Micro/Nano Circuits and Systems Design and Design Automation: Challenges and Opportunities

By GERT CAUWENBERGHS, *Fellow IEEE*

Department of Bioengineering, University of California at San Diego, La Jolla, CA 92093 USA

JASON CONG, *Fellow IEEE*

Department of Computer Science, University of California at Los Angeles, Los Angeles, CA 90095 USA

X. SHARON HU, *Fellow IEEE*, and SIDDHARTH JOSHI, *Member IEEE*

Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556 USA

SUBHASISH MITRA, *Fellow IEEE*

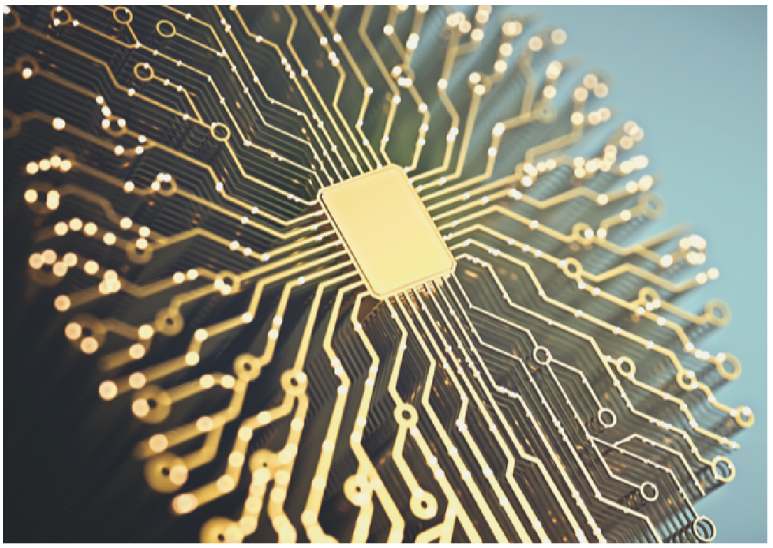
Department of Electrical Engineering and the Department of Computer Science, Stanford University, Stanford, CA 94305 USA

WOLFGANG POROD, *Life Fellow IEEE*

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA

H.-S. PHILIP WONG, *Fellow IEEE*

Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA



Digital Object Identifier 10.1109/JPROC.2023.3276941

I. INTRODUCTION

The field of design and design automation of micro/nano circuits and systems promotes interdisciplinary research spanning computer science, computer engineering, and electrical engineering. This field has created key technologies without which it would be impossible to achieve advances in information processing, which is an inseparable part of our everyday lives. For example, fundamental principles and tools created by this field have empowered Moore's law scaling for over 50 years. Without design and design automation of circuits and systems, it would be impossible to create multibillion transistor integrated circuits (ICs)

that form the foundations of today's information age.

Like other technical fields, existing approaches in the field of design and design automation are facing challenges. For example, traditional ways of improving silicon CMOS technologies or designing, verifying, and testing ICs and systems are approaching various limits such as physical size, power, and reliability limits, as well as complexity limits. At the same time, our dependency on such systems continues to grow. This creates major research opportunities for new approaches beyond conventional paths. Moreover, the recent rise of machine learning and artificial intelligence (ML/AI) applications, the recent trend toward domain-specific accelerators and computing at the edge, and recent progress in NanoSystems enabled by beyond-silicon CMOS technologies create new opportunities for customized solutions to designing electronic systems in contrast to general-purpose processors of the 20th century. By NanoSystems, we refer to systems across multiple scales—from IC chips all the way to very large-scale systems—built on nanotechnology foundations. The overall systems aspects, coupled with nanotechnologies that form the foundation, are emphasized.

The breadth of these challenges spanning multiple domains has spurred a concerted effort of highly interdisciplinary research, crossing boundaries between several traditionally separate fields of investigation. For example, more and more design and design automation researchers are collaborating with (and contributing to) adjacent fields such as ML/AI, cybersecurity, edge computing, and device technologies. Such cross-disciplinary interactions raise several natural questions including: 1) what are the high-risk and high-return research topics? 2) what other adjacent fields should electronic design automation (EDA) research aggressively seek collaborations with? 3) where and how scientific findings should be disseminated in order to have the greatest impact given the interdisciplinary nature of EDA research? and 4) where should research funding come from and how should it be distributed to encourage more transformative research? Answering these questions requires forward thinking and planning.

This article addresses these questions by providing a high-level overview of the current state and future directions/needs along the following five themes[1]: EDA (Section II), foundational technologies and NanoSystems (Section III), ML/AI/brain-inspired (BI) hardware design (Section IV), physics-inspired hardware design (Section V), and application domains beyond circuits and electronic systems (Section VI). Fig. 1 shows these five themes and the interdependencies among them. While much progress on EDA will be in the realm of conventional technology and refinements thereof, we expect to see in the future a heavy emphasis not only on heterogeneous integration and BI hardware but also on emerging technologies, where the computation will take advantage of the specific attributes of the physics-based dynamics of novel materials and structures. Specific examples of the interdependencies shown in Fig. 1 include new EDA tools that translate application needs (e.g., energy, throughput, and security) into technology targets (e.g., improvements in logic, memory, and connectivity), which will guide technology researchers, or new EDA methods to quickly emulate physical systems (e.g., nonlinear dynamical systems) in real time for physical computing applications. Additional details and more examples are elaborated in Sections II–VI. Two critical cross-cutting issues—infrastructures for supporting research and education, and education and workforce training—are summarized in Sections VII and VIII, respectively.

[1]Much of the content of this article is based on the views expressed by the roundtable attendees, speakers, and panelists of the NSF Workshop on Micro/Nano Circuits and Systems design and Design Automation: Challenges and Opportunities, held on December 14–16, 2020.

II. ELECTRONIC DESIGN AUTOMATION

EDA tools and methodologies have played a central role in managing the exponential increase of design complexity due to Moore's law scaling and powered the electronic industry to realize a cost-efficient digital revolution. However, it is significantly underinvested compared to (such as computer networking or ML). Exponential scaling according to Moore's law will continue for at least another decade, despite numerous technical challenges. Beyond that, micro/nano circuits and systems are expected to get even more complex. As a result, the design complexity has also grown exponentially, demanding more efficient and scalable EDA technologies and tools for compute, memory, and interconnect design and optimization.

The fast evolution of integrated computing systems in the late 1980s and 1990s was followed by a revolution in the Internet and wireless communications systems over the last two decades. This has driven the proliferation of wirelessly interconnected distributed computing systems for a range of real-time applications: humanoid and other robots [1], [2], [3], self-driving cars [4], [5], unmanned aerial vehicles [6], [7], and sensor networks [8]. Novel EDA tools are needed to consider computation versus communication costs across distributed compute nodes. The underlying (adaptive) hardware architectures will be highly heterogeneous, consisting of processor cores, GPUs, various kinds of domain-specific accelerators (e.g., those for ML), sensors, mixed-signal components, and wireless interfaces. Verification, post-silicon validation, design debug, manufacturing test and post-manufacture tuning of such systems will pose key challenges since existing design validation and testing techniques are not directed at high levels of heterogeneity and real-time adaptivity.

Safety and dependability of the electronic system (e.g., the control of autonomous vehicles) rely heavily on

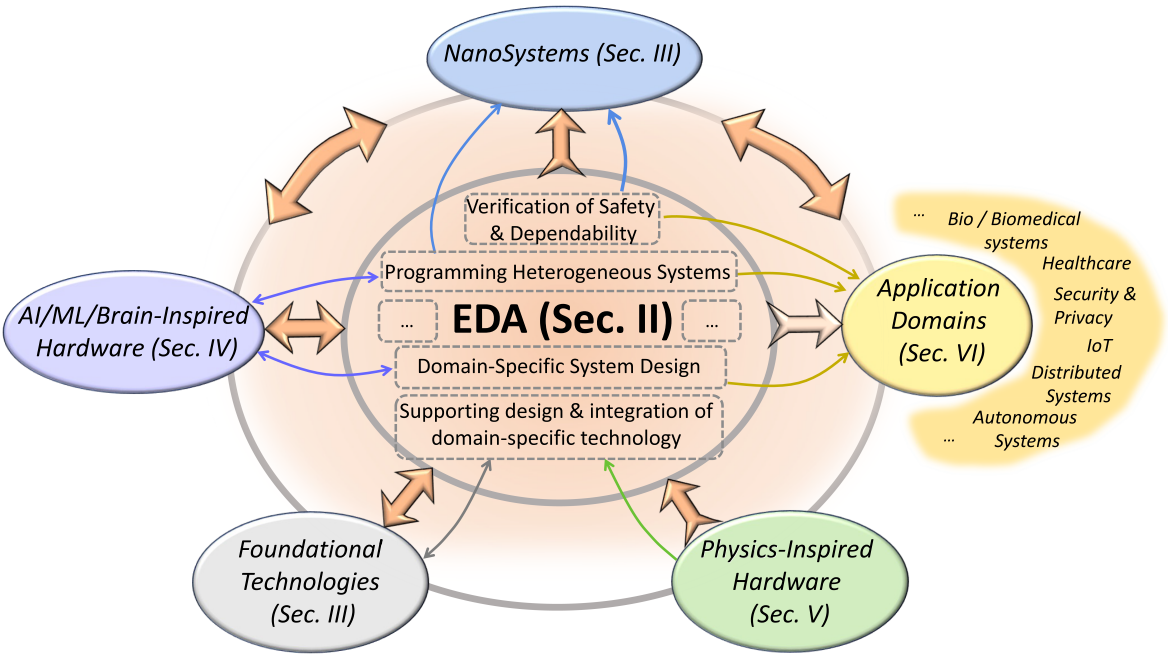


Fig. 1. *Depiction of the five themes covered in Sections II–VI and the interdependencies among the themes, from foundations and advancement of EDA (Section II) to application of EDA in domains ranging from foundational technologies (Section III) to circuits and electronic systems (Sections IV–V) to various application domains (Section VI).*

the verification capability. Verification remains a significant challenge for modern EDA flows. From pre-silicon to post-silicon, all the way to system integration, verification takes a significant and growing amount of the product development cycle. Despite consistent progress in verification (e.g., [9], [10], and [11]), developments to date are far from sufficient to meet the rapid advancement of complex electronic systems. Given the exponential scaling of circuit size, it is not an exaggeration to say that none of the large, industrial-scale designs have been completely formally verified. Most went through bounded checking before the verification time and resources were exhausted. Moreover, as we raise the level of design abstraction to behavioral C/C++ or even domain-specific languages such as TensorFlow and Halide, the growing semantic gap makes verification much more challenging. Verification and validation tend to be largely done on flat designs, which is the primary limitation of its scalability.

While traditional EDA focused mainly on platform creation, we believe that EDA can play an even bigger role in the realization of future heterogeneous platforms. Domain-specific hardware acceleration is one of the most promising approaches to combating the stagnation of single-thread performance scaling [12], [13]. Along this line, EDA has seen enormous success for decades in addressing the challenges of designing and implementing highly complex heterogeneous computing devices that feature massive parallelization and extensive specialization. There are a host of new opportunities for EDA to enable software-inclined developers to productively exploit accelerators for a novel domain. Domain-specific accelerators often evolve rapidly in size, topology, and capability to adapt to changes in application demands [14]. Hence, it is crucial to support quick (in days instead of many months) bring-up of software stacks that can adapt to a moving target "instruction set architecture." To address these challenges, it is necessary to rethink the abstraction and objectives of EDA algorithms and tools by providing domain experts with a better tradeoff between design optimality, agility, and scalability.

Ultimately, EDA is successful if it enables the design process to scale. The key to this is being able to make higher quality decisions earlier. Enabling rapid design and development of complex analog IP remains a grand challenge for current EDA systems [15], [16]. Enabling automated generation and layout of critical components or automatically migrating process nodes can reduce time to product dramatically, saving many engineer months of time [17], [18], [19]. Similar gains can be achieved in the realm of domain-specific accelerators. For example, there are billions of systolic array configurations for a given design [20]. It is not feasible to evaluate every configuration all the way down to the detailed physical design. A good predictive model is desired. When compute and schedule resources can be applied with greater effect, we enable the scaling of solution quality. This highlights several intertwined needs. 1) scaling of the design process requires ability to

"see ahead," i.e., to predict outcomes of downstream design optimization steps (e.g., [21] and [22]); 2) there is a need to recover solution suboptimality that has been left on the table over the course of decades, as the EDA industry and its research were driven by turnaround time requirements [23]; 3) we need to quantitatively measure the suboptimality gap of the EDA solutions and identify opportunities for improvement, e.g., through measuring the quality of leading academic and industrial tools [24]; and 4) we should reach out to other research communities, such as applied mathematics, statistics, operation research, and theoretical computer science to work together to constantly expand the EDA optimization toolbox. Recent work on using graph neural networks for performance prediction provides promising examples [25], [26].

A. Research Needs

We summarize the needs for coordinated research efforts in the field of EDA as follows.

1) New EDA approaches to address exponentially increasing complexity at all stages of design, test, and in-field operation.
2) New EDA approaches for new families of systems enabled by a wide variety of novel 2.5-D and 3-D integration technologies.
3) Special emphasis on robust operation with resilience to functional bugs, manufacturing defects, reliability failures, and security attacks.
4) New EDA methods that can automatically capture complex analog/mixed-signal design constraints and facilitate rapid design including novel circuit topologies.
5) New EDA methods to facilitate NanoSystem designs based on analog subsytems, emerging logic, memory, and integration technologies.
6) New formulation of EDA problems based on theoretical foundations in optimization and ML. In particular, given recent interest in the use of ML for EDA, interactions and collaborations should be promoted between EDA researchers and ML experts to jointly address the challenges of ever-increasing design automation challenges.
7) New EDA tools beyond hardware platform creation. For heterogeneous and accelerator-rich computing systems, programmers must navigate a large design space for performance optimization. This problem gets even more complex as the accelerators evolve rapidly. It is crucial to support efficient compilation and runtime systems that can adapt to new accelerators quickly. By extending the traditional EDA methodology beyond hardware platform creation, the EDA research can benefit not only tens of thousands of hardware designers but also millions of software programmers and data scientists as well.

III. FOUNDATIONAL TECHNOLOGIES AND NANOSYSTEMS

Nanotechnologies are the foundations for building NanoSystems. Coming generations of abundant-data applications will process unprecedented amounts of loosely structured data (such as streaming video, natural language, real-time sensor readings, contextual environments, or even brain signals) to overcome global grand challenges [27]. Yet, at this exact moment, when 21st-century applications are demanding the largest improvements in computing performance, conventional approaches to improving performance are stalling.

The computation demands of 21st-century applications far exceed the capabilities of today's systems, from energy-constrained embedded systems all the way to the cloud, and cannot be met by isolated "business as usual" improvements in technology, circuits, and architectures. Fortunately, there are many innovative research ideas at the level of foundational technologies and also at the level of NanoSystems in response to these computation needs. As discussed in [28], the three pillars for future computing systems include technologies for logic, memory, and the connectivity between memory and logic. A wide variety of foundational technologies and NanoSystems are being pursued by researchers, too many to be covered exhaustively in this article. Several beyond-traditional silicon technologies have been implemented in industrial facilities (e.g., [29], [30], [31], [32], [33], [34], and [35]). NanoSystems that leverage the unique properties of various foundational technologies enable new and transformative architectures (e.g., [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], and [55]).

The combination of foundational technologies and NanoSystems architectures promises to deliver unprecedented functionality, performance, and energy efficiency for future computing systems. Without such continued advances, these next-generation applications cannot be realized.

NanoSystems research focuses on innovations at the circuits and architecture levels (and associated design methodologies) enabled by novel nanomaterial, nanofabrication, and nanodevice concepts. From a hardware demonstration standpoint, we expect that NanoSystems research will strive to build at least medium-scale circuits and systems to demonstrate the effectiveness, practicality, and scalability of new concepts. This is in contrast to traditional nanomaterial, nanofabrication, and nanodevice research efforts that often focus on hardware at the scale of a few transistors or a few memory cells (often fewer than 1000). Medium- and large-scale circuits and systems can be built on top of existing silicon infrastructure, e.g., new nanotechnologies integrated on top of silicon wafers to demonstrate interesting circuit- and system-level capabilities.

Computing systems today are (and will continue to be) dramatically different from the 20th-century ones. Domain-specific accelerators are already rising as the speed and energy benefits of classical technology scaling diminish. The diversity of applications, algorithms, and accelerator hardware architectures is changing very rapidly. For example, more than 200 hardware accelerators for AI inference and training have been published over the past 3–4 years. Beyond AI, hardware accelerators for data analytics, graph processing, genomics, and security are also growing. Similarly, an explosion of new concepts in foundational technologies and NanoSystems is also emerging: not only new transistor technologies but also a wide variety of memory technologies, new sensing technologies, new interconnect technologies, new on-chip and interchip integration technologies, and new thermal technologies. Unlike innovations mostly around the silicon transistor and its miniaturization in the past, there is growing recognition about combining these wide varieties of technologies in innovative ways to create new architectures optimized for various application domains.

A. Research Needs

Based on the above discussions, we identify the following needs for coordinated research efforts in the area of foundational technologies and NanoSystems design:

1) EDA for Foundational Technologies and NanoSystems: There is growing recognition that combining the rapidly growing variety of technologies in innovative ways is crucial to create new architectures/NanoSystems optimized for various application domains. Such approaches require a new set of EDA tools, different from current commercial offerings. Development of such EDA tools must consider not only classical metrics (energy, throughput, and cost) but also emerging metrics (e.g., security, privacy, accuracy of results, robustness to manufacturing, and environmental variations). Many emerging nanotechnologies are expected to be used to realize accelerators, and their ultimate adoption will be decided by strong competition among candidate approaches. EDA will play a crucial role in supporting this competition with ammunition of even strengths for various contenders. There must be tight integration between new technology development and new EDA for the following reasons.

1) New EDA tools must translate application needs (e.g., energy, throughput, and security) into technology targets (e.g., improvements in logic, memory, and connectivity) that will guide technology researchers.
2) EDA acts as a technology enabler (as articulated in Section II) to unlock the potential benefits of new technologies.

2) Codesign for NanoSystems: There is an immediate need for new research efforts focusing on codesign for NanoSystems, connecting hardware circuits and architectures with applications on one end of the spectrum and foundational nanotechnologies on the other end—a codesign approach. Three examples are given as follows.

1) Connect abundant-data workloads (e.g., speech and video processing, graph processing, data analytics, and security) with new nanotechnologies [56].
2) Connect the wide variety of (existing and new) ML/AI models with new nanotechnologies [42].
3) Connect emerging models of computation (stochastic computing and p-bits, approximate computing, Ising, and others) with new nanotechnologies [52], [57], [58] to realize a wide variety of NanoSystems (including digital, analog heavy, low temperature, superconducting, coupled oscillators, thermodynamic, and other implementations).

3) Hardware Prototyping and Benchmarking: Demonstration of medium-to-large-scale hardware prototypes for codesigned NanoSystems using the nanotechnologies established (by leveraging the infrastructures to be discussed in Section VII) should be strongly encouraged. Such projects are expected to be major efforts with high costs, high risks, and high rewards and may be conducted on a five-year scale. Adaptivity is critical as various factors can change in the course of such ambitious projects.

Benchmarking foundational technologies in order to assess their benefits at the application level is essential to evaluate the myriad of technology, circuit, and architecture alternatives. Developing well-accepted benchmarks requires close collaboration among technology, EDA, system, and application researchers. Industry support is also indispensable.

IV. ML/AI/BI HARDWARE DESIGN

The growing impact of ML, AI, and BI algorithms over the past decade has made them a prime target for hardware acceleration. Indeed, as ML/AI/BI algorithms have evolved over time, their deployment has placed ever-increasing demands on the capabilities of the underlying hardware. Ever newer accelerators are being developed to meet these demands while also driving algorithmic progress in the field, as evidenced by recent large-scale models (GPT-4, GPT-3, GLaM, and Megatron) [59], [60], [61], [62]. Further gains in accelerator design might be possible through emerging technologies [63]. Many such technologies offer strikingly different tradeoffs compared to the conventional CMOS paradigm [64]. Extracting efficiencies will require careful explorations of this design space through hardware–software algorithm codesign. Such research can be driven by new algorithmic innovations or through empirical or utilitarian considerations.

Contemporary ML/AI/BI accelerators predominantly leverage digital implementations due to greater design productivity, ease of

verification, improved scalability, and improved resilience when compared to analog systems of similar complexity and scale [65]. In addition, digital accelerators are well positioned to taking advantage of various algorithmic techniques that can improve ML/AI/BI algorithm efficiency such as sparsity, compression, and low-rank approximations [66]. This trend is further reinforced by the need to accommodate rapidly evolving ML algorithms. A concomitant issue is how to verify such ML/AI/BI hardware, perhaps supplanting known formal or semiformal methods. New research is also necessary to identify hardware bugs/faults and methods to mitigate their effect on hardware performance in the context of ML applications.

While the advantages offered by digital accelerators are clear, analog computing solutions may offer an elegant path toward extreme energy efficiencies that rival that of the human brain. Specifically, analog computation has the potential to deliver extreme energy efficiency and computational density, particularly when the application might be amenable to low-precision computation [67]. A particularly promising architecture for analog computation, in-memory computing, has been the focus of recent research in ML/AI/BI accelerators (e.g., [68], [69], and [70]). In addition, analog computation is particularly well suited to application spaces where machine intelligence must be applied, in real time, on signals acquired from physical sensors, such as intelligent imaging, cognitive transceivers, closed-loop implants, and sensors for robotics. Indeed, analog computation near sensors might be pertinent to AI/BI algorithms inspired by embodied intelligence. However, analog computing systems—in-memory computing and near-sensor computing—need more in-depth study before they become practical. The techniques employed for analog computation open-up another avenue for codesign and optimization when deployed as part of sensing and decision systems. Given the unique advantages of analog computing and digital computing, research in the field must embrace a heterogeneous approach where both analog computing and digital computing coexist [71].

Breakthrough ML/AI/BI advances benefit from highly interdisciplinary interactions among computer engineers, algorithm designers, EDA tool developers, and increasingly cognitive scientists, neuroscientists, and bioengineers. Such synergy will elevate the understanding of how the embodied brain computes and interacts with its environment toward more autonomous, effective, efficient, and resilient operation of computing machinery.

A. Research Needs

Based on the above discussions, the needs for coordinated research efforts in the area of ML/AI/BI-inspired hardware design are summarized as follows.

1) Novel Technologies and Architectures for ML/AI/BI: Emerging circuits and technologies promise orders of magnitude improvement in accelerator performance. Hardware-aware codesign of algorithms (discussed later) will be crucial to actualize the gains offered by such systems. Accelerators employing analog computation, e.g., in-memory computing, must overcome analog impairments and nonidealities, which are typically exacerbated in emerging technologies. Tools that improve a designer's productivity for analog computation will also be required to make such systems practical. It is critical that systems employing analog circuits or emerging technologies can be fabricated and prototypes developed, as outlined in Section VII.

2) Codesign for (ML/AI/BI) Systems: More research is required to encapsulate the end-to-end benefits of incorporating different architectures within a system. It is imperative that future research is scaled beyond small-scale macros and capture the complexity of the interplay between various subsystems. Hardware–software codesign—across technology, circuit, architecture, algorithm, system, and application levels—is critical to meet the rapidly growing demands of ML/AI/BI-inspired algorithms. Such codesign efforts are best accompanied by medium- and large-scale hardware and system prototyping of ML/AI/BI systems by leveraging the infrastructures to be discussed in Section VII.

3) Cross-Disciplinary Coexploration of ML/AI/BI Systems: Cross-disciplinary coexploration of ML/AI/BI systems, especially between computer science/engineering, cognitive science, and neuroscience should be strongly encouraged. In particular, collaboration with domain experts in nontraditional application domain fields, e.g., robotics, bioengineering, and high-energy physics, can drive transformative research. These include neuromorphic engineering approaches to modeling brain function in physical systems for sensing, perception, cognition, and action rooted in the biophysics of neural computation further discussed in Section V.

V. PHYSICS-INSPIRED HARDWARE DESIGN

Physical computing combines physics and computation in a complementary and synergistic fashion. On the one hand, one can exploit physics to efficiently perform a computational task, and, on the other hand, one can view computation as emerging from physics. Computing with physics encodes computational variables in physical quantities, and the computation is performed by exploiting the physics of that particular medium. *Physics as Computing* interprets physical state variables as computational quantities, and the time evolution of the physical system (according to the *Laws of Physics*) realizes the computation.

While hardware implementation of all algorithms ultimately involves physical implementations, they do not necessarily mimic natural laws that our models (of physics) are meant to follow. Physics-inspired hardware

design seeks to map the problem at hand to the behavior of a system described by the natural laws of physics. The time evolution of that system should be describable not only by the (state) variables having physical interpretations but should also be constrained by fundamental conservation laws, e.g., energy, momentum, and entropy. Depending on the structure of the computational problem under consideration, the mapping may range from very simple to exceedingly complex, e.g., it could be an elementary one-to-one mapping (in which case a physical medium may be envisaged as a hardware substrate for computation), an isomorphism, or, at an extreme, disparate dimension and be highly complicated and nonlinear. The potentially complex (nonlinear) nature of these mappings can give rise to behaviors of the hardware implementation significantly deviant from the behaviors required by the intended physics of the system (e.g., deviate from meaningful concepts of robustness, stability, dissipativity, and so on), which may involve additional compensation.

Physical computing involves algorithms operating on state variables realized by physical quantities in physical domains (such as electric, magnetic, photonic, and plasmonic) and in time, which may be any combination of continuous and/or discrete representations [72], [73], [74], [75]. A crucial aspect is the mapping of the structure of a computational problem to an appropriate physical system [76]. Generally, physical computing may involve several levels of abstraction to address a wide hierarchy of spatial and temporal scales. In contrast to the conventional digital approach, physical computing even may take advantage of noise (to escape local minima, e.g., in search problems) [77] and may include the phase of a physical quantity (in addition to its amplitude) [78].

Examples of this approach include the large body of work on BI neuromorphic hardware design [79], [80] and more recent approaches based on Ising machines [81], [82] and networks of coupled oscillators [83], [84], [85].

A. Research Needs

In light of the above discussions, we identify the following need for coordinated research efforts in the area of physics-inspired hardware design.

1) EDA Support for Physical Computing: An EDA community ecosystem doing physical computing is needed to accelerate development at manageable timescales, ideally in real time on physical emulation platforms, e.g., [86], [87]. The development of tools and computational framework for physical computing applications becomes essential for its long-term development. This ecosystem means that education of the current and next generation of researchers must happen to empower these physical computing approaches. Given the interdisciplinary nature of these efforts, as well as the need to develop the larger computing stacks for these technologies, we need to raise up tall-thin people characteristic of the early digital VLSI development [88] as well as experts in the various subdomains.

VI. APPLICATION DOMAINS BEYOND CIRCUITS AND ELECTRONIC SYSTEMS

EDA methodologies and tools have achieved great success in managing the enormous complexity in electronic systems by building on the general principles of abstraction (bottom-up) and refinement (top-down) as well as decomposition and composition for both verification and design. Such design automation principles and methodologies are equally applicable to designing other *engineered systems*.

A number of application domains adjacent to electronic circuits and systems have directly benefited from EDA but also present unique challenges. One example is the development of ICs based on many flourishing beyond-CMOS technologies ranging from phase change to resistive switching arrays, from spintronic to ferroelectric devices, and from superconductive Josephson junctions to nanophotonic devices (e.g., [89], [90], [91], [92], and [64] and also see Section III). Such new systems will be heterogeneous in nature and will utilize a variety of emerging technologies while coexisting with deeply scaled CMOS. Another example is designing large-scale systems—systems-of-systems, which requires expanding the scope of traditional EDA in order to consider complex computing platforms, software, and physical plants as well as safety and security issues [93]. Autonomous systems are another example, which are safety-critical and/or demand high availability, thus requiring research into assured autonomy [94].

Beyond the adjacent application domains, a number of other application domains (e.g., smart buildings [95] and electric vehicles [96]) have also embraced systematic design automation processes that bear similarity to EDA approaches. Some representative application domains that are "farther away" from electronic systems include: 1) design of reliable and secure large-scale networks [97]; 2) synthetic biology that is built upon genetic engineering by adding the engineering principles of standards, abstraction, and decoupling [98], [99]; 3) development of microfluidics lab-on-chip technology and its adoption for microbiology [100], [101]; and 4) modeling and design of biomedical systems and drugs [102], [103].

The confluence of design automation in these application domains brings new challenges and opportunities. One could envision a field of "ESDA" as a natural evolution of the current design automation solutions from different application domains. For example, the concept of platform-based design methodology was first introduced for embedded system design [104], and later, it was extended to smart building design [95]. This methodology, including abstraction, formalization, and systematic design flow, can be

further developed into ESDA theories that are readily adopted by other engineered systems. For ESDA to be widely applicable, it will need to deal with the proliferation of complex, large, CPSs where software is as important as hardware in determining the functionality and performance of the target-engineered system and where regulatory requirements and new concerns (such as self-learning, privacy, and sustainability) must be met.

A. Research Needs

Based on the above discussions, we summarize the needs for research efforts in leveraging EDA principles to develop design automation methodologies and techniques for emerging and new application domains as follows.

1) Design Capabilities for Heterogeneous Systems: Integrated design capabilities need to be developed to support heterogeneous systems that combine a variety of CMOS and beyond-CMOS technologies across multiple platforms, including 3-D integration, silicon in a package (SiP), and wafer-scale integration (WSI). Many local and global sources of variability in extreme-scaled CMOS devices and circuits as well as the inherently random nature of many beyond-CMOS technologies demand that EDA platforms and tools properly model and cope with deterministic and stochastic behaviors and noisy inputs/outputs while supporting approximate (imprecise) computations. Next-generation EDA tools will also need to support other types of heterogeneity, such as large-scale multiphysics-based heterogeneous systems, heterogeneous timing paradigms ranging from fully synchronous to self-timed systems, and mixed-signal and RF circuits integrated with on-chip antennas. Designing such heterogeneous systems will require design algorithms, techniques, and tools that support the full spectrum of capabilities, including synthesis, simulation, modeling, verification, and test.

2) Domain-Specific Design Automation: Dedicated interdisciplinary research projects that draw on EDA principles should be supported for exploring design automation of engineered complex systems arising from a variety of application domains, including autonomous systems, biomedicine, drug discovery, and synthetic biology. EDA's deep expertise in creating domain-specific abstractions, algorithm development, and optimization would facilitate the development of tools that can greatly expedite the design and validation of such complex systems.

3) New Design Concerns: New design concerns, such as autonomy, privacy, resiliency, and sustainability of computing systems, can be on par with more traditional design concerns such as area, speed, reliability, and power efficiency. EDA platforms and tools must model and enable meaningful tradeoffs among these often-conflicting concerns in a way that would empower not only the product developers but also the end users of the electronic products. To maintain design quality and to safeguard evolution in the field operation phase, EDA support should be extended to the field with highly automated versions of EDA tool functionality, including the related model base for self-modeling and context modeling. New tools for dependency and automated failure analysis will be needed, as well as for model adaptation and model error detection.

VII. INFRASTRUCTURE

The U.S. government supported the MOSIS program [105] that starting around 1981 unleashed the innovation of circuit designers and enabled circuit research and education to proceed by way of abstractions that decoupled circuits research from device technology research. Fast forward 40 years and the needs of today are drastically different. End-user design innovations are now strongly coupled with chip-/system-architecture innovations. Circuit/architecture innovations often derive from the use of new device and integration technologies, and conversely, device technology innovations are driven by application needs and require circuit/architecture level optimizations and demonstrations to be relevant. In short, codesign across the technology stack is the future of tomorrow's systems, and innovations and investments are needed to push beyond the traditional approaches. University clean rooms (such as those supported by the NSF NNCI [106]) today are missioned to facilitate basic science discoveries and engineering research at the single or few devices level. These facilities, while they are successful in fulfilling their stated missions, do not have the capability to fabricate state-of-the-art transistors that are relevant to practical applications, nor do they have the capability to yield a large enough number of devices for meaningful circuit demonstrations. The ability to demonstrate circuit- and system-level functionality and benefits, using advanced technology nodes, or using emerging not-yet-commercialized technology, or using lab-scale technology developed at universities is the core of research that will break down abstraction boundaries to effect codesign and cooptimization—a technical direction that is highlighted by earlier studies on the subject [107].

The access to semiconductor foundry can be broadly categorized into three areas: 1) foundry access for IC designers to advanced technologies as well as commercial-class mature node technologies that allow significantly sized chips to be built; 2) foundry access for technology developers for creating new technology demonstrators; and 3) access to design ecosystems (EDA tools, design flows, and IP blocks) supporting system-level demonstrations. While such access is available to a small set of select research groups, through personal networks and serendipitous or historical connections, access is spotty across the board for most academic researchers. This has

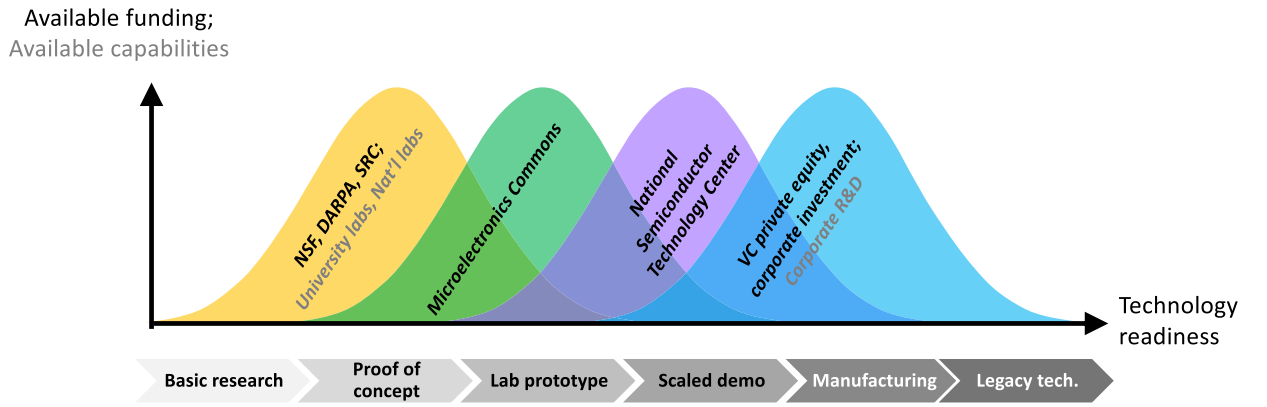


Fig. 2. ***Lab-to-fab translation fills a much-needed gap in advancing innovations along the technology readiness timeline.***

significantly hampered the pace of research and limited the opportunity to innovate to a subset of researchers. In many cases, research ideas simply cannot be executed or have to be abandoned due to the lack of access and sometimes end up being reinvented in other geographies. The net result of this access problem is a severe underutilization of a large group of talented researchers and technology developers.

To make things worse, the time cycle for hardware experimentation is currently much longer than software-only and/or simulation-based studies. The pace of progress in hardware is not keeping up with the pace of advances in software and applications. Yet, we know that the software and the hardware must go hand-in-hand. We cannot run today's software on 20-year-old hardware. More powerful software requires more powerful hardware. If hardware fails to progress, then software will shortly follow.

A. Research Needs

Progress in infrastructure will benefit from addressing the following critical needs.

1) Facilitate Access to Leading-Edge Silicon CMOS Technologies: Currently, only select groups of researchers have access to advanced technologies (silicon CMOS and beyond) and advanced integration technologies. Even those accesses are limited to two generations behind the state-of-the-art. The majority of researchers are still working with technologies that are at least three generations behind. Current practices [e.g., by Defense Advanced Research Project Agency (DARPA) and Intelligence Advanced Project Activity (IARPA)] have shown that access can be arranged when the full support of the U.S. government is brought to bear. We must find ways to broaden such access to a wider academic community. Such access includes leading-edge silicon as well as affordable access to mature nodes. This is as much an issue about access, as it is about the cost of access. Funding mechanisms should be developed that support the cost of chip design tape out in addition to the traditional cost of research in circuit design and computer/system-architecture fields.

2) Support/Establish a National Facility for Prototyping Emerging Technologies At-Scale: The establishment of the multiproject wafer (MPW) service by MOSIS has dramatically changed the landscape of circuit design education, research, and commercialization. We must find ways to demonstrate emerging device technologies at scale, beyond the 1–1000 devices scale that are sufficient for an initial exploration. A national facility should be established. The mission of such a National Facility would: 1) enable fast turn-around experimentation of chip-scale and package-scale systems; 2) achieve flexibility (of material and process technologies) at scale; and 3) facilitate lab-to-fab translation of systems technology, thereby making academic research relevant for advancing foundational microelectronics technology for the country (see Fig. 2). Such a national facility would take foundry wafers as starting materials and integrate various materials and devices on the Si CMOS foundry wafer. Similar to an MPW service, there needs to be a wafer brokerage that defines the technology and the design interface protocols. Partnership with other branches of the U.S. government may leverage the same facility to further lab-to-fab translation to industry.

3) Open Access for Design Ecosystem: EDA tools and design flows are currently proprietary and have become so complex that they are almost a black box. NSF must invest in open-source EDA tools and design flows since industry mainly focuses on proprietary tools. Using a mixture of commercial and research/open-source EDA tools (e.g., in the cloud) as well as advanced and robust IP blocks will enable a vibrant design ecosystem. The learning curve for a tape out is steep; this hampers innovation and turns many students away. There needs to be a concerted effort to make the circuit design process as easy as software development.

4) Design Enablement for Emerging Technologies: While today's advanced

EDA tools will continue to support industrial technology offerings, there needs to be a major emphasis on new design and verification tools to address emerging technologies and their complexities (in tandem with new technology capabilities created as 3) mentioned above). Without this enablement, the exciting promise of system-level demonstrations of emerging technologies cannot be fulfilled.

5) Education and Workforce Development: To remain globally competitive, there is no excuse to teach students using old technology nodes because learning on old technology nodes does not serve the students' needs as they graduate and find jobs in industry that uses leading-edge technologies that have vastly different design constraints. The task of maintaining the infrastructure to support circuit design has grown beyond the capability of individual faculty members. We must find ways to incentivize and assist universities to develop and offer engaging IC design courses using real technologies that are used in practice.

VIII. EDUCATION AND WORKFORCE TRAINING

The size, reach, and impact of design and design automation have grown significantly over the years, powered by the intellectual efforts of highly trained engineers/researchers who have decades of experience in EDA research and development. However, the entering pipeline is notably shallow, with far fewer new students (at both the undergraduate and graduate levels) choosing the technology/circuit/architecture design and design automation profession. In general, undergraduate enrolment in computer engineering and electrical engineering has been declining across universities in the United States. It is imperative that these declining numbers can be reversed to prevent a further decline in critical onshore IC manufacturing and design capabilities. Several possibilities exist: 1) fewer students find semiconductor and related jobs compelling, i.e., there is a perception that these jobs are moving away from the United States; 2) high-school and early undergraduate students migrating away from these disciplines due to a perception of software can do more; 3) outdated curricula that do not excite students; and 4) not enough young faculty and role models. In the following, a more detailed discussion on the current status, challenges, and needs related to education and workforce developments is provided.

A. Core EDA

The early 1980s and 1990s were the golden era for academic and industrial EDA: the combination of a simple set of design rules coupled with structured, hierarchical design abstractions created a level playing field for computer scientists to collaborate fruitfully with circuit designers and electrical engineers, resulting in the development of many sophisticated optimization and synthesis algorithms, as well as complex simulation frameworks that enabled early design space exploration and rapid concept-to-design cycles. This resulted in a wealth of academic research and tools for creating increasingly complex VLSI chips, in EDA areas such as circuit design [108] and physical design [109]; synthesis tools and design flows at the logic [110]; register-transfer level (RTL) and behavioral levels [111]; formal models and equivalence checking to ensure the correctness of designs generated by EDA tools [112]; and testing and validation of VLSI circuits [113]. Industry and academia collaborated actively during this period, with the annual Design Automation Conference (DAC) [114] drawing thousands of academic researchers and practitioners from a diverse set of small, medium, and large EDA companies. Indeed, the Silicon Valley start-up booms in the late 1990s also generated tremendous interest for academics to be active in start-up EDA companies. All of this excitement resulted in great interest in educating students, both at the undergraduate and graduate levels, resulting in the proliferation of many EDA courses and curricula across the country, as well as a strong pipeline of EDA professionals.

At some level, the early EDA academic and research communities became victims of their own successes in the 2000 decade, with academic and EDA research tools transitioned to the EDA companies and large chip design houses (Intel, AMD, IBM, TI, and so on). Increasingly complex and rapid device technology advances coupled with competition/secrecy in the industry with regard to advanced technology nodes and design drivers posed significant barriers for academia to know the real challenges faced by designers and sample datasets that could be used to drive impactful academic EDA innovations. Furthermore, the EDA industry underwent a significant consolidation into a few, large EDA companies, which created further barriers to cooperation between the EDA industry, design houses, and academic research.

The 2010 decade opened the floodgates for excitement and innovations in big data analytics and ML/AI. Many EDA-trained students were swept up, both by generous salaries/stock options, as well as overall excitement in these emerging fields. Government funding for EDA did not grow sufficiently to sustain the growing need for research to meet new challenges arising from advanced technologies, newer applications, and increased design complexity. This resulted in many EDA faculty reorienting their research skills in emerging non-EDA arenas, as well as fewer students pursuing EDA research and careers in EDA.

We believe that it is critical to reinvigorate the excitement of EDA as a vibrant and critical field. This requires cultivation of seed corn for growth of this field through education of the next generation of EDA students and researchers, through a multipronged approach.

B. Beyond Core EDA: Circuit and Systems Design and General Design Automation

Design and design automation go hand-in-hand. Circuits and systems designers are the users of EDA tools, while EDA tool development must respond to the needs of new foundational technologies and new applications (including ML/AI/BI applications). Furthermore, EDA principles and solutions can benefit problem domains beyond traditional semiconductor ICs.

With the conventional CMOS scaling reaching its limit, close interactions between technology and EDA tool development are indispensable. The rapidly moving ML/AI/BI field further increases the needs of connecting device technologies → circuits → architectures → systems → applications. The fields of foundational technologies and NanoSystems for emerging applications create exciting opportunities for students to make a meaningful impact. At the same time, the fields present a number of challenges in attracting undergraduate and graduate students.

1) It is important to understand the interplay between device technologies, circuits, architectures, and applications—a "cross-layer" approach. While the cross-layer approach is exciting, it is also challenging to create a practical curriculum that covers both depth and breadth sufficiently.
2) Access to latest foundational technologies is often limited to a few research groups and even more limited for classroom teaching for various reasons. The gap between industry and academic technology access has only grown. Such limited access severely limits innovations by university students and researchers.
3) The lack of infrastructure and support for student designs results in very few students getting the opportunity to even design in modern processes. This limits the knowledge of IC design to very few students. It may also reduce the attractiveness of circuit and system design courses to students.
4) Unlike many other fields (e.g., those related to the development of application software and algorithms), there can be a long cycle to obtain results and to reach gratification, sometimes spanning several years.
5) Overly simplistic messages (frequently driven by commercial motives) equating the miniaturization wall or the power wall with the end of hardware technology advances often demotivate young students from entering the field (especially in the United States).

System-on-chip (SoC) design and system-level integration skills have emerged as a critical requirement for today's workforce. SoC design and system integration are key to developing hardware-based system that can effectively and efficiently address the requirements of today's complex and multifaceted applications with varying design specifications and demanding performance requirements. Unfortunately, a majority of academic institutions of higher education are not equipped with the resources, know-how, and tools needed to train such a workforce. These institutions are very good at providing deep technical knowledge of a given field (such as networking, computing, signal processing, or communication systems) but fall short when it comes to training a skilled person in the art of combining various point solutions in computing, communication, applications, and so on into a unified hardware–software platform that addresses the applications' needs.

EDA principles and solutions have been applied in recent years to problem domains outside of traditional semiconductor ICs. For example, EDA research has broadened to encompass topics in other fields such as systems biology, lab-on-chip, smart grid, quantum computing, hardware security, AI accelerators, and CPS. There is clearly a need for innovations in education that can prepare the next generation of researchers and practitioners for the new EDA landscape. The traditional curriculum has emphasized semiconductor electronics, chip design, algorithms and formal methods, software engineering, and optimization techniques. Future innovations in curriculum design must go beyond these topics and encompass ML/AI, statistics, data science, physics of new types of devices, and the convergence with the life sciences (microbiology and biochemistry). The key is to abstract out EDA concepts that are presented narrowly in the context of chip design and present them in a broader context so that students can apply these concepts to new domains. In addition, there is an opportunity to integrate EDA concepts in the lifelong learning of working professionals in all these domains, e.g., through training workshops, tutorials, and online courses.

C. Recommendations to the Community

The EDA research community can play an important role in education and workforce training to enable next-generation system design. EDA education should continue to emphasize fundamental principles, not just techniques. In the age of AI, it is critical to educate students in learning when to use which: physical modeling or AI. EDA education should also help train practitioners in the domains where ad hoc design techniques are traditionally used so that new EDA tools and methodologies can be more widely adopted.

To address the dire needs for attracting and educating future engineers and scientists in the field of design and design automation of NanoSystems, the community must come together to work on the following aspects.

1) Create ways to attract high-school and undergraduate students to the critically important field of NanoSystems design

and design automation to advance future computing. Such efforts are essential for the United States but are missing today.

2) Create a community of acceptance and create a community of excitement and innovation around NanoSystems design and design automation, and foundational technologies, which will help attract top diverse candidates to the field.
3) Create fellowships at the undergraduate, master's, and Ph.D. levels for students pursuing research in NanoSystems design and design automation (with an emphasis on diversity as well) can make a tremendously positive impact.
4) Similar to technology access in research, create ways to incentivize and assist universities to develop and offer engaging courses on NanoSystems design and design automation, and foundational technologies (with the possibility of taping out exciting NanoSystems ideas using nanotechnologies as part of course projects).
5) Develop educational materials for the broader definition of system design and design automation. Tools can be developed to support new types of design—educational tools do not need to support everything required for industrial adoption. Benchmarks, datasets, and sample designs can be developed to enhance learning.

IX. CONCLUSION

This article examines the many opportunities and unique challenges presented to the community of micro/nano circuits and systems design and design automation. Besides the research needs and recommendations discussed in Sections II–VIII, there is an immediate need for the community to help organize and coordinate awareness campaigns in the society, at least at the levels of AI, robotics, and quantum computing campaigns, to emphasize: 1) the critical importance and tremendous potential of hardware technologies and NanoSystems to revolutionize almost every aspect of all our lives and 2) the increasingly crucial role of EDA and its growing opportunities in directly impacting hardware and software technologies moving forward. Furthermore, funding is needed to support the EDA community in creating large-scale (both in terms of problem complexity and participating teams) competitions/challenges to ignite the interest of students and young researchers in NanoSystems design and EDA. It is time to user in another golden age for electronic design and design automation. ■

Acknowledgment

This article is based on the two reports from the NSF Workshop on Micro/Nano Circuits and Systems Design and Design Automation: Challenges and Opportunities [115]. This workshop was sponsored by the National Science Foundation CISE/CCF Division under Grant CCF-2041598. The authors are grateful to Dr. Sankar Basu, an NSF Program Director in the CCF Division, for his effort in initiating and organizing the NSF workshop and providing valuable feedback to the draft of this article. They would also like to thank all the speakers, panelists, and roundtable participants of the workshop for their insightful and stimulating presentations and discussions at the workshop. The complete list of the speakers, panelists, and roundtable participants can be found in [116], [117].

REFERENCES

[1] M. Hirose and K. Ogawa, "Honda humanoid robots development," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 365, no. 1850, pp. 11–19, Jan. 2007.
[2] B. Adams, C. Breazeal, R. A. Brooks, and B. Scassellati, "Humanoid robots: A new kind of tool," *IEEE Intell. Syst.*, vol. 15, no. 4, pp. 25–31, Jul. 2000.
[3] E. Guizzo, "By leaps and bounds: An exclusive look at how Boston dynamics is redefining robot agility," *IEEE Spectr.*, vol. 56, no. 12, pp. 34–39, Dec. 2019.
[4] R. Sell, M. Leier, A. Rassõlkin, and J. Ernits, "Self-driving car ISEAUTO for research and education," in *Proc. 19th Int. Conf. Res. Educ. Mechatronics (REM)*, Jun. 2018, pp. 111–116.
[5] A. Eskandarian, *Handbook of Intelligent Vehicles*, vol. 2. Cham, Switzerland: Springer, 2012.
[6] S. Hayat, E. Yanmaz, and R. Muzaffar, "Survey on unmanned aerial vehicle networks for civil applications: A communications viewpoint," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2624–2661, 4th Quart., 2016.
[7] M. Mozaffari, W. Saad, M. Bennis, Y. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.
[8] N. Rathi, J. Saraswat, and P. P. Bhattacharya, "A review on routing protocols for application in wireless sensor networks," 2012, *arXiv:1210.2940*.
[9] F. Lonsing, S. Mitra, and C. Barrett, "A theoretical framework for symbolic quick error detection," in *Proc. Formal Methods Comput. Aided Design (FMCAD)*, Sep. 2020, pp. 1–10.
[10] E. Singh et al., "A-QED verification of hardware accelerators," in *Proc. 57th ACM/IEEE Design Autom. Conf. (DAC)*, Jul. 2020, pp. 1–6.
[11] B.-Y. Huang, H. Zhang, P. Subramanyan, Y. Vizel, A. Gupta, and S. Malik, "Instruction-level abstraction (ILA): A uniform specification for system-on-chip (SoC) verification," *ACM Trans. Design Autom. Electron. Syst.*, vol. 24, no. 1, pp. 1–24, Jan. 2019.
[12] J. Cong, Z. Fang, M. Huang, P. Wei, D. Wu, and C. H. Yu, "Customizable computing—From single chip to datacenters," *Proc. IEEE*, vol. 107, no. 1, pp. 185–203, Jan. 2019.
[13] W. J. Dally, Y. Turakhia, and S. Han, "Domain-specific hardware accelerators," *Commun. ACM*, vol. 63, no. 7, pp. 48–57, Jun. 2020.
[14] E. Chung et al., "Serving DNNs in real time at datacenter scale with project brainwave," *IEEE Micro*, vol. 38, no. 2, pp. 8–20, Mar. 2018.
[15] B. Xu et al., "MAGICAL: Toward fully automated analog IC layout leveraging human and machine intelligence: Invited paper," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2019, pp. 1–8.
[16] K. Hakhamaneshi, M. Nassar, M. Phielipp, P. Abbeel, and V. Stojanovic, "Pretraining graph neural networks for few-shot analog circuit modeling and design," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, early access, Oct. 26, 2022, doi: 10.1109/TCAD.2022.3217421.
[17] J. Han, W. Bae, E. Chang, Z. Wang, B. Nikolic, and E. Alon, "LAYGO: A template-and-grid-based layout generation engine for advanced CMOS technologies," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 3, pp. 1012–1022, Mar. 2021.
[18] T. Dhar et al., "ALIGN: A system for automating analog layout," *IEEE Design Test*, vol. 38, no. 2, pp. 8–18, Apr. 2021.
[19] A. F. Budak et al., "Joint optimization of sizing and layout for AMS designs: Challenges and opportunities," in *Proc. Int. Symp. Phys. Design*, Mar. 2023, pp. 84–92.
[20] J. Wang, L. Guo, and J. Cong, "AutoSA: A polyhedral compiler for high-performance systolic arrays on FPGA," in *Proc. ACM/SIGDA Int. Symp. Field-Program. Gate Arrays*, Feb. 2021, pp. 93–104.
[21] A. B. Kahng, "MLCAD today and tomorrow: Learning, optimization and scaling," in *Proc. ACM/IEEE 2nd Workshop Mach. Learn. CAD (MLCAD)*, Nov. 2020, p. 1.
[22] L. Guo et al., "AutoBridge: Coupling coarse-grained floorplanning and pipelining for high-frequency HLS design on multi-die FPGAs," in *Proc. ACM/SIGDA Int. Symp. Field-Program. Gate Arrays*, Feb. 2021, pp. 81–92.

[23] A. B. Kahng, "Reducing time and effort in IC implementation: A roadmap of challenges and solutions," in *Proc. 55th Annu. Design Autom. Conf.*, Jun. 2018, pp. 1–6.

[24] C.-C. Chang, J. Cong, M. Romesis, and M. Xie, "Optimality and scalability study of existing placement algorithms," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 23, no. 4, pp. 537–549, Apr. 2004.

[25] A. Sohrabizadeh, Y. Bai, Y. Sun, and J. Cong, "Automated accelerator optimization aided by graph neural networks," in *Proc. 59th ACM/IEEE Design Autom. Conf.*, Jul. 2022, pp. 55–60, doi: 10.1145/3489517.3530409.

[26] Z. Guo, M. Liu, J. Gu, S. Zhang, D. Z. Pan, and Y. Lin, "A timing engine inspired graph neural network model for pre-routing slack prediction," in *Proc. 59th ACM/IEEE Design Autom. Conf.*, Jul. 2022, pp. 1207–1212, doi: 10.1145/3489517.3530597.

[27] *NAE Grand Challenges for Engineering*. Accessed: May 25, 2023. [Online]. Available: http://www.engineeringchallenges.org/challenges.aspx

[28] H.-S. Wong et al., "A density metric for semiconductor technology," *Proc. IEEE*, vol. 108, no. 4, pp. 478–482, Apr. 2020.

[29] Z. Krivokapic et al., "14nm ferroelectric FinFET technology with steep subthreshold slope for ultra low power applications," in *IEDM Tech. Dig.*, Dec. 2017, p. 15.

[30] A. Antonyan, S. Pyo, H. Jung, and T. Song, "Embedded MRAM macro for eFlash replacement," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–4.

[31] M. D. Bishop et al., "Fabrication of carbon nanotube field-effect transistors in commercial silicon manufacturing facilities," *Nature Electron.*, vol. 3, no. 8, pp. 492–501, Jun. 2020.

[32] T. Srimani et al., "Heterogeneous integration of BEOL logic and memory in a commercial foundry: Multi-tier complementary carbon nanotube logic and resistive RAM at a 130 nm node," in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2020, pp. 1–2.

[33] C.-C. Chou et al., "A 22nm 96KX144 RRAM macro with a self-tracking reference and a low ripple charge pump to achieve a configurable read window and a wide operating voltage range," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2020, pp. 1–2.

[34] S. Beyer et al., "FeFET: A versatile CMOS compatible device with game-changing potential," in *Proc. IEEE Int. Memory Workshop (IMW)*, May 2020, pp. 1–4.

[35] P. Narayanan et al., "Fully on-chip MAC at 14nm enabled by accurate row-wise programming of PCM-based weights and parallel vector-transport in duration-format," in *Proc. Symp. VLSI Technol.*, Jun. 2021, pp. 1–2.

[36] M. M. S. Aly et al., "Energy-efficient abundant-data computing: The N3XT 1,000x," *Computer*, vol. 48, no. 12, pp. 24–33, Dec. 2015.

[37] M. M. S. Aly et al., "The N3XT approach to energy-efficient abundant-data computing," *Proc. IEEE*, vol. 107, no. 1, pp. 19–48, Jan. 2019.

[38] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, "Integer factorization using stochastic magnetic tunnel junctions," *Nature*, vol. 573, no. 7774, pp. 390–393, Sep. 2019.

[39] M. Alamdar et al., "Spin orbit torque domain wall-magnetic tunnel junction devices and circuits for in-memory and neuromorphic computing," *Bull. Amer. Phys. Soc*, vol. 118, no. 11, p. 112401, 2021.

[40] S. Dutta et al., "Monolithic 3D integration of high endurance multi-bit ferroelectric FET for accelerating compute-in-memory," in *IEDM Tech. Dig.*, Dec. 2020, pp. 36.4.1–36.4.4.

[41] E. Esmanhotto et al., "High-density 3D monolithically integrated multiple 1T1R multi-level-cell for neural networks," in *IEDM Tech. Dig.*, Dec. 2020, p. 36.

[42] M. Giordano et al., "CHIMERA: A 0.92 TOPS, 2.2 TOPS/W edge AI accelerator with 2 MByte on-chip foundry resistive RAM for efficient training and inference," in *Proc. Symp. VLSI Circuits*, Jun. 2021, pp. 1–2.

[43] G. Hills et al., "Modern microprocessor built from complementary carbon nanotube transistors," *Nature*, vol. 572, no. 7771, pp. 595–602, Aug. 2019.

[44] P. S. Kanhaiya, C. Lau, G. Hills, M. D. Bishop, and M. M. Shulaker, "Carbon nanotube-based CMOS SRAM: 1 kbit 6T SRAM arrays and 10T SRAM cells," *IEEE Trans. Electron Devices*, vol. 66, no. 12, pp. 5375–5380, Dec. 2019.

[45] K. Myny, E. van Veenendaal, G. H. Gelinck, J. Genoe, W. Dehaene, and P. Heremans, "An 8-bit, 40-instructions-per-second organic microprocessor on plastic foil," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 284–291, Jan. 2012.

[46] R. M. Radway et al., "Illusion of large on-chip memory by networked computing chips for neural network inference," *Nature Electron.*, vol. 4, no. 1, pp. 71–80, Jan. 2021.

[47] M. Shulaker et al., "Carbon nanotube computer," *Nature*, vol. 501, pp. 526–530, Sep. 2013.

[48] M. M. Shulaker et al., "Three-dimensional integration of nanotechnologies for computing and data storage on a single chip," *Nature*, vol. 547, no. 7661, pp. 74–78, Jul. 2017.

[49] S. Wachter, D. K. Polyushkin, O. Bethge, and T. Mueller, "A microprocessor based on a two-dimensional semiconductor," *Nature Commun.*, vol. 8, no. 1, Apr. 2017, Art. no. 14948.

[50] W. Wan et al., "33.1 A 74 TMACS/W CMOS-RRAM neurosynaptic core with dynamically reconfigurable dataflow and in-situ transposable weights for probabilistic graphical models," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2020, pp. 498–500.

[51] Z. Wang et al., "An all-weights-on-chip DNN accelerator in 22nm ULL featuring 24×1 mb eRRAM," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2020, pp. 1–2.

[52] T. F. Wu et al., "Brain-inspired computing exploiting carbon nanotube FETs and resistive RAM: Hyperdimensional computing case study," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 492–494.

[53] T. F. Wu et al., "14.3 A 43pJ/cycle non-volatile microcontroller with 4.7μs shutdown/wake-up integrating 2.3-bit/cell resistive RAM and resilience techniques," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 226–228.

[54] X. Xu et al., "11 Tops photonic convolutional accelerator for optical neural networks," *Nature*, vol. 589, no. 7840, pp. 44–51, Jan. 2021.

[55] S. Zanjani et al., "3D integrated monolayer graphene–Si CMOS RF gas sensor platform," *Nature 2D Mater. Appl*, vol. 1, no. 1, p. 36, 2017.

[56] R. M. Radway et al., "The future of hardware technologies for computing: N3XT 3D MOSAIC, illusion scaleup, co-design," in *IEDM Tech. Dig.*, Dec. 2021, pp. 4–25.

[57] J. Kaiser and S. Datta, "Probabilistic computing with p-bits," *Appl. Phys. Lett.*, vol. 119, no. 15, Oct. 2021, Art. no. 150503.

[58] N. A. Aadit et al., "Massively parallel probabilistic computing with sparse ising machines," *Nature Electron.*, vol. 5, no. 7, pp. 460–468, Jun. 2022.

[59] Wikipedia Contributors. (2023). *Generative Pre-Trained Transformer 4 Wikipedia, the Free Encyclopedia*. Accessed: May 8, 2023. [Online]. Available: https://en.wikipedia.org/wiki/GPT-4

[60] T. B. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.

[61] M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, and B. Catanzaro, "Megatron-LM: Training multi-billion parameter language models using model parallelism," 2019, *arXiv:1909.08053*.

[62] N. P. Jouppi et al., "In-datacenter performance analysis of a tensor processing unit," *ACM SIGARCH Comput. Archit. News*, vol. 45, no. 2, pp. 1–12, 2017.

[63] H.-S. P. Wong et al., "Metal–oxide RRAM," *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, Jun. 2012.

[64] A. Aziz et al., "Computing with ferroelectric FETs: Devices, models, systems, and applications," in *Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Mar. 2018, pp. 1289–1298.

[65] P. N. Whatmough, M. Donato, G. G. Ko, S. K. Lee, D. Brooks, and G. Wei, "CHIPKIT: An agile, reusable open-source framework for rapid test chip development," *IEEE Micro*, vol. 40, no. 4, pp. 32–40, Jul. 2020.

[66] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," 2017, *arXiv:1710.09282*.

[67] N. R. Shanbhag, N. Verma, Y. Kim, A. D. Patil, and L. R. Varshney, "Shannon-inspired statistical computing for the nanoscale era," *Proc. IEEE*, vol. 107, no. 1, pp. 90–107, Jan. 2019.

[68] N. Verma et al., "In-memory computing: Advances and prospects," *IEEE Solid State Circuits Mag.*, vol. 11, no. 3, pp. 43–55, Summer. 2019.

[69] C.-X. Xue et al., "A CMOS-integrated compute-in-memory macro based on resistive random-access memory for AI edge devices," *Nature Electron.*, vol. 4, no. 1, pp. 81–90, Dec. 2020.

[70] W. Wan et al., "A compute-in-memory chip based on resistive random-access memory," *Nature*, vol. 608, no. 7923, pp. 504–512, Aug. 2022.

[71] S. Joshi, C. Kim, S. Ha, and G. Cauwenberghs, "From algorithms to devices: Enabling machine learning through ultra-low-power VLSI mixed-signal array processing," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2017, pp. 1–9.

[72] J. Hasler and E. Black, "Physical computing: Unifying real number computation to enable energy efficient computing," *J. Low Power Electron. Appl.*, vol. 11, no. 2, p. 14, Mar. 2021.

[73] K. A. Sanni and A. G. Andreou, "A historical perspective on hardware AI inference, charge-based computational circuits and an 8 bit charge-based multiply-add core in 16 nm FinFET CMOS," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 3, pp. 532–543, Sep. 2019.

[74] K. Bernstein, R. K. Cavin, W. Porod, A. Seabaugh, and J. Welser, "Device and architecture outlook for beyond CMOS switches," *Proc. IEEE*, vol. 98, no. 12, pp. 2169–2184, Dec. 2010.

[75] R. Sarpeshkar, "Analog versus digital: Extrapolating from electronics to neurobiology," *Neural Comput.*, vol. 10, no. 7, pp. 1601–1638, Oct. 1998.

[76] X. Yin, B. Sedighi, M. Varga, M. Ercsey-Ravasz, Z. Toroczkai, and X. S. Hu, "Efficient analog circuits for Boolean satisfiability," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 26, no. 1, pp. 155–167, Jan. 2018.

[77] K. Y. Camsari et al., "From charge to spin and spin to charge: Stochastic magnets for probabilistic switching," *Proc. IEEE*, vol. 108, no. 8, pp. 1322–1337, Aug. 2020.

[78] G. Csaba, Á. Papp, and W. Porod, "Perspectives of using spin waves for computing and signal processing," *Phys. Lett. A*, vol. 381, no. 17, pp. 1471–1476, May 2017, doi: 10.1016/j.physleta.2017.02.042.

[79] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990. [Online]. Available: https://web.stanford.edu/group/brainsinsilicon/documents/MeadNeuroMorphElectro.pdf

[80] C. Mead, "How we created neuromorphic engineering," *Nature Electron.*, vol. 3, no. 7, pp. 434–435, Jul. 2020.

[81] Y. Yamamoto, "Optical neural network operating at the quantum limit—Coherent Ising/XY/recurrent neural network machines," *Photon. Switching Comput. (PSC)*, vol. 2018, pp. 1–4, 2018.

[82] T. Wang and J. Roychowdhury, "OIM: Oscillator-based Ising machines for solving combinatorial optimisation problems," in *Proc. Int. Conf. Unconventional Comput., Natural Comput.* Cham, Switzerland, 2019, pp. 232–256.

[83] D. E. Nikonov et al., "Coupled-oscillator associative memory array operation for pattern recognition," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 1, pp. 85–93, 2015.
[84] A. Raychowdhury et al., "Computing with networks of oscillatory dynamical systems," *Proc. IEEE*, vol. 107, no. 1, pp. 73–89, Jan. 2019.
[85] G. Csaba and W. Porod, "Coupled oscillators for computing: A review and perspective," *Appl. Phys. Rev.*, vol. 7, no. 1, Mar. 2020, Art. no. 011302, doi: 10.1063/1.5120412.
[86] C. Pehle et al., "The BrainScaleS-2 accelerated neuromorphic system with hybrid plasticity," *Frontiers Neurosci.*, vol. 16, Feb. 2022, Art. no. 795876, doi: 10.3389/fnins.2022.795876.
[87] T. Yu and G. Cauwenberghs, "Analog VLSI biophysical neurons and synapses with programmable membrane channel kinetics," *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 3, pp. 139–148, Jun. 2010.
[88] C. Mead and L. Conway, *Introduction to VLSI Systems*. Reading, MA, USA: Addison-Wesley, 1980.
[89] C. Batten, A. Joshi, V. Stojanović, and K. Asanović, "Designing chip-level nanophotonic interconnection networks," in *Integrated Optical Interconnect Architectures for Embedded Systems*. Cham, Switzerland: Springer, 2013, pp. 81–135.
[90] M. Kazemi, E. Ipek, and E. G. Friedman, "Adaptive compact magnetic tunnel junction model," *IEEE Trans. Electron Devices*, vol. 61, no. 11, pp. 3883–3891, Nov. 2014.
[91] L. Amarú, P. Gaillardon, S. Mitra, and G. D. Micheli, "New logic synthesis as nanotechnology enabler," *Proc. IEEE*, vol. 103, no. 11, pp. 2168–2195, Nov. 2015.
[92] D. Ielmini and H.-S.-P. Wong, "In-memory computing with resistive switching devices," *Nature Electron.*, vol. 1, no. 6, pp. 333–343, Jun. 2018.
[93] M. Wolf and D. Serpanos, "Safety and security in cyber-physical systems and internet-of-Things systems," *Proc. IEEE*, vol. 106, no. 1, pp. 9–20, Jan. 2018.
[94] S. Saidi, D. Ziegenbein, J. V. Deshmukh, and R. Ernst, "EDA for autonomous behavior assurance," in *Proc. IEEE/ACM Int. Conf. Comput. Aided Design (ICCAD)*, Nov. 2020, pp. 1–3.
[95] R. Jia et al., "Design automation for smart building systems," *Proc. IEEE*, vol. 106, no. 9, pp. 1680–1699, Sep. 2018.
[96] C. Lv, X. Hu, A. Sangiovanni-Vincentelli, Y. Li, C. M. Martinez, and D. Cao, "Driving-style-based codesign optimization of an automated electric vehicle: A cyber-physical system approach," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 2965–2975, Apr. 2019.
[97] G. Varghese, "Network verification—When Clarke meets Cerf," in *Proc. Formal Methods Comput.-Aided Design (FMCAD)*, Oct. 2016, p. 3.
[98] D. Endy, "Foundations for engineering biology," *Nature*, vol. 438, no. 7067, pp. 449–453, Nov. 2005.
[99] J. A. McLaughlin et al., "The synthetic biology open language (SBOL) version 3: Simplified data exchange for bioengineering," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 1009, Sep. 2020.
[100] M. Ibrahim, A. Sridhar, K. Chakrabarty, and U. Schlichtmann, "Synthesis of reconfigurable flow-based biochips for scalable single-cell screening," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 38, no. 12, pp. 2255–2270, Dec. 2019.
[101] T.-C. Liang, Z. Zhong, Y. Bigdeli, T.-Y. Ho, K. Chakrabarty, and R. Fair, "Adaptive droplet routing in digital microfluidic biochips using deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 6050–6060.
[102] D. Sahoo, "The power of Boolean implication networks," *Frontiers Physiol.*, vol. 3, p. 276, Jul. 2012.
[103] A. Zhavoronkov et al., "Deep learning enables rapid identification of potent DDR1 kinase inhibitors," *Nature Biotechnol.*, vol. 37, no. 9, pp. 1038–1040, Sep. 2019.
[104] A. Sangiovanni-Vincentelli and G. Martin, "Platform-based design and software design methodology for embedded systems," *IEEE Design Test Comput.*, vol. 18, no. 6, pp. 23–33, Nov./Dec. 2001.
[105] *MOSIS*. Accessed: May 25, 2023. [Online]. Available: https://www.mosis.com/
[106] *NNCI*. [Online]. Available: https://www.nnci.net/
[107] (2018). *Basic Research Needs for Microelectronics*. DoE Report. [Online]. Available: https://www.osti.gov/biblio/1545772-basic-research-needs-microelectronic
[108] ISSCC. *International Solid-State Circuits Conference (ISSCC)*. Accessed: May 25, 2023. [Online]. Available: http://isscc.org/
[109] *ISPD*. Accessed: May 25, 2023. [Online]. Available: http://www.ispd.cc/
[110] *IWLS*. Accessed: May 25, 2023. [Online]. Available: http://www.iwls.org/
[111] *ESWEEK*. Accessed: May 25, 2023. [Online]. Available: https://esweek.org/
[112] *CAV*. Accessed: May 25, 2023. [Online]. Available: http://i-cav.org/
[113] *ITC*. Accessed: May 25, 2023. [Online]. Available: http://www.itctestweek.org/
[114] *DAC*. Accessed: May 25, 2023. [Online]. Available: https://www.dac.com/
[115] *NSF Workshop on Micro/Nano Circuits and Systems Design and Design Automation*. Accessed: May 25, 2023. [Online]. Available: https://nsfedaworkshop.nd.edu/
[116] *A Report for NSF Workshop on Micro/Nano Circuits and Systems Design and Design Automation*. Accessed: May 25, 2023. [Online]. Available: https://nsfedaworkshop.nd.edu/assets/432289/nsf20_eda_workshop_report.pdf
[117] *A Report on Semiconductor Foundry Access by U. S. Academics*. Accessed: May 25, 2023. [Online]. Available: https://nsfedaworkshop.nd.edu/assets/429148/nsf20_foundry_meeting_report.pdf