A Report on Semiconductor Foundry Access by US Academics

(Discussion held at a meeting virtually held at the National Science Foundation on Dec 16, 2020)

Organizers: Sankar Basu[1], Erik Brunvand[2], Subhasish Mitra[3], H.-S. Philip Wong[4]
Scribes: Sayeef Salahuddin[5], Shimeng Yu[6]

NSF Workshop
Micro/Nano Circuits and Systems Design

UNIVERSITY OF NOTRE DAME

Search this site

Home ›

Home
About
Schedule
Speakers
Foundry Meeting
Schedule
Steering Committee

Foundry Meeting

Semiconductor Foundry Access by US Academic Researchers in Micro- and Nano- Circuits and Systems (NSF workshop), December 16, 2020

Goal

To evaluate academic needs for access to semiconductor foundry and associated support infrastructure for design tools and IP solutions, and brainstorm ways to provide such access and support to US academic researchers.

Background

The MOSIS program that started around 1981 unleashed the innovation of circuit designers and enabled circuit research to proceed in parallel with device technology research. Fast forward 40 years, the needs of today are drastically different from decades ago. Design innovations are now strongly coupled with chip-/system-architecture innovations; circuit innovations often derive from the use of new device technologies; and, device technology innovations are driven by application needs and require circuit/architecture level demonstrations to be relevant.

[1] NSF, sabasu@nsf.gov
[2] NSF, ebrunvan@nsf.gov
[3] Stanford University, subh@stanford.edu
[4] Stanford University, hspwong@stanford.edu
[5] University of California, Berkeley, sayeef@berkeley.edu
[6] Georgia Institute of Technology, shimeng.yu@ece.gatech.edu

Executive Summary

Semiconductor technology and microelectronics[7] is a foundational technology that without its *continued* advancement, the promises of artificial intelligence (AI), 5G/6G communication, and quantum computing will never be realized in practice. Our nation's economic competitiveness, technology leadership, and national security, depend on our staying at the forefront of microelectronics.

We must accelerate the pace of innovation and broaden the pool of researchers who possess research capability in circuit design and device technologies, and provide a pathway to translate these innovations to industry. This meeting has brought to the fore the urgent need for access to semiconductor foundry and design ecosystem to achieve these goals.

Microelectronics is a field that requires sustained and rapid innovations, especially as the historical rate of progress following a predictable path, is no longer guaranteed as it had been in the past. Yet, there are many plausible paths to move forward, and the potential for further advances is immense. There is a future in system integration of heterogeneous technologies that requires end-to-end co-design and innovation. Isolated push along silos, such as miniaturization of components, will be inadequate. It is in this context that large-scale efforts, best coordinated by the National Science Foundation, can make a substantial difference.

Because end-to-end co-design innovation is essential moving forward, there is an urgent need for academia and even companies, to access foundry technologies that will facilitate bridging the lab-to-fab gap, thereby enabling the building of prototypes of new technologies (device and integration technologies) with more than a few component devices (as often happens today in a university lab). There is also a strong need from academia to access advanced technology nodes for circuits and systems research. The design ecosystem must also advance in response to the emerging technologies, and cope with enormous complexity associated with systems built using advanced technologies and new emerging technologies. Furthermore, there is a need to lower the barrier to entry for circuit design and shorten design time.

[7] Semiconductor technology broadly refers to all forms of micro- and nano-electronics, photonics, sensors and actuators, as well as the circuit and system architecture design, manufacturing, and packaging technologies. For simplicity, we use "semiconductor technology" and micro-/nano-electronics synonymously.

To ensure progress, we must lower the barrier for ***all*** US researchers to gain access to state-of-the-art foundry services and the design ecosystem. We recommend the following efforts led by the NSF.

A. **Create a new initiative that connects software and hardware foundations.** This new initiative would specifically address research questions related to the use of new device component technologies.
B. **Facilitate access to leading-edge (silicon CMOS and beyond) technologies**. Currently, only select groups of researchers have access to advanced technologies (silicon CMOS and beyond) and advanced integration technologies. The full support of the US government must be brought to bear to ensure access to a wider academic community. Such access includes leading-edge silicon as well as affordable access to mature nodes. Funding mechanisms should be developed that support the cost of chip design tape out in addition to the traditional cost of research.
C. **Support/establish a national facility for prototyping emerging technologies at-scale.** We must find ways to demonstrate emerging device technologies *at scale*, beyond the 1 to 1,000 devices scale that are sufficient for an initial exploration. A national facility should be established with the mission to enable fast turn-around experimentation of chip-scale, and package-scale systems, achieve flexibility (of material and process technologies) at scale, and facilitate demonstration of *system* technologies.
D. **Open access for design ecosystem.** NSF must invest in open-source electronic design automation (EDA) tools and open EDA design flows. The learning curve for a tape out is steep and this hampers innovation. There needs to be a concerted effort to make the circuit design process as easy as software development.
E. **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 part C above).
F. **Education and workforce development.** The NSF must find ways to incentivize and assist universities to develop and offer engaging integrated circuit (IC) design courses using real technologies that are used in practice.

1. **Introduction**:

NSF organized a (virtual) meeting on December 16, 2020, to evaluate academic needs for access to semiconductor foundry and associated support infrastructure for design tools and IP solutions, and brainstorm ways to provide such access and support to US academic researchers. Representatives from government, academia, industry, and foundry service providers gave opening talks on the current status and needs for semiconductor foundry access by US academics and small businesses in the startup phase. Fifty one attendees from academia, industry, government, and research institutes attended the meeting. Leaders from peer institutions in Europe and Asia provided a broader worldwide perspective that helps in benchmarking and sow the seed for future collaboration.

2. **Background:**

The US government supported MOSIS program [MOSIS] that started around 1981 unleashed the innovation of circuit designers and enabled circuit research and education to proceed by way of abstractions that uncoupled 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, co-design 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 [NNCI]) 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 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 co-design and co-optimization – a technical direction that is highlighted by earlier studies on the subject [DoE18].

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, 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 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 under-utilization of a large group of talented researchers and technology developers.

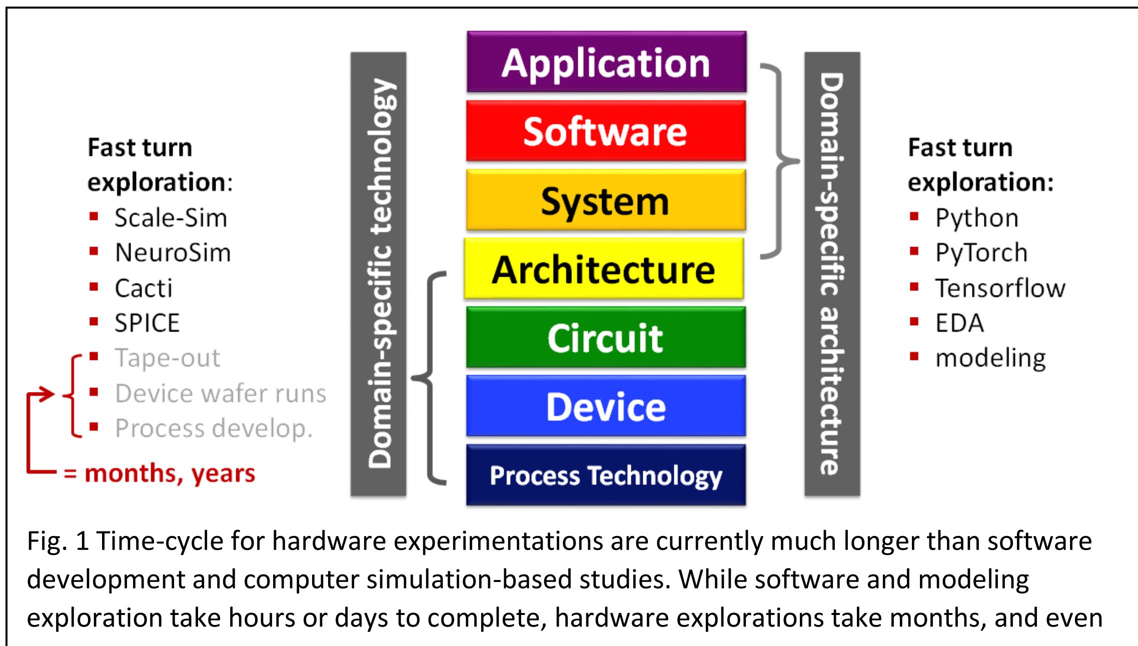

Fig. 1 Time-cycle for hardware experimentations are currently much longer than software development and computer simulation-based studies. While software and modeling exploration take hours or days to complete, hardware explorations take months, and even

To make things worse, the time-cycle for hardware experimentations are currently much longer than software-only and/or simulation-based studies (Fig. 1). 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.

In addition to identifying the needs from stakeholders, this meeting aims to seek possible solutions. Toward that goal, this report collects information from peer institutions in Europe and Asia, with a view to benchmark, leverage, and identify possible collaborations in the future. One of the important goals of academic research is education and workforce training. As some of our participants pointed out, improving foundry access goes a long way toward raising the quality of education and encouraging a broader cross-section of students to be trained in microelectronics – a well-articulated priority of the US government, as we find ways to bolster domestic semiconductor manufacturing capability.

3. **Foundry access to advanced technologies by designers:**

3.1 Leading-edge (advanced) nodes (16-nm FinFET nodes and beyond):

Access to leading-edge (e.g. 7-nm and below) semiconductor technology is needed to build circuits from the densest CMOS to benchmark new ideas against existing approaches and to further attain new performance/power metrics and demonstrators to illustrate concepts that outperform conventional, industry approaches. For instance, if the fundamental device/circuit blocks of a particular technology node changes the power/performance/area/reliability/testability/security of IC's then it is reasonable to assume that access to such a node is needed to demonstrate the merits of the new design. Currently, access to even 16-nm and 22-nm node (two and three generations behind the state-of-the-art) is spotty. Furthermore, advanced technologies does not mean logic technology nodes only – it must cover advanced memory technologies, advanced (beyond-silicon) logic technologies, and, very importantly, advanced integration technologies. Most US researchers have no access to these "advanced" technologies.

Currently, no-cost access to advanced logic technology is mostly acquired through specific government programs (e.g. 22-nm FinFET from Intel or 16-nm FinFET from TSMC through DoD programs) or very special individual relationship with foundries. The access is typically tied to a specific project the funding sponsor is interested in, making it difficult to explore research white spaces. Access through commercial sources are prohibitively expensive for academic researchers[8].

There are generally three challenges for accessing leading-edge technologies: (a) chip fabrication cost is prohibitive (if paid from research funding sources), (b) availability

[8] MUSE MPW offers \$24,000/mm^2 for TSMC 16-nm FinFET technology.

(both cost and NDA) of EDA tools and design IPs, and (c) export control. *The first two may be addressed by a coordinated effort from government funding agencies with appropriate funding, coordination, and/or partnership with foundries and companies that own the IPs*. Concrete examples include new NSF programs that solicit advanced chip design proposals and sponsor the tape-out cost only; and an improved version of MOSIS that provides a common legal framework to ease the negotiation between individual researchers and selected foundries or IP vendors for leading edge node.

NDA and restricted access may be solved using the cloud design environment [VDE] as is the current practice through the MOSIS Service. Yet, the cost of access to the cloud design environment remains an unsolved issue. IP availability may find relief when open-source IP becomes available (see Section 5 below). Today, open source IP is still a vision that has yet to be realized (and it is not clear when that vision can be realized).

The export control issue has already presented difficulty to several universities as leading-edge nodes are classified as export controlled technology. In fact, even for mature nodes (e.g. 28-nm node), technologies from the largest foundry (TSMC) are classified as 3E991 in accordance with the U.S. Department of Commerce Control List. Many universities cannot accept 3E991 information on their campuses.

3.2 Mature nodes (28-nm and older):

There are research problems that may be addressed with mature CMOS nodes. Examples include some circuits and architectures for power management, mm-wave signal processing, internet-of-things (IoT), bioelectronics, and hardware security. Mature nodes can also be cost-effective ways to explore novel computational concepts, new logic, memory, or hybrid integration (see Section 3.3, Chiplets, below). Commercial-class technologies at mature nodes also allow significantly sized chips to be built – these may not need to be the most advanced technologies, but they do need to be available and affordable due to the large chip size required. Data from TSRI (Taiwan) [TSRI] show nodes from 90-nm to 45-nm are still popular among academic researchers. The learning curve for design and EDA tools are easier to climb and IP blocks more readily available. The cost-benefit tradeoff clearly depends on the research questions and if the knowledge from one node is general enough that is amenable to translation to other nodes.

3.3 Chiplets, heterogeneous integration, and advanced packaging:

As noted at the background of this meeting (Section 2), the future is system integration. Future research requires access to heterogeneous technologies including 2.5D and 3D integration/packaging technologies. Typically, the development of packaging technology is quite difficult, it is usually integrated directly with the customer's products. IC-Link from IMEC [IMEC-IC-Link] in Belgium provides services from design to packaging. TSRI in Taiwan has started the development of 2.5D chip-on-chip-on-package, and photonics interposer; but these are still in development and are not generally available. Solutions may be found in establishing a national prototyping facility that offer advanced packaging as a service (see Section 4).

The custom-design nature of packaging is a general issue of the technology in industry. The issue is the lack of a common interface protocol. The DARPA Common Heterogeneous Integration and IP Reuse Strategies (CHIPS) [DARPA-CHIPS] program aims to develop the design tools and integration standards required to demonstrate modular integrated circuit (IC) designs. This vision of an ecosystem of discrete modular, reusable IP blocks, which can be assembled into a system using existing and emerging integration technologies, has yet to be realized. Modularity and reusability of IP blocks will require electrical and physical interface standards to be widely adopted by the community supporting the ecosystem. *Until such an ecosystem exists, and there are foundries that can offer such integration services, chiplets heterogeneous integration and packaging remain inaccessible to academic researchers*.

4. **Foundry access for creating new technology demonstrators**

4.1 *Value proposition*

As much as science likes a simple story, the history of microelectronics advancement has never been a straight line of two-dimensional (2D) scaling down of the device size, and it is certainly not simply about developing the next-generation lithography. New physics (quantum mechanical tunneling, strained silicon), new materials (copper, low-k dielectric isolation, high-k gate dielectrics, metal gate electrodes), new fabrication methods enabled by chemistry and materials fundamentals (chemical-mechanical polishing, atomic layer deposition), new design methodologies (TCAD and EDA tools, design-technology co-optimization that relies on accurate and fast models and simulation methods), all contributed to the phenomenal growth in energy efficiency, speed, and functionality of information and communication technology over the past 50 years.

Yet, microelectronics is at a crossroads today. The maturation of 2D scaling has driven the development of microelectronics in qualitatively different directions that promise dramatically enhanced performance and energy efficiency. In particular, there is an acceleration in adopting new materials and new devices into the broad microelectronics ecosystem. This is prefaced by two decades of investments in nanotechnology by the NNI at the national scale, and sustained investments in basic sciences by federal agencies such as the NSF and the DoE that have created a long and broad research pipeline ready for translation into technologies for microelectronics based on new materials, new physics, new fabrication methods, and new system architectures using new device technologies. The recent resurgence of microelectronics (e.g. the DARPA ERI program) as a focused area of research is indicative of the vast opportunity in front of us.

The highest impact will come from end-to-end, hybrid integration of heterogeneous technologies (logic, memory, interconnect, photonics, spintronics, nanomechanics, sensors/acutators, RF/mm-wave, communication) in all forms of 3D integration. Heterogeneous on-chip integration is facilitated by acquiring Si CMOS wafers from foundries and integrating beyond-CMOS devices on to these wafers. An early example is shown by the IARPA project on Trusted Integrated Circuits (TIC) [IARPA TIC] in which a variety of beyond-CMOS devices are integrated onto foundry CMOS wafers using facilities at universities (e.g. NSF NNCI nodes). These include reconfigurable photonic networks, wavelength tunable hybrid III-V/silicon laser, nanomechanics, piezoelectrics, AlN MEMS, resistive switching memory, SAR ADC, wideband RF receiver, neurocomputing associative memory – integrated onto foundry silicon CMOS, analogous to a "sauce plus pasta" menu that is extremely rich in functionality [IARPA TIC publications]. The CMOS circuits from the foundry wafer provide the control, sense, amplification, and computation for system-level functionality required for lab-to-fab translation demonstrations. Similarly, the DARPA 3DSoC program is pursuing dense monolithic 3D integration of a variety of heterogeneous nanotechnologies (silicon CMOS, carbon nanotube FETs, Resisitive RAM) at SkyWater Technology Foundry.

University cleanrooms (e.g. NSF NNCI facilities) excel in discovery and they are not missioned for larger scale circuit/macro level demonstrations, let alone complete systems that show system-level benefits. For a truly new technology, foundry access is impossible because until the industry itself decides to do internal research, nothing is available. Currently, the value of these discoveries is largely unrealized because there is no means to integrate them into system scale demonstrations to prove their value and to de-risk their adoption by industry. In general, commercial foundries are unable and unwilling to

pursue such transformative work, which is incompatible with their business models. Also, historically, even when industry does decide to invest in such transitions via internal research, they keep the new capabilities as proprietary trade secrets. Industrial fabs are further inhibited in these new domains by their strict contamination controls and regimented process flows, and operate with costly 300 mm tools. Hence, introducing new materials and new devices require extensive initial proof of efficacy before the risk and efforts can be justified. A foundry is not going to create a new device recipe to facilitate creation of a new technology that relies on system demonstration to show potential benefits. Yet, the history of academic research includes many examples of successful systems that were not only designed, but actually built, at universities. This has essentially ended. Academic computing architectural research is essentially all simulation at this point. With better access to fabrication (even “mature nodes”) we could get back to a mode where researchers not only design and simulate, but actually build systems. Therefore it is necessary to create new capabilities that bridge this “innovation lab-to-fab gap” by connecting academic research labs and commercial fabs with new facilities and new kinds of partnerships.

The challenge today is in demonstrating the benefits of technology innovations beyond the laboratory scale of 1 to 1,000 devices. Innovation at system and architecture level are meaningful (that is, solving meaningful problems) only when demonstrated at scale. For example, array-level characterization of memory is critical for 3-sigma statistics including variations, read/write disturb, and aging – issues that require understanding of fundamental physics for solutions. Often, it is possible to mitigate some of the device weaknesses at the circuit level. While one can explore the circuit/system design space using simulation tools (e.g. NeuroSim, Destiny), the underlying models used in these tools must be validated by actual experimental data. Moreover, to keep up with quickly changing applications and architectures, re-engineering existing simulators to support new applications, architectures, and technologies requires significant development time. By the time simulators are re-engineered, the accelerators they target may become obsolete. To keep up with this fast pace, a tight loop between experimental system measurements and simulators is crucial. This creates a critical need to translate application needs into technology targets, build systems using such technologies, and calibrate back system-level models – *very quickly*.

4.2 *Present solutions*

A small subset of university researchers have access to wafers from foundries through ad hoc personal connections (with lengthy legal processes) and post-process these wafers above the back-end-of-the-line (BEOL) layers to add the “sauce on the pasta” in the university cleanrooms. At university cleanrooms, demonstrations at the 1,000 device scale are generally the limit of such experimental settings. This is due in part to the impossible dual tasks of maintaining stable processes while at the same time allowing maximum flexibility for explorations.

In a very small number of cases, where new devices are emerging from commercialization (e.g. RRAM, MRAM), university researchers with special agreements may tape out circuits using those emerging device technologies at foundries. However, even in those cases, foundries only provide macro cells as a black box. A designer neither can try new circuit topologies using individual transistors and memories, nor can she simulate the device using a SPICE model of the unique devices because the device models are not available. Moreover, there is no way to customize and improve the device characteristics and it is challenging to adapt these rigid devices from foundries to the new applications are the inspirations of the innovations.

NIST has been sourcing mature-node CMOS wafers from foundries and developing processes for integrating RRAM on these foundry wafers. NIST has been considering building a standardized set of CMOS wafers that would exist solely for the use of U.S. universities to prototype novel devices and measure them in the back-end-of-line (BEOL). The wafers would have I.P. macros for measurement of arrays of two-, and three- terminal devices, I/V characterization capability, C/V measurement ability, as well as some high frequency oscillator and pulse based measurements, among other possible measurement technologies. The mask sets for this would be owned by the government with permission for U.S. universities to directly source the wafers from a domestic foundry at the cost of wafer fabrication. For integration into research fabrication processes, the wafers would come planarized at an intermediate damascene step and have alignment marks for stepper, contact, and e-beam lithography. A key challenge has been finding foundries that are willing to modify their processes even slightly to allow for subsequent integration (NIST has identified and engaged with a few suitable partners as noted above). There is also the very large cost of the mask sets. However, after the maskset costs are paid, the per-wafer cost is much more manageable. This approach may work best with an open PDK (see Section 5) to facilitate the free and open exchange of information and design I.P. as well as reduce barriers to entry to the foundry space and facilitate technology transfer. If implemented, this may be a useful resource for advanced prototyping and characterization and could grow over time to meet

evolving needs. Additional engagement with other agencies and potential users of this resource is needed to realize this plan.

4.3 *Proposal for a National Facility*

The initial exploration at NIST suggests that it may be possible to establish a National Facility to facilitate demonstration of realistic systems of emerging technologies. Having such lab-to-fab translation of systems technologies makes academic research relevant and goes a long way toward advancing device/process technology as well as architecture innovations in US universities.

The mission of such a National Facility would be (a) fast turn-around experimentation of chip-scale and package-scale systems, (b) achieve flexibility at scale, and (c) facilitate lab-to-fab translation of systems technology, thereby making academic research relevant for advancing foundational microelectronics technology for the country. This national facility will be analogous to a "MOSIS for technologists and system designers" [MOSIS].

The National Facility[9] will consists of well-equipped facility/facilities along with a network of regional hubs across the US. The hubs will have dual functionality: (i) they will put in place a protocol (such as alignment, material compatibility etc.) by which they will take completed wafers with CMOS circuits from foundries as starting material and integrate new functional devices on top at wafer scale (albeit smaller diameter wafers $\leq$ 200 mm) targeted for at-scale demonstrations (ii) they may also have capability to fabricate front end transistors including a simple CMOS baseline. The aim for this CMOS baseline is to not compete with Foundry made circuits but rather (a) provide a facility where advanced research on front-end transistors can be performed at the relevant dimension (L_G~10-20 nm) and (b) provide a pathway to develop small prototype circuits (<1000 devices) where fine-grain interconnection between transistors and other functional devices is required. For example, a number of emerging computing concepts, such as probabilistic computing, hyper-dimensional computing, some of the computing in memory schemes, and also alternative state variable based ideas such as spintronics, rely on architecture where the basic component is not just a transistor but a larger block involving transistors and other functional devices. There will be permanent staff who work with visiting researchers to translate new processes (developed at universities) into capabilities for system prototypes. There will also

[9] In one estimate, a National Facility such as the one described here would require funding for US$ 1.6B over the course of 4 years.

be protocol in place to transfer the successful technology developed at the university hubs to the central core for large scale integration. There is a general agreement now that the computing technology in the next decade and beyond will look very different from what we have today. These hubs will play a critically role for US preparedness for that exciting future.

To make this broadly available to academic researchers, there will be a design ecosystem that consists of a wafer brokerage and well-defined design interfaces that specify the process technology level interfaces (e.g. alignment marks, materials and planarity of the top surfaces) as well as the circuit design protocols. Some level of standardization of the fabrication and design methodology is needed. Design enablement such as PDKs, foundational IPs, parametric test structures and device models should be made available. There will be a one-stop shop for handling legal processes and non-disclosure agreements, collecting and composing multiple design projects into mask sets, distribution of foundry wafers to users, and serve as the interface to foundries.

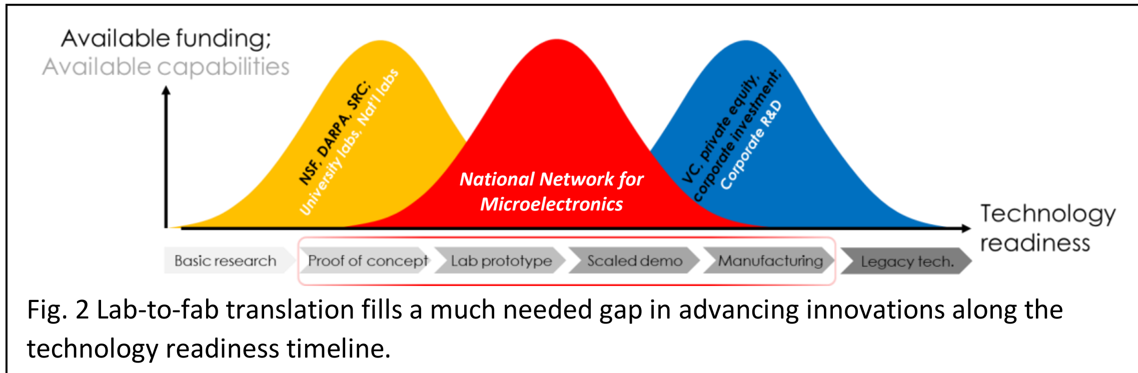

Fig. 2 Lab-to-fab translation fills a much needed gap in advancing innovations along the technology readiness timeline.

Elements of such National Facility exist in other parts of the world. The TSRI in Taiwan provides foundry shuttle wafers for Taiwan university researchers working at the memory array level. TSRI can do post-processing on such shuttle wafers to fabricate and integrate new devices such as MRAM, RRAM, ferroelectrics, and gas sensors. A 248-nm DUV stepper lithography tool will be put into service in 2021, with the goal of fabricating 1 Mb memory array level as a first demonstration target in 2021. In France, CEA-LETI worked with a US university and offered 130-nm CMOS+RRAM/PCM technology for system demonstrations [CEA-LETI projects]. Device fabrication at national facilities at foreign countries (such as TSRI and CEA-LETI) require the researcher to have a collaboration project with the foreign party and is not a simple fee-for-service arrangement. At IMEC (Belgium), architecture-informed technology innovation is critical. IMEC has been facilitating research in this way and has the capability to prototype beyond-CMOS

devices at 200-mm and 300-mm wafers. A US university can participate through joint projects. A path for direct access to IMEC technologies is not available currently. If the US intends to remain competitive in these domains, however, it needs to make investments such as those described in this document.

5. **Access to design ecosystem (EDA tools, design flows, IP blocks):**

5.1 *Current status*:

Not all academic researchers have equal access to commercial EDA tools, especially tools for leading edge technologies. In addition to tools, access to libraries and design IP (also proprietary) is uneven. A pre-condition to foundry access is Non-Disclosure Agreements (NDAs) and satisfying US Export Control regulations. NDA with foundries takes months (or even years) to negotiate with onerous requirements that are difficult to satisfy on university campuses that are open to all students. Export Control constraints often mean some foundry technologies are not accessible or some students cannot participate in certain projects.

With the continual evolution of the semiconductor fabrication process (such as transition from planar bulk transistor to FinFET and the future nanosheet transistors) as well as significant design-technology co-optimization, the complexity of design rules have skyrocketed. The technical threshold of chip design and use of EDA tools has also become much higher than before. A large body of institutional memory and expertise are needed for successful execution of tape-outs, including, for example, tape-out knowledge (top layer routing, pads, pad frames), CAD tool expertise and design flows, IP and memory compilers, verification, test, board design, inputs, and measurements. As the complexity of 2.5D and 3D integration technology increases, more cross-domain interaction effects will affect the characteristics of components and circuit systems.

Tool support has surfaced as a critical need. This is often built up within the graduate student population. But this only works for research groups that have extensive circuit design activities. In some instances, a particular research group might develop expertise on one particular process node over the course of several years and if they were to move to a more advanced node, there is a need to relearn and redo the group’s design infrastructure before productive research results are forthcoming.

It becomes increasingly difficult for the individual researcher to sustain the design-related knowledge and expertise within one's research group. For larger research groups, there may be some sharing of such items on projects. Also, multiple professors working together have made some inroads on this. Those require a level of trust and cooperation that is developed over time. In other words, the institutional memory and expertise for supporting chip design have to be spread across a wide range of faculty. For smaller research groups or academics in smaller academic departments, this is not possible. Ideally, university research staff would be the best resource for such institutional knowledge. However, almost no university, including the largest ones, can afford such staff, let alone individual faculty members. This situation is akin to those experienced by device technology researchers: there is a steep learning curve for students to learn all the tools and processes needed to execute the research idea, and the tool infrastructure is so complex and expensive that sharing of knowledge and costs become necessary.

5.2 *Open design tools and design flows*:

The current movement toward open source EDA tools and open design flows using a mixture of commercial and research EDA tools (e.g., in the cloud) may be a step toward solving some of the current issues outlined above. For example, the DARPA POSH Open Source Hardware [DARPA POSH] program aims to manage the complexity of SoC development, design reuse in the form of Intellectual Property (IP) modules. IP modules are pre-designed, functional circuit blocks that are similar in nature to software library functions and are developed internally by an organization or procured from an external, third-party IP vendor. However, for successful deployment, it is critical that the IP modules are thoroughly verified to ensure their robustness. The OpenRoad and other projects, part of DARPA's IEDA program, aim to enable no-humans, 24-hour design and catalyze open source EDA [OpenROAD]. Cloud-based design flows [Synopsys cloud] [Cadence cloud] [Google cloud] are beginning to emerge and offer new possibilities. For example, in 2019 TSMC hosted the first IC layout contest in Taiwan and students from different campuses participate in the contest remotely using a cloud design environment [Cloud IC Layout Contest]. Wide-spread academic access to these cloud-based design tools is still not available.

While RISC-V provided the free and open RISC instruction set processor core, there is still the need for libraries of modules such as accelerators, networks/buses, serial interfaces, and wireless modules. An answer to the access to design tool and design flow quandary are community-driven cloud-based design frameworks. Setting up, maintaining and updating frameworks and tools is beyond most universities. There is a need to transition

to higher-level abstractions to link to architecture and system communities. Examples of such efforts include CHISEL [CHISEL] and CHIPYARD [CHIPYARD]. An ecosystem for robust IP blocks is critical; fast yet thorough verification and characterization of digital and analog/mixed-signal blocks will be key. It is necessary to ensure robust evolution and maintenance of such design blocks – both from design standpoint as well as from EDA and research standpoint (since it will then free up the legal white space for EDA innovators / startups). Along these lines, Google has collaborated with FOSSI to allow architects to make or access custom hardware, albeit at fairly mature technology nodes.

6. **Current offerings**:

Table 1 is a summary of available offerings from foundries and related resources. The information is not meant to be all encompassing, but it does indicate the degree to which various foundry access is available to academic researchers in the US. Commercial companies that may be amenable to low volume technology development type activities include NHanced Semiconductor, Micross, xFab, and Skywater.

Institution	Technology	Availability	Country
MOSIS	**TSMC**: Up to 12/16 nm FinFET **Globalfoundries**: Up to 12nm FinFET **Intel Custom Foundry**: 22nm FinFET (22FFL)	Fee for service	USA
FOSSI (in collaboration with Google)	Skywater: 130nm PDK: 130nm digital standard cells, I/O, analog, SRAM, Flash	Free, design must be open source	USA
DARPA	Intel 22-nm	Must be part of DARPA program, e.g. FRANC, 3DSOC, T-MUSIC, CHIPS, POSH,	USA

		IDEA, PIPES, RTML	
Efabless (in collaboration with Google)	FOSSI 130 nm PDK	Free. Design must be open source	USA
SRC Texas Analog Center of Excellence	65 nm analog	Must be part of Center	USA
MUSE	TSMC, up to 28-nm, basic TSMC IPs	Fee for service	USA
NIST	User facilities at Boulder, CO, and at Gaithersburg	Boulder: superconducting electronics Gaithersburg: MEMS/photonics	USA
IMEC IC-Link	TSMC	From design services to package. Fee for service.	Belgium
TSRI	TSMC 16-nm	Discounted access through TSRI (CIC) to TSMC 16-nm tape-out. In-house fabricated 16nm FinFET, 8-inch process line (DUV, stepper), Si/Ge/2D, MRAM/RRAM, GaN, silicon photonics, MEMS. Collaboration projects only.	Taiwan

7. **Education and Workforce Development:**

Training students on relevant chip design is becoming highly challenging due to the difficulty and expense to access current technology nodes and the time associated with tape-outs. The lack of relevant design access is eroding students' interest. For example, registration data from Carnegie Mellon University showed a continuous decline since 2011

in the proportion of undergraduate and MS students who have taken either an analog or digital IC design course. By 2019, there were only 2 BS and 2 MS students who had taken IC design courses from a graduating cohort of over 500 students!

In 2020, CMU redesigned the IC design curriculum and added a new VLSI implementation course that included design and realistic system-scale chip tape outs with a follow-on course on the chip testing. Entry into the course required completion of at least one analog or digital design class, and the initial VLSI offering enrollment was 19 students, in large part due to the excitement about actual fabrication and testing of complete systems. This redesigned IC design course was made possible by (a) an industry sponsor that paid for the chip tape-outs (28-nm CMOS at TSMC) and (b) one-off connections that gained access to 28-nm PDK from TSMC for teaching purposes. The course is being offering again in Spring 2021 and the enrollment is at the maximum limit of 30 students with a waitlist.

It is important to note that investments required by a university to *support* IC design are more significant than the *cost* of the silicon tape-out. The investments include commitment to teaching and research faculty with VLSI expertise, staff and teaching assistants for CAD tool installation and support, faculty oversight of NDAs and PDK management, and cost of certain proprietary IP blocks.

While 28-nm and older technologies are still relevant for the exploration of novel computational concepts or hybrid integration, exposure to the most advanced technologies <20-nm (FinFET) is also essential to the education of the next-generation designer. With the current trend of end-user system companies (such as Apple, Amazon, Google, Facebook, and Tesla) doing their own chip designs to augment buying standard products from fabless companies (e.g. Nvidia, AMD) or integrated device manufacturers (e.g. Intel), the demand for circuit design expertise is only going to grow.

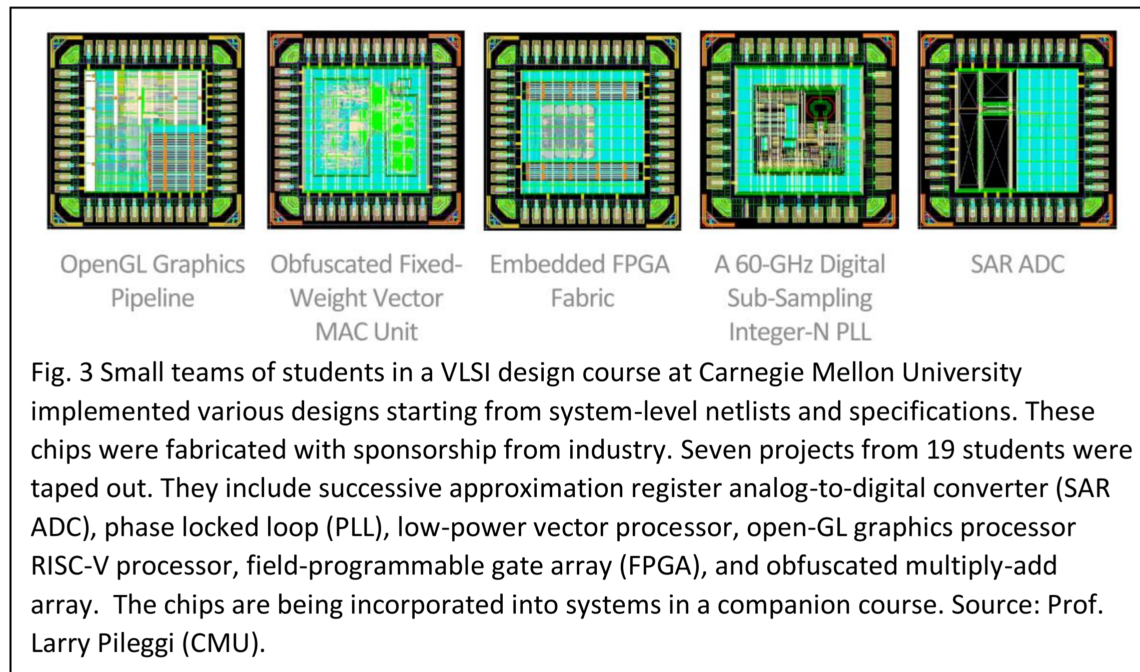

Fig. 3 Small teams of students in a VLSI design course at Carnegie Mellon University implemented various designs starting from system-level netlists and specifications. These chips were fabricated with sponsorship from industry. Seven projects from 19 students were taped out. They include successive approximation register analog-to-digital converter (SAR ADC), phase locked loop (PLL), low-power vector processor, open-GL graphics processor RISC-V processor, field-programmable gate array (FPGA), and obfuscated multiply-add array. The chips are being incorporated into systems in a companion course. Source: Prof. Larry Pileggi (CMU).

8. **Conclusions and Recommendations:**

"*Semiconductors are essential to many products used in modern life, from computers, cellular telephones, and solar panels, to medical diagnostics and self-driving cars. Progress in semiconductors has opened up new frontiers for devices and services that use them, creating new businesses and industries and bringing massive benefits to American workers and consumers and to the global economy*." [PCAST 2017]

From environment and sustainability, to human health, economic growth, and equity of opportunity, almost all societal-scale issues we face today rely on continued advances in information and communication technology. As an example, 12 out of the 17 goals of sustainable development identified by the United Nations [UN] are directly dependent on sustained technology advances.

Semiconductor technology and microelectronics is a foundational technology that without its *continued* advancement, the promises of artificial intelligence (AI), 5G/6G, and quantum computing will never become a reality. Our nation's economic competitiveness, technology leadership, and national security, depend on our staying at the forefront of microelectronics.

Decades of investments in basic nanoscience, by the NSF, DoE, and the National Nanotechnology Initiative, have nurtured many potential ideas ready for translation into technologies. However, the cycle time for hardware experimentations are currently much longer than software development and computer simulation-based studies. We are also not fully utilizing the research brain power due to poor access to semiconductor foundries. We must increase the pace of exploration and broaden the set of researchers who possess the capability to do research in circuit design and device technologies. There is an urgent need for access to semiconductor foundry and design ecosystem to achieve these goals.

Microelectronics is a field that requires sustained, rapid innovations. Semiconductor technology is at a crossroads, and the historical rate of progress is no longer guaranteed as it had in the past. Yet, there are many plausible paths to move forward, and the potential for further advances is immense. The future is system integration [DARPA ERI 2020]. And this requires end-to-end co-design and innovation. Isolated push along silos will be inadequate. It is in this context that large-scale efforts, best coordinated by the National Science Foundation, can make a substantial difference.

Because end-to-end co-design innovation is essential going forward, there is an urgent need for academia and even companies, to access foundry technologies that will facilitate bridging the lab-to-fab gap, thereby enabling the building of prototypes of new technologies (device and integration technologies) with more than a few component devices (as often happens today in a university lab). There is also a strong need from academia to access advanced technology nodes for circuits and systems research. The design ecosystem must also advance in response to the emerging technologies, and cope with enormous complexity associated with systems built using advanced technologies. Furthermore, there is a need to lower the barrier to entry for circuit design and shorten design time.

To ensure progress, we must lower the barrier for ***all*** US researchers for accessing state-of-the-art foundry services and the design ecosystem. We recommend the following efforts led by the NSF.

A. **Create a new initiative that connects software and hardware foundations.** This new initiative would specifically address research questions related to the use of new device component technologies.
B. **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 2 generations behind the state-of-the-art. The majority of researchers are still working with technologies that are at least 3 generations behind. Current practices (e.g. by DARPA, IARPA) have shown that access can be arranged when the full support of the US 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 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.

C. **Support/establish a national facility for prototyping emerging technologies at-scale.** The establishment of the multi-project wafer (MPW) service by MOSIS has dramatically changed the landscape of circuit design education, research and its commercialization. We must find ways to demonstrate emerging device technologies *at scale*, beyond the 1 to 1,000 devices scale that are sufficient for an initial exploration. A national facility should be established with the mission to enable fast turn-around experimentation of chip-scale, and package-scale systems, achieve flexibility (of material and process technologies) at scale, and facilitate demonstration of *system* technologies. 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 a MPW service, there needs to be a wafer brokerage that define the technology and the design interface protocols. Partnership with other branches of the US government may leverage the same facility to further lab-to-fab translation to industry.

D. **Open access for design ecosystem.** Electronic design automation (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, using a mixture of commercial and research/open-source EDA tools (e.g., in the cloud) as well as advanced and robust IP blocks to 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.

E. **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 part of C above). Without this enablement, the exciting promise of system-level demonstrations of emerging technologies cannot be fulfilled.

F. **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. The NSF must find ways to incentivize and assist universities to develop and offer engaging integrated circuit (IC) design courses using real technologies that are used in practice.

ACKNOWLEDGMENTS

Professor Sayeef Salahuddin (UC Berkeley) and Professor Shimeng Yu (Georgia Institute of Technology) kindly served as scribe for the workshop discussion. Professor H.-S. Philip Wong led the effort to draft this report with inputs from all participants.

REFERENCES

[1] [MOSIS] https://www.mosis.com/, accessed January 27, 2021.

[2] [NNCI] https://www.nnci.net/, accessed February 26, 2021.

[3] [DoE18] DoE report: "Basic Research Needs for Microelectronics," Oct 2018. https://www.osti.gov/biblio/1545772-basic-research-needs-microelectronics, accessed January 27, 2021.

[4] [VDE] https://pr.tsmc.com/english/news/1973, accessed January 27, 2021.

[5] [TSRI] https://www.tsri.org.tw/en/index.html, accessed January 27, 2021.

[6] [IMEC-IC-Link] https://www.imec-int.com/en/ic-link, accessed January 27, 2021.

[7] [DARPA-CHIPS] https://www.darpa.mil/program/common-heterogeneous-integration-and-ip-reuse-strategies, accessed January 27, 2021.

[8] [IARPA TIC] https://www.iarpa.gov/index.php/research-programs/tic, accessed January 27, 2021.

[9] [IARPA TIC publications] https://scholar.google.com/scholar?start=0&q=N66001-12-C-2008+OR+N66001-12-C-2013+OR+N66001-12-C-2009+OR+N66001-12-C-2011+OR+N66001-12-C-2010+OR+N66001-12-C-2012&hl=en&as_sdt=0,47, accessed January 27, 2021.

[10] [CEA-LETI projects] T.F. Wu, B.Q. Le, R. Radway, A. Bartolo, W. Hwang, S. Jeong, H. Li, P. Tandon, E. Vianello, P. Vivet, E. Nowak, M.K. Wootters, H.-S. P. Wong, M.M. Sabry

Aly, E. Beigne, S. Mitra, "A 43pJ/Cycle Non-Volatile Microcontroller with 4.7µs Shutdown/Wake-up Integrating 2.3-bit/Cell Resistive RAM and Resilience Techniques," *International Solid State Circuits Conference (ISSCC)*, paper 14.3, San Francisco, CA, February 17 – 21, 2019.

[11] [DARPA POSH] https://www.darpa.mil/program/posh-open-source-hardware, accessed January 27, 2021.

[12] [OpenROAD] T. Ajayi, V. A. Chhabria, M. Fogaça, S. Hashemi, A. Hosny, A. B. Kahng, M. Kim, J. Lee, U. Mallappa, M. Neseem, G. Pradipta, S. Reda, M. Saligane, S. S. Sapatnekar, C. Sechen, M. Shalan, W. Swartz, L. Wang, Z. Wang, M. Woo and B. Xu, "Toward an Open-Source Digital Flow: First Learnings from the OpenROAD Project", Proc. ACM/IEEE Design Automation Conference, 2019, pp. 76:1-76:4. See also https://theopenroadproject.org/, accessed January 27, 2021.

[13] [CHISEL] https://www.chisel-lang.org/, accessed January 27, 2021.

[14] [CHIPYARD] https://chipyard.readthedocs.io/en/latest/, accessed January 27, 2021.

[15] [Synopsys cloud] https://www.synopsys.com/cloud.html, accessed January 27, 2021.

[16] [Cadence cloud] https://www.cadence.com/en_US/home/solutions/cadence-cloud/cloud-hosted-design-solution.html, accessed January 27, 2021.

[17] [Google cloud] https://cloud.google.com/blog/products/compute/scale-up-your-eda-flows-on-google-cloud, accessed January 27, 2021.

[18] [TSMC Cloud IC Layout Contest] https://www.tsmc.com/tsmcdotcom/PRListingNewsArchivesAction.do?action=detail&newsid=THHIPGPGTH&language=E, accessed January 27, 2021.

[19] [DARPA ERI 2020] H.-S. P. Wong, "The Future is System Integration," keynote, *DARPA ERI Summit (ERI Summit)*, Seattle, WA, August 18 – 20, 2020. Video of presentation: https://youtu.be/cJh7EMWa6J0, accessed January 27, 2021.

[20] [PCAST report 2017] President's Council of Advisors on Science and Technology (PCAST), report to the President on "Ensuring Long-Term US Leadership in Semiconductors," https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_ensuring_long-term_us_leadership_in_semiconductors.pdf, accessed January 27, 2021.

[21] [UN] https://www.un.org/sustainabledevelopment/sustainable-development-goals/, accessed January 27, 2021.

Appendix:

A.1 Meeting Participants:

First name	Last name	Institution	Provided written input
Tim	Ansell	Google	x
Sankar	Basu	NSF	
Keren	Bergman	Columbia University	
Erik	Brunvand	NSF	
Gert	Cauwenberghs	UC San Diego	
An	Chen	SRC	
Jason	Cong	UC Los Angeles	x
Damian	Dudek	FDG, Germany	x
Thomas	Ernst	CEA LETI, France	x
Brad	Ferguson	Skywater Technology	
Paul	Franzon	North Carolina State U	x
Lisa	Friedersdorf	NNCO	
Paolo	Gargini	IEEE IRDS	x
Erwin	Gianchandani	NSF	
Tim	Green	SRC	
David	Henshall	SRC	
Richard	Ho	Google	
Brian	Hoskins	NIST	x
Sharon	Hu	U. Notre Dame	
Todd	Hylton	UC San Diego	x
Subu	Iyer	UC Los Angeles	
Alex	Jones	NSF	
Andrew	Kahng	UC San Diego	x
Wei-Shinn	Ku	NSF	
Thomas	Kuech	NSF	
In Hee	Lee	U. Pittsburgh	
James	Liddle	NIST	
Tsu-Jae	Liu	UC Berkeley	
Margaret	Martonosi	NSF	

Pinaki	Mazumder	NSF	x
Thomas	Mikolajick	NamLab, Germany	x
Subhasish	Mitra	Stanford University	
Thyaga	Nandagopal	NSF	
Larry	Pileggi	Carnegie Mellon University	x
Robinson	Pino	Department of Energy	
Wolfgang	Porod	U. Notre Dame	
Geoff	Porter	MUSE Semi	
Jan	Rabaey	UC Berkeley	
Daniel	Radack	IDA	x
Arijit	Raychowdhury	Georgia Institute of Technology	
Elyse	Rosenbaum	U. Illinois, Urbana Champaign	
Tajana	Rosing	UC San Diego	
Mark	Rosker	DARPA	x
Rob	Rutenbar	U. Pittsburgh	
Sayeef	Salahuddin	UC Berkeley	x
Devanand	Shenoy	ISI, University of Southern California	x
H.-S. Philip	Wong	Stanford University	
Wen-kuan	Yeh	TSRI, Taiwan	x
David	Yeh	SRC	x
Todd	Younkin	SRC	
Shimeng	Yu	Georgia Institute of Technology	x

A2. Meeting Agenda:

December 16, 2020
2:15-3:45 pm
Introduction: NSF
Plenary Talks

- Philip Wong, Stanford
- Mark Rosker, DARPA
- Larry Pileggi, CMU
- Richard Ho, Google
- Dev Shenoy, MOSIS

4:00-5:00 pm
Position statements:

- **Theme 1:** Foundry access to advanced technologies by designers (*Moderators:* Erik Brunvand, Subhasish Mitra).
- **Theme 2:** Foundry access for creating new technology demonstrators (*Moderator:* Philip Wong)
- **Theme 3:** Access to design ecosystem - EDA tools, design flows, IP blocks. (*Moderators:* Subhasish Mitra, Erik Brunvand).

5:00-5:30 pm
Further Discussions and Concluding Remarks

A3. List of questions for participants:

Theme 1: Foundry access to advanced technologies by designers
(Moderators: Erik Brunvand, Subhasish Mitra).

Unlike in the past (with access to advanced technology nm nodes), future research might mean access to heterogeneous technologies INCLUDING integration/packaging technologies.

a) Access to advanced technology nodes for standard CMOS, 7-nm and beyond. Why is this essential? Who in research needs this? Why? How do we deal with $$$, NDA, Export control, IP blocks for these? At what point does this stop being an in-house academic exercise and start requiring active foundry participation?

b) Can anything meaningful be done by having access to old (90nm) node silicon only? What are those? What values can they bring? More generally, what are the most useful nodes? Where's the sweet spot? 130nm? 90nm? 65nm? 45nm? 28nm? 22nm? (e.g. cost-benefit, ease of use, access to IP, device characteristics that are suitable, suitability for in-house design by students, etc.).

c) What testing facilities are required for projects fabricated in the various process nodes? Where are the testing facilities/machines and how do academics get access to those? Is there a node below which testing becomes much harder?
d) Access to emerging technologies (in development) such as MRAM, RRAM, or carbon nanotube FETs, Ferroelectric FETs. Do these necessarily have to be at the most advanced nodes? (Some foundries may have these at older nodes.)

e) There may be unique value in access to new integration technologies (such as chiplets with active interposers, 3D using TSVs, bonding, monolithic 3D integration, inter-chip photonics). There is far less streamlined process for access to these integration technologies, and access to these integration technologies may be more important moving forward. How do we enable that?

f) Similar to (d) above, access to sensor technologies or GaN, for example.

Theme 2: Foundry access for creating new technology demonstrators
(Moderator: H.-S. Philip Wong)

a) For the "new" device technology you are working on, is there a need to demonstrate functionality and performance (broadly defined) beyond single device measurements? What can you show if you have more than one device on the chip that you cannot do otherwise?

b) From a circuit/system person point of view, do you need experimental demonstration or is modeling and simulation good enough for you?

c) For the "new" device technology you are working on, how do you currently demonstrate functionality and performance beyond single device measurements?

d) I know that some of the participants are able to work with foundries to do array level demonstrations. Can you tell us how you did it? And what are the limitations, if any?

e) How many devices are needed to do a meaningful exploration in the design/architecture arena?

f) What facilities and technologies are available for “new” device demonstration at TSRI in Taiwan?

g) and in Europe?

h) CEA-LETI has done many exciting research in this arena, do you work with universities in the US? If so, can you describe the mode of operation? For example, do you provide foundry services for a fee? Or do you collaborate in research? When you collaborate, who pays for the wafer runs? What technologies are available?

i) In the ideal world, what should the NSF or the US Federal government do to support this kind of research?

Theme 3: Access to design ecosystem - EDA tools, design flows, IP blocks.
(Moderators: Subhasish Mitra, Erik Brunvand).

a) Does everyone have access to EDA tools? Does everyone have the ability to support those tools at their institution? Ideally this involves personnel to help support the tools, and this type of staff person is expensive and hard to find. How can the EDA industry help with this?

b) What are the open-source alternatives to commercial EDA tools? Why open-source EDA tools alone may not be enough?

c) Access to EDA tools isn't enough. Reference design flows (from front end to back end, including power optimization, verification, board-level) are critical. And these reference flows (especially backend) may be customized for the technology nodes. How can researchers get access to those?

d) Categories (a) and (b) above aren't sufficient. Access to IP blocks (not just standard cells, macro blocks, and RAM but also I/Os, HBM and advanced) is key for any meaningful SoC demonstration. That ecosystem doesn't exist.

e) How do we create design enablement for new technologies (such as 3D, chiplets with active interposers), requiring new tools and design flows (e.g., for partitioning, thermal analysis).

f) Points (a)-(d) above are also important for driving EDA research, and we should perhaps have a discussion on that.

g) How do we revitalize support for semiconductor education? “Back in the day” chip fab by students was supported by MOSIS (through various means). How do today’s students get class-based experience in designing, fabbing, and testing?