Power management can cause latchup in CMOS chips

Paul Rako, Creative Writer, Atmel - February 02, 2016

If you leave signals applied to the inputs of a CMOS chip with the power turned off, the chip might explode when you re-apply power. This is called latchup. Similarly, if you drag the outputs of a CMOS chip above or below the power supply rails you can latchup the part. Latchup is not always destructive. Sometimes the part heats up, but when you remove all power and signals, the chip has survived the latchup condition. It does not matter if the CMOS IC is a microcontroller, an operational amplifier, an analog-to-digital converter (ADC), logic, or analog multiplexor.

Latchup becomes a real problem when you try to power up and down different sections of your design to save power. It is also a problem when you have cables or inputs from other devices going directly to your chip. Another common problem is when a CMOS output is connected to a large capacitive load. The part will go into latchup the moment you turn off its power. As long as you don’t turn the power back on for a few moments you are fine, the energy in the capacitor dissipates and the CMOS part is out of latchup. But if someone cycles power quickly, or if there is a momentary dropout or glitch, boom, the part blows its lid.

I got to thinking about latchup when I heard that the Atmel SAM L21 microcontroller has five separate power domains inside the chip. Rather than just kill the clock to unneeded circuit blocks, the SAM L21 turns off power to the block. This is great, since it removes any leakage current contributed by that block. The IC designers can use tiny fast cheap transistors that tend to have higher leakage, but it does not matter. When you turn off the circuit block, with zero volts across the logic gates there can be no leakage.

The IC designers who did the SAM L21 had to suffer with all the same latchup problems as we systems folks do. One of the tricks I have used is to put high-value resistors in series with the inputs to the device you need to turn on and off (Figure 1a). Another is to use Schottky diode clamps on the inputs and outputs, so they can never get more than a diode drop (0.6V) above or below the rails (Figure 1b).
To prevent latchup in CMOS chips you can put high-value resistors between the inputs and outputs (a). Another solution is to put Schottky diode clamps in the wires to prevent them from going more than 0.3V above or below the switched Vcc net (b). Note that power will still flow into the switched chip via the input pins and internal ESD diodes.

It’s nice to have discrete diode clamps on your schematic to remind you that you already have them inside the chip in the form of ESD (electrostatic discharge) diodes. I have seen engineers struggle for hours trying to figure out why an expensive military electronics box was not working quite right. It was kind of working, but nothing made sense. The problem was they had not turned on power to the box. The CMOS circuitry was being powered by the inputs from another box or the test equipment (Figure 2).
Figure 2 Voltage on the input pin of most any IC will flow through its ESD diode structure and self-power the chip and worse yet, any chips that are connected to the same power net.

Don’t laugh, it has happened to a lot of good engineers, and you might be next. Since CMOS circuitry needs so little power, the inputs can flip the internal ESD diodes around and you will not only power the chip’s Vcc line, that Vcc line connects to all the other chips and they will be powered up too. So then you risk latching up the first IC and you are not really saving power in the first place.

The physics of latchup

Years ago, when I was at National Semiconductor, I came across a nice simplified die diagram that purported to explain latchup (Figure 3). I loved it since the color-coding would relate the die areas with the emitters, bases, and collectors of the parasitic transistors that form when the part goes into latchup. I dragged the diagram out and was going to use it again for this article. Fortunately I went to the third floor to ask IC designer Scott Fritz if the diagram made sense. Scott is a fellow analog aficionado, and was quick to point out many deficiencies and over-simplifications in the diagram. I knew it was off-base when he said, “The substrate is usually P-type silicon with an N-type well.”
A CMOS gate has a P-channel FET on top and an N-channel one on the bottom. The die diagram shows how process engineers make these transistors. But there are also parasitic bipolar transistors formed by the structures, shown on the far right. The colors of the semiconductor structures match the color of the schematic elements. With the connections on those parasitic bipolar transistors, you can see once a little current starts to flow, it will feed off itself until the device melts.

With Scott’s direction, I worked up a much better diagram (Figure 4). I made the substrate of the transistor pair out of P-type material, and put the upper transistor in an N-well. Scott noted that while op-amps will use plus and minus power, it is easier to just think of power and ground for most CMOS circuits. The other major mistake of the first diagram was that it did not show the power and ground connections to the N-well and the P-substrate. Scott explained that for output transistors you might have only one P-channel transistor in the N-well, but typically there are dozens or hundreds. The reason I had misgivings about that first diagram was that it showed the parasitic transistors, but it really did not show the mechanism of how the latchup occurs, how you get those parasitic NPN and PNP transistors to form in your CMOS chips.
An improved diagram shows the wafer substrate as the more typical P-type material. You run most CMOS circuits single-ended, with just power and ground. Also shown are the P+ and N+ connections needed to apply bias between the N-well and the substrate.

After Scott explained that the N-well is hooked to Vcc and the substrate is hooked to ground, he pointed out those connections add two resistors to the equivalent circuit on the right side of the figure. He then came at the problem from an IC designer’s perspective. He said there are two primary things an IC designer wants to do to prevent latchup. One is to make those resistors as low-valued as possible. The other is to make the beta ($\beta$, or current gain) of the parasitic transistors as low as possible.

Note that the bias, the power applied between the N-well and the substrate, means any free electrons in the substrate will seek that ground. Similarly any free holes in the substrate will seek Vcc. If the only path is through those parasitic emitters, then boom, the part will latch. So having those resistors that are formed by the N+ to N and the P+ to P paths be as low value as possible will provide a shunt path around the emitters to let the free carriers seek power and ground.

Since the resistor’s value is directly proportional to the distance the free carriers have to travel. Fritz noted that my improved diagram is drawn for clarity, but the best way to reduce the values of the resistors was to add a second N+ pad to the other side of the N-well and move the P+ pad to the
other side of its transistor (Figure 5). The two resistors in the N-well parallel and reduce the effective resistance. Moving the P+ pad in the substrate puts it very close to the substrate material that acts as the base of the NPN parasitic transistor. I asked why you need N+ and P+ to begin with. Scott explained that if you just hook metallization to the N-type material it forms a Schottky diode. The greater impurities in the N+ and P+ material makes the silicon more metallic and the connection to the metallization becomes galvanic.

![Diagram](image_url)

**Figure 5** An IC designer can ameliorate latchup problems by lowering the equivalent resistors in the right-side diagram. She does this by putting two N+ pads in the N-well, so the two resistors parallel in value. For the lower transistor, she can move the P+ pad where it is closest to the effective base area in the substrate. In addition to lowering the resistances, it is important to make the current gain of the parasitic NPN and PNP transistors as low as possible.

The other way to prevent latchup in CMOS circuits is to reduce the beta of the parasitic transistors. If you remember your transistor physics class, after a few whiteboards’ worth of equations, the beta of a bipolar transistor is directly proportional to the thickness of the base area the free carriers are being swept through by the applied voltages. Looking at Figure 4 again, the easiest way to make the base of the lower NPN parasitic transistor thick is to move the whole CMOS N-channel device away from the upper P-channel device. Now there is a lot of orange-colored substrate between the two devices, and that means the parasitic base is very thick and the beta is very low.

There is no simple IC design trick to reduce the beta of the top parasitic PNP transistor. This is a vertical PNP since the P, the N, and the P areas are arranged vertically in the die (note the figure has the die cross-section standing on edge to mimic the arrangement of the schematic). The beta of the parasitic PNP transistor is fixed by the thickness of the N-well between the P+ source
connection and the substrate. That is a function of the semiconductor process, but Fritz assures me it is well-controlled and predictable, despite being set by diffusion and not lithography.

There is an economic benefit to having thin wafers since you can cut more of them from a silicon crystal ingot. But the practical mechanical limit on wafer thickness still allows process engineers plenty of room to make the N-well deep enough to ensure the beta of the parasitic PNP transistors is very low. Another batch of mathematics will show that you have to keep the product of the two transistors’ betas below 1 to prevent latchup. Since it is easy to make the NPN beta very low, the beta of the PNP can be above 1 but when multiplied by, say, the 0.25 beta of the NPN transistor, the product is still less than one.

Beta increases with temperature, and an IC designer has to factor that into his design. Beta is also higher with low collector currents. A CMOS device on the edge of latchup has miniscule collector current to begin with, so the beta is high. Once the part latches, the currents go way up, but it is too late; once the latchup begins the part will melt unless something external limits the current, perhaps the bond wires melting. Neither situation is a desirable design result.

**The mechanisms of latchup**

Now that you understand the physics of latchup, you can see how it starts. This is what was unsatisfactory in the first diagram, it showed the parasitic transistors you end up with, but not how they get formed out of a CMOS circuit made from MOSFETs. Looking at Figure 4 (shown again below), imagine that you take the output pad and drag it below ground. That blue N+ pad will now sink current once you get past the P-N diode drop voltage, about 0.6V. The orange substrate material will be full of electrons and they will seek ground through the green N+ parasitic NPN emitter pad. An analogous situation happens in the N-well when you pull the output above Vcc. The free carriers in the base regions of the parasitic bipolar transistors are all you need to start the latchup, and once it begins it just feeds on itself.
Figure 4 An improved diagram shows the wafer substrate as the more typical P-type material. You run most CMOS circuits single-ended, with just power and ground. Also shown are the P+ and N+ connections needed to apply bias between the N-well and the substrate.

This is one reason switching a chunk of CMOS circuitry on and off to save power can cause latchup. If you turn off Vcc, a volt applied to the output will cause current to flow, perhaps as it drags the Vcc voltage up, but current flows nevertheless. When your power management system re-applies power, boom, the part is in latchup.

Having voltage on the inputs with power off will cause latchup in a more indirect way. No current can flow in the glass gate insulation. What happens is that the current flows in the internal ESD diodes that are needed to protect the input pin (Figure 6). Here the device is operating as a simple two-transistor buffer. Figure 6 builds off the simplicity of Figure 4, since the IC designer’s tricks are not pertinent to this analysis. The lower ESD protects the part against negative voltage pulses on the input. If you imagine the input pad being drawn negative, you can see current will flow from the P+ pad connected to ground to the N+ pad that has a negative voltage. The current flows directly in the substrate, so the substrate is full of carriers. Once again, if you apply power to the part, those free carriers will seek ground and Vcc through the emitters of the parasitic transistors. With enough free carriers, enough beta, and enough shunt resistance, the part will latch.
The ESD diodes used to protect the pins on a CMOS chip will also cause latchup problems. If you drag the input pin above power or below ground it will inject free carriers into the substrate of the ESD diode structures. When those carriers seek Vcc and ground via the emitters of the parasitic NPN and PNP transistors they can cause the part to enter latchup.

For the case where you drag the input pad to a voltage above Vcc, you are forward biasing that top ESD diode structure. It sits in an N-well and the IC designer will make sure that N-well is separate from the one he used for the P-channel transistor. Nevertheless, there is yet another parasitic vertical PNP formed by the ESD diode structure. So once again, the substrate gets an injection of free carriers. Sure, the ESD diode powers the part as it pulls up the Vcc node, and maybe your power system has low impedance when it is off, so that Vcc node lags a bit, but once you apply real power to the Vcc line, those free carriers in the substrate seek power and ground, and that causes the part to latch.

You can see why latchup is the bane of CMOS IC designers. I once saw a brilliant IC designer struggle for months to keep his op amp design from latching up. He never did solve it. I asked why he was having such problems; after all, the power group had used this exact process for a successful part. He pointed out that a switching regulator was a bandgap reference, a comparator, and an output driver. An op amp has a lot more circuitry and a lot of different structures and transistors and devices in the die. The op amp group had other engineers look at the problem, but ultimately they realized they just could not use that process for amplifiers.

You need to understand the mechanisms and problems of latchup since it is your responsibility to make your designs work. There is nothing the IC designer can do if you abuse his part. It is the IC designer’s job to know the input and output structures, and all the really complicated stuff in
between. But as a system person, you have to at least understand the input and output structures since that is what your discrete components and cables are connected to. Your complexity resides in the interactions between all the input and output structures in your design. [See sidebar, “Looking inside the chip.”]

I once got a call from a famous test equipment company. Their German subsidiary was selling a factory sensor that they would power up, take a measurement, and then power down. The problem was, the op amp in the design would not work for seconds, and then would finally settle down and start working. But the company wanted to take readings in a few milliseconds so they could shut the device down and save power.

I hooked a very fast oscilloscope (made by the company in question) to the circuit. I used JFET probes since they do not load the circuit with as much stray capacitance as a regular probe. It was then apparent that there was an input to the op amp that went to a volt before the op amp Vcc line started to rise. The company had put a very large tantalum capacitor on the power line. I went to that self-same IC designer that had struggled with latchup on his design. He pointed out that the part was well-designed and recovered gracefully from the latchup, but the fact that the ESD diode was forward biased for a few nanoseconds still would put a multitude of free carriers into the substrate. I noted the part did not hard latch, but was just not a functional op amp for a few seconds. He pointed out that modern semiconductor processes are really pure, so that it takes a lot of time for the holes and electrons in the substrate to recombine, and that might well be on the order of a second or two. In the meantime, those free carriers were screwing up the operation of every transistor in the amplifier, so it was no surprise it did not work for a few seconds.

To fix the problem I removed the tantalum decoupling capacitor from the design. Now the op amp Vcc line came up faster than the input to the operational amplifier. There was still a ceramic capacitor to provide decoupling and keep the part out of oscillation. You can imagine the delight of the test equipment company when I told them they could fix the problem by taking out the tantalum capacitors and reducing the cost of their product.

You too can be a hero and miracle-working troubleshooter if you understand just a little of what goes on inside a modern CMOS integrated circuit.

---

**Looking inside the chip**

My mentor, Big John Massa, taught me the importance of knowing what was going on inside the chips I used in my system designs (Figure A). Many times this gives you a way to make sure a chip will work even though it might not meet some spec. One example Massa used was RS-232 receivers. The RS-232 standard is based on the input signal swinging between plus and minus 12V. But Massa would point to the equivalent schematic of the RS-232 receiver chips we used. “Note that input is the base of a Darlington transistor,” he would note. So even though the standard did not recognize an input of plus and minus 5V, or even a single-ended input of zero and 5V, “it would usually work and you can get away with it.”
John Massa knows that electrical engineers have to know what is inside the chips they use so they can properly hook them together.

On the other hand, Massa would point out all the designs where a 74LS244 bus driver chip was powering LEDs. He noted that the output spec of the bus driver really was not sufficient in all cases. Sure, like that RS-232 trick, it would work most times. But when I looked at the guaranteed minimum output voltage and current drive of the chip, and the potential maximum forward voltage of the LEDs we were using, sure enough, there could be a case where the LEDs would not light nearly as brightly as we needed them to.

So this is where engineering judgment and that analog sense of “good enough” is essential to your job. The RS-232 connection was for debugging and it was fine that it was marginal since it would only be used in a nice warm lab with controller conditions. We redesigned the LED drive since that was a production variation problem that might bite us down the road, especially if we changed chip or LED vendors.

So don’t think what is inside the chip is only the IC designer’s problem. It’s also your job to understand the inputs and outputs inside these wonderful magical chips.

Also see:

- Power-supply interrupter fights ESD-induced device latch-up
- Looking for latch-up?