A Roadmap and Vision for Physical Design

Andrew B. Kahng
UCSD CSE and ECE Depts., La Jolla, CA 92093-0114
abk@ucsd.edu, http://vlsicad.ucsd.edu/

ABSTRACT

This invited paper offers "roadmap and vision" for physical design. The main messages are as follows. (1) The high-level roadmap for physical design is static and well-known. (2) Basic problems remain untouched by fundamental research. (3) Academia should not overemphasize back-filling and formulation over innovation and optimization. (4) The physical design field must become more mature and efficient in how it prioritizes research directions and uses its human resources. (5) The scope of physical design must expand (up to package and system, down to manufacturing interfaces, out to novel implementation technologies, etc.), even as renewed focus is placed on basic optimization technology.

Categories and Subject Descriptors

B.7.2 [**Integrated Circuits**]: Design Aids – Layout, Place and Route; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Non-Numerical Algorithms and Problems – Routing and Layout; J.6 [**Computer-Aided Engineering**]: CAD

General Terms

Algorithms, Design

1. PREFACE

This is not a scientific paper - and it is more "review and commentary" than "roadmap and vision". References are not intended to be complete. Many passages are excerpted from material that I have written for the Design and System Drivers chapters of The International Technology Roadmap for Semiconductors (2001 Edition) [9], and several other papers. I am grateful to my coauthors, to the many individuals who contributed to the 2001 ITRS, and to various copyright-holding entities for their indulgence. This paper would not exist but for the kind invitation of Massoud Pedram and the ISPD02 program committee, as well as the infinite forbearance of Lisa Tolles-Efinger at Sheridan Printing.

2. WHAT WE NEED

This section gives PD-relevant excerpts from the 2001 ITRS [9]. It should be kept in mind that the ITRS is a statement of technology needs, made to semiconductor supplier industries (of which EDA - Design - is one) by the semiconductor industry. The excerpts here will draw most heavily from the "logical, physical, and circuit design" section of the 2001 ITRS Design chapter. The ITRS taxonomy, which treats physical design holistically along with logical and circuit design, gives some indication of the appropriate *scope* for PD. The *role* of PD is to serve the physical implementation flow, which serves the system designer: PD enables system-level signoff into reliable, predictable implementation fabrics.

Nanometer-scale silicon complexity makes it difficult to estimate and abstract the effects of physics and embedding on eventual design quality (timing, power, signal integrity, reliability, manufacturability, etc.). Hence, logical design and eventually system-level design become more closely linked with PD. The paradigm of hierarchical, top-down, layout-based implementation planning supported by a tightly integrated, incremental static (power, timing, noise) analysis "backplane" is now well-established. Chip implementation flows will vary according to methodology choices that juggle process/device abstraction, constraint manipulation, analyses and optimizations in the face of exploding complexities and emerging concerns such as error-tolerance, variability and cost.

The following precepts govern future (physical) design technology innovation. Precepts 3 through 6 are aspects of a (top-down, iteration-free, decomposition-oriented) "correct by construction" approach. Precepts 7 and 8 are more suited to "construct by correction", where iterations are expected but made less painful. Ultimately, these lead to the future design system architecture depicted in Figure 1.

1. Exploit reuse.
2. Evolve rapidly. Typical evolutions: (a) analyses evolve into verifications, which evolve into tests, and (b) analyses and simulations evolve into models and verifications, which evolve into either objectives or constraints for synthesis and optimization. A related trend is "bottom-up commoditization" (e.g., of characterization and RLC extraction, then delay calculation, static timing / noise analyses, standard-cell placement, global routing, ...).
3. Avoid iteration. Iteration between levels of design incurs translation and other interfacing costs, and hampers predictability and reliability of the design process.
4. Replace verification by prevention. Lower-level problems (e.g., crosstalk-induced delay uncertainty) are more cheaply addressed by higher-level prevention (e.g., repeater insertion and slew rate control rules).
5. Improve predictability. Constructive estimation does not afford productivity leverage; efficient search for good design so-

ISPD'02, April 7-10, 2002, San Diego, California, USA.

lutions requires prediction-based estimation. Better estimates enable design space exploration at higher levels.

6. Orthogonalize concerns (e.g., congestion from timing, timing from layout, computation from communication). Unrelated issues should remain separate whenever possible.

7. Expand scope. For example, we require greater integration of software and analog/mixed-signal/RF (AMSRF) design with digital flows; this must be supported by modeling, analysis and simulation at multiple levels of the system hierarchy (up to package and board, and down to mask and process).

8. Unify. Silicon complexity induces the unification of previously disparate areas, e.g., synthesis-analysis, logical-physical-timing, or even design-test. Unifications improve the downstream flow of intentions and assumptions, and the upstream flow of estimation/prediction models. Associated frameworks are successive approximation, and incremental optimization. Increasingly, unifications cross the die-package and design-manufacturing boundaries.

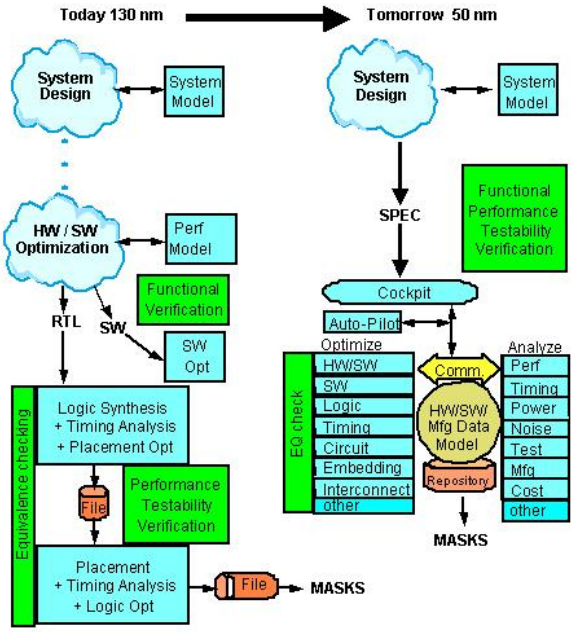

Figure 1: Evolution of design technology.

Specific challenges for logical, circuit and physical design center around (1) predictability and (2) improved capability to model, analyze and leverage nanometer-scale circuit phenomena. The remainder of this section flags some of the highlights.

2.1 Predictable Implementation

Scalable, incremental analyses and optimizations. PD tools operate at low levels of abstraction and face instance complexities that grow by at least 2X per technology node. Scalability will require new ways to manage data, traverse solution spaces, and map optimizations onto distributed/parallel computational resources. To enable construct-by-correction, we require incremental synthesis and analysis capabilities with runtimes proportional to the amount of change made to the input, and with no loss of solution quality. To enable reuse, we require new measures of solution quality (e.g., stability or robustness in new environments or when added into other optimization instances). In addition, chip implementation increasingly entails large-scale, interacting, multi-level and multi-objective global optimizations; tools will need to generate families of solutions that capture complex trade-offs among different objectives. Basic technology needs include techniques for constraint-dominated global optimization, computational resource-bounded optimization, and optimization with partial or probabilistic design information.

Unified planning, estimation and prediction. Today's implementation planning tools create logic and timing structure concurrently with constraint budgets and spatial embedding. The standard approach combines RT-level floorplanning and global interconnect planning to define repeater insertion and pipelining as well as detailed layout of global signals. The result (after appropriate hierarchy reconciliations) is passed as a constraint to logic synthesis and optimization, placement and routing. PD must enable concurrent system architecture optimizations that reduce global wires, and even interconnect architecture optimizations that match the number and dimensions of wiring layers to the given system. Evolutionary unifications will continue (layout-clock-test, detailed-routing driven performance closure, etc.) to enable the co-optimization depicted in Figure 1. Since top-down design depends on downstream predictability, we must more directly seek to define, understand and improve the concept of "predictability" across the entire scope of PD.

Synchronization and global signaling. Across-chip communication requires increasingly many clock cycles; microprocessor global interconnects are already pipelined and are not a limiting factor for clock frequency. However, clock distribution in purely synchronous designs can account for over 40% of dynamic power and is subject to increasing stress (e.g., shielding resource requirements, limits of edge rates and jitter, and parametric yield loss due to variability-induced skew). There is a clear trend to more robust and power-efficient hybridizations of synchronous and asynchronous design, both GALS (globally asynchronous, locally synchronous) and GSLA. On-chip clock frequencies will vary by two orders of magnitude or more, and reach 15+ GHz; multi-cycle interconnect paths will be commonplace. PD must deliver and support new synchronization paradigms for such regimes. PD must also acknowledge timing structures wherein "more paths are critical" in the aftermath of timing and power optimizations. The latter phenomenon is accompanied by quadratically increasing delay sensitivities to process variation; the need for greater margins ultimately limits the return on traditional delay slack optimization, so new performance tuning approaches will be required. On-chip global signaling follows off-chip signaling trends: it increasingly relies on low-swing differential and multi-phase clocking techniques, as well as digital PLL/DLL synchronization. Improved efficiency and signal reliability of traditional buffered global interconnect must be fortified by new syntheses and analyses for boosters, state- and transition-awareness, multilevel encoding, time-sharing of interconnect resources, and other emerging signaling paradigms.

Heterogeneity. Single-die integration of analog, mixed-signal and RF (AMSRF) with digital logic will challenge PD from planning through layout. Tools must handle greater sensitivities of, and interactions between, AMSRF circuits with respect to noise and interference, along with *constraint-dominated* formulations (e.g., matching, symmetry, and electrical rules in layout). Long-term, the integration of MEMS or other technologies will require novel fault avoidance and fault tolerance methods. On another front, system cost optimizations (manufacturing cost, and the cost of communication) must be supported by modeling of integration choices that span multi-die (system-in-package) and stacked-die options. Synthesis and analy-

sis of global signaling must extend to optimization of system-level interconnect from die through package and board.

2.2 The Design-Manufacturing Interface

Another aspect of predictability is *variability*. Atomic-scale effects increasingly govern the statistics of many process steps, and 3-sigma variations of such fundamental parameters as L_{gate}, t_{ox} and interconnect dimensions now exceed 15%. New circuit topologies and logic/layout optimizations are needed to cope with this level of variability.

Statistical timing analysis and performance verification must comprehend parasitics, delays and geometrics that are parameterized by distributions. Design centering must optimize for parametric yield and revenue per wafer, rather than traditional performance metrics. (For example, consider the difference between "L" and "staircase" layouts of a global line with a single repeater, when variabilities are spatially correlated across the die [8].) Manufacturing variability (and the proliferation of new materials and processes) also requires a more extensive design-manufacturing interface that supplies design rules and process abstractions to layout. Richer statistical and electrical/geometric characterization of manufacturing variability sources is needed. As power densities continue to rise, naive guardbanding against thermally induced variability will be costly. More accurate analyses and bounds for local thermal variation are needed to reduce the amount of overdesign.

Reticle enhancement technology (RET) encompasses planarization of multilayer interconnect processes (necessitating layout density control with area fill) and deep-subwavelength optical lithography (necessitating optical proximity corrections (OPC) and layout of alternating-aperture phase-shifting masks (PSM)). RET places a growing burden on physical design with respect to layout design complexity, manufacturing handoff complexity, and manufacturing (mask) NRE cost. With OPC and PSM, layout synthesis productivity is challenged by complex, context-dependent design rules. Layout verification must also handle regimes where "local design rules" no longer exist. Physical verification must accurately understand and model, e.g., the RLC extraction impact of downstream dummy metal insertion in the post-tapeout layout database.

Indiscriminate application of RET explodes data volumes and mask write/inspection costs. RET insertion (and mask inspection) must therefore understand that only certain critical device or interconnect dimensions are worth the expense of careful enforcement, and that some enforcement mechanisms are costlier to implement and verify than others. A data volume- and cost-sensitive PD flow will enable such selectivity by passing detailed functional intent, performance analysis results, and sensitivities forward throughout the layout, verification and mask flows.

2.3 Silicon Complexity and Non-Ideal Scaling

Lower supply voltages, along with larger currents stemming from increased power densities, lead to larger relative supply rail inductive noise. This is exacerbated by less aggressive scaling of bump counts and pitches. Issues such as IR drop and decoupling capacitance have been addressed in the recent PD literature. Near-term open issues include control of temperature variation across the die for package and performance reliability. This entails new tools spanning algorithm development, logic synthesis, and timing/layout optimization that must cooperate to manage both instantaneous and average power. The large estimated "power management gap" for microprocessor and SOC system drivers implies a continual focus on power reduction. At the PD level, library characterization, synthesis, and layout (including power distribution design) must together deliver the roughly 5X available power reduction from fine-grain use of multiple thresholds and supplies (and oxide thicknesses, and biasing) in the same core. Potentially, PD tools must automatically produce structures that enable active thermal management via OS-mediated dynamic frequency and supply scaling.

Reliability and fault-tolerance. Reliability criteria (hot-carrier effect, electromigration, joule self-heating, etc.) have been integrated into implementation flows via simple and transparent abstractions (e.g., upper bounds on gate load capacitance vs. output slew time). Such "methodological" abstractions currently permit correctness by construction with little disruption of traditional flows. However, improved abstractions and analyses that reduce guardbanding will be needed in the future. With respect to single-event upsets (SEU) caused by ionizing radiation, decreasing feature sizes lower Q_{crit} values to such levels that even the noise pulse from an alpha particle can be trapped as a logic fault. Automated methods are needed to modify logical, circuit and physical design (e.g., by automatic introduction of error correction, sizing, etc.) to prevent or manage SEU without violating design constraints.

2.4 Circuit Implementation

Non-ideal scaling impacts (notably from scaling supplies faster than thresholds) include higher gate and drain leakage currents, body effect (making pass gate logic less attractive), and loss of overdrive. In light of power management challenges, past tradeoffs of higher power and noise susceptibility (along with unavailability of automated tools) in return for speed become less attractive. As alternatives to static CMOS are deployed to permit overall speed/power performance gains, layout automation and physical verification (e.g., automated extraction of novel active and passive structures from layout) will be needed. For example, self-sufficient circuits such as clock-delayed domino or delayed-reset domino will become more popular with use of globally synchronous, locally asynchronous architectures. Circuit modeling must be consumable at ever-higher levels, as systems move to compiler- and OS-based control of such parameters as body bias, clock, and supply rails. For SOI, PD has already made progress toward necessary analyses (e.g., history-dependence of timing, coupling dependence of static power) and syntheses (e.g., planning of decoupling capacitance).

Analog synthesis. Scaling of SOC design productivity requires elimination of today's "analog design bottleneck". Existing specialized circuit syntheses for particular classes of circuits (PLL, op-amp, power amplifier, etc.) must be augmented by more general techniques as well as automatic layout syntheses. Analog syntheses must handle increased opportunities for distortion and nonlinearity due to impact ionization, thermal nonlinearity, body contacts acting as bandpass filters, etc. Syntheses must also handle future regimes of increased manufacturing variability, e.g., by hybrid analog-digital compensation for device mismatch. Isolation techniques must be flexibly applied. In the near term, new synthesis tools for optical interface circuits and high-Q tunable CMOS on-chip inductors/resonators are needed. Circuit types of interest in the long term include extremely low-power sensing and sensor interface circuits, as well as micro-optical (e.g., beam-steering) devices. Due to the difficulty of analog verification and testing, BIST circuitry will be increasingly synthesized around, and laid out with, analog circuitry such as high-speed networking interfaces. Such techniques must be nonintrusive, i.e., they cannot degrade performance of embedded analog blocks.

3. WHAT WE DO

According to Dataquest, EDA as a whole has less than 6,000 R&D engineers worldwide. Total tools revenue per designer has

Topic	1996	1997	1998	1999	2000	2001
1	4/4/7	8/10/6	4/5/7.5	5/5.5/3	4.5/6/5	3/4.5/8
2	0/4/3	1/0.5/2	3/2/4	3.5/1/5	4.5/4/4	0/5.5/3.5
3	2/1/3	2/0/1	2/1/3	1.5/2/1	0/0/2	2/0/0
4	1/1/2	1/1/2.5	0/5/2	2.5/2/3	2/0/5	2/0/4
5	4.5/3.5/2.5	0.5/4/3	0/2/2.5	4/4.5/2.5	2/3/2+1	0/0.5/4
6	0/0/1	0/0/1	2/0/1.5	1/0/0	1/0/1+1	0/1/0
7	3/2.5/0	1/1.5/1	0/0/0.5	0.5/0.5/3	1/0/0+1	0/1.5/0
8	0/1.5/2	0.5/2/1	1/0/1.5	1.5/0/0	0/2/4+2	0/1/1+2
9	2.5/0/2	7/0/0	0/6/0	1/1.5/0.5	0/0/0	0/0.5/0
10	0/0/0	0/0/1.5	0/4/0	2/0/1	0/0/0	0/2/0
11	1/2/0	1/0/2	0/0/1	0/0/1	3/1/1+3	0+5/0/1
12	0/0.5/1.5+2	2/0/0	0/0/0	0/0/0	1/2/0	0/0/1
13	0/0/0	0/0/1	1/3/3.5	1/0/4	0/0/0	0/0.5/1.5

Table 1: Distribution of physical design papers among 13 topics, for cycles of DAC(Y)/ICCAD(Y)/ISPD(Y+1) with Y = 1996, ..., 2001. Invited papers are indicated by (+). Data compiled by Bao Liu, Swamy Muddu and Puneet Gupta.

increased at 3.9% per year on average for the past decade [9, 15]. Pure research investment is low relative to other semiconductor supplier industries (my estimate is in the tens of millions of dollars per year).

In the overall history of EDA, physical design has played a very large and critical role. However, most of today's design technology "crises" – verification, embedded software and system-level design, enabling of power management at the architectural and operating-system levels, analog/mixed-signal synthesis and reuse, design for test – lie elsewhere. The ratio of design value over effort is *perceived* to decrease as the level of abstraction moves from behavior down to layout. Thus, PD is now just another piece (say, at most one-sixth, whether by market size or by headcount) of EDA and design technology as a whole. PD researchers, developers and funding are not plentiful resources. With this as backdrop, let us now consider the distribution of PD research activity relative to PD roadmap needs.

Certainly, there are critical design technology requirements that can and must be addressed by physical design research. Examples include analog layout synthesis and reuse; layout-BIST synergies for "deep-submicron fault models"; new paradigms for global signaling, synchronization and system-level interconnect; modeling and simulation; mitigation of increased process variability and non-recurring costs in the mask and foundry flows; and multi-(V_{dd}, V_t, t_{ox}, biasing) performance optimization. Yet, there is a clear mismatch between these needs and the core of PD research activity as defined by the technical program committees of ISPD and other major conferences.

Table 1 shows results of a cursory scan of DAC, ICCAD and ISPD program content to assess the relative prominence of 13 topic areas in physical design.[1] The 13 topic areas are: (1) placement / partitioning; (2) routing / global routing / wireplanning; (3) interconnect tree (buffered / Steiner / RAT / etc.) construction; (4) floorplanning / block packing / macro-cell placement; (5) performance optimization (sizing, etc.); (6) RTL-down methodology / flow; (7) clock; (8) power; (9) custom layout (transistor-level / migration / compaction); (10) analog; (11) manufacturability / yield; (12) logical-physical interactions; and (13) signal integrity.[2] The Table shows that most "classical" (1-5) topic areas still receive the bulk of the community's attention. Custom layout (9) is "classical" as well, but has not been a substantial part of the literature in recent years. ISPD and ICCAD seem to publish more papers on "classical" topics than DAC. Non-classical issues are typically visible only in isolated years.

Looking backward, 1997 and 1998 may have displayed greater shifts in attention, with several paper slots dedicated to emerging topics (methodology, signal integrity, custom layout, analog physical design automation, etc.). However, these topic areas quickly lost steam, and today analog design, clocking, custom layout, and manufacturability are all research holes that have become critical to industry. (In 2000 and 2001, the emerging topic appears to be (11) manufacturability and yield. But despite PDW96's special session on yield and ISPD99's invited tutorial on subwavelength lithography [14], ISPD01 invited three of its four papers on this topic, and DAC01 invited all five of its manufacturability papers.[3]) It is reasonable to conclude from the Table that the scope of PD has remained fairly stable, with certain semiconductor industry needs remaining under-resourced. In light of ITRS needs, PD researchers should expand the scope of physical design up to package and system, down to manufacturing interfaces, out to novel device/circuit implementation technologies, and back to analog and custom layout.

What Is Going On Here?

Table 1 is, by itself, not surprising. Consider the following explanations. (1) There are long lead times and latencies as research moves from problem formulation to solution to technology transfer to marketplace. (2) Critical research topics can have greater startup costs or other barriers to entry. (3) Any research field will tend to recreate itself in its own image and perpetuate its prevailing standards (e.g., via program committees and following of research threads). *These are inefficiencies which neither the PD field nor the semiconductor industry can afford.* The first two inefficiencies are highly pernicious, but can be addressed by a "coopetitive" mindset change as outlined in the next section. The third inefficiency merits a few comments in the remainder of this section.

- It's not as if we are failing to hit a moving target: the target is in the open and it's standing still.[4] More specifically, the physical design roadmap has been *static* for quite a few years. Key goals such as (1) convergent integration of logical, timing and spatial embedding, or (2) unification of incremental timing/SI closure with a performance analysis backplane, have been in the NTRS/ITRS since 1997 (and in Sematech's CHDS requirements as of 1995). Most of the methodology precepts and physical design requirements have been stated since the 1997 NTRS, and arguably since the 1994 NTRS. The modern methodology and routing contexts are quite close to what was described in [3, 10]. The 2001 ITRS litany of manufacturability challenges reiterates an ISPD99 presentation [14].

- We may be spending too much effort on "back-filling" of well-known issues and already-commercialized techniques into the public literature. In my past work, [7] is an example, albeit made with good intentions. Today's standardized planning framework (cf. [9, 10, 3]) has commercial implementations [26, 27, 28] that - in terms of infrastructure, functionality and *evaluatability* - are superior to what academia can likely manage. "Classic" multi-stage approaches to large-scale fixed-die

[1] Other analyses are also possible, e.g., of paper submission statistics or measures of "scientific health" (cf., e.g., [21, 20]).

[2] In recent years, papers on logical-physical interactions (12) have shifted to, and are tallied in, the more specific planning (2) and RTL-down methodology (6) topic areas. Papers on signal integrity (13) have for the most part shifted to the global and detailed routing (2) topic area. Each topic area makes its own advances.

[3] Guess who organized these sessions :-) .

[4] It's not as if I am saying anything new: Lou Scheffer has provided a similar roadmap and message 5-6 years ago [22, 23].

routing are also better implemented in industry. Melding hierarchical graph-based layout resource abstractions [24, 18, 25] with shape-based detailed routing [3] is now standard in industry, but may not be worth developing in academia. The value of "back-filling", aside from impact on IP, is unclear to me; I believe that more industry-strength shared infrastructures should have higher priority.

- Is the goal of research novelty in "formulation", or novelty in "optimization"? To be provocative, I will claim that PD is tending to focus more on new problem statements, while only transferring or reusing core optimization techniques.[5] However, representation and formulation are not the end goal (e.g., [10] noted a "packing obsession" (really, representation obsession) which has been worked-around in industry by hierarchical coarse placement). We have no shortage of problems; we have a shortage of optimization tools from which we can create solutions.

4. WHAT WE NEED TO DO

This section describes two "mindset changes". The first, "coopetition", addresses two inefficiencies mentioned in the previous section and opens the door to greater effectiveness and impact of the PD field. The second, "shared red bricks", opens the door to greater rewards for greater impact of the PD field in furthering the semiconductor roadmap.

4.1 Coopetition

Phrases such as "design productivity gap" (stemming from a 1994 Sematech presentation, and propagated through several editions of the NTRS/ITRS) and the existence of the MARCO Design and Test Focus Research Center (www.gigascale.org) reflect a perceived need to improve the effectiveness of CAD research. In 1999, the Design Automation Conference instituted a new topic area: "Fundamental CAD Algorithms". In the area of hypergraph bipartitioning, [5] noted that silent, undocumented implementation decisions could change result quality by more than 400% - and that contemporaneous papers reported more than 1,000% differences in solution costs returned by implementations of the same well-known algorithm. To help the CAD community address these problems, a new medium for CAD-IP reuse - thus far centered in the PD domain - has been established under the auspices of the MARCO Gigascale Silicon Research Center (http://www.gigascale.org/) [6]. Called the MARCO GSRC Bookshelf, it serves as a clearinghouse and a repository for intellectual property in CAD.[6]

Motivations for CAD-IP reuse include the following. (1) Better software development processes can improve productivity. Expected time to completion grows with project risk, which in turn grows with the amount of code that must be written from scratch. Cheaply reusing existing codes with documented performance and a history of successful reuse can reduce risk and shorten time to market even if new features or interfaces must be added. (2) Evaluation of result quality requires thorough understanding of algorithms and software tools, including common benchmarks, evaluation methodologies, and known-good performance results. Episodes in hypergraph partitioning cited in [5] show the risks when any aspect of a sophisticated technology's leading edge is unclear. Such risks increase as a research domain matures and the literature expands - in other words, as a problem gains importance and attention. When individual researchers can no longer keep track of all relevant research, there is a risk of poor reinvention of the wheel, and slower overall progress in the field. A partial remedy is to leverage community resources in verifying performance claims and implementations. (3) EDA research tends to focus on narrow optimizations, e.g., netlist hypergraph partitioning in physical design. If this narrowness is accompanied by the inability to evaluate new results in the context of full design flows, the utility of research decreases. (4) For mature, nondifferentiating technologies whose details do not affect competitiveness in the marketplace, reimplementation incurs tremendous waste. Resource imperatives eventually force data models, polygon database implementations, placers and routers, etc. to evolve into nondifferentiating, commoditized (but nearly free) "foundation IP" (following the course of operating systems, data structures, and GUI components before them). This evolution is consistent with the culture of "coopetition" (collaboration among competitors) exemplified by consortia such as Sematech.

Designers and researchers need an infrastructure that clearly identifies the best results in the CAD field at any given time. Such an infrastructure should reuse accumulated knowledge about standard benchmarks, evaluation methodologies, and performance comparisons. This goes beyond mere code reuse and entails more general reuse of CAD-IP. The GSRC Bookshelf to date includes 28 distinct "slots" (areas of CAD-IP) ranging from single interconnect tree synthesis to standard-cell placement to vertical benchmark designs. In effect, the Bookshelf is a new "publication medium" that offers a unique combination of emphases: algorithmic CAD IP, advancement of leading-edge design technology, free reuse, comparison and evaluation methodologies, and common data modeling. Ongoing work[7] under the auspices of the MARCO GSRC aims to build the Bookshelf's content and breadth, and to make the Bookshelf an integral part of the research process within the design technology community.

4.2 Shared Red Bricks

As noted in [11, 12, 13] the ITRS defines a "red brick" as a "technology requirement for which no known solution exists". Solving any given red brick is expensive, and requires large R&D investments. The ITRS is now full of red bricks, to the extent that these red bricks seem to form a "red brick wall" in the not too distant future.[8] My contention is that many red bricks stem from trying to continue old ways or old trends without seeking synergy with other parts of the semiconductor supply chain.[9] We need a more globally

[5] Fifteen or more years ago, PD was the source of simulated annealing, the LP relaxation and rounding framework, hierarchical routing [4], etc. Force-directed placement is a decades-old framework [19]. In recent years - and I again look to my own works for examples - I feel that PD has been more in the mode of transferring methods (e.g., multilevel in [1] and then [2, 16], various numerical methods, etc.). Fundamental new optimization approaches have not appeared in recent years, but seem overdue in light of new computational platforms, new instance scales, and new objectives.

[6] The Bookshelf is one of three initiatives - along with technology extrapolation (a "living roadmap") and measurement infrastructure to support design process improvement ("metrics") - in the GSRC's "Calibrating Achievable Design" ("C.A.D.") research theme (http://vlsicad.ucsd.edu/GSRC/) that together seek to improve design technology productivity.

[7] E.g., building new linkages among Bookshelf slots, common data models, and industrial data interchange formats to enable vertical benchmarking and more complete evaluation of algorithm innovations.

[8] In the 1999 ITRS, the red brick wall with respect to overlay accuracy, CD control, equivalent oxide thickness, and ILD permittivity appeared to be between six and nine years away. In the 2001 ITRS, these red bricks appear to be between four and six years away.

[9] The following metaphor may help to clarify this point. Think of the ITRS - the semiconductor industry's technology foundations -

optimized allocation of R&D investments, i.e., "shared red bricks".

Three examples of potential shared red bricks are as follows. (1) Must lithography, front-end processes, and interconnect technologies continue to push for 10% tolerances in critical dimensions? This would mean gate length and oxide thickness tolerances in the range of a single atomic monolayer by the end of the roadmap. Or are there design-for-variability solutions that share the red brick of variability between Design and these other industries? The first generation of variability-aware analysis tools is available now. However, variability-aware synthesis tools (centering for robustness under variability, or for maximum $/wafer) are a long ways off. Appropriate (and standardized) characterizations of variability sources in manufacturing equipment and processes also appear to be a long ways off. Circuit and layout techniques for high-variability regimes must also be explored. (2) Should the industry build new, faster mask writers that can handle 250 Gbytes of data for a single mask layer, after optical proximity correction (OPC) and fracturing? Or, should the industry reduce data volumes and relax inspection tolerances - thus improving mask throughput, yield, and cost - by exploiting design hierarchy and an awareness of which features are functionally critical? (3) Do we really need dielectric permittivities below 2.0 or copper interconnect resistivities below 1.8 mΩ-cm, as specified in the ITRS? Is the latter even possible? Would developing better circuit and interconnect architectures, and better layout techniques (cf. the X Initiative [29]), more cost-effectively share performance and noise management red bricks between Design and Interconnect technologies?

These examples highlight the potential of deeper partnerships between design technology and other ITRS technology industries. The key observation is that PD is the interface to circuits, devices, materials, and packaging; PD is therefore at the heart of nearly all such potential partnerships. Since such partnerships can potentially resolve key red bricks at greatly reduced cost to the semiconductor industry, it would be natural for PD to pursue not only shared solutions but also a share of the R&D investment reward.

5. WHAT WE NEED TO DO, II

This concluding section of "key directions" will need to be short. (0) Sensible unifications (sub-flows) that truly co-optimize global signaling, manufacturability enhancement, and clock/test/power distribution. (1) Fundamental new combinatorial optimization technologies, and possibly geometry engines, for future constraint-dominated layout regimes. (2) New decomposition schemes for physical design. (3) Global routing that is truly path-timing aware, truly combinatorial, and able to invoke "atomistic" interconnect synthesis. (4) In-context layout synthesis that maximizes process window while meeting electrical (functional) spec, and is able to handle restricted geometries (forbidden widths/pitches, on-grid constraints [17], halation rules). (5) Efficient analog and mixed-signal layout synthesis. (6) Methods for synchronization and global signaling at multi-GHz or -Gbps, extending to system-level integration. (7) New analysis, modeling and simulation methods that are tied more closely to PD syntheses, and that adapt to resource and accuracy / fidelity constraints. (8) Revival of platform-specific (parallel, distributed, hardware-accelerated) algorithm implementations. (9) Mindset changes as described in Section 4, and a new culture of "duplicating, deconstructing and debunking" [30].

as a car. The supplier industries (packaging, lithography, design, etc.) are the parts of the car. The car must continue to be driven along the Moore's Law road, e.g., if the car goes 150mph today then four years from now we require the car to reach speeds of 600mph. It is absurd to think that super tires alone, or super seats alone, will get us to 600mph. However, the seat industry might specify its requirements, and the concomitant levels of R&D investment, from the perspective that super seats alone must enable the 600mph car! It is economically wasteful and technologically impossible for each supplier industry to attempt to continue Moore's Law all by itself.

6. REFERENCES

[1] C. J. Alpert, L. Hagen and A. B. Kahng, "A Hybrid Multilevel/Genetic Approach for Circuit Partitioning", *Proc. ACM SIGDA Physical Design Workshop*, April 1996, pp. 100-105.

[2] C. J. Alpert, J. H. Huang and A. B. Kahng, "Multilevel Circuit Partitioning", *Proc. ACM/IEEE Design Automation Conf.*, 1997, pp. 530-533.

[3] R. Brashears and A. B. Kahng, "Advanced Routing for Deep Submicron technologies", *Computer Design*, May 1997. http://www.computer-design.com/Editorial/1997/05/supplement/597suprouting.html

[4] M. Burstein and R. Pelavin, "Hierarchical Wire Routing", *IEEE Trans. CAD* 2(4) (1983), pp. 223-34.

[5] A. E. Caldwell, A. B. Kahng, A. A. Kennings and I. L. Markov, "Hypergraph Partitioning for VLSI CAD: Methodology for Reporting, and New Results", *Proc. ACM/IEEE Design Automation Conf.*, June 1999, pp. 349-354.

[6] A. E. Caldwell, A. B. Kahng and I. L. Markov, "Toward CAD-IP Reuse: A Web Bookshelf of Fundamental CAD Algorithms", *IEEE Design and Test*, May 2002.

[7] A. E. Caldwell, A. B. Kahng and I. L. Markov, "Can Recursive Bisection Alone Produce Routable Placements?", *Proc. ACM/IEEE Design Automation Conf.*, 2000, pp. 477-482. (Cf. the "quadratic placement revisited" work as well.)

[8] Y. Cao, P. Gupta, A. B. Kahng, D. Sylvester and J. Yang, "Toward a Framework for Assessing the Impact of Variability in Nanometer Design", *manuscript*, 2001.

[9] *The International Technology Roadmap for Semiconductors*, 2001 edition, International Sematech, Austin, Texas, December 2001. http://public.itrs.net/

[10] A. B. Kahng, "Classical Floorplanning Harmful?", *Proc. Intl. Symp. on Physical Design*, April 2000, pp. 207-213. See talk slides at http://vlsicad.ucsd.edu/papers/slides/ispd00-cfh.ppt

[11] A. B. Kahng, "Design-Process Integration and Shared Red Bricks", *Proc. Design and Process Integration for Microelectronic Manufacturing*, SPIE vol. 4692, March 2002.

[12] A. B. Kahng, "The Road Ahead" (column), *IEEE Design and Test*, 2002.

[13] A. B. Kahng, *various talks*, http://vlsicad.ucsd.edu/~abk/TALKS/ .

[14] A. B. Kahng and Y. C. Pati, "Subwavelength Optical Lithography: Challenges and Impact on Physical Design", *Proc. Intl. Symp. on Physical Design*, April 1999, pp. 112-119. See http://vlsicad.ucsd.edu/ISPD99TUTORIAL/ispdtutorial.ppt (similar presentation material is in http://vlsicad.ucsd.edu/ISQED00TUTORIAL/isqed2000tutorial-abk.ppt).

[15] A. B. Kahng and G. Smith, "A New Design Cost Model for the 2001 ITRS", *Proc. ISQED*, March 2002.

[16] G. Karypis, R. Aggarwal, V. Kumar and S. Shekhar, "Multilevel Hypergraph Partitioning: Application in VLSI Domain", *Proc. ACM/IEEE Design Automation Conference*, 1997, pp. 526-529.

[17] M. D. Levenson, J. S. Petersen, D. G. Gerold and C. A. Mack, "Phase Phirst! An Improved Strong-PSM Paradigm", *Proc. 20th Annual BACUS Symposium on Photomask Technology*, Monterey, CA, USA, 13-15 Sept. 2000, SPIE vol. 4186, pp. 395-404.

[18] R. Nair, "A Simple Yet Effective Technique for Global Wiring", *IEEE Trans. on CAD* 6(6) (1987), pp. 165-172.

[19] N. R. Quinn and M. A. Breuer, "A Force-Directed Component Placement Procedure for Printed Circuit Boards", *IEEE Trans. on Circuits and Systems* 26(6) (1979), pp. 377-88.

[20] S. Savage, "On the Biodiversity of SIGCOMM", *2000 ACM SIGCOMM Outrageous Opinions*, Stockholm, Sweden, August 2000.

[21] S. Savage, "On the Caching and Prefetching of Program Committees", *1999 ACM SIGCOMM Outrageous Opinions*, Cambridge, MA, September 1999.

[22] L. Scheffer, "We're Solving the Wrong Problems", *Proc. 6th Physical Design Workshop*, April 1996, Reston, Virginia, pp. 89-91.

[23] L. Scheffer, "A Roadmap of CAD Tool Changes for Sub-Micron Interconnect Problems", *Proc. Intl. Symp. on Physical Design*, 1997, pp. 104-109.

[24] B. S. Ting and B. N. Tien, "Routing Techniques for Gate Array", *IEEE Trans. on CAD* 2(4) (1983), pp. 301-312.

[25] P.-S. Tseng and C. Sequin, "Codar: A Congestion-Directed General Area Router", *Proc. ICCAD*, 1988, pp. 30-33.

[26] http://www.chipcenter.com/eda/dac_files/ppt/DAC99-Avant3.ppt

[27] http://www.eetimes.com/dac98/news_tera.html

[28] http://www.synopsys.com/products/layout/floorplan_cs.html (last modified October 13, 1997).

[29] http://www.xinitiative.org

[30] Workshop on Duplicating, Deconstructing, and Debunking (held in association with the Intl. Symp. on Computer Architecture), http://www.ece.wisc.edu/~wddd .