Design Challenges at 65nm and Beyond

Andrew B. Kahng, University of California at San Diego
abk@ucsd.edu http://vlsicad.ucsd.edu/

1. Introduction.

Semiconductor manufacturing technology faces ever-greater challenges of pitch, mobility, variability, leakage, and reliability. To enable cost-effective continuation of the semiconductor roadmap, there is greater need for *design technology* to provide "equivalent scaling", and for product-specific *design innovation* (multi-core architecture, software support, beyond-die integration, etc.) to provide "more than Moore" scaling. Design challenges along the road to 45nm include variability and power management, and leverage of design-manufacturing synergies. Potential solutions include "design for manufacturability" bridges between chip implementation and manufacturing know-how.

2. Variability and leakage.

Critical dimension (CD) control and process variations will challenge both manufacturing and design until the end of the CMOS roadmap. In mature 45nm products, local pattern and pitch dependencies in resist and etch processes will be ameliorated by restricted layout rules and improved dummy structure methodologies; these issues are nevertheless still problematic at 65nm. In the back end of the line (BEOL), metal interconnect performance exhibits greater variability in today's copper / low-k processes, even if "percentage-wise" control of the chemical-mechanical polishing (CMP) process remains the same. This is due to such factors as non-scaling of the barrier layers that prevent interaction of copper with surrounding materials: e.g., a constant 5nm barrier layer magnifies CMP-induced thickness variation, or etch-induced width variation, of a metal wire in terms of RC performance.

Leakage currents (subthreshold / gate / junction leakage, band-to-band tunneling, etc.) are arguably the industry's dominant concern at 65nm and 45nm. Scaling device performance in the absence of high-k materials requires thinner gate oxides, and leakage current per unit gate width rises by up to an order of magnitude per node. Leakage power is not only "wasted"; it also compromises achievable form factor, integration density, packaging choice, reliability, and other product metrics. Further, the impact of process variability on leakage is very costly: leakage scales inversely and exponentially with respect to the critical dimension of the transistor gate (i.e., channel length), and total leakage may vary by up to 5X to 20X across chips from the same wafer lot. Mitigation of leakage through multi-Vt, MTCMOS, or higher-level design techniques incurs area overhead and design process complexity, along with added variability (random dopant fluctuations and reduced supply voltage headroom make today's triple-Vt strategies less viable in future nodes).

3. Stress and reliability and more...

Today's scaling of on-current and device speed is based on stress engineering, e.g., through embedded SiGe. Stress due to shallow trench isolation (STI) can affect device on-current by up to 40%. In the front end of the line (FEOL), stress changes mobility and threshold voltage of transistors; in the BEOL, stress changes the integration and reliability of interconnects. Stress affects electron and hole mobilities with impact proportional to 1/LOD, where LOD (length of oxide definition) is the length of the active region. Change in mobility due to stress is also a function of transistor width and length. Stress also impacts threshold voltage due to enhancement and suppression of dopant diffusion, again with impact proportional to 1/LOD, and again a function of transistor width and length. By the late 45nm node, design tools and methodologies must actively modulate stress to improve timing (mobility change) as well as leakage (threshold voltage change). The semiconductor industry must holistically comprehend mitigation of *unpredictability*, whether for wearout (NBTI, TDDB, electromigration, etc.), parametric variation (line-edge roughness, random dopant fluctuation), or transient phenomena (particle strikes, supply noise). Academic research efforts, such as the ElastIC project at the University of Michigan, are developing new paradigms of self-diagnosis, adaptivity and self-healing that provide flexibility and resiliency at multiple levels of abstraction: system, architecture, and circuit. As such ideas reach production, the challenge for design and CAD is to minimize the cost of on-chip monitoring, redundancy, and reconfiguration structures.

Many other technical and business challenges cloud the future of CMOS beyond 45nm, ranging from next-generation lithography and consensus on radical layout

978-3-9810801-2-4/DATE07 © 2007 EDAA

restrictions, to software development for the highly concurrent, multi-core SOCs to which many application spaces have converged. With this in mind, we now discuss several solution elements that can help address these challenges.

4. Future scaling.

It is well-understood that Moore's Law trajectories of performance, density and cost depend on *equivalent scaling* via design technologies that reduce power or improve density without requiring any innovation in the process technology. Up to half of a process node of power, a third of a node of area, and one full node of performance can be gained – with the only question being whether the industry will recognize and invest sufficiently in this opportunity. *Design innovation* will be the workhorse for "more than Moore" scaling that goes beyond what underlying process and design technologies can achieve. We thus can envision a roadmap aligned to the continued delivery of semiconductor *value* – with that value arising from a combination of manufacturing technology, design technology, and design innovation. Finally, the balance of these contributors to scaling will be determined by a new understanding of *system scaling*. Dennard's scaling theory no longer holds, and neither does ITRS-scale scaling in terms of such parameters as A factors or FO4 delays or CV/I metrics. The future of scaling will be dominated by application-specific and product-specific *system constraints* on reliability, adaptivity, cost, reusability, software support, etc.

5. Clean abstractions.

With each new process generation, design rules have become more numerous, complex, and even conflicting. Today, designers face a Moore's Law corollary explosion of absolute, context-dependent and recommended rules that appear (usually without any explanations) in an ever-thicker design rule manual. Making detailed process statistics available to designers is not necessarily the right solution. (Imagine if a foundry had to sign up to the exact process statistics to which a design was optimized. Or, consider that process models may be obsolete before the design is completed – and design optimizations targeted to early models may actually be harmful in the matured process.) How should statistics of Vt variation due to random dopant fluctuation, or a chemical mechanical polishing (CMP) model, affect the way the chip designer performs synthesis, place and route? Designers have enough to worry about without having to become process experts as well. Separation of concerns between design and manufacturing is a fact of life even in IDMs, and is pivotal to the preservation of the fabless / foundry model.

6. Design for Manufacturability (DFM).

As we move into 65nm, parametric failures – i.e. chips that fail to meet power and timing requirements – become a dominant yield-limiting mechanism. Parametric yield loss continues to grow in significance at the 45nm node and beyond. In this context, there are many opportunities for DFM technology to bridge design and process, and to deliver high-value equivalent scaling advances. Three fundamental precepts – (1) drive design requirements into manufacturing, (2) bring manufacturing awareness into design, and (3) work within existing design environments without requiring major changes to design flow, design signoff, handoff to manufacturing, or fab equipment line – must be applied to afford end customers a true "design for value" (maximizing profit per wafer) capability.

Of particular interest is the notion of "electrical DFM", which focuses on objectives that the designer or product engineer cares about: leakage power, dynamic power, timing, timing and power variability, process window, and even reliability. The drivers for such optimizations consist of analysis engines that comprehend a full spectrum of physical and electrical implications of manufacturing. The "knobs" or degrees of freedom to achieve the optimization goals include changes to placement, wiring, vias – even the dimensions of individual transistors. Examples of how electrical DFM solutions take into account design-specific information include (1) iso-dense awareness of pitch-dependent through-focus CD variation, to reduce timing guardbands and improve timing robustness; (2) post-layout transistor gate-length biasing, specified at tapeout but realized in the foundry's OPC flow, to reduce leakage and leakage variability; (3) "self-compensating design" techniques that minimize the inherent sensitivity of critical paths to various sources of process variation (dopant density, oxide thickness, Leff, etc.); and (4) timing- and SI-driven CMP fill that maximizes both timing robustness and post-CMP wafer uniformity. All of these techniques are in production tools today.

At 65nm and beyond, chip designers' power / timing requirements will be used to tailor the manufacturing line for each individual transistor of each individual design – without any changes or adjustments to the fab equipment. Chip designers will be able to take advantage of available entitlement or process margin so that the process delivers significantly improved parametric quality of the silicon. Not only will designs be driven toward a sweet spot for the process, but the process can correspondingly be driven toward a sweet spot for the design. These are just some of the ways in which DFM offers a path by which design technology, after decades in the shadow of process technology, can emerge as a key enabler of continued value for the semiconductor and electronics ecosystem.