14-1

Agile Pathfinding Technology Prototyping: the Hunt for Directional Correctness

Daniel Chanemougame, Jeffrey Smith, Paul Gutwin, Brandon Byrns and Lars Liebmann

TEL Technology Center America, Albany, NY, USA

***Abstract*—New tools and methodologies are fused with conventional elements of the process-design-kit (PDK) and design enablement to introduce a rigorous yet fast and agile technology prototyping platform. This design technology co-optimization (DTCO) solution replaces the rigid, time and resources consuming PDK to enhance the core functions critical to evaluate power, performance area and cost (PPAC). Any technology definition with any device or process integration innovation can be evaluated at the standard cell level first, and then at the block level to explore and understand the requirements of different design applications. The flexibility and fast turn-around make it practical to imagine, test and compare many technology prototypes. From simple evolutions to innovative disruptions, the feasibility and value of the technology choices and hardware tools required can be identified early with great detail, significantly accelerating the development of future process tools. To illustrate the efficiency of the platform, complementary-FET (CFET) [1] technologies are compared to reference finFET technologies. As we approach the fundamental limits of dimensional scaling, with so many choices ahead of us including 3D constructs, we need efficient technology prototyping to navigate and steer in the right direction.**

Keywords—Pathfinding DTCO, PDK, PPAC, LGAA, CFET, technology prototyping, compact modeling, parasitic extraction, timing characterization, synthesis place and route

I. Introduction

The bleeding edge of the semiconductor industry will likely transition to the lateral gate all around (LGAA) architecture. Beyond, the semiconductor scaling roadmap is uncertain. While the device architecture remains important, advanced technologies are susceptible to other critical components, which ultimately manifest themselves as parasitic resistance and capacitance (RC) loads. As technologies scale, conductors become smaller and closer to each other, resulting in the challenge of ever-increasing parasitic RC, wasting precious energy and performance during circuit operations. In the tremendous effort to limit waste and extend the roadmap, engineers are facing an explosion of critical PPAC technology options at all levels from device architecture and process integration to materials. As the challenges become harder than ever, more disruptive options must be considered as well. With so many combinations, the risk of implementing the wrong technology definition is significant with serious cost consequences.

Logic standard cells have long been the industry benchmark to model, predict and adjust the PPA value of a technology early in the definition. While ring oscillator (RO) circuit simulations are easy to run, major inputs specific to that technology are required which are typically part of a PDK. Building a PDK for advanced technologies, even a lighter evaluation version, is a multi-months affair, involving many experts from large design enablement teams (fig. 1). The concept of digital twins is common in the semiconductor industry. However imperfect, the PDK is a digital twin of a silicon technology: its ultimate goal is to provide the necessary inputs to predict the PPA so designers can fully leverage the technology. A collection of multiple software elements, the PDK is a remarkable and capable piece of modeling engineering, but extremely customized and rigid by definition. In the early stages of development, months before circuit simulations can reveal the PPA entitlement and potential roadblocks, major technology options have to be anchored to prevent development delays, making any significant corrections that much harder and leaving behind many options. These challenges have been quite severe and steadily growing since 20nm despite DTCO efforts.

Additionally, a classic PDK approach cannot practically address the speed, agility and variety of scope required in Pathfinding, as any process integration changes will necessitate multiple slow PDK updates. For this discussion, using CFET as a demonstration, we will present our work on an agile technology prototyping platform (fig. 2), built to increase the breadth of DTCO with reasonable resources, and accelerate the quantitative PPAC exploration of various technology definitions, ranging from evolutionary to disruptive, leaving no idea unexplored. To limit the scope and size of this paper, we will focus on the critical design enablement components leading up to synthesis place and route (SPnR) experiments for design applications exploration, and then discuss some of the results.

II. Agile Technology Prototyping

A modern production PDK has many components, but the transistor compact models and the parasitic RC extraction deck (fig. 1) represent the core functions essential to circuit simulations and therefore are the first steps in technology prototyping. The first two sections will review how we enabled these two critical capabilities without a PDK.

Ironically, as DTCO efforts in recent nodes managed to extend the scaling of standard cell height [2], cell routability started to suffer [3]. As a result, the scaling benefits achieved at the block level are reduced compared to what was initially observed at the standard cell level. This is a major problem for modern designs and technologies in their attempts to continue Moore's law. It is therefore necessary that the PPAC work extends to block level for a comprehensive analysis. Enabling block level SPnR experiments represents a significant step up in the design enablement efforts and requires a third critical component, characterization, which we will also describe in the third section.

By embracing a broad scope, ranging from process assumptions and critical design rules, to device models, parasitic extraction and characterization, and ultimately to block-level place and route, we can efficiently identify and quantify the complex PPAC tradeoffs associated with the process disruptions and innovations needed.

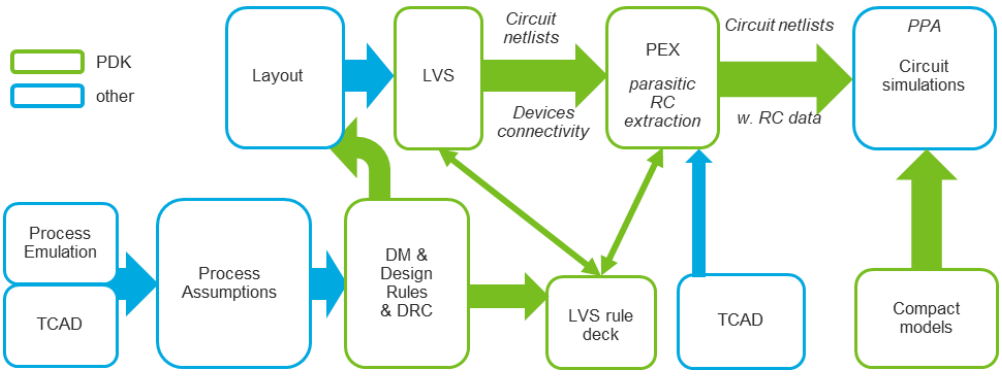


Fig. 1. High level description of the conventional PDK. PDK elements are represented in green. The goal is to enable circuit simulations to evaluate a technology's PPA. In addition to the time and resources involved, any significant change in the definition can easily break the flow, making this approach not suitable for the speed and agility required in Pathfinding.

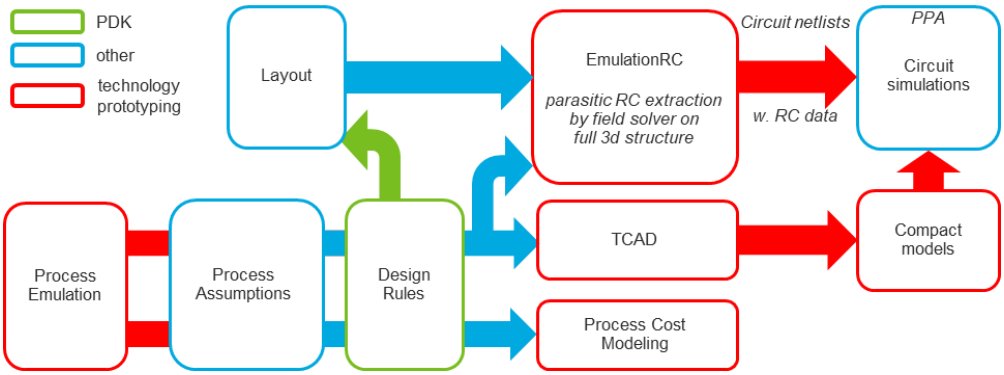
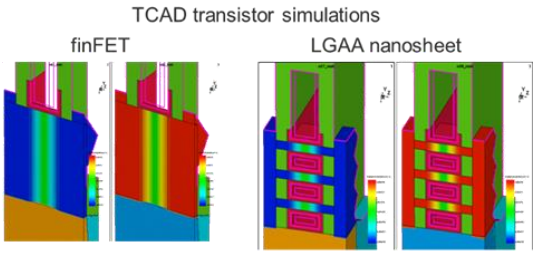


Fig. 2. High level description of the Agile Technology Prototyping platform developed and used in this study. The platform is designed to provide the core functions essential to circuit simulations without developing a PDK. This results in a consolidation of tools closer to the typical TCAD environment (red boxes). The tools involved are more rigorous and less dependent on expert user settings, therefore more flexible and robust.

A. *Device compact models*

Compact models such as BSIM-CMG [4] have come a long way to capture the behavior of advanced transistors by modeling more microscopic device phenomena [5]. An evaluation compact model still represents a significant effort, involving quality silicon and experienced modeling engineers. In Pathfinding, even if there is preliminary silicon device data, it is likely not good enough. Already used in early stages of development, the quality and accuracy of TCAD device simulations provide a great alternative to silicon measurements and a good starting point.



critical parameters	N3 finFET	CFET LGAA
CPP, nm	45	45
nominal Vdd, V	0.7	0.7
Lg, nm	18.4	15
Lmetal, nm	15	15
EOT, nm	1.11	1.11
gate overlap, nm	1	0
channel thickness, nm	5	5
channel width, nm	50	36

Fig. 3. Critical device geometries and parameters simulated for finFET and CFET LGAA.

Fig. 3 shows examples of the devices that were simulated for our N3 reference finFET and the CFET LGAA devices, as well as some critical dimensions and parameters. For the CFET, the NMOS and PMOS are simulated as individual LGAA devices, each with 3 stacked nanosheets, allowing more manageable computation time.

device		GPa	Idsat uA	Idsat µA/um	N/P Idsat	Idsat LGAA vs finFET	
finFET	P	-1.5	54	1070	1.3	/device	/um
	N	0	71	1418			
LGAA	P	0	95	881	2.0	1.8	-18%
	P	-0.75	141	1302	1.3	2.6	22%
	N	0	188	1739		2.6	23%

Fig. 4. FinFET Idsat vs CFET from fig. 3 devices, with different stress for CFET PMOS. The unstressed LGAA PMOS is 2x weaker than the NMOS. With PMOS stress, the N/P balance can be improved. The last 2 columns show the Idsat ratio between LGAA and finFET per device and then normalized by the total width of the device.

Additionally, a hybrid simulation approach was adopted to capture properly the electrostatics and transport effects in these advanced devices: the sub-threshold part of the IdVg curves is based on drift-diffusion (DD) to capture the leakage, while the above-threshold part is based on Monte-Carlo to capture transport effects and on-currents. IdVd curves are based on Monte-Carlo while CV curves are based on DD with calibrated density-gradient. An interesting point of note on Fig. 4 is that very much like planar devices on (100)/[110] with beta ratio of 2, the LGAA NMOS is twice as strong as the PMOS. If the PMOS can be strained, the N/P balance is significantly improved. Finally, Idsat is

compared both per device and normalized per total width, between the single-fin finFET and the 3-nanosheet LGAA.

Fig. 5 shows the device geometries that were simulated to enable our CFET LGAA technology exploration. Only one geometry was done for the finFET reference. Different stress options were also applied for a total of 20 devices. This TCAD reference data provides solid foundational grounds. If a good model fit can be achieved, we will have a rigorous description of these advanced devices, and this is the primary goal: achieve accurate and pertinent Pathfinding circuit simulations.

LGAA	NMOS						PMOS											
stress Gpa	0						0						-0.75					
Lm nm	12			15			12			15			12			15		
NSW nm	10	24	36	10	24	36	10	24	36	10	24	36	10	24	36	10	24	36

Fig. 5. CFET LGAA devices simulated to extract compact models. Lm is Lmetal, and NSW is the nanosheet width. 2 reference finFET devices (N/PMOS) were simulated with Lm=15nm.

Once the reference TCAD data is generated for all the devices, the task of optimizing and fitting the compact model remains. We leveraged new tools and methodologies to automate the optimization strategies to best fit the TCAD reference curves. An initial BSIM-CMG model card is provided to get the optimization going. This new TCAD to compact model methodology is very effective in handling all the reference data and quickly generated more than 20 different model cards. Fig. 6 shows some example of fit achieved across different LGAA geometries. This methodology is the first significant design enablement improvement, as generating the many high quality compact models necessary for comprehensive advanced Pathfinding becomes practical. The TCAD becomes by far the slowest portion of the cycle, but Monte-Carlo tools are improving and make efficient use of parallel processing.

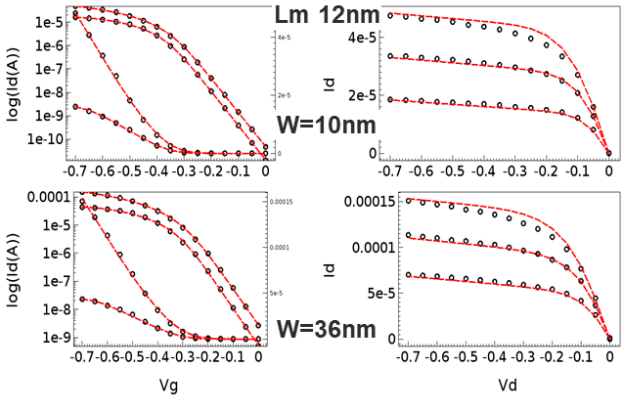


Fig. 6. Examples of TCAD to compact model extraction and fit on IdVg (both log and linear Id shown) and IdVd curves, shown for a PMOS LGAA (Lm=12nm, 2 nanosheet widths). Black circles: TCAD reference data, red curves: compact model. A similar quality of fit is achieved across all geometries.

B. *Emulation-based parasitic extraction*

Equipped with analytical compact models, we now need a circuit to perform SPICE simulations [6]. A standard cell is a circuit formed by transistors connected together to achieve its function through shared diffusions, gates and interconnects. To simulate this circuit, we therefore need a complete and accurate electrical description including all interconnects represented as parasitic resistance and capacitance elements, and all transistors. This description is known as a netlist and is a list of all components linked together by their associated electrical nodes. Each netlist is specific to a standard cell and its layout implementation.

All transistors must be identified and, despite the 2D representation of the layout, the 3D nature of interconnects must be captured. In a conventional PDK, the netlist generation is handled by the parasitic extraction deck (PEX), as shown in fig. 1. Building a PEX initially takes time and expert resources, but then running extraction is quite fast. The drawbacks are the lack of flexibility and accuracy which is continuously challenged by the complex 3D shapes of advanced technologies.

Highlighting the need for flexibility, the CFET architecture shown on fig. 7 represents a significant modeling challenge for the PEX. Process emulation is an excellent way to combine a cell layout with an integration flow to output a high fidelity, detailed 3D structure. Initially conceived for process integration development only, research very quickly focused on electrical RC extraction as well. This solution reached maturity after many years of development. The resulting netlist of emulation-based extraction is on the right of fig. 7.

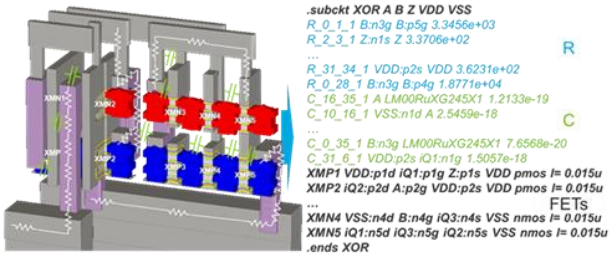


Fig. 7. On the left, CFET architecture with buried power rail. PMOS in blue at the bottom, NMOS in red on top. Parasitic RC elements originating from the interconnects are shown for illustration. On the right, an abbreviated example of emulation-based netlist extracted through rigorous field solver simulation.

The complex emulated 3D structure is directly meshed and exported to a rigorous field solver for extraction. The results depend very little on the level of tool setting expertise, and although running field solver simulations is much slower than running a traditional PEX, runtimes are largely manageable. The flexibility to accurately tackle any integration flow for any technology definition more than make up for longer runtime. With about 100 CPUs and 2TB of memory, a hundred cells can be extracted in about 24h, including cells with more than 30 transistors, making running multiple libraries or technology iterations possible within a week.

C. *Logic library timing characterization*

RO simulations provide meaningful PPA results with limited efforts on a few simple standard cells. Only a few layouts have to be created per technology definition, and correspondingly, only a few netlists need to be extracted. Block level experiments on the other hand require a much larger library of standard cells. Typical product libraries have hundreds, even thousands of standard cells, and each library is specific to a technology. Thankfully, for Pathfinding, the library size can be significantly reduced and still provide meaningful and pertinent results. For this work, we used about 80 cells. This is still a sizeable effort as the

number of libraries or technology prototypes multiplies the total number of layouts that have to be created.

Logic designs at the block level are assembled by SPnR tools. The synthesis tool determines what standard cells are needed to achieve the block function. Through iterations and complex optimizations, all cells are physically placed and routed together to hit a combination of target density and speed. Evaluating the speed and power of a block-level circuit cannot use the circuit simulation methodology we described so far. Typical blocks have tens of thousands of standard cells, which means simulating hundreds of thousands of transistors together. SPICE simulations are not that efficient. Instead, the SPnR relies on a higher abstraction level: each standard cell is a black box with only input and output pins, with its complete power performance or timing profile attached to it. To generate this timing profile, each netlist of the library is run through many SPICE simulations where the output response is measured in terms of delay and power usage, while input signal conditions and output load capacitances are varied (fig. 8).

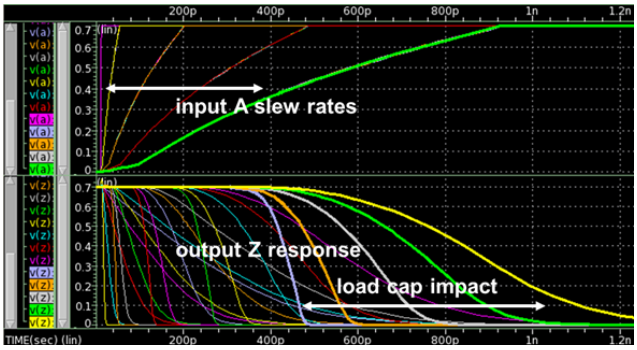


Fig. 8. SPICE simulations performed during timing characterization of a standard cell. As the input signal conditions and load capacitances are varied, the output response is measured.

The range of input signals is designed to represent the variety, shape and spread of signals that cells actually receive during operation in real designs. The SPnR tools will know precisely the cell's behavior and how and where to use it best in a design. The full library is characterized by running these simulations on all cells . The power and delay data collected on all cells is recorded and organized in a library timing file using the liberty format [7].

The challenge of characterization for Pathfinding is to strike the right balance between reasonable runtimes and sufficient accuracy and quality to achieve directional correctness in our results. In comparison, product characterization not only deals with much larger libraries, but must also cover the technology corners (multiple Vt flavors, process variability) and noise as well.

III. Results and Discussion

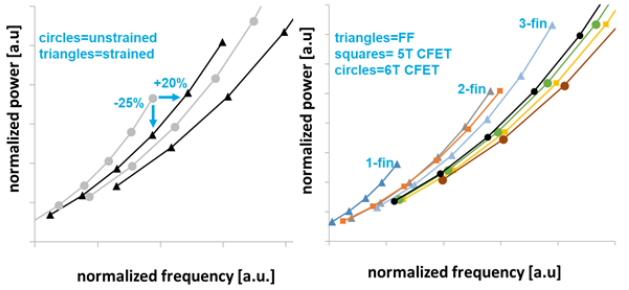
Leveraging this platform to compare CFET to a reference finFET baseline, 18 technology prototypes were built out of different integration flows and design rules, each prototype containing a 80-cell library. 8 prototypes for finFET to provide a solid and relevant baseline, and 10 for CFET, all mixing driving strength, cell height, integration flow and power delivery options either conventionally through a M0 power rail (PR) or through a buried power rail (BPR).

A. Ring Oscillator simulations

The first step of PPA benchmarking of different technology prototypes is a 15-stage RO circuit with a fan-out of 3 to measure dynamic power usage and stage delay. For each prototype, 3 standard cells (INV, NAND2, NOR2) are run and their results averaged, resulting in more than 50 base RO simulations, multiplied by different compact model options.

We discuss here a small subset shown on fig. 9. On the left, the power-performance chart shows the impact of the LGAA PMOS device strength on circuit performance. Illustrated by 2 CFET libraries with different nanosheet widths, 2 PMOS stress conditions are compared: unstrained where the LGAA PMOS is 2x weaker than the NMOS, and strained where both are matched. When both devices are matched, we observe a 20% speed gain at iso-power, or 35% power reduction at iso-speed. In other words, if the CFET PMOS channel cannot be strained, significant power and performance are left on the table.

On the right of fig. 9, 5 CFET libraries with strained PMOS are compared to 3 reference finFET libraries (PMOS also strained, fig. 4), all mixing cell heights, drive strength and power rail options (M0 vs BPR). The first observation is that BPR offers only slightly better power-performance than M0 PR (black vs green circles). While BPR has a much lower resistance than M0 PR in these scaled libraries, the power vias associated with BPR are more resistive due to higher aspect ratios, limiting BPR advantage at least at the cell level. Next, we compare CFET to finFET as summarized in the table of fig. 9.



isoP/isoF		5T CFET low	6T CFET mid	5T CFET mid	6T CFET high
6.5T FF 1-fin	freq	28%	47%	51%	56%
	power	-35%	-47%	-50%	-53%
	area	-23%	-8%	-23%	-8%
6.5T FF 2-fin	freq	5%	22%	26%	31%
	power	-8%	-30%	-34%	-41%
	area	-23%	-8%	-23%	-8%
8.5T FF 3-fin	freq	-5%	13%	17%	21%
	power	10%	-20%	-25%	-29%
	area	-41%	-29%	-41%	-29%

Fig. 9. On the left: impact of the LGAA PMOS device strength illustrated on 2 CFET libraries with different nanosheet widths, with and without PMOS stress. On the right: 5 CFET libraries are compared to 3 reference finFET libraries. The table summarizes the PPA differences between CFET and finFET. Low, mid and high refer to the CFET drive strength, as 1-fin, 2-fin and 3-fin for finFET.

As expected, the CFET libraries with a stack of 3 nanosheets outperform the finFET in almost all cases while always being smaller. When comparing the 5T CFET libraries to the single-fin 6.5T finFET, the CFET ranges from 28 to 51% faster at iso-power, or 35 to 50% power savings at iso-speed, for the low and mid-drive strength respectively, while being 23% smaller (5 vs 6.5T). Compared to the 2-fin, this now

ranges from 5 to 26% faster, or 8 to 34% power savings, for the low and mid-drive respectively, and 23% smaller. The 6T CFET with high-drive offers 31% more performance or 41% power savings, with an 8% area benefit (6 vs 6.5T). Finally, if compared to the 3-fin 8.5T finFET, the 5T CFET low-drive is slower by only 5%, or uses 10% more power, while the 5T mid-drive is 17% faster, or uses 25% less power, with a 41% area benefit (5 vs 8.5T). The 6T high-drive offers 21% more performance or 29% less power usage, with 29% area benefit (6 vs 8.5T).

CFET can provide area gains at the cell level and the potential to significantly outperform finFET. Nevertheless, if CFET is a roadmap solution, the industry will adopt it in 2 or 3 nodes after the transition from finFET to LGAA. The PPA gains must then reflect that. As shown here, the device still matters, finding ways to effectively balance N and PMOS drive strength will be critical. On the other hand, the capability to finely and broadly tune the drive strength within the same cell height is a very interesting and an important feature for designers. A library with high drive strength will be fast but tends to be big and burn a lot of power when speed is not required. Inversely, one with low drive strength will be power-efficient, but will fail to boost speed where needed. Designers typically push the technology through different libraries to provide the flexibility to cover a large PPA space, and achieve the best PPA tradeoffs in the various parts of their designs. CFET enables that design flexibility with finer granularity than finFET.

B. Place and route enabled

Very little difference is observed at the standard cell level between M0 and BPR, and that can be expected. Power delivery network (PDN) and other elements like the clock signal tree are evaluated properly only at block-level. The work presented here provides the necessary inputs to enable these experiments as illustrated in fig. 10. A CFET library with BPR and a finFET library with M0 are compared on a wire-dominated design made of around ten thousand cells. As the tool improves the density of the placement, violations of design rules occur, indicating the best area achieved. The results indicate a potential 29% density gain for CFET BPR compared to finFET M0. For scaled libraries, CFET allows for better integration of BPR than finFET, and BPR is critical for back-side PDN [8].

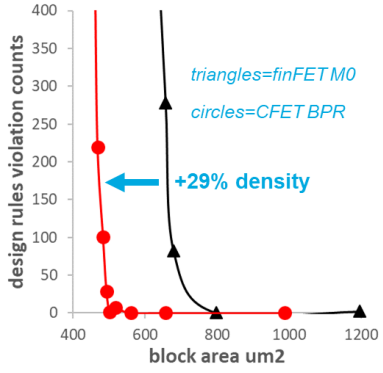


Fig. 10. Design rules violations as a function of the achieved block area after place and route. The block area achieved with CFET BPR is potentially 29% better than finFET M0.

C. Cost modeling

Wafer cost (including cycle time, depreciation, materials, facilities and labor) is just as important in Pathfinding as integration choices and PPA analyses. Our efforts have focused on automating the cost evaluation of any process integration flow to perform relative comparisons. Each process step of an emulation integration flow is linked to an industry-calibrated cost model, delivering accurate, fast and highly flexible relative cost comparisons. Any integration flows and variants can be compared on the fly, providing directional correctness for cost as well. A preliminary assessment of the front-end-of-line up through M1 indicates a 16% cost increase for CFET compared to the reference finFET. Accounting for the dilution of this added cost across a common back-end-of-line, and considering the higher transistor density achievable with CFET, the cost-per-transistor for CFET is anticipated to be competitive.

IV. Summary

We developed a rigorous yet fast and agile technology prototyping platform dedicated to embrace the uncertainty of the scaling roadmap. Fueled by many new innovative and disruptive ideas, the platform efficiency makes it practical with limited resources to start tackling the explosion of technology choices by quantifying their PPAC impact at the standard cell level, but also by enabling block-level explorations for a more comprehensive holistic approach. SPnR is a complex field rich in expertise, and exploring this space extensively as well in Pathfinding will uncover many more technology innovations.

Acknowledgements

The authors acknowledge the valuable collaborations with Synopsys and IC Knowledge. This work would not have been possible without their support.

References

[1] J. Ryckaert et al., “The Complementary FET (CFET) for CMOS scaling beyond N3” 2018 IEEE Symposium on VLSI Technology

[2] M. Garcia Bardon, Y. Sherazi, P. Schuddinck, D. Jang, D. Yakimets, P. Debacker et al, “Extreme scaling enabled by 5 tracks cells: Holistic design-device co-optimization for FinFETs and lateral nanowires”, IEDM, 2016, pp. 28.2.1-28.2.4

[3] L. Liebmann, V. Gerousis, P. Gutwin, X. Zhuc, and J. Petykiewiczc, “Exploiting Regularity: Breakthroughs in Sub-7nm Place-and-Route” Proceeding SPIE, Design-Process-Technology Co-optimization for Manufacturability XI, vol. 10148, May 2017

[4] N. Paydavosi, S. Venugopalan, Y. S. Chauhan, J. Duarte, S. Jandhyala, A. Niknejad, et al., "Bsim-spice models enable finfet and utb ic designs", IEEE Access, vol. 1, pp. 201, 2013

[5] M. Miura-Mattausch, “Compact Modeling Perspective – Bridge to Industrial Applications”, International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), 2019, pp. 319-322

[6] Nagel, L. W, and Pederson, D. O., “SPICE (Simulation Program with Integrated Circuit Emphasis)”, Memorandum No. ERL-M382, University of California, Berkeley, Apr. 1973

[7] Open Source Liberty. http://www.opensourceliberty.org

[8] D. Prasad et al., "Buried Power Rails and Back-side Power Grids: Arm® CPU Power Delivery Network Design Beyond 5nm," 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2019, pp. 19.1.1-19.1.4