

## **Asynchronous circuits: better power by design**

By Andrew Lines (contributed article)

May 1, 2003

*Asynchronous circuits feature lower overall power consumption and reduce noise and electromagnetic radiation by more uniformly consuming power. Adopting asynchronous-circuit design poses many challenges, but solving these challenges is a compelling reason to do it.*

Managing the power consumption of ICs is an increasingly difficult challenge, because each new generation of portable device includes expanded features and demands longer battery lives. Very-high-speed semiconductor designs now face power consumption as a primary limiting factor as power distribution and heat removal become prohibitively difficult. Process-technology improvements and design changes are addressing this power-consumption challenge. The drive to rein in power consumption has sparked a renewed interest in asynchronous-circuit design.

Although their designers did not necessarily intend them as a "low-power" technology, asynchronous chips feature many "free" low-power capabilities and present an alternative to the difficult power-optimization techniques that synchronous designs use. Asynchronous circuits dissipate power only when they are active. Their power consumption may be comparable with that of synchronous circuits at peak activity, but it may also dramatically decrease during normal operation. Asynchronous circuits eliminate many of the headaches you associate with simultaneous switching noise and electromagnetic radiation. And some styles of asynchronous-circuit design make it easier to bias the physical properties of the transistors to optimize power consumption during continuous operation.

### **An asynchronous primer**

Today, most circuits are synchronous, although in the early days of digital-circuit design, designers gave little regard to the difference between asynchronous and synchronous circuits. Using a system clock simplified development as IC designs became more complex and served as the primary style for digital circuits. Although asynchronous designs became less mainstream, research continued through the years at the California Institute of Technology and other universities and research centers around the world. This research resulted in advances that enable asynchronous chips to perform competitively with synchronous ICs—especially in recent years, when higher frequencies and larger chips began demanding complex clock-generation and distribution schemes. The clocking scheme that once simplified chip design has now become one of the most difficult aspects of design.

Asynchronous logic (also known as event-driven or self-timed logic) is essentially any circuit-design style that uses a sequencing mechanism other than a global clock. Sequencing a system without a global clock still requires a strict ordering of events as they happen within the system, but the availability of valid data from previous blocks triggers logical computation rather than a clock signal. Substantial research in the last several years has gone into optimizing the method of communicating data and timing information between logic blocks in an asynchronous system. One approach, known as dual-rail encoding, uses two wires for each data bit, so that



circuits transmit logic validity that is intertwined with the data value. An alternative approach is bundled data, in which each data bit uses one wire, and a separate wire contains the timing information for one or more data bits. Both styles require an acknowledge wire going backward for flow control.

A circuit that sequences data flow with dual-rail encoding using domino logic can use a pipelining mechanism to control when to compute and when to precharge the domino logic. In its simplest form, one circuit can pass a 0 or a 1 bit to an adjoining circuit by raising the appropriate 0 or 1 wire. The receiving circuit then raises the acknowledge wire to the sending circuit. Once the sending circuit receives the acknowledgment, it lowers its data wires, and the receiving circuit lowers its acknowledge wire (Figure 1). This method allows circuit computation and communication to be insensitive to delays; that is, computations proceed regardless of the delays in the transistors or the wires and without waiting for a clock or the worst possible delay path. In contrast, other asynchronous-design styles, such as bundled data, rely on predictable timing between the datapath and the control circuitry.

### **No idle power**

In a typical synchronous IC, the clock and latches account for 30 to 40% of the system power consumption. The continuous clock pulses to all logic blocks in the chip mean that these blocks are consuming energy whether or not they are doing useful work. Clock gating can minimize this consumption by stopping these wasted oscillations in inactive blocks. Although this situation is common in full-custom silicon, such as microprocessors, it is less common in ASICs because of the added design complexity.

Asynchronous designs have "perfect" clock gating down to the smallest circuit without additional design complexity, but it would be an overstatement to say that this clock gating results in power savings at all usage levels (Figure 2). The handshaking protocol in an asynchronous chip can draw nearly the same amount of power as the synchronous clock structure at peak usage. The transistors themselves do not necessarily offer an inherent power savings; when they are all in use, the amount of power they burn is comparable with the amount that synchronous circuits burn. The low-standby-power advantage becomes most apparent in designs in which a fraction of the circuitry is active—especially for processors, chip interconnects, and memories with a multibank structure.

Data gating is another power-saving technique that can help you avoid wasting datapath oscillations rather than clock oscillations. Such designs can use simple routing primitives for conditional communication and can avoid using glitching logic. For example, in a synchronous ALU, it is common to compute the sum, shift, XOR, and comparison results in parallel and then multiplex the desired output to the latch. This computation consumes power for all four operations, even though the processor discards three of the results. The normal asynchronous option would be to conditionally route the operands and results to and from these subunits, so that only the desired operation consumes any power.

### **Power-related noise and uncertainty**

Simultaneously switching signals in synchronous circuits creates noise problems because the signals unevenly draw current over a clock cycle from almost no power right before latching to peak power between clock edges.



This unevenness results in huge current surges, especially at the lower voltage supplies of modern CMOS processes. The inductance of the power-distribution wires results in power-supply ringing (ground bounce). Large on-chip decoupling capacitors buffer the power-supply fluctuation to acceptable levels to manage this ringing. Even in system-on-chip designs with multiple clock domains, the possibility exists for all the clock edges of different domains to align, thus causing the worst-case current spikes.

Asynchronous circuits more smoothly draw their current. Each stage of an asynchronous pipeline computes slightly out of phase with the previous one, thus uniformly spreading power consumption over time. Even with perfect clock gating an asynchronous pipeline cannot "instantly" go from no power consumption to peak consumption. Only during the latency cycle of the pipeline can an empty pipeline be filled, thus drawing its peak power. And if data stops arriving, another latency cycle of the pipeline is necessary for it to drain and return to no power. This situation results in current surges that are orders of magnitude smaller than synchronous circuits.

Modern synchronous chips also radiate microwaves at their clock frequency. These microwaves can interfere with other components in the same system or even with a TV across the street. To meet FCC standards, many systems are shielded with metal, which adds weight and cost. An asynchronous circuit lacks the coherent oscillation required to radiate, thus it creates less EMI noise.

Another challenge for synchronous designers is precisely specifying a chip's maximum power consumption. With synchronous chips (and bundled-data asynchronous chips), different data sequences can consume different amounts of power. For example, a synchronous bus transferring all zeros alternating with all ones dissipates much more power than one constantly sending zeros. You must carefully characterize the worst case to report the power bounds to system designers. A miscalculation may result in overheating and failure.

A dual-rail asynchronous design uses approximately the same power to process data regardless of its value, which results in simple and predictable power specifications for a given application. These tighter bounds on power variation let you lower the worst-case power for system design.

### **Scaling, leakage, and environment**

Voltage scaling is a popular method of battery conservation in mobile devices. Power is roughly proportional to the frequency times the switching capacitance times the supply voltage squared ( $P=fCV^2$ ). Scaling back the clock frequency results in a linear savings in power. Simultaneously scaling the voltage results in a cubic savings of power, because frequency is roughly linear in voltage. However, in a synchronous design, dynamically scaling the voltage requires the use of several clock frequencies, which can exacerbate race conditions and requires additional timing verification and design effort.

With delay-insensitive asynchronous circuits, you can adjust the supply voltage over a wide range without creating race conditions that could cause a chip to fail. Lab characterizations show that circuits usually work with a VDD of slightly more than the threshold voltage to well past the punch-through voltage of a particular fabrication process. (In TSMC's 180-nm process, the operating range is roughly 0.6 to 2.7V.) You can expect



vendors to deliver asynchronous chips characterized over a range of supply voltages, depending on the power or speed you desire.

You can also adjust the supply voltage on the fly, based on the load. For example, when a user types on a laptop or PDA, an asynchronous processor could ramp down to a very low voltage to conserve power. But if the user runs a computationally intensive application, the voltage would rise to increase speed at the expense of battery life. Although you can design synchronous chips to operate at several voltages, asynchronous chips more easily achieve this feature.

Leakage current is consuming more power in the new process technologies—by some estimates, rising by an order of magnitude with each new manufacturing node. Several process techniques, including dual-threshold transistors, can help with this leakage. Another new technique, biasing the substrate voltage, reduces leakage currents by raising the threshold voltage of the transistors. However, it also slows transistor operation.

Delay-insensitive asynchronous circuits operate correctly even if the device properties, such as the threshold voltage, change dramatically. As with voltage scaling, you can expect that asynchronous circuits will continue operating even if you adjust the body bias on the fly. This continued operation allows fine-grained dynamic control of both the supply voltage that determines active power and the substrate voltage that influences leakage power. A portable device could monitor its usage and decide how to set both voltages for the lowest power consumption.

Finally, speed and power are fundamentally exchangeable via voltage scaling. If a circuit has a 30% timing margin on its clock period, then it is running 30% faster than it needs to run. Synchronous circuits need this timing margin for reliable operation. Delay-insensitive asynchronous circuits do not. With adaptive control of the supply voltage, an asynchronous circuit can reduce its voltage and, hence, its power to the bare minimum to sustain the desired performance. Instead of failing from insufficient timing margin, it could nudge the voltage supply to catch up when it lags behind, thus sustaining nearly 0% timing margin.

The demand for lower power chips will accelerate, and the drive to improve the power consumption of the next-generation semiconductors will impact the functions and time to market of many of these chips. Asynchronous circuits feature lower overall power consumption by using power only when needed. They consume power more uniformly, reducing noise and electromagnetic radiation. Asynchronous circuits will reach their true potential when the circuits dynamically adjust the transistor properties to optimize power consumption at the desired level of performance. Adopting asynchronous-circuit design poses challenges, but solving power-related problems is an important reason to try it.