The Road Ahead

The Future of Signoff

Andrew B. Kahng
University of California, San Diego

"Signoff" is a near-magical word. It connotes a design state that is—with respect to the designer's functional constraints and the manufacturer's modeling requirements—ready for fabrication. Signoff spans timing, signal integrity, and power integrity analyses, and integrates margins for process, reliability (e.g., NBTI aging), voltage noise, and operating temperature. Design teams today complain bitterly about the rapidly increasing difficulty of signoff (or "design closure"). Moving to 40- or 28-nm process technology demands new methodologies that can scale productivity and quality in the face of hundreds of mode-corner combinations, complicated on-chip variation (OCV) deratings, growing best-worst margins in foundry models, new accounting for dynamic power supply noise effects, and more—not to mention traditional challenges ranging from bogus constraints to weak tools.

What is the future of signoff? Will it move up into the physical design flow, become statistical, revert to *k*-factor deratings, expand its scope, limp along on more licenses and larger server farms, or—? Here, I focus on electrical-functional (stemming from the foundry's Spice and extraction models), as opposed to geometric (stemming from the foundry's layout ground rules), aspects of signoff. I thank the organizers of the recent TAU 2011 Workshop for inviting a talk on this topic, from which this column is derived.

Where will signoff be embedded?

A key contributor to IC cost is design turnaround time. In today's physical design (PD) and signoff flow, designers often confront timing problems at the signoff stage that are unseen at the PD stage. The Achilles' heel of traditional "signoff after PD" is miscorrelation between the two stages' assumptions regarding, for instance, interconnect models, accuracy of delay calculation, or abstractions with respect to operating conditions. Even as new and more accurate analyses become available at signoff, PD tools use only limited information in their optimization, due to practical runtime considerations. Miscorrelations result in crucial unresolved timing problems remaining at the signoff stage, so designers must iterate PD and signoff stages until the timing converges. Further, as one or more humans form the interface between PD and signoff, errors are unavoidable. The number of PD-signoff iterations, as well as the time required per iteration, has been increasing for many technology generations with the explosion of design complexities. Future PD implementation tools must comprehend all information that is used at signoff, and smoothly span implementation, optimization, and signoff use models (to this end, scalable multicore engines may be enabling). Eventually, signoff must move inside the PD tools to prevent human errors at the implementation-signoff interface.

What should be modeled?

Model guardbands are used to guarantee parametric yield, considering both variability in process and operating conditions as well as the cost of managing such variability. Critical dimension (CD) variability is one of the most significant factors in process variation. For deeply subwavelength patterning (e.g., sub-80-nm local metal pitch using 193-nm-wavelength light at the 20-nm half-node), double-patterning lithography (DPL) is the current plan of record. DPL enables twice the pattern density by printing alternate patterns in each of two exposure steps. For example, "even" tracks might be printed with one exposure, and "odd" tracks with a second exposure. Although DPL solves the fundamental pitch limitation of the

0740-7475/11/$26.00 © 2011 IEEE Copublished by the IEEE CS and the IEEE CASS

traditional single-exposure lithography, it brings a new challenge of "bimodal" CD distribution: the CD of transistor gates have two uncorrelated distributions, corresponding to the two exposure steps used to pattern the transistors. Bimodality means that two instances of the same standard-cell master can have significantly different electrical characteristics depending on how layout is decomposed into the masks for the two exposures, and on the actual sequence (e.g., odds before evens, or evens before odds) of the two exposures.

Traditional signoff methodology (along with upstream Spice modeling, library characterization, and PD analysis engines) would handle a bimodal CD distribution as a unimodal distribution with larger sigma. However, the unimodal representation would incur unnecessary guardband and diminish the expected benefit from adoption of double patterning. Hence, future signoff must comprehend the precise decomposition and assignment of patterns to masks, so as to apply the correct (i.e., narrower) guardband of each transistor corresponding to the mask assignment. Two notes: first, extreme ultraviolet (EUV) lithography will not arrive in time to avoid this issue for logic design, and second, bimodality and pattern decomposition also impact interconnect delay analyses.

Limits of modeling

Two trends are of interest in modeling for signoff. The first trend is margin reduction through more accurate modeling of manufacturing and process. Consider three examples.

- A fixed CD variation is assumed in the Spice model corners, but this variation actually depends on product engineering and manufacturing choices (e.g., CD variation will decrease if a smaller mask field is used).
- Spice model corners are fixed for a given transistor gate length and width, but a number of proximity effects (reticle enhancement, litho, etch, stress, well proximity, etc.) can lead to very different electrical characteristics for identically drawn transistors.
- Random variation has a spatial correlation that is captured through OCV options in modern signoff timing tools. However, convenient tool choices such as use of minimum bounding boxes of timing paths instead of minimum enclosing circles can cost 1-2 ps of margin, which may one day be important enough to recapture.

When benefits sufficiently outweigh costs is a business decision. That said, bimodality may be just the first of several future margin reductions that bring "the manufacturing flow" into signoff.

The second trend is the increasingly varied and detailed accounting of scenarios of process, voltage, temperature, aging, and functional modes. This trend only accelerates with deployment of low-power techniques beyond voltage scaling and power gating—and places huge burdens on IP library management and signoff analyses. New approaches will be needed to stem the tide of corner libraries. For example, a new library scaling methodology such as instance-based, multidimensional derating tables may restore productivity in signoff.

From the designer side

The future of signoff will include enhancements from designers as well. Here again, two trends pertain. First, what is signed off will change. For example, better communication of design intent can lead to more effective representation and handling of timing constraints; this will avoid the hundreds of thousands of exceptions that ultimately do not change the signoff, yet disturb optimization quality and efficiency. Or, understanding a given block's usage (say, 2-year product lifetime, 5% duty cycle) can dramatically reduce the aging guardband needed in signoff.

Second, the meaning of signoff will also change. For example, it is well-recognized that traditional worst-case signoff can overly increase area and power dissipation. The emerging paradigm of resilient design allows optimization for the typical rather than the (rare) worst case, with variability-induced errors being tolerated through redundancy techniques such as Razor, avoided with performance-monitoring sensors, or even intentionally allowed if the design can operate approximately (e.g., in human sense-related applications). With such paradigms, we can even contemplate the synthesis of special monitoring circuits, such as "design dependent" ring oscillators, as part of the signoff.

And more . . .

No doubt there will also be a place for statistics in future signoff, and for dynamic voltage fluctuations and thermal gradients as well—when the benefit-cost ratio is compelling. For example, statistical static timing analysis (SSTA) has been used since the 45-nm node as an adjunct to standard corner-based

analyses, but the methodology is still applied mostly to limited numbers of paths after OCV-based signoff. There are two main obstacles to SSTA's becoming mainstream. First, the SoC design team would require complete statistical views of all IP blocks in the chip. Second, the net benefit of SSTA is kept in check by OCV add-ons to traditional STA, by improved identification and modeling of systematic variations (i.e., less variation is "random"), and by variation-reducing design practices such as regular layout.

Who will realize the future of signoff? From the scenarios I've depicted here, we see that responsibilities lie with tool developers, the foundry interface, and designers alike. Whether these constituencies will deliver remains to be seen—on the road ahead. ■

■ Direct questions and comments about this column to Andrew B. Kahng, University of California at San Diego, Dept. of Computer Science and Engineering, 9500 Gilman Dr., MC-0404, La Jolla, CA 92093-0404; abk@ucsd.edu.