynamic gate.

NORA (NO RAce) logic [1] replaces the static inverter with another precharged gate. The replacement gate is precharged low and evaluates high through PMOS transistors, as shown in Figure 2. Though such PMOS dynamic gates are not as fast as normal NMOS dynamic gates, they are still faster than inverters.

Unfortunately, now the noise margins of the gates are greatly reduced. The PMOS precharge block only has a noise margin of $V_t$, like the NMOS precharged block. Unfortunately, its inputs now come from noise-prone dynamic nodes rather than from statically driven gates. DEC tried the technique on the J-11 chip and “it was a disaster.” AT&T tried a related technique, Zipper Domino [2], on their CRISP microprocessor. Zipper swings the clocks less than rail-to-rail so the precharge transistors remain partially on and serve as weak keepers to improve noise margins. Unfortunately, Rev 1 of the CRISP ALU did not work.

NORA and its variants are therefore generally regarded as too risky for general use.
4.2 BiCMOS

BiCMOS logic takes advantage of the excellent current drive characteristics of bipolar transistors to build buffers that can quickly drive large fanouts. A typical BiCMOS inverters shown in Figure 3.

**FIGURE 3. BiCMOS Inverter**

![BiCMOS Inverter Circuit](image)

Figure 4 shows the relative delays of BiCMOS and regular CMOS inverters. For the same input capacitance, a BiCMOS driver has a larger intrinsic delay but a smaller fanout-dependent slope. This is the principle argument for using BiCMOS drivers.

**FIGURE 4. CMOS and BiCMOS inverter delays**

![CMOS and BiCMOS Inverter Delays](image)

The flaw in this argument is that good CMOS designers would never drive huge fanouts with an inverter anyway. Instead, they would construct an exponential horn of inverters with a fanout of around 4 per stage to drive a large load in time logarithmically related to fanout. Such a delay curve is shown in Figure 5. Note that these delays are not to scale, but are meant to illustrate the general concepts.
Although BiCMOS is not better than CMOS by a huge factor, it still can be somewhat faster. Bipolar transistors take a large amount of area and require extra processing steps, but the costs can be justified where speed benefits sufficiently. Intel used BiCMOS gates with success on Pentium and Pentium Pro in their 0.8 and 0.6 micron processes, and in a slow version of their 0.35 micron process.

Bipolar transistors finally passed out of style with the scaling of VDD. Power supplies are falling low enough that the $V_{BE}$ drop can no longer be tolerated. A 0.8 volt drop on a 5 volt supply is tolerable, but a 0.8 volt drop on a 1.6 volt supply produces an output that is very hard to use. Intel no longer uses bipolar transistors on its 0.35 micron and below processes intended for low power supplies.

Exponential made a valiant attempt to use bipolar transistors in a PowerPC implementation [3]. Using bipolar ECL logic and BiCMOS caches, they originally targeted 533 MHz operation at under 85 watts of power. Unfortunately, they never managed to ship parts above mid-400 MHz. Since bipolar logic is less dense than CMOS, they were also forced to use a simpler microarchitecture delivering fewer SpecInt/MHz, so the overall performance was scarcely better than Motorola/IBM PowerPC chips and the price was not competitive.

### 4.3 Low-Threshold Circuits

Some processes feature both high and low threshold devices. These can be used to trade between speed and power consumption. For example, NMOS-only pass-transistor circuits can be built with low-threshold transistors. If they drive gates with high thresholds, the threshold drop from the NMOS pass-transistor may be small enough not to turn on the next gate. Thus, level-restoring circuits are not needed. Another example is to build the evaluation transistor of dynamic gates with a high threshold device, but other logic transistors with a low threshold. This reduces subthreshold conduction and power consumption during precharge because the clocked evaluation transistor is off, while preserving the speed benefits of the low-threshold devices in the remainder of the pulldown stack. Unfortunately, noise margins of the dynamic gate are just the threshold of the low-threshold transistor, which can be very small. Also, a keeper is still necessary to fight leakage through the low-threshold transistors while the gate is in evaluation.
4.4 Partial Swing Circuits

The time required for a gate to charge a load capacitance is $\Delta t = C/I \Delta v$. Sizing transistors optimizes the $C/I$ ratio of the load to the driver to minimize $\Delta t$. However, reducing the output swing $\Delta v$ can also reduce $\Delta t$. Circuits which capitalize on this idea fit in the class of partial swing circuits.

One example of partial swing logic is DSL, Differential Split-Level logic (DSL). DSL clamps the internal swings of gates to between 0 and VDD/2 by using a cascoded NMOS transistor tied to VDD/2 + Vt. The output voltages can pull up to VDD but not quite all the way down to 0. As a result, there is some DC power consumption.

The original paper on DSL [4] obtained good results because it built the pulldown NMOS trees with 1 micron long transistors in a 2 micron process! The paper claimed that the reduced $V_{DS}$ of the pulldown transistors allowed reliable operation. This reduction in channel length more than compensated for the increase in delay due to the cascode transistor. In general, reducing transistor length below the minimum for the process is not wise and/or possible.

DSL also suffers from several other problems. One is that if the gate idles for long enough, subthreshold conduction may pull internal nodes above VDD/2. This will slow the gate the next time it switches because the internal voltage swing is large. Another problem is generating the reference voltage and routing it around the chip without much noise.

4.5 Sense Amplifier Circuits

An extension of the partial swing circuit idea is to build a sense amplifier into a gate which detects a small swing and quickly amplifies it. A classic example of a good application of sense amplifiers is in SRAM arrays. Amplifiers have also been used in a variety of esoteric differential dynamic logic families, mostly with poor results.
Lecture 8: Circuits & Pitfalls

In a SRAM array such as in a cache, a small SRAM cell activates when its word line is asserted and must read out onto differential precharged bitlines. Since there may be 127 or more other SRAM cells on the same line and since the SRAM transistors are sized very small for good density, the delay for one bitline to drop is very large. Instead of waiting for the bitlines to take on valid logic levels, a sense amplifier is used to amplify small differences between the two bitline voltages. Self-timed delays may be used to activate the sense amplifier when sufficient there is sufficient voltage difference. Examples of such circuits will come in a future SRAM lecture.

Three versions of dynamic differential logic using amplifiers are Sample-Set Differential Logic (SSDL) [5], Enabled/disabled CMOS Differential Logic (ECDL) [6], and Differential Current Switch Logic (DCSL) [7]. All are prone to charge sharing noise getting amplified and causing incorrect evaluation. The performance is often no better, or even worse, than normal dual-rail domino, so such logic families are not widely used.

4.6 Conclusions

As design requirements and tools change and implementation technologies evolve, the set of useful circuit families also evolves. Thus, it is somewhat dangerous to make blanket recommendations of circuit families. Nevertheless, there are clear trends that just a few circuit families are used for the great majority of designs. Let us summarize these trends, keeping in mind the reasons as well as the results.

Most circuits are built using static CMOS, transmission gates, and dynamic logic. Dynamic logic is used for speed. Transmission gates may be used for multiplexors or XOR structures. Static CMOS is used everywhere else for simplicity and robustness. Some circuit styles allow occasional use of pseudo-NMOS gates when DC power consumption can be tolerated and wide NOR structures are needed. Specialized circuits are used in certain well-understood custom blocks, such as small-swing sense amplifiers in SRAM and special drivers on low-voltage I/O pads.

BiCMOS has been popular in the Intel Pentium and Pentium Pro processors, but is no longer used in Pentium II because the power supply is low enough that the VBE drop cannot be tolerated. Exponential made a valiant attempt to use bipolar transistors in their high-speed PowerPC implementation, but sadly failed to meet performance targets. As power supplies continue to fall, bipolar transistors will probably vanish from the digital design scene.

Most other circuit families suffer from pitfalls that doom them to remain academic curiosities. When evaluating proposed new circuit families, keep a critical eye toward potential pitfalls.

5.0 References


