Dark Silicon

A method of conserving power in ICs by powering down segments of a chip when they are not in use.

DESCRIPTION

With advances in technology nodes and the ability to pack more and more transistors on the same die, design engineers are reaching a wall where only a fraction of a design can be powered on due to power and thermal implications.

Moreover, the challenges that force this kind of complex power management will only continue to grow at each new process node. Current estimates are that the dark silicon fraction will be about one-third of total area in the 20nm technology node (including 16/14nm finFETs), increasing to as much as 80% by the 5nm node. Real products are likely to achieve better results, but clearly power consumption imposes an increasingly severe design constraint.

For the last several decades, integrated circuit manufacturers have focused their efforts on Moore's Law, increasing transistor density at constant cost. For much of that time, Dennard's Law also held: As the dimensions of a device go down, so does power consumption. Smaller transistors ran faster, used less power, and cost less.

However, there was a limit. Smaller devices with thinner dielectrics and shorter channels are more prone to leakage. Indeed, leakage, negligible for much of the industry's history and ignored in Dennard's original paper, now approaches the same order of magnitude as the circuit's dynamic power. Advances such as the introduction of high dielectric constant gate dielectric materials helped, but leakage-limited transistor structures are now a fact of life. Switching a transistor at a lower threshold voltage requires a thinner gate dielectric, but leakage constraints place a lower bound on dielectric thickness. As a result, while feature sizes have continued to shrink, threshold voltage has not.

This failure of Dennard scaling has introduced the era of what designers call "dark silicon." If the number of transistors doubles, but the power budget for the circuit as a whole stays the same – or goes down, thanks to the proliferation of mobile devices – then the available power for each transistor is cut in half. If threshold voltage stays the same, then the number of transistors that can operate at one time is also cut in half. These non-operational transistors are dark silicon, measured as a fraction of the chip's total area.

While “dark” silicon is not “useless” or “wasted” silicon, there is ongoing debate about how much dark silicon can be eliminated through better design. In many of today’s designs, many circuit paths will be “dark” at any given moment.

Some elements, such as specialized logic and cache memory, are particularly “dark-silicon friendly,” in that they contribute to overall IC performance while consuming power only in special situations. In fact, this insight led to the integrated circuit industry’s current focus on multicore designs. If a problem can be broken into parallel components, then several cores running at a relatively low speed can still deliver better overall performance than a single core running at high speed. Many problems, and in particular many computation-intensive problems – digital photography, video rendering, database searching, etc. – are readily parallelizable. Moreover, the availability of parallel processing allows designers and software engineers to address larger problems with larger data sets in the same amount of time.