DA Systemization of Knowledge: A Catalog of Prior Forward-Looking Initiatives

(Invited Paper)

Farinaz Koushanfar
Department of Electrical and Computer Engineering
Rice University
Email: farinaz@rice.edu

Azalia Mirhoseini
Department of Electrical and Computer Engineering
Rice University
Email: azalia@rice.edu

Gang Qu
Department of Electrical and Computer Engineering
University of Maryland
Email: gangqu@umd.edu

Zhiru Zhang
School of Electrical and Computer Engineering
Cornell University
Email: zhiruz@cornell.edu

***Abstract*—Electronic Design Automation (EDA) has had a profound impact on the development of modern computing and information technology which in turn has transformed our lives and society. Despite its dominant focus on electronics, EDA is one of the first fields in engineering that has taken a truly interdisciplinary route: several abstractions, computational models, algorithms, methodologies, and tools have been jointly developed by the chemists, device physicists, electrical engineers, computer scientists, applied mathematicians, and optimization experts. These EDA tools are capable of not only synthesizing and optimizing design from its high-level functional description to physical entity, but also performing the complex tasks of simulation and verification. In recent years, with the Moore's law approaching it's near-end, a number of studies and new-initiatives have been focusing on more contemporary problems and novel application domains for the field. This paper provides an overview of knowledge gathered by the prior forward-looking efforts pursued by the EDA community. Our goal is to systemize the knowledge, trends, and visions that can help DA community to move beyond its traditional boundaries.**

I. Introduction

Over the past five decades, Integrated Circuits (ICs), the computational engine of virtually every digital design today, have relied on the continuous downsizing of the their transistor components to deliver exponential productivity as predicted by the Moore's law. A major component of the successful scaling of the ICs to deliver complex designs and applications has been the development of sophisticated models, abstractions, algorithms, and automated tools created by the EDA research and industry. Without the design productivity and short time to market provided by the EDA methodologies, the unprecedented speed, efficiency, and cost reduction of ICs would not have been possible. With Moore's law approaching its near end-of-life, there is a prevailing fear among researchers, industries, and government that both the traditional IC design/manufacturing and its fueling EDA industry are on a decline.

To address the challenges faced by the end of scaling, industry, government, and research sectors have taken steps to assure profit and growth of the digital systems in the new era. For example, billions of dollars are being spent on new and pending nano-scale technologies that could replace traditional silicon. As another example, the ongoing More than Moore trend attempts to add value to the digital devices by incorporating components and functionalities that do not follow the Moore's law scaling. Traditional IC design and EDA companies have been diversifying their product and service portfolios to mitigate the impact of the saturation of the conventional IC technology as well as its supporting methodologies and tools. On the research front, there is an increasing focus on the enabling (promising) technologies, methodologies, tools, and applications that have a great potential to ensure a continuous growth and impact of the modern digital industry.

The EDA community has pursued a number of initiatives to explore challenges and opportunities for the field in the upcoming decade. Prominent examples include two workshop series sponsored by the National Science Foundation (NSF) and Computing Community Consortium (CCC). In particular, these workshops assess the inadequacies of current EDA funding and education, technological challenges facing the CMOS industry, and new markets to which EDA methods can be exported. Complementary to these efforts, this article is a study for the new IEEE CEDA technical activity group which further examines the research directions taken by prominent centers funded by the Semiconductor Research Corporation (SRC) and the NSF. As EDA is arguably at a crossroads, it is critical to examine the plethora of new ideas and technologies produced by the academic and research community. This article aims to systemize the knowledge, trends, and visions created as a result of several forward-looking initiatives within EDA.

TABLE I. Catalog of prior and ongoing forward-looking initiatives pursued by the EDA community

	Section
Recent Roadmaps	
NSF Workshop on EDA: Past, Present, and Future	II
CCC Workshop on Extreme Scale Design Automation	III
International Technology Roadmap for Semiconductors (ITRS)	IV
SRC Research Centers	
Global Research Collaboration (GRC)	V.A
Center for Low Energy Systems Technology (LEAST)	
Center for Spintronic Materials, Interfaces and Novel Architectures (C-SPIN)	
Center for Future Architectures Research (C-FAR)	
Center for Systems On Nanoscale Information fabrics (SONIC)	
TerraSwarm Research Center (TerraSwarm)	
NSF Expeditions in Computing	
Center for Domain-Specific Computing (CDSC)	V.B
Variability-Aware Software for Efficient Computing (VE)	

Table I provides a helpful catalog of roadmaps and research

978-1-4673-8388-2/15/$31.00 ©2015 IEEE

centers covered in this article. We begin by summarizing the 2009 NSF workshop on EDA's past present and future in Section II and the 2013-2014 CCC workshop series in Section III. Section IV describes the most recent roadmap released by the International Technology Roadmap for Semiconductors (ITRS) on the needs and challenges facing the semiconductor industry over the upcoming 15 years. Section V outlines the major centers funded by the SRC multi-year efforts in the first subsection, and those funded by the NSF Expedition in Computing grants relevant to EDA in the second subsection. Section VI provides a summary based on classifying the knowledge from the previous efforts and research centers. Finally, we conclude in Section VII.

II. NSF Workshop on EDA: Past, Present, and Future

In 2009, a workshop on EDA was sponsored by NSF to achieve two main objectives: first, to reflect on EDA's success and determine whether EDA practices and methodology can influence other fields of computer science so they can be applied to other application domains; and second, to review the progress made under the National Design Initiative established in 2006 and evaluate what new directions and topics should be added to it. The workshop participants included a number of top EDA researchers from both academia and industry. Our summary is based upon two articles published and publicly available as the workshop report [1], [2].

The workshop report presents an overview of EDA, its funding history, a discussion of major challenges and related emerging technologies, and a look at how EDA experience might help in developing these technologies, along with associated educational aspects and challenges. The workshop also considers EDA's relation with computer science theorists and how to revive this collaboration. Finally, it presents recommendations on how to promote EDA and help it meet the serious challenges it faces in the future. The recommendations are divided into three parts: promoting research, supporting educational programs, and encouraging enhanced collaboration with industry.

Part 1 of the workshop report discusses the workshop objective, EDA definition and history, and EDA funding sources at the time of the workshop. The report argued that a proper definition of EDA shall equally emphasize three aspects: (i) EDA consists of a collection of methodologies, algorithms, and tools that assist and automate the design, verification, and testing of electronic systems; (ii) EDA embodies a general methodology that seeks to successively refine a high-level description to a low-level detailed physical implementation for designs ranging from ICs (including SoCs), to PCBs and electronic systems; and, (iii) EDA involves modeling, synthesis, and verification at every level of abstraction. They argued that the second and third aspects of EDA in this definition can be applied easily to application fields other than electronic system design.

On the funding front, the report presents numbers that are relatively low compared with other computer science and electrical engineering fields at the time of writing. The total NSF funding for EDA was estimated to be between \$8-12 million, with the number of CAREER awardees in the field on the decline. The SRC support for EDA was estimated to be about \$5 million, but largely focused on the design aspects. Other major funding source in EDA comes from large centers funded collaboratively by DoD, NSF, SRC and industry which amount to about \$5 million per year funding. Thus, the average total funding for EDA is estimated to be about \$20 million annually which the report indicated is very low compared with European and Taiwanese competition.

Part 2 of the workshop report outlines the following as the major foundational areas for future EDA support: (i) Verification and model checking, with particular emphasize on scalability, embedded systems, analog and mixed signals, assertions, and tools; (ii) Synthesis research, with specific focus on variability, higher levels of abstractions, interactions with verification, and multiple objectives; (iii) Programming languages that are more tailored to the contemporary design needs than the current practice; (iv) Analog and mixed-signal design, and (v) Nonlinear model reduction. The key challenges that face EDA were also outlined in the report as follows: (i) Scalable design methodologies; (ii) Scalable design synthesis and validation/verification; (iii) Dealing with new technology; (iv) Designing with uncertainty and fragility; and (v) New classes of algorithms.

The report highlights the following as the emerging areas relevant to EDA: (i) Biology systems including system biology and synthetic biology; (ii) Emerging computing, communications, and storage fabrics and manufacturing substrates; (iii) Analysis, characterization, and potential design of hybrid electronic and biological systems; (iv) Cyber-physical systems; (v) Data centers design and optimization; and (vi) Software systems. The report also acknowledges that the above list is not exhaustive, and further joint investigations would be needed with other domain experts from physics, information science, and so on to develop a sharper focus and a more convincing justification, as well as to establish a consensus on feasibility and identify verifiable order-of-magnitude improvements potentially achievable by exploiting the DA methodologies.

Lastly, the report focuses on the EDA education. At the undergraduate level, they recommended the development of a good senior-level introductory CAD class to be offered more broadly in the US. At the graduate level, they depict a broad chart to demonstrate the foundational areas and domain knowledge in EDA and subsequently argue that exactly what subset to teach is a challenge because EDA is a broad, interdisciplinary field that continues to expand. They discuss the possibility of teaching a number of fundamental courses to cover the diversity in the EDA subareas, such as discrete optimization, model order reduction, and machine learning. They emphasize that training in EDA fundamentals and experience with EDA-scale problems are required for continuing technological innovations. One concern pointed out is the number of declining summer internships, jobs, and academic positions in the field. They are hopeful that this trend would not last, as EDA continues to evolve, expanding into several exciting areas with new technologies on the horizon. The workshop conclusion is that tremendous but exciting challenges are looming which could pave the way to a brighter future if correct support, funding, and educational models are adopted.

III. CCC/SIGDA Workshops on Extreme Scale Design Automation

The CCC and ACM Special Interest Group for Design Automation (SIGDA) sponsored a series of three workshops in 2013 and 2014 to examine the current difficulties faced by the EDA industry in educating, and funding the next generation of EDA professionals as well as explore new directions for the EDA field.

The first workshop focused on the future of EDA education and workforce, and was summarized in an MSE'2013 paper [3]. The key issue discussed was how to raise interest for EDA-related jobs and careers in undergraduates and new graduates as well as how to connect classroom material in EDA courses to the impact made by the field in the real world. Other topics included what EDA industry leaders can do to engage aspiring ECE students compared to the wide media presence and social appeal of companies such as Apple, Facebook, and Google. Finally, the paper also examines ways to get students interested in graduate-level EDA research. Strategies to improve the current situation include the use of massively open online courses (MOOCs) to reach a broader audience, emphasizing widely applicable EDA methodologies over low-level design skills, and creating more social media presence.

The second and third workshops looked at challenges and opportunities for the EDA field over the next decade, with topics such as traditional EDA, hybrid post-CMOS electronics, and design automation of things. Additional details from these workshops are summarized in a recently published CCC report [4], which we briefly describe below.

Many real-world designs today still utilize older technology nodes due to cost, yield, and reliability concerns. Academic research in mature flows continue to improve tools and algorithms, but few of these innovations make the leap into a commercial product, suggesting the need for a greater level of technical interaction between academia and industry. In cutting-edge flows ($<$10 nm), the workshop participants emphasized effective abstractions, design metrics, and practical benchmarks to guide research and the need to accurately model reliability issues and physical phenomenon. Existing EDA tools also face difficulties interacting with emerging technologies such as memristors, spintronics, and graphene devices. While a large-scale obsolescence of silicon is unlikely to occur, these technologies will have be integrated with existing electronic devices. It is important for researchers to rethink the EDA stack and collaborate with device experts to lay the groundwork for an ecosystem of new automation tools, which has the potential to reshape many new markets including medical devices, synthetic biology, smart grid, and cloud computing. According to the workshop participants, the most critical markets currently lie in cyber-physical systems (e.g., wearables and Internet of things) and cyber-security systems (e.g., home monitoring and remote medical diagnostics). New conferences and workshops geared towards expanding EDA to these new markets will be necessary to help steer the community towards the most fruitful directions.

IV. ITRS Roadmap

The ITRS is perhaps the most well-recognized example of a forward-looking effort in the semiconductor industry after the famous Moore's Law. As its name suggests, the mission of ITRS is to ensure the continued cost and performance scaling of IC technology by studying key challenges and innovative ideas, and then forming a roadmap to guide future academic research and industry investment. Since its formation in 1998, ITRS has become the de facto guide for the continuing growth of the semiconductor industry.

The *Design* and *System Drivers* chapters of the ITRS are the most relevant International Technology Working Groups (ITWGs) to EDA. The design chapter has existed since the first ITRS (indeed it was in NRTS before) and the system driver chapter was added in 2001. They have been an important part of the ITRS' vision to "provide a view into critical design technology challenges and required solutions across various market domains and abstraction levels" [5]. In the following, we summarize some of the important contributions of these two ITWGs and their predictions in the latest 2013 edition.

In the 80s, new memory devices were introduced and adopted by IC makers roughly every three years. This pace was increased from three years to two years in the 90s with the help of many EDA tools. This started the era of "More Moore" where design (and DA) provides "equivalent scaling" to the geometric scaling predicted in Moore's Law. For example, "better crosstalk-aware routing is equivalent to reduction of dielectric permittivity. Better model- and timing-driven dummy fill is equivalent to a better CMP module spec". The 2005 edition captured this by defining the term "More Than Moore", which means that besides the geometric scaling, heterogeneous integration of new non-digital functionalities into smart systems has become a driving factor for the technology roadmap. DA is not only providing design and verification tools, it has also become "a key enabler of the overall semiconductor roadmap, and is increasingly aligned and mutually dependent with the manufacturing and device aspects of the ITRS" [5].

Therefore, a significant amount of efforts have been invested on (1) quantifying the design technology roadmap with precisely defined metrics, predicting design requirements and challenges that have motivated many design solution inventory and design-driven semiconductor innovations such as design for manufacturability, and (2) introducing new system drivers that are aligned to key segments of semiconductor industry with the perception that "each system driver must be driving something in the ITRS". The networking system driver in 2007, the automotive driver in 2008, and the medical and defense drivers in the later editions all have been fundamental in shaping today's IC industry.

In the past decade, we have witnessed a new ecosystem featuring devices known as system-on-chips (SoCs), with fabless design companies and intellectual property (IP) providers, as well as integration of new system components such as specialized hardware accelerators, MEMS sensors, radios, and passives. This current decade is the start of the era of the Internet of things (IoT), where traditional users of IC such as telecommunication companies, data and information distributors, and content providers are playing a more active role in the semiconductor industry. The number of cores in mobile and networking SoCs is growing rapidly. Power will remain as one of the most dominant factors in the roadmap. SoC-CP's power consumption estimation is suggested to follow the scenario-based approach because "key functional blocks

(GPUs, RFs, and multimedia IPs) have very extreme switching activity discrepancies in different scenarios". DA community possesses the necessary skill sets to solve these challenges.

Another challenge valuable for DA is how to close the so-called design capability gap: the scaling rates of die contents and device/interconnect geometries, or the 2X per node available Moore's Law geometric scaling vs. the 1.6X per node of the realized transistor density scaling. With a projected stall of M1HP scaling in the next technology node, design-based equivalent scaling (DES) methodologies (e.g., error correcting codes, clock gating, adaptive voltage and frequency scaling) need to be integrated into the roadmap to mitigate this scaling crisis and it is predicted that DES will be capable of closing the gap for the next generation node [6]. This shows again that design technology has become an enabling technology for the scaling.

V. Major DA Centers and Expeditions

In this section, we outline major centers funded by the SRC multi-year efforts and also cover the large NSF expedition grants with DA themes.

A. SRC Centers

SRC is a university research management consortium for semiconductor technologies, founded in 1982. SRC's mission is to manage a range of consortial university research programs, some of which are worldwide, each matching the needs of its sponsoring entity [7].

While a mainstream focus of SRC remains on research that supports physical scaling of CMOS technology to its limits, new unprecedented research directions are being pursued. SRC is now expanding its horizon to novel research areas including identifying new technologies that provide increased value without explicit dependence on scaling; new materials, processes, devices, and interconnect technologies that can boost or substitute extremely-scaled CMOS. In the following we outline the major centers that are initiated by SRC.

1) Global Research Collaboration (GRC): GRC drives near-term materials, interconnect, devices, design, and tools progress [8]. Although GRC heavily focuses on the current priorities of the semiconductor industry, including the continued scaling of semiconductor, since 2014 it has launched multiple efforts for expanding its application space and time horizon. New research directions include the targeted topics of cyber-security, semiconductors and biology, advanced connectivity, and the IoT.

GRC's trustworthy and secure semiconductors and systems (T3S) research, has a goal to develop cost-effective strategies and tools to design and manufacture chips and systems that are reliable, trustworthy, secure, and resistant to attack or counterfeiting. GRC's semiconductors and biology (SemiSynBio) research seeks to explore synergies between synthetic biology and semiconductor technologies. Research results along this direction will lead to more energy-efficient devices, improved manufacturing techniques, and enhanced architectural approaches to computational capability.

GRC's advanced connectivity (EP3C) research aims to identify architectures and associated interconnect technology that minimize the energy constraints on future computing performance improvements. GRC's innovative and intelligent Internet of things (I3T) research, explores energy-constrained computation and connectivity with an emphasis on sensing and actuation, integration and packaging, bringing together specific industry needs that enable breakthrough technologies for the next generation of intelligent, connected and autonomous devices.

2) Center for Low Energy Systems Technology (LEAST): LEAST is a five-year multi-university research center with a focus on developing low-voltage and steep subthreshold swing components for beyond-CMOS electronic systems [9]. Based upon the observation that today's ICs are limited by power dissipation and heat, which limits the transistor chip packing density, LEAST aims to enable devices that will run cooler and pack tighter. The mission of LEAST is to discover transistors that outperform CMOS at low voltage (less than 0.4V).

LEAST was founded in 2013, following the success of MIND (a previous SRC center). MIND led the development of the tunnel field-effect transistor (TFET), a device which can outperform current transistors at low voltages. LEAST now aims to construct transistors at even smaller sizes and still lower voltages. With an emphasizes on steep device technology, LEAST broadens the search for physical mechanisms leading to abrupt transistor switching including the tunnel field-effect transistor and collective switching mechanisms associated with ferroelectric gates and phase change materials. The center has a strong focus on materials growth, modeling, characterization, and physical understanding. Device benchmarking, circuit development, including memory, and new system architectures are also under investigation to drive applications.

LEAST employs four inter-dependent threads to accomplish its mission. Theme 1 focuses on materials, interfaces, and surfaces, and investigates fundamental challenges associated with growth, interface control, and surfaces for steep devices. Theme 2 is a device-related topic that focuses on quantum engineered steep transistors and aims to understand and realize the potentials of steep slope tunneling devices in 2D graphene and dichalcogenide crystals, III-Nitrides, and complex oxides. Theme 3 explores transduction mechanisms beyond tunneling to further lower subthreshold swing and add new functionality to steep devices. Finally, theme 4 provides the benchmarking and explores applications for steep technologies including low-power digital logic, low-power analog, high-frequency mixed signal, security, non-von-Neumann machines, and non-Boolean computing.

3) Center for Spintronic Materials, Interfaces and Novel Architectures (C-SPIN): Spin-based memory and computation has the potential to overcome the power, performance and architectural constraints of conventional CMOS-based devices [10]. Spin-based logic devices based on the hybridization of magnetic materials and semiconductors can fundamentally outperform their charge-based counterparts because of their unique scalability and compatibility with well-developed spin transfer torque (STT) memory technology.

Ground-breaking experimental and theoretical investigations performed in the past decade have cleared the pathway to realizing spin-based computation. These include successful

demonstrations of perpendicular STT devices, spin injection and detection into semiconductors, switching ferromagnets using pure spin currents, voltage-controlled switching of magnetic tunnel junctions (MTJs), spin Hall torque devices, and novel spin logic architectures.

To realize a practical spin-based memory and logic technology, C-SPIN focuses on five research themes based on new findings in magnetic/spin transport material, spintronic devices, circuits and novel architectures. Theme 1 aims to push the state of the art in magnetic materials synthesis to obtain low damping, high spin polarization and high anisotropy and address the need for extremely small junction dimensions ($<$10 nm). Theme 2 investigates novel materials for spin channels with a focus on emerging two-dimensional material systems, including graphene, single-layer MoS2 and topological insulators. Theme 3 addresses the critical problem of interface engineering between magnetic materials and spin channels and the need for new designs of tunnel barriers. Theme 4 focuses on the design, fabrication and modeling of spin-based logic and memory devices and novel magneto-optical transducers for spin information transfer. Theme 5 addresses the design of spin-based circuits and architectures for low-power and high-performance spin-based computation.

4) Center for Future Architectures Research (C-FAR): C-FAR conducts research on scalable future computing architectures with emphasis on the efficient integration of new devices and circuit fabrics, as well as emerging application domains [11]. The current compute-centric architectures were established when data was small and compute was expensive. A reversal of this context demands a complete rethink of computer architectures to focus on data from acquisition to result delivery. This rethinking has the potential to deliver orders-of-magnitude energy and performance improvements. Future architectures must maximally exploit the headroom offered by novel logic, interconnects (including 3D), and memory technologies, which provide non-volatility, density, power and performance advantages. The benefits of homogeneous parallelism are already being heavily mined through current research investments in parallel computing, meaning that the next challenge of Amdahl's constraint must be addressed by reducing computation, communication and storage latencies. Given the significant design and deployment costs for applications and platforms, the solutions must amortize this cost over generations, over a range of capabilities in a generation, and possess a manageable migration path from current systems.

C-FAR's mission is to sustain the performance, power, and cost scaling of CMOS devices via solutions that span multiple layers of the computing stack, including circuits, architectures, compilers, and languages. C-FAR research themes include non-traditional computational paradigms, data-centric architectures, and novel system integration paradigms.

5) Center for Systems On Nanoscale Information fabriCs (SONIC): SONIC is a five-year multi-university interdisciplinary research center funded in 2013 that emphasizes information rather than data processing in the design of a robust, energy efficient new computing paradigm to enable the continuation of technology scaling on post-CMOS nanoscale fabrics [12]. Based on the observation that both information and post-CMOS fabrics are inherently statistical, SONIC employs three top-down systems-driven themes (themes 1, 2, and 3) anchored on an experimental bottom-up hybrid devices and circuits theme (theme 4). While the first three themes emphasize statistical inference-based computation models of information processing and their associated processing architectures, the last theme develops nanoprimitives and nanofunctions to support these models and architectures. Drawing on inspirations from communication theory and neuromorphic computing, SONIC aims to build an error-tolerant information processing system that achieves high performance and low energy.

Theme 1 focuses on developing statistical, error-tolerant computational models based on the stochastic characteristics of applications and post-CMOS computational fabrics. Recent efforts focus on low-complexity and energy-efficient statistical error compensation techniques at the microarchitectural level and strive to derive the minimum level of protection required for useful application output. Theme 2 provides support for the core computational fabric with compatible mixed-signal circuits from physical sensors to RF processors to ADC. Recent works range from statistical data reconstruction in embedded sensing to energy harvesting with mm-size nodes to high-speed, low-energy, error-tolerant I/Os. The goal of theme 3 is to discover, design, and demonstrate cognitive information processing architectures that provide 100X enhancement in energy efficiency using beyond-CMOS nanoscale fabrics. Ongoing efforts include theoretical studies on neuro-principled computational models as well as practical implementations of neuro-architectures.

As a foundation for the first three themes, theme 4 designs novel nanofunctions and nanoprimitives using beyond-CMOS technologies. SONIC researchers plan to investigate other nanofunctions such as CNT-based stochastic machine learning accelerators and nanoprimitives such as CNFET and RRAM.

Based on the overall themes of SONIC center, we envision that design automation will provide a systematic method for evaluating the statistical characteristics of both application and substrate to enable intelligent customization of both the software and hardware for optimal error compensation. Based on the systematic models, stochastic characteristics of algorithms can be effectively matched to those of the fabric to achieve the next leap in performance and energy efficiency. With innovations in nanofunctions and nanoprimitives, high-quality design automation techniques tailored to emerging fabrics and systems will play an important role in guaranteeing the robustness of beyond-CMOS computing platforms.

6) TerraSwarm Research Center: The TerraSwarm Research Center [13] was launched in 2013 with the vision to address the huge potential and risks of pervasive integration of smart, networked sensors and actuators into our connected world. It aims to enable simple, reliable, and secure deployment of a multiplicity of advanced distributed sense and control applications on shared, massively distributed, heterogeneous, and mostly uncoordinated platforms through an open and universal systems architecture.

The term TerraSwarm comes from the prediction that by 2020 there will be thousands of smart sensing devices per person on the planet [14]. The TerraSwarm applications will pose a unique combination of challenges: large scale, distributed, cyber-physical, dynamic, adaptive, and heterogeneous. To reach its ambitious goals, researchers from industry

and ten universities join forces to investigate the following four themes:

Theme 1. Smart cities, which is the testbed of the TerraSwarm technologies and the center integrator of the other themes from a systems perspective. Both cities during normal operation and during natural or man-made disasters will be considered. The critical research issues to be addressed include how to recruit and compose heterogeneous resources, how to dynamically adapt applications to changing resources and contention for resources, and how to share resources without compromising safety, security, or privacy.

Theme 2. Platform architecture and operating systems. The approach, called SwarmOS, is a highly distributed infrastructure to efficiently allocate resources (sensors, actuators, networks, energy, storage, computing devices, and wireless spectrum) in order to provide optimized service with appropriate security and privacy. The SwarmOS is anticipated to support continuous reconfiguration, heterogeneous components featuring energy efficiency and build-in security.

Theme 3. Services and cloud interaction. This effort focuses on technologies for scalable, adaptive composition of heterogeneous services. The vision is to have control as a service (where the design of a TerraSwarm system is decentralized to make it more robust, adaptable, and opportunistic) and the cloud as a companion (such that resources can be recruited opportunistically to deliver the right data at the right time and the right place).

Theme 4. Methodologies, models, and tools. Although design automation plays an important role in all the themes in the TerraSwarm project, the last theme is centered on DA. The applications and infrastructure in TerraSwarm system features high heterogeneity and availability of resources and dynamic reconfiguration and requirements of applications. This makes the distinction between design-time and run-time blurred and creates the need of advanced modeling, verification and adaption approaches.

B. Major NSF Expeditions

The NSF Expeditions in Computing award targets ambitious and transformative research agendas which have the potential to redefine computing and information for years to come. With one of the largest grants in computer science and engineering, the expeditions usually go towards the funding of entirely new inter-disciplinary research centers. In the following we summarize two such centers whose research agendas are closely related to the DA field.

1) Customizable Domain-Specific Computing (CDSC): CDSC was established in 2009 with the support of an NSF expeditions award to look beyond homogeneous multicore scaling by focusing on domain-specific customization as the next disruptive technology to bring orders-of-magnitude computing efficiency improvement to important classes of applications [15]. Specifically, medical image processing was selected as the main driver applications at the CDSC, owing to the computationally demanding nature of medical imaging algorithms and their high importance in the healthcare domain.

The CDSC project was carried out as a collaborative effort between four universities: UCLA (the lead institution), Rice, UCSB, and Ohio State. This research team has explored a broad range of topics on customizable computing, including (1) customizable computing engines that incorporate a heterogeneous mix of coarse-grain processor cores and fine-grain reconfigurable fabrics, (2) customizable RF-interconnect for high-performance on-chip communication, (3) application modeling through domain-specific language specifications, and (4) compilers and runtime systems for automated application mapping to customizable heterogeneous platforms.

In particular, accelerator-rich architectures are proposed and identified as a promising architectural template for building high-performance yet energy-efficient computer systems. The main idea is to make extensive use of specialized hardware accelerators to significantly reduce energy consumption of key algorithms/applications in the target domain. To mitigate the additional design complexity incurred by architectural heterogeneity, the CDSC team has proposed a set of accelerator-centric hardware/software co-optimization techniques such as accelerator synthesis, runtime accelerator allocation and composition, memory hierarchy optimization, etc. On a set of medical imaging algorithms, employing accelerator-rich architectures has led to at least one order of magnitude improvement in both performance and energy compared to optimized software execution on high-end CPUs.

More complete references to the research outcomes of the CDSC project are available in a recently published synthesis lecture on "Customized Computing" [16]. While the CDSC project primarily focused on design customization at the chip level and server-node level, a new line of research on customizable computing is being pursued at the cluster and datacenter level. Clearly, enabling productive design and implementation of accelerator-rich systems is increasingly important to computing community and also gives rise to a host of new challenges and opportunities to the field of design automation.

2) Variability-Aware Software for Efficient Computing (VE): VE envisions computing systems that can be constantly monitored, predicted, and adapted to system and component variability by relying on novel proactive software-dominated solutions [17]. As semiconductor manufacturers build ever-smaller components, circuits and chips at the nanometer scale become less reliable and more expensive to produce — they no longer behave like precisely chiseled machines with tight tolerances. Modern computing tends to ignore the variability in the behavior of underlying system components from device to device, their wear out over time, or the environment in which the computing system is placed. This makes them expensive, fragile and vulnerable to even the smallest changes in the environment or component failures.

Based on the above observation, VE proposes a new class of computing machines that are adaptive and highly energy efficient. They will continue working while using components that vary in performance or grow less reliable over time and across technology generations. A fluid software-hardware interface is a key to mitigating the variability of manufactured systems and making machines robust, reliable, and responsive to changing operating conditions while achieving fundamental gains in computing performance.

VE is addressing major technical challenges across five general research thrusts, including (i) measurement and mod-

eling; (ii) design tools and testing; (iii) micro architecture and compilers; (iv) runtime support; and (v) applications and testbeds.

(i) Based upon the observation that technology scaling increases the variability in hardware, VE aims to identify, quantify, and eventually model this variability so that it can be exposed to higher layers of the software stack. Doing so enables the reduction of vulnerabilities and improves fault tolerance, efficiency, and power consumption.

(ii) VE aims towards dramatically reducing hardware design and test complexity for computing systems, while achieving maximum performance potential at minimum energy and total costs. To achieve this goal, hardware design objectives need to shift from optimization for fixed specifications to designs with well-behaved variability characteristics.

(iii) The third thrust of VE focuses on the software stack that can respond to the application needs based on changing data and environment (e.g., platform characteristics and behaviors). The eventual goal is to model those needs through a responsive architecture and programmer interface that will extend the traditionally fixed instruction set architecture (ISA) specification of a computing machine to one where the ISA functionality and performance are mutable across different instances of the hardware, different invocations of an application, and within the entire lifetime of an application. The compilers developed by this project will combine semantic analysis of the application behavior with compiler strategies exploiting both static and lightweight on-the-fly techniques to adapt applications that can leverage the underlying variability in hardware.

(iv) The software stack can take several different types of run-time actions in response to hardware variability, such as altering the workload, using alternative hardware resources, changing the algorithm, or altering the operational setting of the hardware. A key difference from software mechanisms developed for fault-tolerant computing is that a variability-aware software stack can take actions preemptively, based on statistically predicted variability behavior of the hardware, and consider not simply functional and temporal correctness but also factors such as energy efficiency. The diverse forms in which variability occurs, and the strong dependence of the response on the current application and system context, present challenges to the software stack.

(v) The last thrust of VE focuses on domain-specific opportunities and versatile testbeds. Many modern-world applications allow trading off output quality with system performance, resource cost, and energy usage by adaptive duty-cycling. Examples of such applications are search engines, machine learning, speech recognition, and computer vision. VE aims to test new solutions and observe hardware variability while experimenting with different systems (e.g., embedded, mobile, servers, etc.) and platforms such as off-the-shelf integrated circuits, FPGA-based, and custom-designed ICs.

VI. Summary and Directions

A number of insights can be derived from the CCC and NSF workshops on the subject of EDA funding, education, technological challenges, and opportunities. Overall funding in the EDA field is lower than other computer science and electrical engineering related fields, and government funding especially is low compared to what is available in Europe and Taiwan. Undergraduate and graduate interest in EDA has decreased due to competition from trendier software companies, but the introduction of new courses emphasizing a broader technical scope could engage students and new graduates. Current EDA methodologies face key challenges in extreme-scale design, validation and verification, and integration of post-CMOS technologies. Nevertheless, a number of emerging markets can benefit from DA knowledge including biomedical engineering, cyber-physical systems, security, and emerging compute and storage media. Leveraging these markets will require collaboration with outside domain experts.

ITRS has been very successful in the past 15+ years in making, monitoring, and updating the semiconductor technology roadmap as well as identifying the near-term and long-term challenges. Design and DA have played a vital role in answering these challenges and become an enabling technology in the era of “More Moore” and “More than Moore”. The latest edition pointed out two novel directions for device scaling to go beyond the fundamental limits of horizontal dimensional scaling — heterogeneous integration of new technologies and invention of devices for the new information processing paradigms, and identified that “continuing Moore’s Law functional density benefits and managing power and performance tradeoffs remain as the key drivers for the Roadmap grand challenges and potential solutions” [6]. The expertise of EDA community from the past five decades will be valuable for the technology scaling along these two directions. More importantly, EDA has played the leading role in balancing performance and power; it is perhaps the most multi-disciplinary group, not only interacting with other semiconductor communities, but also open to societies such as telecommunication, software, medical devices, power grid, and automotive industry; DA community welcomes the technology roadmap challenges in the era of IoT and will thrive in solving these challenges.

Based on the information we have collected, the aforementioned DA centers and expeditions are primarily pursuing four major research directions spanning from devices to systems, whose coverage is categorized in Table II. The first direction focuses on *hybrid and post-CMOS technologies* that provide new materials, processes, devices, and mixed-signal components to complement and substitute current CMOS devices. Post-CMOS technologies serve as the physical foundation for potential breakthroughs in next-generation high-performance, energy-efficient computing, and are heavily pursued by at least four research centers. For example, SONIC devotes an entire theme to developing nanofunctions and nanoprimitives using beyond-CMOS technologies, while C-SPIN focuses on spin-based logic devices for computation and memory.

Diminishing benefits of technology scaling and challenges in physical design over the past decade have also prompted extensive research and development into alternatives to traditional general-purpose von Neumann machines. For example, CDSC and C-FAR centers have been investigating specialized hardware architectures as a means for sustaining improved performance and energy efficiency with post-Dennard scaling. Concurrently, VE explores a more adaptive computing

TABLE II. MAJOR RESEARCH DIRECTIONS OF DA CENTERS – Post CMOS = Hybrid and post-CMOS technologies; Alternative Compute = Alternative computing, communication, and storage paradigms; Methodologies & Tools = New design methodologies and tools; Systems of Things = Complex systems of things.

	Post CMOS	Alternative Compute	Methodologies & Tools	Systems of Things
GRC	●	●	●	●
LEAST	●	○		
C-SPIN	●	○		
C-FAR	○	●	○	
SONIC	●	●	○	○
TerraSwarm		○	●	●
CDSC		●	●	
VE	○	●	●	

A filled circle indicates that the center has one or more dedicated research themes on the subject; An empty circle indicates that the subject is partially covered by the research themes of the center.

paradigm that exploits variability inherent to smaller technology nodes. In addition, tremendous effort are warranted in the development of *alternative computing, communication, and storage paradigms* in order to fully exploit emerging devices developed as part of the post-CMOS direction. These alternative paradigms provide the means for applications to efficiently execute on post-CMOS fabrics. Not surprisingly, nearly all centers are investing efforts along this research direction. For instance, LEAST is investigating architectures for non-Boolean computing, while SONIC is exploring stochastic and cognitive information processing. TerraSwarm works on continuous reconfiguration of heterogeneous computing components.

New design methodologies and tools targeting the alternative paradigms enable productive modeling and design of efficient, reliable, and secure systems that are robust to uncertainty and can be run-time reconfigured to adapt to the application characteristics and fabric variability. For instance, CDSC develops compilers and runtime systems for customizable compute and communication. VE further explores modeling and design and testing tools. TerraSwarm dedicates one theme for tools and models that manage its highly heterogeneous, dynamically reconfigurable resources to serve a large variety of real-time applications. GRC develops tools to address emerging design objectives to anticipate future and post-CMOS challenges. Both TerraSwarm and GRC include design for security as one of their key objectives. As we see, efforts in the direction of methodologies and tools provide ways to accelerate the development of post-CMOS technologies and increase productivity in achieving higher performance and lower energy on top of alternative paradigms.

An ultimate goal of new technologies, paradigms, methodologies, and tools is to enable more *complex cyber-physical systems of things* in which each device, interconnected but also autonomous, can perform sensing and actuation while being energy constrained. Two centers currently focus on studying these complex systems and how to leverage new technologies to build these systems. GRC specifically explores the energy constraints in computation and connectivity in such systems, while TerraSwarm aims to develop massively distributed, interconnected systems of heterogeneous smart devices with shared resources and run-time reconfigurable computing fabrics. Overall, new DA techniques and toolflows are expected to emerge to establish a systematic approach for modeling tiny devices to massive systems and enable the level of productivity, optimization, and customization that are otherwise infeasible.

VII. CONCLUSION

The EDA industry currently faces a plethora of challenges ranging from managing the extreme scales of leading-edge devices to integrating emerging post-CMOS technologies to productive and trustworthy design for IoT. At the same time, the research community faces difficulties in finding, educating, and funding the next generation of EDA professionals. This article systemizes the knowledge, trends, and visions produced by a variety of forward-looking initiatives related to these challenges. We summarize two recent workshops which assess the funding and education situation, as well as emerging markets for DA methodologies and tools. We then describe the challenges facing the traditional semiconductor industry as determined by the well-known ITRS roadmap. We finally examine the research directions of a number of DA-related research centers funded by the SRC and the NSF expedition grants. It is our hope that this information will serve as a useful resource that facilitates moving the DA community beyond its traditional boundaries in the upcoming years.

REFERENCES

[1] R. Brayton and J. Cong, "NSF Workshop on EDA: Past, Present, and Future (Part 1)," *IEEE Design & Test of Computers*, vol. 27, no. 2, pp. 68–74, 2010.

[2] ——, "NSF Workshop on EDA: Past, Present, and Future (Part 2)," *IEEE Design & Test of Computers*, vol. 27, no. 3, pp. 62–74, 2010.

[3] I. Bahar, A. K. Jones, S. Katkoori, P. H. Madden, D. Marculescu, and I. L. Markov, "'Scaling' the Impact of EDA Education, Preliminary Findings from the CCC Workshop Series on Extreme Scale Design Automation," *Int'l Conf. on Microelectronic Systems Education (MSE)*, pp. 64–67, 2013.

[4] ——, "Workshops on Extreme Scale Design Automation (ESDA) Challenges and Opportunities for 2025 and Beyond," 2014.

[5] J.-A. Carballo and A. B. Kahng, "ITRS Chapters: Design and System Drivers," *Future Fab Special ITRS Focus*, pp. 52–56, 2013.

[6] "International technology roadmap for semiconductors (2013 edition)," *Semiconductor Industry Association*, 2013.

[7] "Semiconductor Research Corporation *(SRC)*," https://www.src.org/.

[8] "Global Research Collaboration *(GRC)*," https://www.src.org/program/grc/.

[9] "Center for Low Energy Systems Technology *(LEAST)*," https://www.src.org/program/starnet/least/.

[10] "Center for Spintronic Materials, Interfaces and Novel Architectures *(C-SPIN)*," https://www.src.org/program/starnet/c-spin/.

[11] "Center for Future Architectures Research *(C-FAR)*," https://www.src.org/program/starnet/c-far/.

[12] "Center For Systems On Nanoscale Information fabriCs Center *(SONIC)*," https://www.src.org/program/starnet/sonic/.

[13] "The TerraSwarm Research Center *(TerraSwarm)*," https://www.src.org/program/starnet/terraswarm/.

[14] E. A. Lee and et al., "The Swarm at the Edge of the Cloud," *IEEE Design & Test of Computers*, pp. 8–20, 2014.

[15] J. Cong, V. Sarkar, G. Reinman, and A. Bui, "Customizable Domain-Specific Computing," *IEEE Design & Test of Computers*, vol. 28, no. 2, pp. 6–15, 2010.

[16] Y.-T. Chen, J. Cong, M. Gill, G. Reinman, and B. Xiao, "Customizable Computing, Synthesis Lectures on Computer Architecture," *Morgan & Claypool Publishers*, 2015.

[17] "Variability Expedition *(VE)*," http://www.variability.org/.