Measuring Progress and Value of IC Implementation Technology (invited paper)

Andrew B. Kahng^{†‡}, Hyein Lee[‡] and Jiajia Li[‡]

[†]CSE and [‡]ECE Departments, UC San Diego, La Jolla, CA, USA

{abk, hyeinlee, jil150}@ucsd.edu

ABSTRACT

Over the past decade, "Moore's Law" has become increasingly well-understood as being a law of "value scaling": success of new electronics- and semiconductor-based products depends on improved cost-efficiency, utility, and value. Design Automation (DA) provides fundamental tools and methodologies that glue together disparate technological advances - across architectures, circuits, process and integration - into actuallyrealized product and system benefits. However, it is still unknown how to measure and credit "progress" of DA with respect to realized product- and system-level benefits. Thus, it is also challenging for industry, academia and funding entities to envision, and define, R&D objectives and prioritizations. In this paper, we contend that "measuring progress and value" is tractable at the level of EDA technology and R&D efforts. We describe four example assessments of progress and value of IC implementation technology: (i) assessment of progress of EDA tools (e.g., for P&R and STA) across multiple releases; (ii) establishing "upper bounds" on future progress (e.g., for 3DIC layout); (iii) robust rank-ordering of alternative design enablements (e.g., including routing tools and interconnect stack options); and (iv) lowering barriers to measuring progress and value of academic research results in "more real-world" contexts.

1. MEASURING PROGRESS AND VALUE

Systems and system design are rapidly evolving across many levels, including (i) emerging cloud, biomedical, autonomous cyberphysical and other system applications; (ii) emerging nonvon Neumann, quantum, nano-crossbar based, neuromorphic and other architectures; (iii) emerging beyond-CMOS next switches, storage elements, and interconnects; and (iv) emerging 3D/heterogeneous integration paradigms. Within this context, Design Automation (DA) technology glues progress made at each of these levels into actually-realized system and product benefits. An ability to measure the progress and value of DA with respect to holistic, productlevel benefits would enable industry, academia and funding entities to define and prioritize R&D objectives for the field. Even more, such assessments can potentially help clarify the appropriate credit and value accorded to EDA technology.

1.1 Design Cost and Low-Power Roadmaps

The ITRS (International Technology Roadmap for Semiconductors) *Design Cost Model* [32] [45] is a prominent, longstanding (2001-2013) effort to quantify both progress and value of IC implementation technology. The Design Cost Model measures progress of design technology according to how well the *cost of design* is kept under control. Design productivity, expressed as the number of transistors designed per engineermonth, is central to the model.¹ According to the model,

ICCAD'16, November 07-10, 2016, Austin, TX, USA.

Copyright 2016 ACM 978-1-4503-4466-1/16/11\$15.00.

specific design technologies (RTL methodology, silicon virtual prototyping, asymmetric multiprocessing, electronic systemlevel design automation, etc.) are each associated with forecast or calibrated productivity improvements – for both hardware and software design – when introduced. Thus, as long as the anticipated design technology innovations are delivered on time to the industry, design productivity will scale sufficiently to manage design costs.

Figure 1 shows the 2013 ITRS design cost projection for a consumer portable system-on-chip (SOC-CP) product, a "system driver" roadmapped in the ITRS System Drivers Chapter [46]. The sum of hardware design cost (blue region) and software design cost (red region) is the total design cost. The Design Cost Model shows how design productivity improvements mitigate the Moore's-Law increase in transistor count of the SOC-CP driver. Furthermore, as recounted in [45] and [32], had design technology innovations after 2000 not occurred, the total SOC-CP design cost would have been at \$1B in 2013, reaching \$70B in 2028. Or, in the absence of design technology innovations after 2013, the total SOC-CP design cost would grow from \$45.4M in 2013 to \$3.4B in 2028. In this way, the ITRS Design Cost Model captures both progress and value of design technology innovation.

Power and energy are well-understood as the ultimate "Grand Challenge" for the semiconductor industry. The 2011 ITRS Design Chapter [44] provides a roadmap of low-power design technology innovations before and after 2011. Examples include low-power physical libraries, adaptive body biasing (ABB), power gating, and dynamic voltage/frequency scaling (DVFS). Static and dynamic power improvement factors for each innovation (e.g., ABB improves static power by $2 \times$ and dynamic power by $1.2\times$, DVFS improves static power by $1.5\times$, etc.) are presented in a similar manner as the design productivity improvements in the Design Cost Model. Figure 2 reproduces the list of low-power design technology improvements, and their impacts on static and dynamic power, from [44].² Progress and value (reduced costs of packaging and cooling, reduced wearout, etc.) are implicit in the low-power design roadmap.

Design Technology Improvement	Vear	Improvements			
Design reennology improvement	1007	Dynamic	Static		
Low Power Physical Libraries		1.50	1.50		
Back Biasing	1	1.00	1.35		
Adaptive Body Biasing (ABB)	1	1.20	2.00		
Power Gating	1	0.90	10.00		
Dynamic Voltage/Frequency Scaling (DVFS)	D.6 2011	1.50	1.00		
Multilevel Cache Architecture	Before 2011	1.00	1.20		
Hardware Multithreading	1	1.00	1.30		
Hardware Virtualization	1	1.00	1.20		
Superscalar Architecture	1	1.00	2.00		
Symmetric Multiple Processing (SMP)	1	1.50	1.00		
Software Virtual Prototype	2011	1.23	1.20		
Frequency Islands	2013	1.26	1.00		
Near-Threshold Computing	2015	1.23	0.80		
Hardware/Software Co-Partitioning	2017	1.18	1.00		
Heterogeneous Parallel Processing (AMP)	2019	1.18	1.00		
Many Core Software Development Tools	2021	1.20	1.00		
Power-Aware Software	2023	1.21	1.00		
Asynchronous Design	2025	1.21	1.00		
Total		4 66	0.96		

Figure 2: Low-power design technology improvements and respective impacts on static and dynamic power, per [44].

 $^{^1\}mathrm{Engineer}$ salaries, server and tool license costs, costs of hardware design and software design, etc. are all elements of the Design Cost Model.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

http://dx.doi.org/10.1145/2966986.2980069.

²The "Total" improvement factors are for all low-power design technology improvements from 2011 onward.



Figure 1: Hardware and software design cost projections in the 2013 ITRS Design Cost Model [45], for the consumer portable system-on-chip (SOC-CP) "system driver" product [46].

The value and progress of design technology and EDA tools may also be seen at a "macro level". Kahng et al. [22] observe that industry revenues and valuations are, at some level, measures of DA research impact and technological progress. Figure 3 shows EDA and semiconductor industry revenues over the past two decades: EDA has been stable at just over 2% of semiconductor billings.³



Figure 3: Annual semiconductor industry revenues (blue bars) [43], and EDA industry total and per-segment revenues [42], reproduced from [22].

1.2 Calls to Action

Measuring the progress of IC implementation technology can enable definition of specific targets for future progress – i.e., "calls to action". For example, even as the semiconductor industry has continued to "race toward the end of the (CMOS, Moore's Law) roadmap", the ability of designers to leverage the "available scaling" afforded by process and device has become notably weaker in the past decade.



Figure 4: Gap between "available" density scaling (gray arrow) and "actual density" scaling in MPU products (red squares), adapted from [16].

As depicted in Figure 4, in advanced technology nodes a significant "design capability gap" [46] exists between the "realized" and "available" benefits of technology scaling. That is, since 2007 the "realized" density scaling in leading-edge products has slowed down to $1.6 \times$ per node, in contrast to the "available" $2 \times$ per node density scaling [15]. One way to compensate such a gap is *design-based equivalent scaling* (DES), as described in [5]. The 2013 edition of the ITRS Design Chapter identifies the need for DES to take on more of the burden of Moore's-Law value scaling; it projects that for server and desktop processors (MPU), DES will need recover one entire node of Moore's-Law scaling from 2013 to 2019, and that for processors in SOCs, DES must recover one node of scaling from 2013 to 2020 [16].

Another call to action is seen in the recent DARPA "Circuit Realization at Faster Timescales (CRAFT)" program [39], which aims to reduce the design time of complex mixed-signal SOCs in sub-20nm technologies from years to months without any loss of "performance at power" (PAP) metric. Figure 5, adapted from [14] (see commentary at [28]), illustrates three hypothesized steps toward such a design time reduction, from 130 weeks to 30 weeks. Given that Moore's Law classically corresponds to "one week = one percent", the targeted 100-week reduction of design time (iso-PAP) is an enormously valuable "moonshot". Toward this goal, a requisite design technology capability noted by [14] is the machine learningbased prediction of tool outcomes and identification of sweet spots for various tools and flows. Such a "big data" approach would enable setting of design-specific tool and flow knobs with maximum predictability of outcomes and minimum guardbanding of design requirements. The hypothesized end result is a fully predictive, one-pass flow with "optimal" tool usage for a given (datacenter, licenses) enablement.



#1. Bespoke, design-specific flow (predictive one-pass flow; optimal tool usage)

Figure 5: Hypothesized reduction of mixed-signal SOC design time from 130 weeks to 30 weeks without loss of "performance at power" metric, in sub-20nm technology nodes [39][14].

Many other "calls to action" vis-a-vis the progress of IC implementation technology can be inferred from various efforts over the years, e.g., as reviewed in [26][20]. For example, the 2009 and 2010 EDA Roadmap Workshops [40][41] noted a number of challenges (with required progress) for EDA. These include (i) shared development efforts (e.g., common design and library library environments), (ii) higher design productivity, (ii) better power management, (iii) design for manufacturing, (iv) more efficient system-level validation, (v)

³This begs the question of whether EDA revenues and valuations are "fair" with respect to the value that design technology provides to the semiconductor and electronics industries. As a point of reference, the sum of "big 3" EDA market caps has risen from \$8.35B 10 years ago to \$18.87B today $(2.26\times)$. The semiconductor SOX index has risen from 454 to 797 $(1.76\times)$ in the same period.

capability to optimize system designs with high complexity and heterogeneity, (vi) adaptivity to new computing (e.g., parallel computing, cloud), and (vii) earlier access to leading-edge process technology.

1.3 This Paper

The above approaches to assessment of DA technology progress and value are only the tip of the iceberg: more holistic, system- and product-oriented metrics of design technology impacts are yet to be developed. In the following, we focus on lower-level, "tactical" methodologies with which the DA and design communities may assess progress while building up to higher-level assessments.⁴ We present four concrete examples to measure or bound progress and value of IC implementation (i) direct, "vertical" tracking of EDA tools' technology: performance across versions; (ii) establishing an upper bound on the value afforded by future technologies (e.g., for 3DIC layout); (iii) a framework for systematic, "universal" rankordering of design enablement (e.g., back-end-of-line (BEOL) stack options); and (iv) means of measuring value and progress of academic tools in more realistic ways. In this paper, we do not achieve a direct metric of value and progress at the system or product level of impact. Yet, while our measures of progress focus on lower levels of the implementation stack, the power, performance and area improvements that we track at lower levels also have significant impacts on system-level metrics such as battery life or cost.

2. ASSESSING EDA TOOL PROGRESS

We now discuss the feasibility of, and potential learnings from, "longitudinal" studies of EDA tool performance across multiple versions. A given EDA company will conduct internal evaluations of its tools, and issue marketing claims for every new release (typically as compared to the immediately previous release). However, we are not aware of any long-term studies of important tool attributes (QOR, capacity, TAT, accuracy, violations, etc.) over time. We believe that such studies can help project future requirements for EDA tools, and guide how tools evolve to handle new technologies or functional requirements.⁵ Furthermore, the value of EDA technology innovation can be seen from such studies. For example, at the P&R level, tool QOR improvements might account for a significant fraction of node-to-node design implementation QOR gains. Or, analysis correlation (accuracy) improvements might account for significant analysis guardband reductions, which in turn lead to QOR gains [13].

2.1 A P&R Tool

We study versions v10, v12, v14 and v15 of a leading P&R tool, released between 2010 and 2015. We synthesize three designs (AES, VGA and LEON3MP) from OpenCores [48] using Synopsys Design Compiler K-2015.06-SP4 [49] with three foundry libraries (28LP dual-Vt, 8-track; 45GS triple-Vt, 9-track; and 65GP triple-Vt, 9-track). The P&R flow uses Multi-Corner Multi-Mode (MCMM) setup with SS and FF corner libraries, and applies power planning, clock tree synthesis (CTS), and optimization steps after each of the placement, CTS and routing stages. We also perform hold time optimization. Timing and power results are measured by Cadence Tempus V15.2 [38]. Crosstalk-aware analysis and optimization are not included in the P&R / signoff flow that we discuss here.

Figure 6 shows v10, v12, v14 and v15 results for different designs in 45GS. All numbers are normalized to v15 results.⁶ In Figure 6(a), total wirelength (NormWL) values of v10, v12 and v14 are respectively larger by 13%, 11% and 10% on

⁶Here, we present only examples of metrics obtainable from "longitudinal" studies. Our presented data and discussion are for small testcases and are by no means comprehensive.



Figure 6: Results of v10, v12 and v14 for AES, VGA and LEON3MP in 45GS, normalized to results of v15. (a) Normalized total wirelength (NormWL); (b) Δ final utilization (Δ FinalUtil); (c) normalized total power (NormPower); (d) Δ number of design rule violations (DRVs) (Δ #DRV); (e) Δ setup slack (ns) (Δ SetupSlack); and (f) Δ hold slack (ns) (Δ HoldSlack). In (a) and (c), the results of v10, v12 and v14 are divided by the results of v15. In (b), (d), (e) and (f), the results of v15 are subtracted from, respectively, the results of v10, v12 and v14.

average when compared to v15 values. Δ FinalUtil values are shown in Figure 6(b). In this experiment, initial utilization is fixed; hence, a higher final utilization indicates that the tool ends up with more buffers inserted (and/or, larger gate sizes on average) to meet timing constraints. We observe that Δ FinalUtil values do not vary much across different tool Average Δ FinalUtil values of v10, v12 and v14 versions. are 3%, 2% and 1%, respectively. In terms of total power, v10, v12 and v14 respectively show 9%, 6% and 3% larger NormPower than v15 (Figure 6(c)). For the number of DRVs, v10 shows fewer DRVs, while v12 and v14 show 30-120 more DRVs, compared to v15 (Figure 6(d)). Earlier versions achieve "better" Δ SetupSlack and Δ HoldSlack (Figures 6(e) and (f)). However, considering that all final designs have positive setup and hold slack values, this could indicate that later tool versions exploit positive slacks more efficiently for other optimizations (e.g., power and area recovery). Internal timing engine accuracy and/or correlation with Cadence Tempus (alternatively, Tempus evolution) may also be a factor in the trajectory of Δ slack values.



Figure 7: (a) Runtime and (b) memory usage for different tool versions (normalized to v15) for AES, VGA and LEON3MP designs in 45GS.

Figure 7 shows runtime and memory usage of v10, v12, v14 and v15 for the three designs in 45GS. On average, the runtimes of v10, v12 and v14 are larger by 31%, 90% and 3% compared to v15, respectively. Earlier versions consume less memory (-40%, -12% and -12% for v10, v12 and v14, respectively).

Figure 8 shows v10, v12, v14 and v15 results for the VGA testcase in 28LP, 45GS and 65GP. We see that v10 and v12 are not able to close the design in 28LP; Figure 8(d) shows > 5000 DRVs for v10 and v12 (we cap y-axis values at 5000 for visualization purposes). The change in

⁴We give a "personalized" perspective that samples from our group's recent efforts. Many other works have without doubt also addressed the fundamental question of "measuring progress".

⁵We study leading EDA tools as identified by such sources as [36]. We do not perform any benchmarking of the tools that we study. No value judgment is intended by, or to be inferred from, our discussion here.



Figure 8: Results of v10, v12 and v14 for the VGA testcase in 28LP, 45GS and 65GP technologies, normalized to results of v15. (a) NormWL; (b) Δ FinalUtil; (c) NormPower; (d) Δ #DRV; (e) Δ SetupSlack (ns); and (f) Δ HoldSlack (ns). In (a) and (c), the results of v10, v12 and v14 are divided by the results of v15. In (b), (d), (e) and (f), the results of v15 are subtracted from, respectively, the results of v10, v12 and v14.



Figure 9: (a) Runtime and (b) memory usage for different tool versions (normalized to v15) for the VGA testcase in 28LP, 45GS and 65GP.

tool behavior with different technologies may indicate, e.g., improved understanding of 28nm pin access and local routing congestion as tools mature beyond the 28nm node introduction.

Figure 9 shows runtime and memory usage of v10, v12, v14 and v15 for VGA in 28LP, 45GS and 65GP. In 28LP, the runtime and memory usage for v12 are higher than for the other tool versions; in 45GS and 65GP, the runtime and memory usage values are comparable across all tool versions.

2.2 A Signoff STA Tool

We have similarly made longitudinal studies of a leading signoff STA tool. We study versions v9, v10, v11, v12, v13, v14, v15 and v16, released between 2009 and 2016. Here, we give sample results for the LEON3MP 28LP testcase, with implementation and extracted RC parasitics obtained using Cadence Innovus V15.2 [37]. We run the signoff STA tool in multi-scenario mode (i.e., MCMM) with two, four, eight, 16 and 32 corners, and we collect runtime and memory usage data, along with reported timing and power. When running the signoff tool, we use averaged power for power analysis, and we enable clock reconvergence pessimism removal (synonymous with common path pessimism removal).

From timing reports, we see that setup and hold slack values are similar between earlier versions (e.g., v10 and v11) and between later versions (e.g., v12 – v16). However, between the two groups (i.e., v10 and v11 versus v12 – v16), differences in reported setup and hold slack values are up to 73ps and 549ps, respectively. The root cause of the setup slack difference is a difference in path delay calculation. The root cause of the hold slack difference is a change in default constraints (involving PI-to-register paths). Differences in reported power across all tool versions are negligible (< 0.05%).

Figure 10 shows (a) runtime and (b) memory usage statistics for different tool versions with various numbers of analysis



Figure 10: (a) Runtime and (b) memory usage for different tool versions (normalized to v16), with various numbers of modes (timing views).

modes (i.e., timing views). We normalize to v16 results. Earlier versions use less memory (v10 did not afford a memory usage reporting mechanism comparable to that of other versions). When the number of modes is 32, v14 shows the minimum memory usage with minimum runtime.

To summarize, we have demonstrated the potential of assessing progress and value of EDA tool improvements, using example leading P&R and signoff STA tool versions since 2009. For the example tools studied, we find that recent P&R tool versions achieve better QOR with reduced runtime and increased memory usage; recent signoff tool versions achieve reduced runtime with increased memory usage. The increased memory usage in recent P&R and signoff tool versions is likely associated with more complex design rules and timing constraints in advanced-node designs.

3. BOUNDING FUTURE PROGRESS

A number of previous works [6][8][10][11][18] have described constructions of benchmarks with known optimal or "good" solutions. In domains such as floorplanning, placement and gate sizing, the use of such benchmarks has led to empirical *lower bounds* on heuristic suboptimality. By contrast, [4] proposes a methodology to estimate an *upper bound* on remaining future benefits from a given technology – specifically, 3D integration). Estimation of maximum benefits from given technologies can also guide R&D objectives and prioritizations. In this section, we review the methodology used by [4] to estimate upper bounds on power and area benefits from 3DIC integration. (A more detailed account is available in [4].)

The authors of [4] propose that *implementation in "infinite dimension*", where all gates can be placed as close as possible (essentially, adjacent) to each other, can be used to derive an upper bound on 3D power and area benefits for a given design, technology node, and tool/flow. Moreover, an implementation in infinite dimension, along with its standard-cell area and power attributes, can be estimated by performing synthesis and netlist optimization with a zero wireload model (0-WLM).

Figure 11, from [4], compares design power and total cell area across various implementation dimensions (pseudo-1D, 2D, 3D (with 2, 3 and 4 tiers), and infinite-dimension) and different clock periods in a 28nm FDSOI, 12-track foundry technology. The results for the various implementation dimensions are obtained as follows. (1) Pseudo-1D implementation indicates design implementation with high aspect ratio layout (i.e., with aspect ratio = 0.1 or block height equal to block width / 10). The pseudo-1D implementation estimates the power penalty of design implementation in "less than two" dimensions, within the limits of what P&R tools can practically handle. (2) 2D implementation is the conventional planar implementation. Here, the authors of [4] empirically seek "optimal" 2D implementations as a baseline for accurate quantification of benefits from 3D (with multiple tiers) and infinite-dimension implementation. To achieve this, they obtain multiple conventional planar (2D) implementations by sweeping several key parameters such as synthesis clock period, placement utilization and BEOL stack options. They then select the best (e.g., minimum power) outcome. (3) 3D implementation with multiple tiers is challenging due to the lack of any "golden" 3DIC implementation flow. The proposed flow in [4] uses the Shrunk2D flow of [31] as a starting point. It then performs min-cut partitioning to divide the cells within each grid of the "Shrunk2D placement" (a commercial placer's results, with scaled LEF for cells) into T clusters which are assumed to be placed on a given T number of tiers. Parasitics are annotated to nets which have cells on different tiers, to



Figure 11: Design power and total cell area evaluated across various implementation dimensions [4].

model the impact of vertical interconnects; this is followed by an incremental optimization based on commercial 2D P&R tools. (4) Last, *infinite dimension* implementation ignores wire parasitics during the implementation. To achieve this, netlist optimization is performed with zero wireload model (0-WLM). Given that benefits from 3D integrations are mainly due to reduced wire parasitics in a shrunk footprint area, the infinitedimension implementation is able to provide an upper bound on 3DIC benefits.

In Figure 11, all the implemented designs have no hold violation and a setup violation less than 10ps. We observe that the maximum power benefits (i.e., the gap between the red curve versus the orange curve) from implementations in infinite dimension are respectively 36%, 39%, 20%, 18% and 26% for CORTEXMO, AES, JPEG, VGA and LEON3MP. The results show a large variation of 3D benefits across different designs. In addition, the power benefits from 3D integration with two, three and four tiers are less than 10% for designs JPEG, VGA and LEON3MP. Area benefits are small (i.e., < 10% for all designs, and < 4% for designs JPEG and VGA), possibly as a consequence of fixing a given standard-cell library as the implementation fabric.

To summarize, this section has shown an example of *upper-bounding* future benefits that can be obtained from an IC implementation technology. Specifically, through the "infinite dimension" concept, the work in [4] estimates upper bounds on future power and area benefits from (improved) 3DIC implementation of standard cell-based blocks (and a fixed library of standard cells). The results show that, e.g., the maximum power benefit according to the infinite-dimension assessment is 18% for particular designs.

4. RANK-ORDERING OF DESIGN ENABLE-MENTS

In this section, we give an overview of a recently-proposed framework [17] to assess alternative design enablements (e.g., alternative back-end-of-line (BEOL) interconnect stacks with a given router, placer-router combinations with a given BEOL stack, etc.). The key outcome is the ability to rank-order alternatives for overall design enablement, or for elements of design enablement. Figure 12 describes the overall flow of the framework proposed in [17]. For a given netlist, the routed layout outcome for a given combination of BEOL stack option, placer and router is evaluated according to robustness with respect to a threshold of detailed-routing design rule violations (DRVs). The framework is able to obtain, e.g., the ranking of routing capacities across different BEOL stacks for a given router. The authors of [17] also confirm that assessments (e.g., routing capacity ranking of BEOL stacks) based on mesh-like placement instances are consistent with those based on real placement instances (i.e., placement solutions of real designs, as output by commercial placers). Thus, it is suggested that the proposed framework can be used to rank placers and routers as well [17].

4.1 Methodology

The basic idea of the proposed methodology is to start with a placed netlist that is "simple" in terms of routability, in that routing can be completed with no DRVs. More specifically, in the proposed framework, placement is performed for a given netlist N_h using a given placer P_k . Then, the placement is gradually perturbed by swapping random pairs of horizontally- or vertically- adjacent cell instances, to increase the routing difficulty. After some number K of swaps, the routing would become infeasible (i.e., the number of DRVs exceeds a predefined threshold). In the proposed framework, three assessments are available, all in terms of routability and DRVs: (i) assessment of BEOL stack options, (ii) assessment of placers, and (iii) assessment of routers. For each assessment, the authors propose various approaches. Of particular interest is the use of mesh-like placement and of real placement with bloated cells

Assessment of BEOL stack options and routers with mesh-like placement. To assess BEOL stack options and routers, in [17], the K values corresponding to placement instances with routing failure are recorded as "K threshold", to measure routing capacities of BEOL stacks $\{B_1, B_2, ..., B_I\}$ for a given router R_j .⁷ [17] uses mesh-like placement (a netlist with mesh topology based on a given 2-input or 3-input cell) for its assessment of BEOL stack options for a given router in order to remove possible dependencies on input placement instances.⁸

Assessment of placers with real placement with bloated cells. In [17], real placement instances (i.e., placement solutions of real designs using a commercial placer) are used for assessment of placers. [17] uses bloated standard cells to avoid placement legalization after each swapping move.⁹ Then, the given commercial placer is used to place the netlist with the bloated cells to obtain an initial placement. Iterative swapping of adjacent cells is performed starting from the initial placement until the placement becomes unroutable. The minimum number of swap moves (K) leading to an unroutable placement is recorded and used as an indicator of the routing capacity of the given BEOL stack, and of the performance of the given placer and router.

⁷The authors of [17] denote the ranking of BEOL stacks in terms of routing capacity as $\Pi_B^{R_j}(N_h, P_k)$. Their goal is to determine a "universal" ranking of BEOL stacks (in terms of routing capacity) for a given router, that is, $\Pi_B^{R_j}(N_h, P_k) = \Pi_B^{R_j}(N_{h'}, P_{k'}) \forall h, k, h', k'$, as shown in Figure 12.

⁸ For 3-input cells, [17] connects the output pin of the gate instance with index (p, q) to input pins of the gate instances with indices (p + 1, q), (p, q + 1) and (p + 1, q + 1). The netlist is placed (accordingly to its mesh topology) uniformly, and all gate instances have the same size.

 $^{^9{\}rm In}$ [17], the cell LEF [47] is modified such that all gate instances in the netlist have the same size (i.e., physically occupy the same



Figure 12: Overall flow of the framework to determine ranking of BEOL stack options, placers and routers in terms of routing capacity [17].

The number of DRVs is measured to K threshold. characterize the K threshold. Figure 13 shows the number of DRVs versus the number of perturbations (K) measured as % of total instances, for mesh-like placement. For each K value, five different trials are made with random sequences of swaps, to mitigate noise from tools and randomness of perturbations. Each dot corresponds to a (number of perturbations, BEOL option) pair. Average values from the five trials are marked as solid lines. In [17], K threshold is defined as the points where the average number of DRVs > 150. In Figure 13, the K threshold values are 500%, 800%, 900% for B_1 B_2 and B_3 , respectively. The authors of [17] observe that the K threshold can be used as an indicator of routability. A higher K threshold value for a particular BEOL stack means that the BEOL stack is more robust in terms of routing more highly perturbed placements (i.e., placements that have more hotspots from "tangling").



Figure 13: Number of DRVs versus perturbation K, for a mesh-like placement implemented with 5000 AOI21 cells and row utilization of 90%. Adapted from [17].

4.2 Experiments and Discussion

Figure 14 shows K threshold values for different BEOL stack options with the same nominal amounts of routing resources. Recent versions of two commercial placers and two commercial routers are used in the study. Results are shown for (i) mesh-like placement with router R_1 (mesh- R_1), (ii) mesh-like placement with router R_2 (mesh- R_2), (iii) real placement from placer P_1 with router R_1 (real- P_1 - R_1), and (iv) real placement from placer P_2 with router R_2 (real- P_2 - R_2). Table 1 gives details of the BEOL stack options and the rank-ordering per the experiments. The results show that BEOL stack options with the same nominal routing resources can have very different K threshold values. Furthermore, the orderings of BEOL stack options are similar across experimental setups (i) through (iv). This suggests that the ordering of BEOL stack options may be "universal" across different routers for the same placement instances.

Table 1: BEOL stack options. $\#1\times$, $\#1.5\times$ and $\#2\times$ respectively indicate the number of $1\times$ layers, the number of $1.5\times$ layers and the number of $2\times$ layers.

Name	Rank	#1×	$\#1.5 \times$	$\#2\times$
B_1	4	3	0	4
B_2	4	2	3	2
B_3	2	3	3	0
B_4	1	4	0	2
B_5	3	5	0	0

The authors of [17] observe that 'counting routing tracks' is not an accurate measurement of the routing capacity of a BEOL stack option. Rather, the measurement of routing capacity is highly nontrivial: gear ratios of metal pitches and via blockages affect routability; there may be effects of 'height' of layers (lower layers being more valuable than upper layers due to vias); etc. Figure 14 also demonstrates that different BEOL stack options with the same nominal routing resources (defined by routing track counts) have different routing capacity with mesh-like placements and real placements from the two placers $(P_1 \text{ and } P_2)$, and the two routers $(R_1 \text{ and } R_2)$. Thus, Kahng et al. [17] also point out that the proposed framework may enable a rank-ordering of placers and routers. In Figure 14, R_2 shows better performance than R_1 in terms of routability (mesh- R_1 versus mesh- R_2). For P&R, the (P_1, R_1) pair shows better routing capability than the (P_2, R_2) pair $(real-P_1-R_1 \text{ versus } real-P_2-R_2)$.



Figure 14: K threshold values for different BEOL stack options. Adapted from [17].

To summarize, in this section, we describe a technique [17] that enables *rank-ordering* of design implementation enablements spanning different BEOL stack options, placers and routers. The results indicate that there may be a *universal* ordering of routing capabilities of BEOL stack options; they also confirm that simply counting the number of tracks to measure routing capability of a BEOL stack option is insufficient. In addition, the proposed framework may enable the ranking of placers and routers.

5. MEASUREMENT OF ACADEMIC PROGRESS AND VALUE

Mismatches in data models, benchmark formats, technology files, library granularity, etc. have for many years precluded assessment of academic tools' progress and value across a wider range of benchmarks and technologies. For example, it is difficult to assess the performance of academic tools using realistic industrial designs and foundry technologies when the tools are hard-coded for a particular contest benchmark format, and/or "tuned" to particular contest metrics and testcases. Hence, there is a high barrier to assessing academic tools' value in the context of real-world design challenges - and, conversely, commercial tools cannot be directly juxtaposed or integrated with academic tools. The net result is hampered overall progress of the field; cf. [3] [23]. In this section, we overview the benchmark generation framework of [21][35], which constructs connections between academic tools and industrial benchmarks and formats, thus enabling more realistic assessments of academic tools. The proposed *horizontal benchmarks* and benchmark extension together seek to maximize "apples-toapples" assessment at specific design stages, across different benchmarks, technologies, and tools.

The scope of the work in [21] is depicted in Figure 15. The authors use sizing and P&R (placement and routing), two key steps in IC implementation, to illustrate their methodologies for horizontal benchmark enablement. By applying the proposed horizontal benchmark extension, [21] demonstrates the feasibility of "apples-to-apples" assessment of four academic tools and three commercial tools in the P&R domain: (i) across ISPD12/13 sizing-oriented benchmarks [29] [30], ISPD11 placement-oriented benchmarks [34], and real designs from OpenCores [48]; and (ii) across ISPD12/13 contest and 28/45/65/90nm foundry technologies. The benchmark generation methodology of [21] has been recently applied to support the ICCAD-15 placement contest [25].



Figure 15: Scope of the work in [21] to extend assessments across different technologies, benchmarks and tools.

The most obvious challenge to benchmark extension arises from IP protection and the limited scope of target problem formulations: benchmarks typically omit information. For instance, partitioning instances (ISPD98) omit cell sizes and signal directions; placement instances (ISPD06/11) omit or obfuscate cell functions and combinational-sequential distinctions; global routing instances (ISPD07/08) omit cell functions and pin locations; etc. Thus, a number of judgment calls must be made to best fill in missing information when performing "benchmark extension". For example, one resolution of missing information in [21] maps nodes of a placement benchmark to cells in a given Liberty/LEF pair, based on cell pin count and cell width. (For details, please refer to [21].)

Another challenge in horizontal extension is that many academic tools are "hard-wired" to particular technology definitions. When assessing "legacy" tools that are no longer under active development, extra enablements are required to migrate benchmarks across multiple technologies. For example, different cell libraries might vary in granularity (number of cell sizes, number of Vt flavors), available logic functions, or naming conventions, and this makes technology migrations not so straightforward. For example, [21] increases library granularity to match the number of cell variants; this allows gate-sizing optimizers to have the same solution space and fair assessment across different technologies. The new cells are generated using interpolation/extrapolation from timing information (cell delay, output transition time) of existing cells, along with logical effort analysis for cells of each given type. Leakage power and pin capacitance values are approximated by fitting second-order models to attributes of existing cells.



Figure 16: Flows to enable horizontal sizer assessment [21].

Figure 16 illustrates the enablement of horizontal evaluation across academic and industry tools for gate sizing (i.e., postrouting leakage reduction). The cell sizing/Vt-swapping optimization reduces leakage *while preserving a timing signoff*. Inputs to sizing tools are netlist (.v), interconnect parasitics (SPEF), timing constraints (.sdc) and timing/power Liberty (.lib). Table 2 from [21] shows that, as of 2014, the academic sizers could achieve similar (Trident [19]) or even larger leakage power reduction (UFRGS [27]) compared to commercial sizers (cSizer1, cSizer2), but with larger runtime. At the same time, the differences in ranking between the ISPD technology and industry technologies across different sizers may indicate the potential for improvement of academic tools' robustness.

A recent "existence proof" that assessments can span academic tools and industrial contexts is seen in [7]. This work demonstrates that a state-of-the-art academic placer, ePlace2.0, can be evaluated in practical IC implementation contexts with a final-routed wirelength criterion, by incorporating an improved (Steiner) wirelength objective and a routability-driven technique. As reported in [7], ePlace2.0 achieves 3.3% routed wirelength reduction (using a commercial router) and 28% fewer overflowed gcells at maximum utilization, compared to a leading-edge commercial placer in a foundry 28LP technology.

To summarize, the above discussion shows that improved assessment across academic and industrial tools and design enablements can be achieved through *horizontal benchmark* and *benchmark extension* methodologies. Such enablement of better-targeted academic research and faster technology transfer into real-world practice seems to be a "no-brainer" for the DA community.

6. FUTURE DIRECTIONS

We conclude by proposing several improvements toward future "measurement of progress and value".

1. Open up commercial EDA tools to benchmarking. An old adage tells us, "Measure to improve". Indeed, benchmarking and measurement are engrained throughout the culture and practice of engineering, the electronics industry, and the semiconductor industry – but not the EDA industry. Direct measurement of EDA tools' value and progress through benchmarking is impossible today, since benchmarking is prevented by explicit language in EDA tool licenses. Opening up both academic [3] and industrial EDA technology to direct comparison and benchmarking will not only enable more precise assessment of value and progress, but can also accelerate the development of technology that is critical to the IC industry's progress.

2. Develop meaningful, scalable, unified benchmarks and metrics. Meaningful and scalable benchmarks and metrics are required for well-targeted (and, accurately assessed) academic research that is in synch with industry needs. Benchmarks must be realistic and able to reflect real design challenges. A unified, standard benchmark format will lower barriers to assessment across various EDA tool chains and design enablements. Particularly with so many new "firstclass careabouts" in leading-edge IC design (signoff across many wide corners and modes, dynamic power integrity, thermal and reliability constraints, advanced patterning rules, etc.), meaningful criteria must be established and communicated so that new DA technology has a chance at delivering "real" value.¹⁰

3. Establish an "Underwriters Laboratories" [50] for EDA. An "Underwriters Laboratories" for EDA could provide a concrete realization and delivery mechanism for (1) and (2) above. Such an entity could help implement and execute standard assessments for progress and value of EDA

¹⁰Two comments. (1) Scalable benchmarks that are both realistic and have known optimal/good solution quality are still an open question for the field. Upper-bounding remaining future improvements is also crucial to better exploit available R&D bandwidth. (2) A recent "DA Futures" initiative by IEEE CEDA may help address the need for improved research enablement in the DA field.

Table 2: Comparison across sizers. Benchmark: ISPD13 NETCARD.

	cSizer1			cSizer2		Trident			UFRGS			
Tech	Leak	WNS	Runtime	Leak	WNS	Runtime	Leak	WNS	Runtime	Leak	WNS	Runtime
	(mW)	(ns)	(\min)	(mW)	(ns)	(\min)	(mW)	(ns)	(\min)	(mW)	(ns)	(\min)
ISPD	5231.6	-0.01	55.0	5591.5	0.0	31.6	5233.1	0.0	179.8	5184.1	-0.2	46.0
28nm	27.8	0.5	64.0	27.8	0.7	35.0	29.4	1.4	43.7	27.7	-3.7	73.5
65nm	45.8	0.4	49.5	45.9	0.5	34.0	46.0	1.2	46.8	45.4	-2.6	77.3

technology.¹¹ Going further, an "Underwriters Laboratories" for EDA might also provide additional guidance on where to seek progress (e.g., toward support of particular IC fabrics or product types, toward particular compute platforms, etc.).

4. Assess costs of, and provide solutions to, noninteroperability. Non-interoperability across different EDA tools, benchmarks and libraries limits the progress and value of IC implementation technology; it also necessitates such ad hoc efforts as "horizontal benchmarking" [21]. It is well-known (at least, as folklore) that CAD integration costs can easily be on par with tool license costs. In an era that has seen a decade-long "design capability gap", and design turnaround time specifically targeted by a recent DARPA program, we believe that interoperability merits renewed attention from the semiconductor and EDA industries.

REFERENCES 7.

- S. N. Adya, M. C. Yildiz, I. L. Markov, P. G. Villarrubia, P. N. Parakh and P. H. Madden, "Benchmarking for Large-Scale Placement and Beyond", *IEEE TCAD* 23(4) (2004), pp. 472-487. [1]
- F. Brylez, D. Bryan and K. Koźmiński, "Combinational Profile of Sequential Benchmark Circuits", *Proc. ISCAS*, 1989, pp. [2]1929-1933.
- A. E. Caldwell, A. B. Kahng and I. L. Markov, "Toward CAD-IP Reuse: The MARCO GSRC Bookshelf of Fundamental CAD [3] Algorithms", IEEE Design and Test of Computers 19(3) (2002),
- W.-T. J. Chan, A. B. Kahng and J. Li, "Revisiting 3DIC Benefit [4]with Multiple Tiers", Proc. SLIP, 2016, pp. 6:1-6:8.
- W.-T. J. Chan, A. B. Kahng, S. Nath and I. Yamamoto, "The ITRS MPU and SOC System Drivers: Calibration and Implications for Design-Based Equivalent Scaling in the Roadmap", Proc. ICCD, 2014, pp. 153-160.C. C. Chang, J. Cong and M. Xie, "Optimality and Scalability
- [6] Study of Existing Placement Algorithms", Proc. ASP-DAC, 2003, pp. 621-627.
- C.-K. Cheng, A. B. Kahng, I. Kang and L. Wang, "ePlace2.0: Improved Solution Quality and Validation of Routable Placements", *manuscript in submission*, June 2016.
- J. Cong, M. Romesis and M. Xie, "Optimality, Scalability and Stability Study of Partitioning and Placement Algorithms", *Proc.* [8] *ISPD*, 2003, pp. 88-94.
- J. Darnauer and W. W.-M. Dai, "A Method for Generating [9] Random Circuits and Its Application to Routability Measurement", *Proc. FPGA*, 1996, pp. 66-72.
- P. Gupta, A. B. Kahng, A. Kasibhatla and P. Sharma, "Eyecharts: Constructive Benchmarking of Gate Sizing Heuristics", *Proc.* [10]DAC, 2010, pp. 592-602.
- L. W. Hagen, D. J.-H. Huang and A. B. Kahng, "Quantified Suboptimality of VLSI Layout Heuristics", *Proc. DAC*, 1995, pp. [11] 216 - 221
- [12] M. D. Hutton, J. Rose, J. P. Grossman and D. Corneil, "Characterization and Parameterized Generation of Synthetic Combinational Circuits", *IEEE TCAD* 17(10) (1998), pp. 985-996.
- [13] K. Jeong, A. B. Kahng and K. Samadi, "Impacts of Guardband Reduction on Design Process Outcomes: A Quantitative Approach", *IEEE SM* 22(4) (2009), pp. 552-565.
- A. B. Kahng, "PPAC Scaling at 7nm and Below", Cadence [14]
- Distinguished Speaker Series talk, San Jose, CA, April 7, 2016. [15] A. B. Kahng, "The ITRS Design Technology and System Drivers Roadmap: Process and Status", Proc. DAC, 2013, pp. 1-6.
- A. B. Kahng, "Lithography-Induced Limits to Scaling of Design Quality", Proc. SPIE, 2014, pp. 905302-1-905302-14. [16]
- A. B. Kahng, A. B. Kahng, H. Lee and J. Li, "PROBE: A Placement, ROuting, Back-End-of-line Measurement Utility", manuscript, August 2016. [17]
- A. B. Kahng and S. Kang, "Construction of Realistic Gate Sizing Benchmarks With Known Optimal Solutions", *Proc. ISPD*, 2012, 18 pp. 153-160.
- A. B. Kahng, S. Kang, H. Lee, I. L. Markov and P. Thapar, "High-Performance Gate Sizing with a Signoff Timer", *Proc.* [19]
- ICCAD, 2013, pp. 450-457. A. B. Kahng and F. Koushanfar, "Evolving EDA Beyond its E-Roots: An Overview", *Proc. ICCAD*, 2015, pp. 247-254. [20]

- [21] A. B. Kahng, H. Lee and J. Li, "Horizontal Benchmark Extension for Improved Assessment of Physical CAD Research", *Proc. GLSVLSI*, 2014, pp. 27-32.
 [22] A. B. Kahng, M. Luo, G.-J. Nam, S. Nath, D. Z. Pan and G. Pabing "Transd Matrices of Design Actions the Designal Actions of Physical Contents of P
- Robins, "Toward Metrics of Design Automation Research Impact", Proc. ICCAD, 2015, pp. 263-270.
- [23]A. B. Kahng and I. L. Markov, "Impact of Interoperability on CAD-IP Reuse: An Academic Viewpoint", *Proc. ISQED*, 2003, pp. 208-213.
- A. B. Kahng and S. Reda, "Evaluation of Placer Suboptimality Via Zero-Change Netlist Transformations", *Proc. ISPD*, 2005, pp. [24]208-215.
- [25]M.-C. Kim, J. Hu, J. Li and N. Viswanathan, "ICCAD-2015 CAD Contest in Incremental Timing-Driven Placement and Benchmark Suite", Proc. ICCAD, 2015, pp. 921-926.
- F. Koushanfar, A. Mirhoseini, G. Qu and Z. Zhang, "DA [26]
- F. Koushanfar, A. Mirhosenni, G. Qu and Z. Zhang, "DA Systemization of Knowledge: A Catalog of Prior Forward-Looking Initiatives", *Proc. ICCAD*, 2015, pp. 255-262.
 V. S. Livertamento, C. Guth, J. L. Güntzel and M. O. Johann, "Fast and Efficient Lagrangian Relaxation-Based Discrete Gate Sizing", *Proc. DATE*, 2013, pp. 1855-1860.
 Paul McLellan, "Andrew Kahng on Industry-Academia Cooperation" [27]
- [28]Cooperation". https://community.cadence.com/cadence_blogs_8/b/
- breakfast-bytes/archive/2016/04/22/andrew-kahng
- M. M. Ozdal, C. Amin, A. Ayupov, S. M. Burns, G. R. Wilke and C. Zhuo, "ISPD-2012 Discrete Cell Sizing Contest and Benchmark Suite", *Proc. ISPD*, 2012, pp. 161–164. http:
- //archive.sigda.org/ispd/contests/12/ispd2012_contest.html. M. M. Ozdal, C. Amin, A. Ayupov, S. M. Burns, G. R. Wilke and C. Zhuo, "An Improved Benchmark Suite for the ISPD-2013 [30]
- C. Endo, An Improved Benchmark Suite for the ISPD-2013
 Discrete Cell Sizing Contest", Proc. ISPD, 2013, pp. 168–170.
 http://www.ispd.cc/contests/13/ispd2013_contest.html.
 S. Panth, K. Samadi, Y. Du and S. K. Lim, "Design and CAD
 Methodologies for Low Power Gate-Level Monolithic 3D ICs",
- [31]Proc. ISLPED, 2014, pp. 171-176.
- G. Smith, "Updates of the ITRS Design Cost and Power Models", *Proc. ICCD*, 2014, pp. 161-165. [32]
- L. Srivani and V. Kamakoti, "Synthetic Benchmark Digital [33] Circuits: A Survey", IETE Technical Review 29(6) (2012), pp. 442-448.
- N. Viswanathan, C. J. Alpert, C. Sze, Z. Li, G.-J. Nam and J. A. Roy, "The ISPD-2011 Routability-Driven Placement Contest and Benchmark Suite", *Proc. ISPD*, 2011, pp. 141-146. [34]http://www.ispd.cc/contests/11/ispd2011_contest.html.
- Horizontal Benchmarks Project Website. [35]
- http://vlsicad.ucsd.edu/A2A
- CAD/CAM/CAE Wallchart. http://www.garysmitheda.com/wp-content/uploads/2015/05/All_WC-15.pdf [36]
- Cadence Innovus User Guide. http://www.cadence.com [37]
- Cadence Tempus User Guide. http://www.cadence.com [38]
- CRAFT Program Aims for Affordable Designer Circuits that Do [39] More with Less Power.
- http://www.darpa.mil/news-events/2015-08-17 [40]
- EDA Roadmap Workshop. http://vlsicad.ucsd.edu/EDARoadmapWorkshop/
- EDA Roadmap Workshop. [41]
- http://vlsicad.ucsd.edu/EDARoadmapWorkshop/ [42]
- EDAC Market Statistics Service. http://edac.org/initiatives/mss/newsletter_q1_2015 (see also: http://www.edac.org/sites/default/files/users/mss/MSS_2015_ Category_Definitions_FINAL.pdf)
- [43]SIA Semiconductor Industry Billing History http://www.semiconductors.org/industry_statistics/historical_ billing_reports/
- ITRS 2011 Design Chapter. http://www.itrs2.net/2011-itrs.html [44]
- ITRS 2013 Design Chapter. http://www.itrs2.net/2013-itrs.html [45]ITRS 2013 System Drivers Chapter. [46]
- http://www.itrs2.net/2013-itrs.html LEF/DEF 5.7 reference. [47]
- http://www.si2.org/openeda.si2.org/projects/lefdef
- OpenCores. http://opencores.org
- Synopsys Design Compiler User's Manual. [49]
- http://www.synopsys.com
- [50]Underwriters Laboratories, Inc. http://www.ul.com

 $^{^{11}\}mathrm{Our}$ research group would gladly take on this basic role, were it to be permitted by the industry.