
COMPUTING
edge
High-Performance Computing
Cybersecurity
Ethics
History
OCTOBER 2023
www.computer.org
IEEE COMPUTER SOCIETY
IEEE

IEEE COMPUTER SOCIETY D&I FUND

Drive Diversity & Inclusion in Computing

Supporting projects and programs that positively impact diversity, equity, and inclusion throughout the computing community.

DONATE TODAY!

IEEE COMPUTER SOCIETY

IEEE Foundation

COMPUTING edge

IEEE COMPUTER SOCIETY computer.org

IEEE COMPUTER SOCIETY | IEEE | Printed with inks containing soy and/or vegetable oils

STAFF

Circulation: *ComputingEdge* (ISSN 2469-7087) is published monthly by the IEEE Computer Society. IEEE Headquarters, Three Park Avenue, 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720; voice +1 714 821 8380; fax +1 714 821 4010; IEEE Computer Society Headquarters, 2001 L Street NW, Suite 700, Washington, DC 20036.

Postmaster: Send address changes to *ComputingEdge*-IEEE Membership Processing Dept., 445 Hoes Lane, Piscataway, NJ 08855. Periodicals Postage Paid at New York, New York, and at additional mailing offices. Printed in USA.

IEEE Computer Society Magazine Editors in Chief

OCTOBER 2023 · VOLUME 9 · NUMBER 10

COMPUTING edge

16 The COVID-19 High-Performance Computing Consortium

34 Gender Asymmetry in Cybersecurity: Socioeconomic Causes and Consequences

50 Early History of Texas Instrument's Digital Signal Processor

High-Performance Computing

8 More Real Than Real: The Race to Simulate Everything
ROSA M. BADIA, IAN FOSTER, AND DEJAN MILOJICIC

16 The COVID-19 High-Performance Computing Consortium
JIM BRASE, NANCY CAMPBELL, BARBARA HELLAND, THUC HOANG, MANISH PARASHAR, MICHAEL ROSENFIELD, JAMES SEXTON, AND JOHN TOWNS

Cybersecurity

24 Paul Butcher on Fuzz Testing
PHILIP WINSTON

28 Unsafe at Any Clock Speed: The Insecurity of Computer System Design, Implementation, and Operation
SEAN PEISERT

Ethics

34 Gender Asymmetry in Cybersecurity: Socioeconomic Causes and Consequences
NIR KSHETRI AND MAYA CHHETRI

40 DiVRsify: Break the Cycle and Develop VR for Everyone
TABITHA C. PECK, KYLA A. MCMULLEN, AND JOHN QUARLES

History

50 Early History of Texas Instrument's Digital Signal Processor
WANDA GASS

52 On Logistical Histories of Computing
MATTHEW HOCKENBERRY

Departments

4 Magazine Roundup
7 Editor's Note: Problem-Solving with HPC
60 Conference Calendar

Magazine Roundup

The IEEE Computer Society's lineup of 12 peer-reviewed technical magazines covers cutting-edge topics ranging from software design and computer graphics to Internet computing and security, from scientific applications and machine intelligence to visualization and microchip design. Here are highlights from recent issues.

Computer

Ethics for Digital Medicine: A Path for Ethical Emerging Medical IoT Design

The authors of this article from the July 2023 issue of *Computer* reflect on the ethical challenges facing digital medicine. They discuss the perils of ethical oversights in medical devices and the role of professional codes and regulatory oversight toward the ethical design, deployment, and operation of digital medicine devices that safely and effectively meet the needs of patients.

Computing in Science & Engineering

Destination Earth: High-Performance Computing for Weather and Climate

Destination Earth is the first grand effort to define and deploy digital twins of the Earth system. The European Commission is making this important, multiyear investment to develop this new type of information system—blending the physical and digital worlds. The scale of computational resources and data flows is unprecedented, and so are the challenges and the opportunities. Digital twins of Earth will support decision-making faced with weather extremes and climate change adaptation as well as provide the means to interact, modify, and create tailored information. Building on the latest science and technology advances, this article from the November/December 2022 issue of *Computing in Science & Engineering* describes the steps to realize the dream of preparing a more resilient society faced with unprecedented climate change in the decades to come.

IEEE Annals of the History of Computing

The Cognitive Being as the User at Project Mac

This article from the April–June 2023 issue of *IEEE Annals of the History of Computing* traces the evolution of time-sharing systems from the late 1950s to the mid-1960s to demonstrate how time-sharing systems reflect a larger epistemological shift in the history of human-computer interaction. By placing time-sharing systems within their sociohistorical context, this essay demonstrates that time-sharing was instrumental for not only the progress of interactive computing but also the development of concepts around the modern computer user. This article revolves around Project MAC, a large time-sharing project founded at MIT in 1963. Started through funding by computer scientist J.C.R. Licklider, Project MAC brought together various academics from across the university for the sole purpose of building a time-sharing system. Within this intellectually diverse community, researchers quickly found that by issuing each user a unique share, time-sharing afforded users the sensation of control.

IEEE Computer Graphics and Applications

Mobile Augmented Reality for Adding Detailed Multimedia Content to Historical Physicalizations

Combining augmented reality (AR) and physicalization offers both opportunities and challenges when representing detailed historical data. In this article from the May/June 2023 issue of *IEEE Computer Graphics and Applications*, the authors describe a framework where mobile AR supplements views of 3D prints of historical locations with interactive functionality and small

 2469-7087/23

visual details that the prints alone cannot display. Since seeing certain details requires bringing the camera close to the physical objects, the resulting camera frames may lack the visual information necessary to determine objects' positions and accurately superimpose the overlay. The authors address this by enhancing tracking of 3D prints at close distances and employing visualization techniques that allow viewing small details in ways that do not interfere with tracking.

IEEE Intelligent Systems

Generating Emotion Descriptions for Fine Art Paintings Via Multiple Painting Representations

The task of generating emotion descriptions for fine art paintings using machine learning is gaining increasing attention. However, captioning the emotions depicted in paintings is challenging due to the artistic and subtle nature of the relied-upon visual clues. Previous studies on painting emotion captioning mainly focus on content-oriented semantic features, resulting in limited performance. Recognizing that facial expressions and body language can reflect human emotions, the authors of this *IEEE Intelligent Systems* May/June 2023 article propose a novel painting emotion captioning model that incorporates two additional features: facial expression feature and human pose feature. Our model includes a feature fusion method to incorporate these features with commonly used object features. The experiment results on public datasets demonstrate that the proposed model outperforms the baseline.

IEEE Internet Computing

Model Ensemble for Predicting Heart and Respiration Rate From Speech

Stress levels are a significant source of information in assessing human well-being, including both mental and physical health. Interestingly, speech signals can be indicative of stress and may be used to infer related physiological markers, such as heart rate and respiration cycles. To this end, this article from *IEEE Internet Computing*'s May/June 2023 issue proposes a nonintrusive, low-cost, and automatic stress monitoring framework facilitating timely activation of stress relief methods and/or stress prevention. The authors design a multidomain speech feature extraction scheme able to reveal complementary stress-related characteristics. Subsequently, these are modeled by a synergistic framework able to encode both linear and nonlinear relationships via suitably learned support vectors and a recurrent neural network. They employ an appropriate corpus encompassing recordings of job interviews, constructed based on a standardized experimental protocol.

IEEE micro

A Mobile 3-D Object Recognition Processor With Deep-Learning-Based Monocular Depth Estimation

A 3D object recognition system is a heavy task that consumes high sensor power and requires complex 3D data processing. In this May/June 2023 *IEEE Micro* article, the proposed processor produces 3D red, green, blue, and depth (RGB-D) data from an RGB image through a deep learning-based monocular depth estimation, and then its RGB-D data are sporadically calibrated with low-resolution depth data from a low-power depth sensor, lowering the sensor power by 27.3 times. Then, the proposed processor accelerates various convolution operations in the system by integrating the in-out skipping-based bit-slice-level computing processing elements and flexibly allocating workloads considering data properties. Moreover, the point feature (PF) aggregator is designed close to the global memory to support the PF reuse algorithm's data aggregation. Additionally, the window-based search algorithm and

its memory management are presented for efficient point processing in the point processing unit.

IEEE MultiMedia

Learning 3-D Face Shape From Diverse Sources With Cross-Domain Face Synthesis

Monocular face reconstruction is a significant task in many multimedia applications. However, learning-based methods unequivocally suffer from the lack of large datasets annotated with 3D ground truth. To tackle this problem, the authors of this article from *IEEE MultiMedia*'s January–March 2023 issue propose a novel end-to-end 3D face reconstruction network consisting of a domain-transfer conditional GAN (cGAN) and a face reconstruction network. The method first uses cGAN to translate the realistic face images to the specific rendered style, with a novel 2D facial edge consistency loss function to exploit in-the-wild images. The domain-transferred images are then fed into a 3D face reconstruction network. They propose a novel reprojection consistency loss to restrict the 3D face reconstruction network in a self-supervised way.

IEEE pervasive COMPUTING
MOBILE SYSTEMS | UBIQUITOUS COMPUTING | INTERNET OF THINGS

Lasers on the Moon: Recommendations for Pioneering Lunar Communication Infrastructure

Future missions carrying humans to the moon will require fast and resilient communication infrastructure to allow occupants to communicate between various orbiters, rovers, and stations—and to relay valuable data back to Earth. This article from *IEEE Pervasive Computing*'s April–June 2023 issue begins by examining the two core enabling technologies for space communications: radiofrequency and optical links. These approaches are compared in the context of recent and ongoing missions by NASA and private entities. The authors propose a set of recommendations for lunar communications infrastructure to enable habitation. Three pillars unify the approach: scalable design and capabilities, resilience to environmental and cyber threats, and universal applicability to a wide range of mission profiles. The engineering efforts and areas of focus necessary to mitigate identified weaknesses and achieve the described capabilities are discussed.

IEEE SECURITY & PRIVACY

Memory Protection Keys: Facts, Key Extension Perspectives, and Discussions

Memory Protection Keys (MPK) offers per-thread memory protection with an affordable overhead, prompting many new studies. With protection key extension, MPK provides more fine-grained protection and better functionality. MPK can be an attractive option for memory protection in industry. Read more in this *IEEE Security & Privacy* article from the May/June 2023 issue.

IEEE Software

Responsible-AI-by-Design: A Pattern Collection for Designing Responsible Artificial Intelligence Systems

Responsible artificial intelligence (AI) issues often occur at the system level, crosscutting many system components and the entire software engineering lifecycle. The authors of this article from the May/June 2023 issue of *IEEE Software* summarize design patterns that can be embedded into AI systems as product features to contribute to responsible-AI-by-design.

IT Professional

Deeper Understanding of Software Change

Managing change is a challenging task in today's complex software engineering. Understanding the diversity of changes and their relationship to current technologies is critical for dealing with volatile business systems. This March/April 2023 *IT Professional* article aims to identify and assess the state of the art toward understanding software change for the sake of providing a deeper understanding of its causes, mechanisms, and effects.

Editor's Note

Problem-Solving with HPC

High-performance computing (HPC) is more powerful than ever thanks to new artificial intelligence and data analytics techniques. Some researchers are asking, "Can we employ HPC to help address global challenges like climate change and pandemics?" This *ComputingEdge* issue explores how HPC resources are being utilized around the world to help tackle large-scale problems.

Computer's "More Real Than Real: The Race to Simulate Everything" presents an interview with two HPC experts who discuss the current state of HPC, as well as where the technology might be headed. "The COVID-19 High-Performance Computing Consortium," from *Computing in Science & Engineering*, describes how the HPC community responded to the pandemic and how their efforts could be used as a model for future global crises.

Today, insecure software can cause major disruptions, which is why effective cybersecurity is paramount. *IEEE Software*'s "Paul Butcher on Fuzz Testing" features insights on an automated technique for finding vulnerabilities in software. In "Unsafe at Any Clock Speed: The Insecurity of Computer System Design, Implementation, and Operation," from *IEEE Security & Privacy*, the author argues for prioritizing secure products over corporate profits.

The cybersecurity community and other computing fields are striving for increased diversity. *Computer*'s "Gender Asymmetry in Cybersecurity: Socioeconomic Causes and Consequences" describes methods for improving women's participation in the cybersecurity workforce. The authors of "DiVRsify: Break the Cycle and Develop VR for Everyone," from *IEEE Computer Graphics and Applications*, urge virtual reality developers to engage with more diverse participant populations, to use more inclusive imagery, and to collaborate with researchers from diverse backgrounds.

This *ComputingEdge* issue concludes with two articles about the history of computer hardware. *IEEE Micro*'s "Early History of Texas Instrument's Digital Signal Processor" recounts the development of the Texas Instruments microprocessor in the 1970s and 80s. In "On Logistical Histories of Computing," from *IEEE Annals of the History of Computing*, the author discusses elements of hardware production, including labor, supply chains, and mass manufacture.

EDITOR: Dejan Milojicic, Hewlett Packard Labs, dejan.milojicic@hpe.com

This article originally appeared in
Computer
vol. 55, no. 7, 2022

DEPARTMENT: PREDICTIONS

More Real Than Real: The Race to Simulate Everything

Rosa M. Badia, *Barcelona Supercomputing Center*
Ian Foster, *Argonne National Laboratory and University of Chicago*
Dejan Milojicic, *Hewlett Packard Labs*

Extreme times require extreme measures. In this column, we discuss how high-performance computing embraces artificial intelligence and data analytics to address global challenges.

Humanity is confronted with unprecedented challenges, such as pandemics and global warming. Our future is at stake, and we require powerful tools to explore solutions to these challenges. High-performance computing (HPC) is needed more than ever to guide us through the many possible paths that are in front of us. Exascale computers are being deployed around the world, and new artificial intelligence (AI) and data analytics techniques are being adopted to complement and enhance traditional HPC techniques focused on number crunching.

Suddenly, using HPC, we are simulating and predicting many aspects of our lives: the spread of pandemics, climate change, gaming, the metaverse, digital twins, supply chains, and much more. We are almost in a race to simulate anything and everything, and that new reality is threatening to become even more real than our real life. Demands are becoming much higher; tolerance for latency to get results much smaller.

In its quest, HPC has many challenges, such as the end of Moore's law, the cost of moving data, energy limitations, and legacy programming models. Cloud computing offers some advantages but comes with its own challenges for HPC. It offers substantially higher (auto)scaling, modern development tools, and natural integration points with AI. However, today's cloud platforms cannot support large-scale, tightly coupled applications, due to jitter and unoptimized interconnects.

Digital Object Identifier 10.1109/MC.2022.3173359
Date of current version: 4 July 2022

To further discuss opportunities and challenges of HPC in modern science, we have engaged two veterans from the HPC and parallel/distributed computing world: Rosa M. Badia, of the Barcelona Supercomputing Center (BSC), and Ian Foster, of Argonne National Laboratory and the University of Chicago. Well traveled across HPC and world locations, with broad and deep perspectives on HPC from an international point of view, Rosa and Ian will help us predict the future of HPC.

Dejan Milojicic: What are the most challenging problems of HPC today?

Rosa M. Badia: In terms of the infrastructure, in my opinion, the challenges are in coping with the new technologies for memory and in the heterogeneity of the new devices in the memory storage. While these new devices have the opportunity of solving the input-output (I/O) bandwidth challenge, they also introduce new problems. In the software area, in my opinion, the challenges come from complexity of the applications that try to leverage the heterogeneous infrastructure. New solutions to develop dynamic workflows and complex applications that involve huge simulations are necessary.

Ian Foster: Everything is changing all at once: hardware, software, and applications. And more things are happening at once within HPC systems, both in the sense of there being more processors (which brings its own challenges) and due to the coupling of more

 Published by the IEEE Computer Society

and different physics, new data-driven applications, AI agents, and online data analysis and reduction, among other factors. Add in growing demand from industry for precisely the skills that HPC specialists have spent so long developing, and you have a difficult situation.

Milojicic: How do you see regional competition in various races in HPC (petascale, exascale, and so on)? Are they helpful in advancing the technology?

Badia: The European High Performance Computing Joint Undertaking (EuroHPC JU) aims to coordinate efforts and pool resources to make Europe a world leader in supercomputing. This will boost Europe's scientific excellence and industrial strength, and support the digital transformation of its economy while ensuring its technological sovereignty. The road map includes the deployment of two exascale systems, three pre-exascale systems (two already under construction) and five petascale systems (four already deployed and one under construction). I do not see this as much as a race but as each region having its own infrastructures and not needing to rely on the others.

Foster: The aggressive development of extreme-scale systems in multiple regions is great for science and engineering and great for technology. Furthermore, competition and diversity of ideas are clearly driving progress. However, I am also concerned that we are missing opportunities for collaboration. The vast scope of the problems to be solved means that every region should be looking carefully at what it must build itself versus what it can build in collaboration with, or borrow from, others. One wrinkle here, discussed by Moshe Vardi in a recent Communications of the ACM column,[2] is that HPC may be perceived as so strategic that useful cooperation is hindered. We'll see how that plays out.

Milojicic: Accelerators helped overcome and at least offset the slowdown if not the end of Moore's law. This is especially true for AI/machine learning (ML)/deep learning (DL). How important are they in HPC?

Badia: If you look at the TOP500 list, most systems are equipped with GPUs. They have been very important to sustain the end of Moore's law, and have been very important for workloads that include some form of AI and, especially, DL. It is clear that accelerators are very important for HPC, but still there are challenges in the programmability of these devices and their integration in the overall system architecture.

Foster: When GPUs first appeared, I expected them soon to be subsumed by more usable extensions to conventional processors. But I was not accounting for DL and the resulting immense demand for linear algebra accelerators. It is ironic that those developments occurred in parallel with computational science moving beyond dense matrices. Now DL is rediscovering sparsity and adaptivity, so perhaps the next generation of accelerators will be more usable for sophisticated HPC algorithms. Or maybe we'll work out how to leverage accelerators better, for example, by rethinking algorithms and repurposing AI/ML methods?

Milojicic: Rumor has it that there are over 100 accelerator start-ups in the world and at least 60 in Silicon Valley. How do these 1,000 blooming flowers help or not help our community? For example, driving innovation versus increasing efforts into system integration and programming them due to nonstandard solutions.

Badia: I think that at this moment, all these start-ups are driving innovation but also making more diverse the programming environments. At some moment, the number of alternatives will get reduced. The efforts for converging into standard solutions for programming have been there for a long time, but still

there is work to do to offer high-level, simple interfaces empowered by toolchains to support performance portability.

Foster: I'm with Rosa that high-quality toolchains and standardized interfaces are the keys to broad impact. And while TensorFlow and PyTorch are suitable interfaces for DL, we need far more for HPC applications.

Milojicic: There is a lot of discussion about quantum computing and a lot of money being invested in this area. What is the take of the HPC community on quantum computing?

Badia: BSC is also investing efforts in quantum computing. Currently, we are coordinating Quantum Spain, a project that includes the construction and commissioning of the first quantum computer in southern Europe. We are also involved in initiatives at the European level. In particular, my group is involved in the parallelization of a classical quantum computing simulator aiming to perform large-scale simulations of quantum systems, what is called *classical simulation*. Since quantum computers are neither cheap nor easy to build, classical simulation is a valuable method for efficient simulation of quantum algorithms. Classical simulation tools are a must to understand noise sources and improve the performance of quantum algorithms. These are hungry simulations both in terms of data and compute, able to fill an exascale supercomputer.

Foster: There is much exciting work underway in quantum computing, communications, and sensing. At Argonne and the University of Chicago, for example, there are efforts ranging from physical testbeds (for example, for quantum networking) to foundational work in algorithms and materials. However, I am not betting on quantum computing playing an important role in HPC any time soon. I suspect that view is widely shared within the community.

Milojicic: What is your take on the confluence of HPC with high-performance data analytics (HPDA) and AI?

Badia: The community has identified the convergence of HPC, HPDA, and AI as critical for the new type of applications that involve the three aspects. I am personally the principal investigator of the eFlows4HPC EuroHPC JU project that aims to provide a software stack to enable the development of workflows that include HPC simulations or modeling together with AI and data analytics. In the project, we have use cases that base their workflows in simulations for manufacturing, climate prediction and modeling, and urgent computing for natural hazards (for instance, earthquakes and their subsequent tsunamis). We aim at providing tools to make the development of such workflows easier. At the core of the software stack, we find the environment developed by my team, PyCOMPSs, which provides the glue for integrating the HPC, HPDA, and AI components. The project also aims at making the deployment and execution of such workflows easier in HPC infrastructures through the new methodology of HPC workflows as a service, inspired in cloud practices.

Foster: It's so important and, indeed, overdue. I'll return later to the opportunities inherent in the integration of HPC and AI, but a key point is that this convergence greatly expands the number of people that can benefit from HPC technologies. And it has implications for just about every aspect of our computing infrastructure, not least in programming models and tools.

Milojicic: How important are novel interconnects in large-scale supercomputers? With the increasing adoption of AI and less tightly coupled code, how are interconnects emerging?

Badia: Simulations based on the message passing interface (MPI) are still dominant in HPC. For this type of application, an efficient, low-latency, and high-bandwidth interconnect is a must. In addition, in my opinion, with larger data sets and more demanding workloads, AI will evolve to distributed computing also, where efficient interconnects will also be needed. While Infiniband is present in a large number of systems, specialized networks, such as the Tofu D interconnect in the Top1 Fugaku, have appeared.

Foster: What Rosa said.

Milojicic: Storage-class memories and high-bandwidth memories are finding their adoption in many areas,

including in HPC. Where do you see additional needs in memories for HPC?

Badia: As I answered in question 1, I think this is one of the challenges of HPC today. These new types of memories are needed due to the demands in terms of latency and bandwidth of the data-intensive applications, such as AI workloads or those that combine AI and HPC. However, at the same time, this new type of memory and the new memory hierarchy is a challenge.

Foster: These are really hard problems. More dynamic and data-intensive applications need places to stash increasing volumes of input, intermediate, and output data on widely varying timescales. New memory technologies certainly can help, but much work will be needed to integrate them efficiently with the full HPC stack, from applications to schedulers and beyond. Maybe we need new data abstractions, like policy-aware key value stores, to enable integration with more complex application flows. What might we do with such memories if they could also perform certain types of computation?

Milojicic: How is adoption of AI changing HPC? Will it enable new applications, new verticals? Broader adoption of HPC beyond science and core engineering?

Foster: The opportunities here are numerous and exciting. First, HPC is, of course, fundamental to modern AI. In particular, DL depends heavily on massive floating-point computation, large-scale linear algebra, high-speed communications, and parallel I/O—all things that the HPC community has been working on for years. (It's no coincidence that Jack Dongarra won the 2021 Turing Award.) So HPC has a new role. These new "HPC for AI" applications are proving to be a fascinating source of new problems that will drive many advances in algorithms, software, and hardware, for example, accelerators and reduced-precision arithmetic.

Meanwhile, "AI for HPC" is enabling a generational revolution in how HPC is employed to solve intractable problems. HPC had become almost dull as researchers worked increasingly to achieve only incremental improvements in model resolution or physics fidelity. Now we see researchers developing new AI methods that learn from computations, for example, to generate fast surrogate climate models, construct machine-learned force fields in computational chemistry, or choose the next region to explore in a molecular dynamics simulation. These new methods, when combined with extreme-scale HPC, can deliver results that are transformative rather than incremental. We're still in the early stages of exploring these opportunities. Not everything will work as expected, but there will also be unanticipated discoveries. For example, what will happen when exascale computers are used to train foundation models on data from large numbers of simulations and experiments?

Badia: I agree with all that Ian said. In fact, in my group, we are interested in providing solutions for applications or workflows that combine, at the same time, the traditional HPC simulation and modeling with AI models. Some of the use cases of eFlows4HPC are leveraging this idea, for example, by performing AI-driven pruning of ensemble members in a large Earth System Models simulation experiment to better use the computational and storage resources and releasing computational resources accordingly or by applying reduced order modeling techniques to a large number of HPC simulation results to generate a digital twin that can be deployed in edge devices to be applied in manufacturing scenarios.

Milojicic: AI has increased adoption of Python and new software frameworks. Are they penetrating the HPC community?

Badia: Yes, sure. Python is the dominating language for AI and data analytics software. In this sense, it is used in HPC systems for these types of workloads but also as programming environments for workloads that combine more traditional HPC with AI and HPDA. An example of this type of environment would be PyCOMPSs from my group, which supports the development of workflows that include MPI simulations and invocations to AI components.

Foster: I love the trailer for Geert Jan Bex's massive open online course[1] on Fortran for scientific computing, in which (at 1:30) the monster Fortran drags away a

lifeless Python as its breakfast. But in practice, Python is being used at ever-larger scales, driven, in part, by the more dynamic and heterogeneous task-parallel applications that we have already discussed. Packages like Parsl, PyCOMPSs, Radical Pilot, Colmena, and DeepDriveMD make it easy to implement such applications, including those in which computational structures evolve over time, for example, by starting, monitoring, reconfiguring, and stopping subcomputations as a simulation proceeds. The subcomputations themselves are typically coded in low-level languages, but the flexibility of Python, if used appropriately, increases overall productivity.

Milojicic: Virtualization has helped in both developer productivity and provider efficiency. After virtual machines and containers, we now have serverless, also known as *function as a service (FaaS)*. Are they convenient for HPC, and what are their advantages versus opportunities for improvement?

Badia: Containers have been largely adopted in HPC. However, the concept of serverless, or FaaS, requires some openness in the systems that are possible. For example, most HPC systems will not easily enable communication between the computing nodes and external servers. In this sense, then the FaaS is limited to the login nodes and, most of the time, under certain conditions.

Foster: Serverless is part of a move to a world in which tasks can flow readily to wherever is most accessible, convenient, or efficient. In such a computing continuum, HPC systems can serve as high-powered compute engines, enabling new applications, such as smart experimental facilities that engage large-scale, on-demand computing to drive decisions. We've also found serverless to be a useful abstraction within HPC systems. We've had success building both classes of applications (for example, via use of the funcX federated FaaS system), but it is certainly true, as Rosa notes, that aspects of current HPC architectures can get in the way.

Milojicic: Will there be a new HPC programming model introduced [for instance, in addition to MPI and partitioned global address space (PGAS)/multithreading]?

Badia: BSC has been proposing a task-based programming model as an alternative to MPI for around 15 years now. BSC has fostered the adoption of tasks and tasks with dependencies in OpenMP as well as its use with accelerators. We have promoted portable solutions with simple, high-level interfaces. In addition to the multicore/accelerator solution that is OmpSs and its successor OmpSs-2, we have also been working on solutions for distributed computing (from the grid to the cloud and now in the continuous edge to cloud, or HPC). This distributed solution has been based on the COMPSs runtime, which recently has been reengineered from a centralized to a distributed design to cope with the needs of the computing continuum.

Foster: As Rosa says, there already is. Task-parallel programming, once a niche alternative to single program, multiple data (whether MPI or PGAS) suitable for a few specialized applications, is increasingly mainstream and, indeed, fundamental to important new applications. There remain barriers to the most effective use of this new model, some of which We've discussed. I expect that we'll be working for most of the next decade to make such applications truly first-class citizens on HPC platforms.

Milojicic: Do you expect the continued growth of commercial cloud providers ultimately to subsume conventional HPC? If so, what are the pros and cons of that happening?

Badia: You can run some MPI workloads in clouds if the latency and number of communications is not very high. Other workloads that do not require this level of interconnection should be able to run well. I personally do not think that security is an issue here.

Foster: If by HPC we mean "computers with hardware and software that support large-scale, fine-grained, often data-intensive parallel computations," then the question is, "Will commercial cloud providers provide computers with such characteristics?" They certainly could, given their vast resources. Whether they will is largely a question of economics. High-end computation is a niche market that HPC centers support very well, so my expectation is that while we'll see increasing use for small-to-medium-scale science and

engineering applications, clouds won't be a factor for extreme-scale computations in the immediate future.

If we take a broader view of HPC as "the computing infrastructure used to solve challenging problems in science and engineering," then there is a big, largely unexploited, opportunity, namely, to use the cloud to host "science services" that, for example, reduce data-sharing friction, enable collaborative analysis of large data sets, and train scientific foundation models. As an example, the Globus research data management service (https://globus.org), running on Amazon Web Services (AWS) for more than a decade, today supports more than 300,000 registered users at thousands of institutions. We can and should be building many more such services.

Milojicic: What are the key benefits of HPC's cloud deployment? Cost/affordability, DevOps (for example, toolchains), something else?

Foster: For HPC as access to specialized computers, I'd point to the elastic capacity as a key potential benefit of cloud deployment. So much of computational science demands rapid response, whether to align with human thinking processes or to meet external demands for decisions. And particularly for smaller computations, commercial cloud can provide more rapid response than conventional HPC. Various HPC centers are deploying on-demand Kubernetes clusters to provide similar capabilities; it remains to be seen whether such deployments can support growing on-demand workloads cost effectively.

For HPC as "scientific computing infrastructure," the incredible richness of cloud ecosystems—far exceeding anything that can reasonably be deployed and operated at an HPC center—can be a major benefit. Using Globus as an example once again, AWS support for geographical replication, scalable logging, elastic scaling, reliable storage, security controls, and many other features has made it possible for a relatively small team to construct and operate a service that could never have been built at an HPC center.

Badia: I agree with what Ian has said, although with COMPSs, we are able to apply elasticity in Slurm-based clusters. I think HPC is also learning from these practices in clouds, like the use of containers to ease and enable automatic deployment. Adopting these methodologies will widen the use of HPC resources in new user communities.

Milojicic: Workflows are an increasingly important abstraction in HPC. As HPC reaches out of supercomputer centers and into the cloud, how are workflows evolving?

Badia: Depending on the nature of the workflows, they can naturally run in clouds the same way that they can run in HPC systems. If the components of the workflow are not very demanding in terms of the interconnection network, they can run well in clouds. In case of the type of workflows I was describing earlier, some of the components of the workflow can be MPI simulation or modeling codes that require the interconnection network of an HPC system.

Foster: We've already talked about the growing importance of task-parallel computing, which engages some aspects of workflow, within HPC systems. Also of growing importance are workflows that link scientific instrumentation with HPC to cope with exploding data rates that far exceed human cognitive capacities. Recurring variants of this pattern include on-demand HPC for data analysis (for example, to reconstruct large data sets), HPC for training ML models that are then deployed at the instrument, and integration of experimental data with simulations to form digital twins of experimental processes.

Milojicic: Is there any other prediction in HPC you would like to make, Rosa and Ian?

Badia: There is a huge potential of almost-here exascale systems that will enable us to tackle grand challenges in science and engineering. These infrastructures will be integrated in the so-called computing continuum, with instruments and sensors in the edge and Internet of Things devices. In this sense, HPC will evolve toward a more usable instrument to the general scientist through the adoption of cloud-inspired methodologies for deployment and execution.

Foster: I've already noted the profound transformation that I see happening in the nature and role of HPC.

In the grand scheme of things, it isn't so long ago that HPC was a niche technology, important only for a few specialized applications. Now HPC technologies and applications are becoming quasi-universal. It seems likely that this trend will accelerate.

I also posit that AI and HPC will combine in more surprising ways. Specifically, I think that scientists and engineers are likely to learn, sooner rather than later, how to construct AI "foundation models" that capture the current state of understanding in specific domains (for example, the climate system, soft materials, and cellular biology). As such models emerge, we will find that simulations and experiments are increasingly driven by observations of internal inconsistencies, uncertainties, and gaps in the models' representations of the world (and associated data) rather than by leaps of human intuition. This move to model-driven simulations and experiments (and simulation- and experiment-driven models) will be immensely powerful, if disconcerting, for individual researchers and will place new demands on both HPC and experimental systems.

REFERENCES

1. G. J. Bex, *Trailer for the MOOC on Fortran for Scientific Programming*. (Jul. 7, 2021). Accessed: May 19, 2022. [Online Video]. https://www.youtube.com/watch?v=I6pEaUttWo8
2. M. Vardi, "War and tech (and ACM)," *Commun. ACM*, vol. 65, no. 5, p. 9, May 2022, doi: 10.1145/3528570.

ROSA M. BADIA is a group manager at the Barcelona Supercomputing Center, Barcelona, 08034, Spain. Contact her at rosa.m.badia@bsc.es.

IAN FOSTER is a director of the Data Science and Learning Division and distinguished fellow at Argonne National Laboratory, Lemont, 60439, Illinois, USA, and a professor of computer science at the University of Chicago, Chicago, Illinois, 60637, USA. Contact him at foster@anl.gov.

DEJAN MILOJICIC is a distinguished technologist at Hewlett Packard Labs, Palo Alto, California, 94306, USA. Contact him at dejan.milojicic@hpe.com.

IEEE COMPUTER SOCIETY

www.computer.org

PURPOSE: Engaging professionals from all areas of computing, the IEEE Computer Society sets the standard for education and engagement that fuels global technological advancement. Through conferences, publications, and programs, IEEE CS empowers, guides, and shapes the future of its members, and the greater industry, enabling new opportunities to better serve our world.

OMBUDSMAN: Direct unresolved complaints to ombudsman@computer.org.

CHAPTERS: Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.

AVAILABLE INFORMATION: To check membership status, report an address change, or obtain more information on any of the following, email Customer Service at help@computer.org or call +1 714 821 8380 (international) or our toll-free number, +1 800 272 6657 (US):

- Membership applications
- Publications catalog
- Draft standards and order forms
- Technical committee list
- Technical committee application
- Chapter start-up procedures
- Student scholarship information
- Volunteer leaders/staff directory
- IEEE senior member grade application (requires 10 years practice and significant performance in five of those 10)

PUBLICATIONS AND ACTIVITIES

***Computer*:** The flagship publication of the IEEE Computer Society, *Computer*, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.

Periodicals: The Society publishes 12 magazines, 19 journals.

Conference Proceedings & Books: Conference Publishing Services publishes more than 275 titles every year.

Standards Working Groups: More than 150 groups produce IEEE standards used throughout the world.

Technical Communities: TCs provide professional interaction in more than 30 technical areas and directly influence computer engineering conferences and publications.

Conferences/Education: The society holds more than 215 conferences each year and sponsors many educational activities, including computing science accreditation.

Certifications: The society offers three software developer credentials.

COMPUTER SOCIETY OFFICES

Washington, D.C.:
2001 L St., Ste. 700,
Washington, D.C. 20036-4928;
Phone: +1 202 371 0101;
Fax: +1 202 728 9614;
Email: help@computer.org

Los Alamitos:
10662 Los Vaqueros Cir.,
Los Alamitos, CA 90720;
Phone: +1 714 821 8380;
Email: help@computer.org

MEMBERSHIP & PUBLICATION ORDERS
Phone: +1 800 272 6657; **Fax:** +1 714 816 2121;
Email: help@computer.org

EXECUTIVE COMMITTEE

President:	Nita Patel
President-Elect:	Jyotika Athavale
Past President:	William D. Gropp
First VP:	Hironori Washizaki
Second VP:	Grace A. Lewis
Secretary:	Carolyn McGregor
Treasurer:	Michela Taufer
VP, Membership & Geographic Activities:	Fernando Bouche
VP, Professional & Educational Activities:	Deborah Silver
Interim VP, Publications:	Greg Byrd
VP, Standards Activities:	Annette Reilly
VP, Technical & Conference Activities:	Grace A. Lewis
2023–2024 IEEE Division VIII Director:	Leila De Floriani
2022–2023 IEEE Division V Director:	Cecilia Metra
2023 IEEE Division V Director-Elect:	Christina M. Schober

BOARD OF GOVERNORS

Term Expiring 2023:
Jyotika Athavale, Terry Benzel, Takako Hashimoto, Irene Pazos Viana, Annette Reilly, Deborah Silver

Term Expiring 2024:
Saurabh Bagchi, Charles (Chuck) Hansen, Carlos E. Jimenez-Gomez, Daniel S. Katz, Shixia Liu, Cyril Onwubiko

Term Expiring 2025:
İlkay Altintaş, Nils Aschenbruck, Mike Hinchey, Joaquim Jorge, Rick Kazman, Carolyn McGregor

EXECUTIVE STAFF

Executive Director:	Melissa Russell
Director, Governance & Associate Executive Director:	Anne Marie Kelly
Director, Conference Operations:	Silvia Ceballos
Director, Information Technology & Services:	Sumit Kacker
Director, Marketing & Sales:	Michelle Tubb
Director, Membership Development:	Eric Berkowitz
Director, Periodicals & Special Projects:	Robin Baldwin

IEEE BOARD OF DIRECTORS

President & CEO:	Saifur Rahman
President-Elect:	Thomas M. Coughlin
Director & Secretary:	Forrest (Don) Wright
Director & Treasurer:	Mary Ellen Randall
Past President:	K. J. Ray Liu
Director & VP, Educational Activities:	Rabab Ward
Director & VP, Publication Services & Products:	Sergio Benedetto
Director & VP, Member & Geographic Activities:	Jill Gostin
Director & President, Standards Association:	Yu Yuan
Director & VP, Technical Activities:	John Verboncoeur
Director & President, IEEE-USA:	Eduardo Palacio

IEEE

revised 25 July 2023

EDITORS: Kathryn Mohror, mohror1@llnl.gov, John M. Shalf, JShalf@lbl.gov

DEPARTMENT: LEADERSHIP COMPUTING

This article originally appeared in
Computing in Science & Engineering
vol. 24, no. 1, 2022

The COVID-19 High-Performance Computing Consortium

Jim Brase, *Lawrence Livermore National Laboratory, Livermore, CA, 94550, USA*
Nancy Campbell, *IBM Research, Yorktown Heights, NY, 10598, USA*
Barbara Helland and Thuc Hoang, *Department of Energy, Washington, DC, 20585, USA*
Manish Parashar, *University of Utah, Salt Lake City, UT, 84112, USA*
Michael Rosenfield and James Sexton, *IBM Research, Yorktown Heights, NY, 10598, USA*
John Towns, *University of Illinois, Urbana, IL, 61801, USA*

In March of 2020, recognizing the potential of High Performance Computing (HPC) to accelerate understanding and the pace of scientific discovery in the fight to stop COVID-19, the HPC community assembled the largest collection of worldwide HPC resources to enable COVID-19 researchers worldwide to advance their critical efforts. Amazingly, the COVID-19 HPC Consortium was formed within one week through the joint effort of the Office of Science and Technology Policy (OSTP), the U.S. Department of Energy (DOE), the National Science Foundation (NSF), and IBM to create a unique public–private partnership between government, industry, and academic leaders. This article is the Consortium's story—how the Consortium was created, its founding members, what it provides, how it works, and its accomplishments. We will reflect on the lessons learned from the creation and operation of the Consortium and describe how the features of the Consortium could be sustained as a National Strategic Computing Reserve to ensure the nation is prepared for future crises.

In March of 2020, recognizing the potential of High-Performance Computing (HPC) to accelerate understanding and the pace of scientific discovery in the fight to stop COVID-19, the HPC community assembled the largest collection of worldwide HPC resources to enable COVID-19 researchers worldwide to advance their critical efforts. Amazingly, the COVID-19 HPC Consortium was formed within one week through the joint effort of the Office of Science and Technology Policy (OSTP), the U.S. Department of Energy (DOE), the National Science Foundation (NSF), and IBM. The Consortium created a unique public–private partnership between government, industry, and academic leaders to provide access to advanced HPC and cloud computing systems and data resources, along with critical associated technical expertise and support, at no cost to researchers in the fight against COVID-19. The Consortium created a single point of access for COVID researchers. This article is the Consortium's story—how the Consortium was created, its founding members, what it provides, how it works, and its accomplishments. We will reflect on the lessons learned from the creation and operation of the Consortium and describe how the features of the Consortium could be sustained as a *National Strategic Computing Reserve* (NSCR) to ensure the nation is prepared for future crises.

Digital Object Identifier 10.1109/MCSE.2022.3145608
Date of current version 14 March 2022.

CREATION OF THE CONSORTIUM

As the pandemic began to significantly accelerate in the United States, on March 11 and 12, 2020, IBM and the HPC community started to explore ways to organize efforts to help in the fight against COVID-19. IBM had years of

Members and Affiliates

Academia
- Massachusetts Institute of Technology
- MGHPCC
- Rensselaer Polytechnic Institute
- University of Illinois
- University of Texas at Austin
- University of California - San Diego
- Carnegie Mellon University
- University of Pittsburgh
- Indiana University
- University of Wisconsin-Madison
- Ohio Supercomputing Center
- UK Digital Research Infrastructure
- CSCS – Swiss National Supercomputing Centre
- SNIC PDC – Swedish National Infrastructure for Computing, Center for High Performance Computing

Federal Agencies
- NASA
- National Science Foundation
  - XSEDE
  - Pittsburgh Supercomputing Center
  - Texas Advanced Computing Center (TACC)
  - San Diego Supercomputer Center (SDSC)
  - National Center for Supercomputing Applications (NCSA)
  - Indiana University Pervasive Technology Institute (IUPTI)
  - Open Science Grid (OSG)
  - National Center for Atmospheric Research (NCAR)

Industry
- IBM
- Amazon Web Services
- AMD
- D.E.Shaw Research
- Dell
- Google Cloud
- Hewlett Packard Enterprise
- Intel
- Microsoft
- NVIDIA

Affiliates
- Atrio
- Data Expedition
- Flatiron
- Fluid Numerics
- Immortal Hyperscale InterPlanetary Fabrics
- MathWorks
- Raptor Computer Systems
- SAS
- The HDF Group

Department of Energy National Laboratories
- Argonne National Laboratory
- Idaho National Laboratory
- Lawrence Berkeley National Laboratory
- Oak Ridge National Laboratory
- Lawrence Livermore National Laboratory
- Los Alamos National Laboratory
- Sandia National Laboratories

International
- Korea Institute of Science and Technology Information (KISTI)
- Ministry of Education, Culture, Sports, Science and Technology (MEXT) - Japan
  - RIKEN Center for Computational Science (R-CCS)

FIGURE 1. Consortium members and affiliates as of July 7, 2021.

experience with HPC, knew its capabilities to help solve hard problems, and had the vision of organizing the HPC community to leverage its substantial computing capabilities and resources to accelerate progress and understanding in the fight against COVID-19 by connecting COVID-19 researchers with organizations that had significant HPC resources. At this point in the pandemic, the efforts in the DOE, NSF, and other organizations within the U.S. Government, as well as around the world, were independent and *ad hoc* in nature.

It was clear very early on that a broader and more coordinated effort was needed to leverage existing efforts and relationships to create a unique HPC collaboration.

Early in the week of March 15, 2020, leadership at the DOE Labs and at key academic institutions were supportive of the vision: very quickly create a public–private consortium between government, industry, and academic leaders to aggregate compute time and resources on their supercomputers and to make them freely available to aid in the battle against the virus. On March 17, the White House OSTP began to actively support the creation of the Consortium, along with DOE and NSF leadership. The NSF recommended leveraging their Extreme Science and Engineering Discovery Environment (XSEDE) Project[1] and its XSEDE Resource Allocations System (XRAS) that handles nearly 2000 allocation requests annually[2] to serve as the access point for the proposals. Recognizing that time was critical, a team, now comprising IBM, DOE, OSTP, and NSF, had been formed with the goal of creating the Consortium in less than a week! Remarkably, the Consortium met that goal without formal legal agreements. Essentially, all potential members agreed to a simple statement of intent that they would provide their computing facilities' capabilities and expertise at no cost to COVID-19 researchers, that all parties in this effort would be participating at risk and without liability to each other, and without any intent to influence or otherwise restrict one another.

From the beginning, it was recognized that communication and expedient creation of a community around the Consortium would be key. Work began on the Consortium website[a] the following day. The Consortium Executive Committee was formed to lay the groundwork for the operations of the Consortium. By Sunday, March 22, the XSEDE Team instantiated a complete proposal submission and review process that was hosted under the XSEDE website[b] and provided direct access to the XRAS submission system, which was ready to accept proposal submissions the very next day.

Luckily, the Consortium assembled swiftly because OSTP announced that the President would introduce the concept of the Consortium at a news conference

[a]https://covid19-hpc-consortium.org
[b]https://www.xsede.org/covid19-hpc-consortium

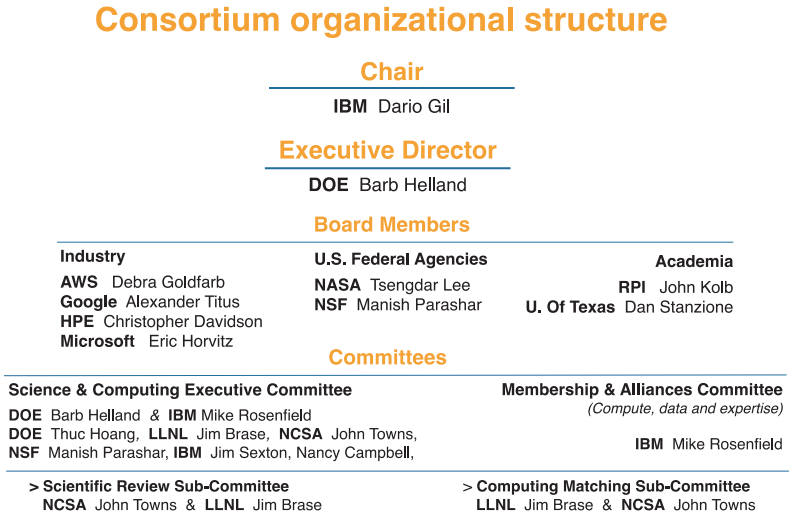


FIGURE 2. Consortium organizational structure as of July 7, 2021.

on March 22. Numerous news articles came out after the announcement that evening. The Consortium became a reality when the website[c] went live the next day, followed by additional press releases and news articles. The researchers were ready—the first proposal was submitted on March 24, and the first project was started on March 26, demonstrating our ability to connect researchers with resources in a matter of days—an exceptionally short time for such processes typically. Subsequently, 50 proposals were submitted by April 15 and 100 by May 9.

A more detailed description of the Consortium's creation can be found in the IEEE Computer Society Digital Library at https://doi.ieeecomputersociety.org/10.1109/MCSE.2022.3145608. An extended version of this article can be found on the Consortium website.[a]

CONSORTIUM MEMBERS AND CAPABILITIES

The Consortium initially provided access to over 300 petaflops of supercomputing capacity provided by the founding members: IBM; Amazon Web Services; Google Cloud; Microsoft; MIT; RPI; DOE's Argonne, Lawrence Livermore, Los Alamos, Oak Ridge, and Sandia National Laboratories; NSF and its supported advanced computing resources, advanced cyberinfrastructure, services, and expertise; and NASA.

Within several months, the Consortium grew to 43 members (see Figure 1) from the United States, and around the world (the complete list can be found at https://covid19-hpc-consortium.org/) representing access to over 600 petaflops of supercomputing systems, over 165,000 compute nodes, more than 6.8 million compute processor cores, and over 50,000 GPUs, representing access to systems worth billions of dollars. In addition, the Consortium collaborated with two other worldwide initiatives: The EU PRACE COVID-19 Initiative and a COVID-19 initiative at the National Computational Infrastructure Australia and Pawsey Supercomputing Centre.[d] The Consortium also added nine affiliates (also listed and described at websites[a,c]) who provided expertise and supporting services to enable researchers to start up quickly and run more efficiently.

GOVERNANCE AND OPERATIONS

Even though there were no formal agreements between the Consortium members, an agile governance model was developed as shown in Figure 2. An Executive Board, comprised of a subset of the founding members, oversees all aspects of the Consortium and is the final decision-making authority. Initially, the Executive Board met weekly and now meets monthly. The Board reviews progress, reviews recommendations for new members and affiliates, and provides guidance on future directions and activities of the Consortium to the Executive Committee. The Science and Computing Executive Committee, which reports to the Executive Board, (see also Figure 2) is responsible for day-to-day operations of the Consortium, overseeing the review and computer matching process, tracking project progress, maintaining/updating the website, highlighting the Consortium results (for

[c]https://covid19-hpc-consortium.org/news

[d]https://covid19-hpc-consortium.org/collaborations

example, with blogs and webinars), and determining/proposing next steps for Consortium activities.

The Scientific Review and the Computing Matching Sub-Committees play a crucial role in the success of the Consortium. The Scientific Review team—comprised of subject matter experts from members of the research community and coming from many organizations[e]—reviews proposals for merit based on the review criteria and guidance[b] provided to proposers, and recommends appropriate proposals to the Computing Matching Sub-Committee. The Computing Matching Sub-Committee team, comprised of representatives of Consortium members providing resources, matches the computer needs from recommended proposals with either the proposer's requested site or other appropriate resources. Once matched, the researcher needs to go through the standard onboarding/approval process at the host site to gain access to the system. Initially, we expected that the onboarding/approval process would be time consuming (since this was the only time where actual agreements had to be signed), but those executing the onboarding processes with the various member compute providers worked diligently to prioritize these requests, and thus, it typically takes only a day or two. As a result, once approved, projects are up and running very rapidly.

The Membership Committee reviews requests for organizations and individuals to become members or affiliates to provide additional resources to the Consortium. These requests are in turn sent to OSTP for vetting, with the Executive Committee making final recommendations to the Executive Board for approval.

PROJECT HIGHLIGHTS

The goal of the Consortium is to provide state-of-the-art HPC resources to scientists all over the world to accelerate and enable R&D that can contribute to pandemic response. Over 115 projects have been supported, covering a broad spectrum of technical areas ranging from understanding the SARS-CoV-2 virus and its human interaction to optimizing medical supply chains and resource allocations, and have been organized into a taxonomy of areas consisting of basic science, therapeutic development, and patients.

Consortium projects have produced a broad range of scientific advances. The projects have collectively produced a growing number of publications, datasets, and other products (more than 70 as of the end of calendar year 2021), including two journal covers.[f] A more detailed description of the Consortium's Project Highlights and Operational Results can be found at https://covid19-hpc-consortium/projects and https://covid19-hpc-consortium.org/blog, respectively.

While Consortium projects have contributed significantly to scientific understanding of the virus and its potential therapeutics, direct and near-term impact on the course of the pandemic has been mixed. There are cases of significant impact, but, overall, the patient-related applications that have the most direct path to near-term impact have been less successful. It may be possible to attribute this to the lower level of experience in HPC that is typical of these groups, but patient data availability and use restrictions and the lack of connection to front-line medical and response efforts are also important factors. These are issues that will need to be addressed in planning for future pandemics or other crisis response programs.

LESSONS LEARNED FROM THE COVID-19 HPC CONSORTIUM

The COVID-19 pandemic has shown that the existence of an advanced computing infrastructure is not sufficient on its own to effectively support the national and international response to a crisis. There must also be mechanisms in place to rapidly make this infrastructure broadly accessible, which includes not only the computing systems themselves, but also the human expertise, software, and relevant data to rapidly enable a comprehensive and effective response.

The following are the key lessons learned.

- The ability to leverage existing processes and tools (e.g., XSEDE) was critical and should be considered for future responses.
- Engagement with the stakeholder community is an area that should be improved based on the COVID-19 experience. For example, early collaboration with the NIH, FEMA, CDC, and medical provider community could have significantly increased impact in the patient care and epidemiology areas. Having prenegotiated agreements with these and similar stakeholders will be important going forward.
- Substantial time and effort are required to make resources and services available to researchers so that they can do their work. A standing capability to support the proposal submission and review process, as well as coordinating with service providers to provide the necessary access to resources and services, would have been helpful.
- It would have been beneficial to have had use authorizations in place for the supercomputers

[e]https://covid19-hpc-consortium.org/who-we-are
[f]http://doi.org/10.1021/acs.jcim.0c00929 and http://doi.org/10.1039/d0cp03145c

and resources provided by U.S. Government organizations.

- While the proposal review and award process ran sufficiently well, there was no integration of the resources being provided and the associated institutions into an accounting and account management system. Though XSEDE also operates such a system, there was no time to integrate the resources into that system. This would have greatly facilitated the matching and onboarding processes. It also would have provided usage data and insight into resource utilization.
- Given the absence of formal operating and partnership agreements in the Consortium and the mix of public and private computing resources, the work supported was limited to open, publishable activities. This inability to support proprietary work likely reduced the effectiveness and impact of the Consortium, particularly in support for private-sector work on therapeutics and patient care. A lightweight framework for supporting proprietary work and associated intellectual property requirements would increase the utility of responses for similar future crises.

NEXT STEP: THE NSCR

Increasingly, the nation's advanced computing infrastructure—and access to this infrastructure, along with critical scientific and technical support in times of crisis—is important to the nation's safety and security.[g,h] Computing is playing an important role in addressing the COVID-19 pandemic and has, similarly, assisted in national emergencies of the recent past, from hurricanes, earthquakes, and oil spills, to pandemics, wildfires, and even rapid turnaround modeling when space missions have been in jeopardy. To improve the effectiveness and timeliness of these responses, we should draw on the experience and the lessons learned from the Consortium in developing an organized and sustainable approach for applying the nation's computing capability to future national needs.

We agree with the rationale behind the creation of an NSCR as outlined in the recently published OSTP Blueprint to protect our national safety and security by establishing a new public–private partnership, the NSCR: a coalition of experts and resource providers (compute, software, data, and technical expertise) spanning government, academia, nonprofits/foundations, and industry supported by appropriate coordination structures and mechanisms that can be mobilized quickly and efficiently to provide critical computing capabilities and services in times of urgent needs.

Figure 3 shows a transition from a pre-COVID *ad hoc* response to crises to the Consortium and then to an NSCR.[i]

Principal Functions of the NSCR

In much the same way as the Merchant Marine[j] maintains a set of "ready reserve" resources that can be put to use in wartime, the NSCR would maintain reserve computing capabilities for urgent national needs. Like the Merchant Marine, this effort would involve building and maintaining sufficient infrastructure and human capabilities, while also ensuring that these capabilities are organized, trained, and ready in the event of activation. The principal functions of the NSCR are proposed to be as follows:

- recruit and sustain a group of advanced computing and data resource and service provider members in government, industry, and academia;
- develop relevant agreements with members, including provisions for augmented capacity or cost reimbursement for deployable resources, for the urgent deployment of computing and supporting resources and services, and for provision of incentives for nonemergency participation;
- develop a set of agreements to enable the Reserve to collaborate with domain agencies and industries in preparation for and execution of Reserve deployments;
- execute a series of preparedness exercises on some frequency basis to test and maintain the Reserve;
- establish processes and procedures for activating and operating the national computing reserve in times of crisis;
- during a crisis,
  - execute procedures to review and prioritize projects and to allocate computing resources to approved projects;
  - track project progress and disseminate products and outputs to ensure effective use and impact;
  - participate in the broader national response as an active partner.

[g]The U.S. needs a National Strategic Computing Reserve, Scientific American, June 2, 2021. [Online]. Available: https://www.scientificamerican.com/article/the-u-s-needs-a-national-strategic-computing-reserve/

[h]https://covid19-hpc-consortium.org/blog/national-strategic-computing-reserve

[i]https://covid19-hpc-consortium.org/blog/national-strategic-computing-reserve

[j]United States Merchant Marine – 46 U.S.C. §§ 861-889 Merchant Marine Act.

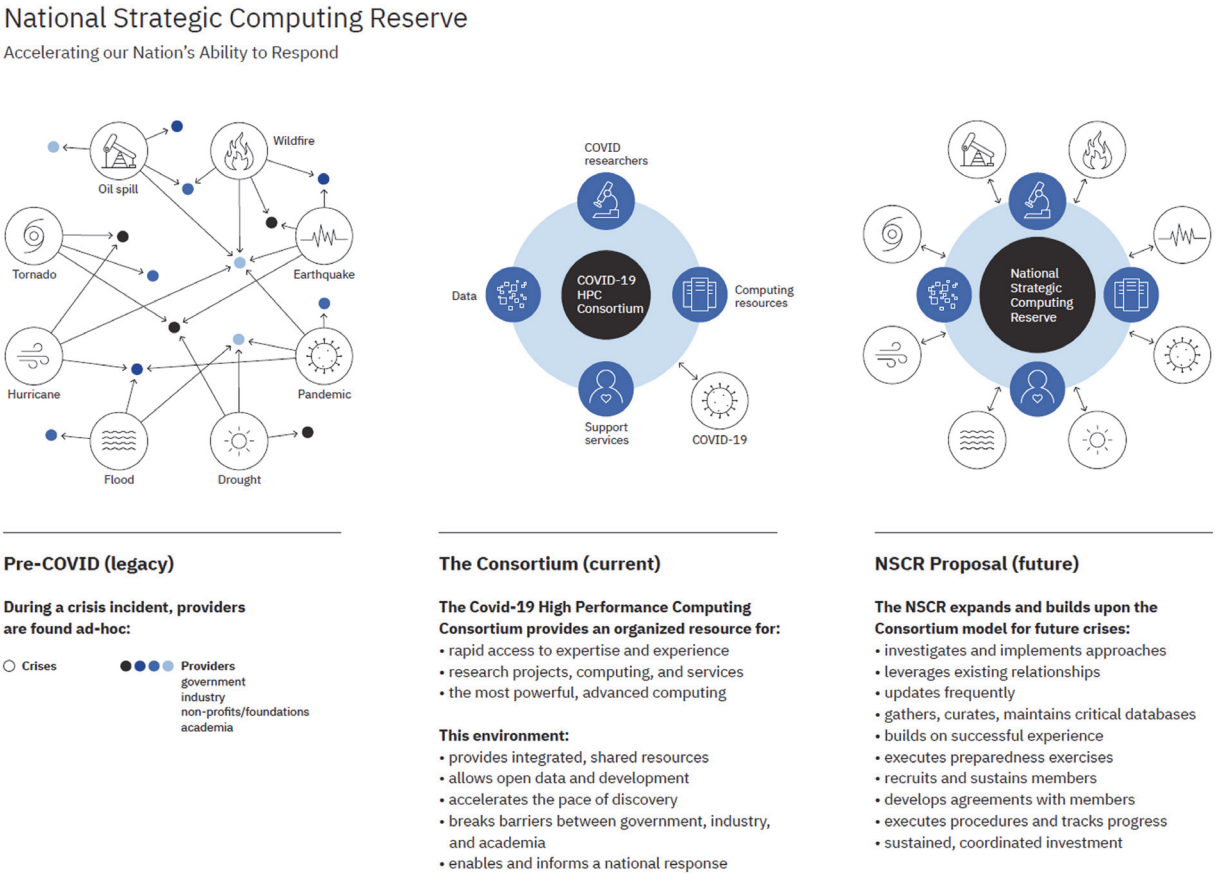


FIGURE 3. Potential path from a pre-COVID to the NSCR.

CONCLUSION

The COVID-19 HPC Consortium has been in operation for almost two years[k] and has enabled over 115 research projects investigating multiple aspects of COVID-19 and the SARS-CoV-2 coronavirus. To maximize impact going forward, the Consortium has transitioned to a focus on the following:

1) proposals in specific targeted areas;
2) gathering and socializing results from current projects;
3) driving the establishment of an NSCR.

New project focus areas target having an impact in a six-month time period and the Consortium is particularly, though not exclusively, interested in projects focused on understanding and modeling patient response to the virus using large clinical datasets; learning and validating vaccine response models from multiple clinical trials; evaluating combination therapies using repurposed molecules; mutation understanding and mitigation methods; and epidemiological models driven by large multimodal datasets.

We have drawn on our experience and lessons learned through the COVID-19 HPC Consortium, and on our observation of how the scientific community, federal agencies, and healthcare professionals came together in short order to allow computing to play an important role in addressing the COVID-19 pandemic. We have also proposed a possible path forward, the NSCR, for being better prepared to respond to future national emergencies that require urgent computing, ranging from hurricanes and earthquakes to pandemics and wildfires. Increasingly, the nation's computing infrastructure—and access to this infrastructure along with critical scientific and technical support in times of crisis—is important to the nation's safety and security, and its response to natural disasters, public health emergencies, and other crises.

[k]https://covid19-hpc-consortium.org/blog/a-year-on-hpc-consortium-national-strategic-computing-reserve

ACKNOWLEDGMENTS

The authors would like to thank the past and present members of the Consortium Executive Board for their guidance and leadership. In addition, the authors would like to thank Jake Taylor and Michael Kratsios formerly from OSTP, Dario Gil from IBM, and Paul Dabbar formerly from DOE, for their key roles in helping make the creation and operation of the Consortium possible. The authors also would like to thank Corey Stambaugh from OSTP for his leadership role on the Consortium membership committee. Furthermore, the authors would also like to thank all the members and affiliate organizations from academia, government, and industry who contributed countless hours of their time along with their compute resources. In addition, the service provided by researchers across many institutions as scientific reviewers are critical is selecting appropriate projects and their time and efforts are greatly appreciated, and, of course, they also want to thank the many researchers who did such outstanding work, leveraging the Consortium, in the fight against COVID-19.

REFERENCES

1. J. Towns *et al.*, "XSEDE: Accelerating scientific discovery," *Comput. Sci. Eng.*, vol. 16, no. 5, pp. 62–74, Sep./Oct. 2014, doi: 10.1109/MCSE.2014.80.
2. R. Light, A. Banerjee, and E. Soriano, "Managing allocations on your research resources? XRAS is here to help!," in *Proc. Pract. Experience Adv. Res. Comput.*, 2018, pp. 1–4, doi: 10.1145/3219104.3229238.

JIM BRASE is currently a Deputy Associate Director for Computing with Lawrence Livermore National Laboratory (LLNL), Livermore, CA, USA. He leads LLNL research in the application of high-performance computing, large-scale data science, and simulation to a broad range of national security and science missions. Jim is a co-lead of the ATOM Consortium for computational acceleration of drug discovery, and on the leadership team of the COVID-19 HPC Consortium. He is currently leading efforts on large-scale computing for life science, biosecurity, and nuclear security applications. In his previous position as an LLNL's deputy program director for intelligence, he led efforts in intelligence and cybersecurity R&D. His research interests focus on the intersection of machine learning, simulation, and HPC. Contact him at brase1@llnl.gov.

NANCY CAMPBELL is responsible for the coordinated execution of the IBM Research Director's government engagement agenda and resulting strategic partnerships within and across industry, academia, and government, including the COVID-19 HPC Consortium and International Science Reserve. Prior to this role, she was the program director for IBM's COVID-19 Technology Task Force, responsible for developing and delivering technology-based solutions to address the consequences of COVID-19 for IBM's employees, clients, and society-at-large. Previously, she led large multidisciplinary teams in closing IBM's two largest software divestitures for an aggregate value in excess of $2.3 billion, and numerous strategic intellectual property partnerships for an aggregate value in excess of $3 billion. Prior to joining IBM, she was a CEO for one of Selby Venture Partners portfolio companies and facilitated the successful sale of that company to its largest channel partner. She attended the University of Southern California, Los Angeles, CA, USA, and serves as an IBM's executive sponsor for the USC Master of Business for Veterans program. Contact her at nncampbe@us.ibm.com.

BARBARA HELLAND is currently an Associate Director of the Office of Science's Advanced Scientific Computing Research (ASCR) program. In addition to her associate director duties, she is leading the development of the Department's Exascale Computing Initiative to deliver a capable exascale system by 2021. She was also an executive director of the COVID-19 High-Performance Computing Consortium since its inception in March, 2020. She previously was an ASCR's facilities division director. She was also responsible for the opening ASCR's facilities to national researchers, including those in industry, through the expansion of the Department's Innovative and Novel Computational Impact on Theory and Experiment program. Prior to DOE, she developed and managed computational science educational programs at Krell Institute, Ames, IA, USA. She also spent 25 years at Ames Laboratory working closely with nuclear physicists and physical chemists to develop real-time operating systems and software tools to automate experimental data collection and analysis, and in the deployment and management of lab-wide computational resources. Helland received the B.S. degree in computer science and the M.Ed. degree in organizational learning and human resource development from Iowa State University, Ames. In recognition for her work on the Exascale Computing Initiative and with the COVID-19 HPC Consortium, she was named to the 2021 Agile 50 list of the world's 50 most influential people navigating disruption. Contact her at barbara.helland@science.doe.gov.

THUC HOANG is currently the Director of the Office of Advanced Simulation and Computing (ASC), and Institutional Research and Development Programs in the Office of Defense Programs, within the DOE National Nuclear Security Administration (NNSA), Washington, DC, USA. The ASC

program develops and deploys high-performance simulation capabilities and computational resources to support the NNSA annual stockpile assessment and certification process, and other nuclear security missions. She manages ASC's research, development, acquisition and operation of HPC systems, in addition to the NNSA Exascale Computing Initiative and future computing technology portfolio. She was on proposal review panels and advisory committees for the NSF, Department of Defense, and DOE Office of Science, as well as for some other international HPC programs. Hoang received the B.S. degree from Virginia Tech, Blacksburg, VA, USA, and the M.S. degree from Johns Hopkins University, Baltimore, MD, USA, both in electrical engineering. Contact her at thuc.hoang@nnsa.doe.gov.

MANISH PARASHAR is the Director of the Scientific Computing and Imaging (SCI) Institute, the Chair in Computational Science and Engineering, and a Professor with the School of Computing, University of Utah, Salt Lake City, UT, USA. He is currently on an IPA appointment at the National Science Foundation where he is serving as the Office Director of the NSF Office of Advanced Cyberinfrastructure. He is the Founding Chair of the IEEE Technical Consortium on High Performance Computing, the Editor-in-Chief of *IEEE Transactions on Parallel and Distributed Systems*, and serves on the editorial boards and organizing committees of several journals and international conferences and workshops. He is a Fellow of AAAS, ACM, and IEEE. For more information, please visit http://manishparashar.org. Contact him at mparasha@nsf.gov.

MICHAEL ROSENFIELD is currently a Vice President of strategic partnerships with the IBM Research Division, Yorktown Heights, NY, USA. Previously, he was a vice president of Data Centric Solutions, Indianapolis, IN, USA. His research interests include the development and operation of new collaborations, such as the COVID-19 HPC Consortium and the Hartree National Centre for Digital Innovation as well as future computing architectures and enabling accelerated discovery. Prior work in Data Centric Solutions included current and future system, and processor architecture and design including CORAL and exascale systems, system software, workflow performance analysis, the convergence of Big Data, AI, analytics, modeling, and simulation, and the use of these advanced systems to solve real-world problems as part of the collaboration with the Science and Technology Facility Council's Hartree Centre in the U.K. He has held several other executive-level positions in IBM Research including Director Smarter Energy, Director of VLSI Systems, and Director of the IBM Austin Research Lab. He started his career at IBM working on electron-beam lithography modeling and proximity correction techniques. Rosenfield received the B.S. degree in physics from the University of Vermont, Burlington, VT, USA, and the M.S. and Ph.D. degrees from the University of California, Berkeley, CA, USA. Contact him at mgrosen@us.ibm.com.

JAMES SEXTON is currently an IBM Fellow with IBM T. J. Watson Research Center, New York, NY, USA. Prior to joining IBM, he held appointments as a lecturer, and then as a professor with Trinity College Dublin, Dublin, Ireland, and as a postdoctoral fellow with IBM T. J. Watson Research Center, at the Institute for Advanced Study at Princeton and at Fermi National Accelerator Laboratory. His research interests include span high-performance computing, computational science, and applied mathematics and analytics. Sexton received the Ph.D. degree in theoretical physics from Columbia University, New York, NY, USA. Contact him at sextonjc@us.ibm.com.

JOHN TOWNS is currently an Executive Associate Director for engagement with the National Center for Supercomputing Applications, and a deputy CIO for Research IT in the Office of the CIO, Champaign, IL, USA. He is also a PI and Project Director for the NSF-funded XSEDE project (the Extreme Science and Engineering Discovery Environment. He holds two appointments with the University of Illinois at Urbana-Champaign, Champaign. He provides leadership and direction in the development, deployment, and operation of advanced computing resources and services in support of a broad range of research activities. In addition, he is the founding chair of the Steering Committee of PEARC (Practice and Experience in Advanced Research Computing, New York, NY, USA. Towns received the B.S. degree from the University of Missouri-Rolla, Rolla, MO, USA, and the M.S. degree and astronomy from the University of Illinois, Ames, IL, both in physics. Contact him at jtowns@ncsa.illinois.edu.

EDITOR: Robert Blumen, Katana Graph, robert@robertblumen.com

This article originally appeared in
IEEE Software
vol. 39, no. 1, 2022

DEPARTMENT:
SOFTWARE ENGINEERING RADIO

Paul Butcher on Fuzz Testing

Philip Winston

FROM THE EDITOR

Paul Butcher, senior software engineer, AdaCore, and lead U.K. engineer for the company's High-Integrity, Complex, Large Software and Electronic Systems initiative, discusses fuzz testing, an automated technique to find security vulnerabilities and other software flaws. Host Philip Winston speaks with Butcher about positive and negative testing, how fuzz testing fits into software development, brute force and blunt force fuzz testing, the American Fuzzy Lop fuzzer from Google, and how fuzz testing works for Ada. We provide summary excerpts below; to hear the full interview, visit http://www.se-radio.net or access our archives via RSS at http://feeds.feedburner.com/se-radio.—*Robert Blumen*

Philip Winston: What is fuzz testing?

Paul Butcher: Fuzz testing differs from standard verification testing. We typically test a software application with a test case that has inputs and a correlating expected output. We execute the test and check that the actual output matched the expected output. With fuzz testing, we're more interested in the behavior than the output. We subject the behavior of the application to inputs it may not be used to dealing with.

Is that considered negative testing?

It is related. Positive testing is more verification based. Our system must satisfy some requirements, and we build evidence that we've met those requirements. Suppose we have a secure room, and we want to test that the entrance to this room works correctly. Positive testing would be to see that all authorized personnel could get through this door. Negative testing would be to test whether anyone who isn't authorized to enter this room can get in. So, with negative testing, the number of potential test cases can be indefinitely large.

What is safety engineering and how does it relate to fuzz testing?

Safety engineering in software is producing a system that is responsible for a critical aspect of the application that could lead to loss of life if it failed. This is prominent in the aerospace industry. Safety engineering plays a key role in the flight management system, the avionics of the aircraft. Fuzz testing is more traditionally associated with security testing, but the disciplines overlap. A vulnerability within your system can be exploited and lead to a safety implication or trigger a sequence of events that could lead to a safety hazard. Fuzz testing could identify that vulnerability. If systems are security critical, they must be free of exploitable vulnerabilities.

Do you employ fuzz testing early or late in the process?

It tends to be later. Some companies have fuzz testing teams that are given a software system before deployment, and they do their security analysis then. Like all

Digital Object Identifier 10.1109/MS.2021.3118906
Date of current version: 23 December 2021

 Published by the IEEE Computer Society

testing, the earlier you can get it into the development lifecycle, the more benefit you will get.

What is High-Integrity, Complex, Large Software and Electronic Systems? Does it relate to fuzz testing, or is it bigger than that?

It's bigger. It's a U.K. government-sponsored research and development program focused on cybersecurity within civilian aerospace. It aims to bring together tier 1 aerospace manufacturers within the U.K. with tool providers like AdaCore but also universities. The research is split into work packages that create standards and guidelines for things like cybersecurity for aerospace, such as how to implement security measures for the industry and how to detect vulnerabilities. This is where things like fuzz testing come in.

How does the term *compiler hardening* relate to these efforts?

Compilers are one kind of development tool that AdaCore works in; we produce compilers for Ada, C, and C++ across multiple platforms. Compiler hardening is a mechanism for accepting that the compiled software system could be subjected to hardware attacks and things like side channel attacks. We're looking into whether we can add security measures at the compiler level to combat these types of attacks. An example is clearing the stack on a function exit and ensuring that valuable data can't be read from those memory areas.

Where is the Ada programming language still used today? Is it still under active development, or is it in a different stage of its lifecycle?

It's still in active development. I studied Ada at university. My first job was an Ada-based program working on the Eurofighter program. I worked on many different programming languages before coming back to working on Ada again. It's a fantastic language, but you tend to see it in software applications that have either a mission-, safety-, or security-critical need. It's widely used within the defense industry within Europe and the U.S. and all over the world in the automotive sector and rail industries and in the nuclear space. The reference manual is an International Organization for Standardization (ISO) standard, and there's an ISO working group that is constantly working on improvements to the language to ensure that it stays competitive with the capabilities of other existing languages.

How do fuzz testing strategies or capabilities differ between Ada and other languages?

Ada has a rich runtime testing capability that's built into the semantics of the language. If you're going to write an Ada compiler, you need runtime constraint checking for buffer overflows. It's a strongly typed language, so if you try to mix types in the assignment calls, the runtime will pick up on that. This runtime raises exceptions

SOFTWARE ENGINEERING RADIO

Visit www.se-radio.net to listen to these and other insightful hour-long podcasts.

RECENT EPISODES

- 481 — Ipek Ozkaya talks with host Jeff Doolittle about her recent book on managing technical debt.
- 480 — Venky Naganathan discusses enterprise chatbots with host Kanchan Chringi.
- 479 — Luis Ceze joins host Akshay Manchale for a conversation about the Apache TVM machine learning compiler.

UPCOMING EPISODES

- Luke Hoban discusses infrastructure as code with host Jeff Doolittle.
- Host Gavin Henry talks with Howard Chu on the B+ tree data structure.
- Audrey Lawrence speaks about time series databases with host Philip Winston.

I WORKED ON MANY DIFFERENT PROGRAMMING LANGUAGES BEFORE COMING BACK TO WORKING ON ADA AGAIN.

IN ADA, YOU CAN WRITE PRE- AND POSTCONDITION CONTRACTS ON YOUR SUBPROGRAMS WITHIN YOUR APPLICATION.

if any of the runtime checks fail. We can capture these within the test harness, and we can indicate to the fuzzing application that a runtime check has failed. So, you tend to switch on as many of the runtime checks as you can before executing the fuzz test. This is where we can not only just check for crashes but also check for the system having entered an unknown operating state.

Programming languages that support design by contracts are interesting for fuzz testing. If you're fuzz testing a C application, you can put assertions in there to say, "If this happens, raise this assertion." The fuzz test picks that up as an anomaly and gives you the test case to reproduce it. In Ada, you can write pre- and postcondition contracts on your subprograms within your application. So, a typical example is, if you have a subprogram that takes two parameters and returns that summation, your post condition would say the output has to be equal to the two inputs when they're added together. If that contract fails, the runtime check will detect that, and the fuzz test will pick up on it. This is where we can start to move toward fuzz testing for functional correctness, as well. Some aspects of Ada are there to ensure that the developer writes code in a structured way. The semantics of the language stop you from doing things that you may not have known could be nonsecure or unsafe.

PHILIP WINSTON is a software engineering consultant and contractor through his company, Tobeva Software, Winchester, Virginia, 22602, USA. Contact him at https://tobeva.com or philip@tobeva.com.

IEEE Computer Society Has You Covered!

WORLD-CLASS CONFERENCES — Stay ahead of the curve by attending one of our 189+ globally recognized conferences.

DIGITAL LIBRARY — Easily access over 893k articles covering world-class peer-reviewed content in the IEEE Computer Society Digital Library.

CALLS FOR PAPERS — Discover opportunities to write and present your ground-breaking accomplishments.

EDUCATION — Strengthen your resume with the IEEE Computer Society Course Catalog and its range of offerings.

ADVANCE YOUR CAREER — Search the new positions posted in the IEEE Computer Society Career Center.

NETWORK — Make connections that count by participating in local Region, Section, and Chapter activities.

Explore all of the member benefits at www.computer.org today!

IEEE COMPUTER SOCIETY

IEEE

This article originally appeared in
IEEE SECURITY & PRIVACY
vol. 20, no. 1, 2022

Unsafe at Any Clock Speed: The Insecurity of Computer System Design, Implementation, and Operation

Sean Peisert, *Editor in Chief*

It appears that there are enormous differences of opinion as to the probability of a [system failure]. The higher figures come from the working engineers, and the very low figures from management. When playing Russian roulette the fact that the first shot got off safely is little comfort for the next. [T]here have been recent suggestions by management to curtail elaborate and expensive tests as being unnecessary. This must be resisted. The proper way to save money is to curtail the number of requested changes, not the quality of testing for each.

Let us make recommendations to ensure that [management deals] in a world of reality in understanding technological weaknesses and imperfections well enough to be actively trying to eliminate them. They must live in reality in comparing the costs and utility. Only realistic schedules should be proposed, schedules that have a reasonable chance of being met. If in this way support [would not exist], then so be it. For a successful technology, reality must take precedence over public relations, for nature cannot be fooled. (Author's Note: ellipses omitted for readability.)

One could be forgiven for thinking that this text came from a critique of SolarWinds Orion, Adobe Flash, or Microsoft Office or Internet Explorer, or from a recent report that led to a set of strong recommendations contained in a recent White House Executive Order on Cybersecurity. As most readers of this magazine likely recognize, that would be wrong, as this is of course excerpted and edited text written by Richard Feynman, in his appendix to the 1986 Rogers Commission report studying the *Challenger* disaster, "Personal Observations on Reliability of Shuttle."[1]

I quoted portions of Feynman's report here because I believe that we have a similar problem in computer software for similar reasons: companies developing software prioritize maximum shareholder profit and productivity over software safety, robustness, and security. It is not unreasonable or unexpected that companies prioritize profit. At the same time, many companies have embraced "corporate social responsibility," having recognized that supporting employees, customers, and the broader world can positively impact both reputation and profit. As just one example, we see health care organizations balance profit with patient safety, because not doing so would lead to public outrage, which in turn would impact profits. However, with only rare exceptions do we see a similar effort to balance shareholder primacy with software security. The consequences of this lack of balance range from events like the major breaches, ransomware, and attacks against critical systems like hospitals and utilitiesthe NotPetya attacks affecting Maersk's shipping and port operations worldwide, the WannaCry ransomware attacks against U.K. National Health Service hospitals, and the Colonial Pipeline attack in the United States.

So where is the public outrage? And how did we get to this state, and why it is acceptable to so many organizations to live with this level of vulnerability and compromise? These incidents are not mere annoyances. Real people are affected in real ways. Given this, how is it possible that this is not a virtually identical moment to automobile safety before Ralph Nader's *Unsafe at Any Speed*[2] demonstrated the need

Digital Object Identifier 10.1109/MSEC.2021.3127086
Date of current version: 25 January 2022

 Published by the IEEE Computer Society

for and barriers to mandating safety improvements in cars, and led directly to seat belts and other safety advances? Or public and agricultural safety before Rachel Carson's *Silent Spring*[3] exposed the toxicity of the chemical DDT and led directly to its ban? Or medical safety before John Snow's *On the Mode of Communication of Cholera*[4] exposing that germs, not "miasma," cause disease, which led directly to water safety and sewage improvements in London and beyond? Or the Flexner Report's[5] impact on bringing mainstream scientific protocols to medical education? Or the Institute of Medicine's *To Err is Human*[6] exposing that the same number of daily deaths from medical errors in the United States is equivalent to the number of deaths from a jumbo jet crashing each day, leading directly to a fundamental change in the approach to quality of care, and the renaming of an agency to the "U.S. Agency for Healthcare Research and Quality"? It is an inconvenient truth that software and hardware engineers make mistakes, those mistakes can become "bugs," some of those bugs represent vulnerabilities that can be attacked, and that, at times that are unpredictable, some of those vulnerabilities are attacked. So where is the equivalent response for software quality?

In fact, the reason for this situation is essentially identical as what Feynman indicated more than three decades ago: profit, expediency, and succumbing to the requests for "changes" (usually "features"). The answer as to why there isn't public outrage surely cannot be because we accept that shareholder profit should be prioritized over software quality. In fact, I would argue that it even does a disservice in the long term to shareholder value to prioritize short-term profit over software quality. At some point, companies that allow enough vulnerability will see the impact in their profits. At the same time, it isn't like we haven't advocated substantially more secure systems, and even "clean slate" solutions before—certainly, with Multics[7] and the aspirations for Orange Book A1-certified computer systems,[8] there were goals to meet provably secure operational requirements. Indeed, 46 years ago, in 1974, Karger and Schell pointed out "Multics is not Now Secure" but went on to suggest essentially that it could be made secure if we worked just a little bit harder.[9] However, writing in 2002 on their observations in the 28 years since the original paper, they note:[10]

> *In the nearly thirty years since the report, it has been demonstrated that the technology direction that was speculative at the time can actually be implemented and provides an effective solution to the problem of malicious software employed by well-motivated professionals. Unfortunately, the mainstream products of major vendors largely ignore these demonstrated technologies.*

A decade later, beginning in 2012, two DARPA programs run by Howie Shrobe, "Clean-Slate Design of Resilient, Adaptive, Secure Hosts" (CRASH) and "Mission-Oriented Resilient Clouds" (MRC)[11] sought to draw inspiration from "visionary ideas of the past" to develop and demonstrate secure and resilient systems. The "Turtles All the Way Down" piece that my colleagues Matt Bishop, Ed Talbot, and I wrote in 2012, advocated building and rebuilding systems with pervasive use of formal methods, diversity, and Byzantine fault tolerance[12] "from atoms to eyeballs" in a 13-level stack.

Fast forward to this past year when Paul van Oorschot noted in this magazine that the C language lacks type and memory safety, "... having learned our lesson from 45 years of use, surely we do not still use C in new projects and in building brand new systems, do we? As it turns out, the evidence suggests we do."[13] Van Oorschot continued, noting that in the past, even though type-safe languages are available for use, such as Java, Go, and Apple's Swift, the fact that those languages have not been appropriate for systems development may have prolonged the use of C and C++. As van Oorschot writes, performance languages appropriate for systems work now exist, but perhaps it will take something like requirements for government procurement to see languages like Rust adopted at scale. (As a side note, it is insufficient to leverage type-safe languages if the runtimes for those languages are also written in C/C++, as the runtimes for Java and Ruby are, for example.) The wonderful "Cyber Moonshot" piece in the very next issue of *IEEE Security & Privacy* by Hamed Okhravi, also advocates the use of semantically rich processors, type and memory-safe systems languages, and fine-grained operating system compartmentalization.[14]

It is probably unreasonable to expect that these examples that I have given of attempts to radically

improve computer security would have the effects of the clarion calls in *Silent Spring* or *Unsafe at Any Speed*—both books specifically aimed at the general public. However, scholarly writings in the medical domain, including *On the Mode of Communication*, *To Err is Human*, and the Flexner Report, have been transformative, whereas despite 46 years of efforts, from Karger and Schell to the present day, I don't believe that we've seen similar effects in transforming computer security.

What I believe has changed since Karger and Schell, and perhaps even since the DARPA CRASH and MRC programs is that technology and techniques have improved to the point that we are now finally at a place where we can actually, practically do something about this situation. In fact, in the same *Challenger* report, Feynman again even gave us a portion of the solutions—bottom-up engineering:

> *The software is checked very carefully in a bottom-up fashion. First, each new line of code is checked, then sections of code or modules with special functions are verified. The scope is increased step by step until the new changes are incorporated into a complete system and checked. But completely independently there is an independent verification group, that takes an adversary attitude to the software development group, and tests and verifies the software as if it were a customer of the delivered product. A discovery of an error during verification testing is considered very serious, and its origin studied very carefully to avoid such mistakes in the future. The principle that is followed is that all the verification is a test of that safety, in a noncatastrophic verification. A failure here generates considerable concern. (Author's Note: ellipses omitted for readability.)*

Yet, regardless of the actual approach—top-down, bottom-up, or some combination of the two—in the past, we have found Feynman's prescription regarding the degree of assurance required utterly untenable for all but the most critical systems. Times have changed in at least two ways: one is that we have gone from a world in which computer-controlled systems were mostly only running commercial and military aircraft and NASA's space shuttles to a world in which dozens or hundreds of processors exist in the modern automobile, building "control systems," and numerous other domains in life in which humans are dependent. A second and vital change is that technology useful for safety and security has advanced profoundly in the past 25 years since the Rogers Commission report was released. Let's take a look at some of those advances:

Consider type-safe languages: buffer overruns have been the "most dangerous" software weakness for years. Why should the public put up with something that is exposed as public enemy number one year after year with little progress? In contrast, Rust has emerged as a type and memory-safe language suitable for systems programming. Mozilla's Servo browser engine is being written in the Rust, and numerous Linux libraries and utilities are being rewritten in Rust. Rewriting old code in Rust can be a tough sell although Google's recent effort to implement site isolation in Chrome, and Mozilla's development and application of RLBox to Firefox—both significant manual efforts—show progress can be made when the needed resources are devoted. This will also become easier as more third-party libraries are developed for Rust and more new developers learn Rust in computer science courses.

Consider formal methods today: there exist many software elements that underlie the modern Internet and its usage that have been revealed as substantially lacking in security rigor for years, such as the vulnerabilities that plagued OpenSSL until organizations like Google, Microsoft, and OpenBSD stepped in. Why is it that the public is so forgiving of the reliance on such blatantly problematic software by major companies? And many other examples of such software certainly still remain. In contrast, seL4 is a formally verified microkernel, CertiKOS is a formally verified kernel, the Linux KVM hypervisor has been formally verified, and DARPA's "Little Bird" is a formally verified autonomous helicopter, having survived hacking contests as part of the DARPA HACMS program,[15] run by Kathleen Fisher, John Launchbury, and Raymond Richards, in 2017, and again at DEFCON this past year. this past year. In addition, numerous key elements of Amazon Web Services have been formally verified, Facebook leverages the Infer system to continuously verify code, and Microsofts Project Everest is developing a formally verified stack to improve secure web communications. Not every formal verification is as useful

as another and it may never be tenable to formally verify all code, but the DARPA exercises alone seem to have demonstrated considerable value. At the very least, there is strong evidence that building systems on top of formally verified elements that are now available and usable could substantially ameliorate a large swath of security problems. Having even more verified systems that provide support for additional functionality would help encourage broader adoption of assured systems.

Consider security-enhanced hardware today: as discussed earlier, security weaknesses often result from the use of "unsafe" languages and shared infrastructure. In contrast, the University of Cambridge and SRI's Capability Hardware Enhanced RISC Instructions (CHERI)[16] provides a capability-based system that provides fine-grained memory protection and software compartmentalization, thereby protecting against a host of weaknesses exposed by the use of unsafe languages, code injection attacks, and more. This is particularly valuable protection when existing software cannot easily be rewritten in type and memory-safe languages, for example, due to the vast amount of existing libraries written in C/C++. CHERI also now has numerous formally verified elements. In addition, Arm's forthcoming CHERI-extended Morello prototype CPU, system-on-a-chip, and board will ship early next year, and will consist of a full industrial quality and high-performance adaptation of Arm's Neoverse N1 CPU design. This prototype is in fact the culmination of a kind of "moonshot" that has been developed over 10 years and with US$250 million of DARPA, United Kingdom government, and in-kind industry funding, and seems like it could serve as a model for advancing other security techniques and technologies.

In addition to capability-enhanced hardware, consider hardware trusted execution environments (TEEs). Running on traditional servers, including those in the cloud, requires complete trust of the system administrator as well as the numerous levels of the stack that seek to mitigate attempts by one user to attack another. Who is really happy about putting complete and unquestioning trust regarding data and computation in giant corporations? In contrast, TEEs provide strong isolation properties, sometimes even from system administrators and physical attacks. they are available or announced from every major CPU platform, including AMD's SEV; ARM's v9's Confidential Compute Architecture, and Intel's SGX, alongside open source TEEs, such as the RISC-V-based Keystone. Further, some form of TEE-like "confidential computing" service is available from the three major commercial cloud providers: AWS Nitro Enclaves, GCP Confidential Computing, and Azure Confidential Computing. The Linux Foundation also hosts the Confidential Computing Consortium. The cloud and community efforts provide the software model and cryptographic infrastructure to make the use of confidential computing more straightforward. For usability and performance reasons, not all of these architectures are useful for general-purpose computing. However, for certain workloads running on single nodes, SEV can carry little overhead beyond that of virtualization itself and is readily available in cloud environments.

The reluctance of organizations to adopt some of these techniques and technologies has echoes of the White Queen informing Alice about the (lack of) availability of jam today.[17] However, despite past failures to make significant progress toward securing systems via Multics and the Orange Book, all of this recent progress has shown what is possible with the tools that we have today. Organizations can use type and memory safety (Rust), formally verified components (seL4, CertiKOS, Linux KVM), and obtain strong hardware isolation (AMD's SEV) today. At least in the case of the Rust language as well as cloud environments that have broad frameworks supporting confidential computing on top of AMD's SEV and related technologies, this can entail little extra effort. Organizations can use Facebook's Infer automated reasoning system for static analysis today. Prototype hardware supporting CHERI will be available roughly at the time this piece goes to press and may well be in broader production in not too many more years.

I've enumerated a nonexhaustive list of numerous techniques and technologies here that could all represent elements of this improvement I speak of. Not all will be part of the final solutions, and undoubtedly there are others that I haven't covered, such as the automated verification tools available that showed such success during DARPA's Cyber Grand Challenge[18] and advances in the "grand challenges" of user-centered security that make it harder for users to make decisions in a way that will lead to security

failures.[19] Furthermore, the solutions that I have discussed will also not solve all problems, and will not be adopted everywhere—for example, consider all of the software written by "citizen developers" or are outsourced to the lowest bidder. But if enough of the important software leverages these solutions, it would seem that doing so could solve a substantial number of problems, thereby enabling security researchers and engineers to focus on the problems for which we do not yet have solutions.

Portions of this cybersecurity vision likely are a "moonshot." But I think it would be a misrepresentation to characterize the entire endeavor as such. So, what's in the way? We've already pointed to cost, so how do we lower that cost or overcome that barrier? Existing public sentiment about every new security breach that takes place clearly hasn't been enough. Perhaps the public has just been convinced it has no choice but to accept the status quo. In contrast, I think the public has a right to be outraged about computer security. Further, even though the government itself suffers computer security failures large and small on an ongoing basis, the appetite for significant regulation (e.g., liability for insecure software, substantially increasing requirements for software and hardware security in government procurement) in this space seems not to exist. Thus, in the face of evidence that there are in many cases relatively low bars to much safer systems, the reason for the continued prevalence of low adoption of the components that would could systems much safer remains something of a mystery, given that numerous key components are here now, and the rest may well not be that far in the future from being deployed.

The barriers to large-scale adoption of emerging security techniques and technologies urgently need to be investigated. This investigation should include a focus on technical issues, but should also include experts who can illuminate usability, education, economic, policy, and social issues, and other systematic barriers to technology transition for innovation. At least on a technical level, there are few excuses not to be embracing many of the approaches that I've illustrated here. There are few excuses for not writing most or all new systems code in Rust; for systems, where appropriate, to be built using verified components and/or on top of security-enhanced hardware, and for applications to be run on those systems wherever possible; and for the most important source code to leverage modern, automated program verification tools and possibly formal methods.

In another passage from their 2002 piece, Karger and Schell[10] write:

> *In our opinion this is an unstable state of affairs. It is unthinkable that another thirty years will go by without one of two occurrences: either there will be horrific cyber disasters that will deprive society of much of the value computers can provide, or the available technology will be delivered, and hopefully enhanced, in products that provide effective security. We hope it will be the latter.*

Computer systems and networks have become "unsafe at any speed." The time to change that is now. The future is here. There is no further room for excuse, ignorance of reality, or fooling of nature.

REFERENCES

1. R. P. Feynman, "The presidential commission on the space shuttle challenger accident report," *Appendix F, Personal Observations Rel. Shuttle*, vol. 2, Jun. 6, 1986.
2. R. Nader, *Unsafe at Any Speed: The Designed-In Dangers of the American Automobile*. New York, NY, USA: Grossman Publishers, 1965.
3. R. Carson, *Silent Spring*. Boston, MA, USA: Houghton Mifflin, 1962.
4. J. Snow, *On the Mode of Communication of Cholera*. London, U.K.: John Churchill, 1855.
5. A. Flexner, "Medical education in the United States and Canada: A report to the Carnegie foundation for the advancement of teaching," *Bull. Carnegie Found. Adv. Teach.*, vol. 4, 1910.
6. Institute of Medicine, *To Err Is Human: Building a Safer Health System*. Washington, DC, USA: National Academies Press, 2000.
7. E. Organick, *The Multics System: An Examination of Its Structure*. Boston, MA, USA: MIT Press, 1972.
8. "Trusted computer system evaluation criteria ['Orange Book']," United States Department of Defense, Arlington, VA, USA, Tech. Rep. DoD 5200.28-STD, Dec. 26, 1985.
9. P. A. Karger and R. R. Schell, "Multics security evaluation: Vulnerability analysis," Electronic Systems Division, Hanscom, MA, USA, 1974. [Online]. Available:

http://csrc.nist.gov/publications/history/karg74.pdf
10. P. A. Karger and R. R. Schell, "Thirty years later: Lessons from the Multics security evaluation" in *Proc. 18th Annu. Comput. Security Appl. Conf. (ACSAC)*, 2002, pp. 119–126, doi: 10.1109/CSAC.2002.1176285.
11. H. Shrobe and D. Adams, "Suppose we got a do-over: A revolution for secure computing," *IEEE Security & Privacy*, vol. 10, no. 6, pp. 36–39, Nov./Dec. 2012, doi: 10.1109/MSP.2012.84.
12. S. Peisert, E. Talbot, and M. Bishop, "Turtles all the way down: A clean-slate, ground-up, first-principles approach to secure systems," in *Proc. New Security Paradigms Workshop (NSPW)*, Bertinoro, Italy, Sep. 19–21, 2012, pp. 15–26, doi: 10.1145/2413296.2413299.
13. P. C. van Oorschot, "Toward unseating the unsafe C programming language," *IEEE Security Privacy*, vol. 19, no. 2, pp. 4–6, Mar./Apr. 2021, doi: 10.1109/MSEC.2020.3048766.
14. H. Okhravi, "A cybersecurity moonshot", *IEEE Security Privacy*, vol. 19, no. 3, pp. 8–16, May/Jun. 2021, doi: 10.1109/MSEC.2021.3059438.
15. K. Fisher, J. Launchbury, and R. Richards. The HACMS program: Using formal methods to eliminate exploitable bugs, *Philos. Trans. A Math. Phys. Eng. Sci.* vol. 375, no. 2014, Art no. 20150401. doi: 10.1098/rsta.2015.0401.
16. R. N. M. Watson et al., "CHERI: A hybrid capability-system architecture for scalable software compartmentalization," in *Proc. 36th IEEE Symp. Security Privacy*, 2015, pp. 20–37, doi: 10.1109/SP.2015.9.
17. L. Carroll, "The rule is, jam to-morrow and jam yesterday—But never jam to-day... It's jam every other day: To-day isn't any other day, you know," in *Through the Looking-Glass, and What Alice Found There*. London, U.K.: Macmillan, 1872, p. 94.
18. M. Walker, Machine vs. machine: Lessons from the first year of cyber grand challenge, in *Proc. 24th USENIX Security Symp.*, Aug. 12, 2015.
19. M. E. Zurko, "User-centered security: Stepping up to the grand challenge," in *Proc. 21st Annu. Comput. Secur. Appl. Conf. (ACSAC)*, 2005, pp. 14–202, doi: 10.1109/CSAC.2005.60.

SEAN PEISERT, Editor in Chief

EDITOR: Nir Kshetri, University of North Carolina at Greensboro, nbkshetr@uncg.edu

This article originally appeared in
Computer
vol. 55, no. 2, 2022

DEPARTMENT: COMPUTING'S ECONOMICS

Gender Asymmetry in Cybersecurity: Socioeconomic Causes and Consequences

Nir Kshetri and Maya Chhetri, *University of North Carolina at Greensboro*

This article reviews the causes of gender asymmetry in cybersecurity and argues that women's increased participation can strengthen the industry and improve business outcomes. It also discusses ways to attract and retain women in the field.

Women are highly underrepresented in the field of cybersecurity. In 2019, their share of the worldwide cybersecurity workforce was 20%, compared to 38.9% in the general workforce (Figure 1). In all the economies presented in Figure 1, there are significantly lower proportions of females in the cybersecurity workforce than in the total labor force. Women have even less representation in cybersecurity leadership roles at larger U.S. corporations such as Fortune 500 companies. For instance, according to Cybersecurity Ventures, only 70 of the Fortune 500 companies, or 14%, had female chief information security officers in 2020,[1] which was lower than the proportion of females in the cybersecurity workforce (Figure 1). Likewise, while 27% of the programmers in the Israeli army are women, the proportion is 12% in cyberunits and only about 3% in the top cyberunits.[2]

Cybersecurity requires strategies beyond technical solutions. Women's representation is important because they tend to offer viewpoints and perspectives that are different from men's, and these underrepresented perspectives are critical in addressing cyber-risks. This article highlights the causes of gender asymmetry in cybersecurity and discusses how women's increased participation can strengthen the field and improve business outcomes. It also looks at some possible ways to attract and retain women in cybersecurity.

Digital Object Identifier 10.1109/MC.2021.3127992
Date of current version: 14 February 2022

CAUSES OF GENDER ASYMMETRY IN CYBERSECURITY

Table 1 lists the major sources of gender disparity in cybersecurity. Some are more general barriers such as those related to political, legal, and cultural factors that are encountered in all types of jobs, while others are specific to technology-related careers. First, in some countries, women's participation in economic activities is hindered by defective legal and regulatory systems. In a related statistic, women worldwide have only three-quarters of the legal rights that men have.[12] For instance, 18 countries are reported to require women to have their husband's permission to work outside the home.[13] Yemen is one such country, where females account for only 7.9% of the labor force.[11]. Likewise, 17 countries restrict women from traveling without permission from a guardian.[14]

Second, in some societies, cultural barriers prevent women from participating in formal labor markets. In Israel, among women with high test scores on psychometric exams, which are standardized tests used in admission to institutions of higher education, a greater proportion of Jewish women than Arab–Israeli women were reported to pursue technology-related careers. According to the chief economist of the country's Finance Ministry, in 2017, only 10% of Arab–Israeli women in the high-test-score category worked in technology, compared to 30% of Jewish women.[2] It is argued that cultural expectations and practices work against Arab women's involvement in the workforce. The culture encourages women to stay home to care for their children.[15]

 Published by the IEEE Computer Society

Third, the societal view is that cybersecurity is a job that men do,[16] though there is nothing inherent to gender that predisposes men to be more interested in or more adept at the work. The low number of women in Internet security is linked to the broader problem of their poor representation in the science, technology, engineering, and mathematics (STEM) fields. While women make up half of the U.S. college-educated workforce, they account for only 30% of the science and engineering workforce[17] and 26% of the professionals in the computer and mathematical sciences.[18]

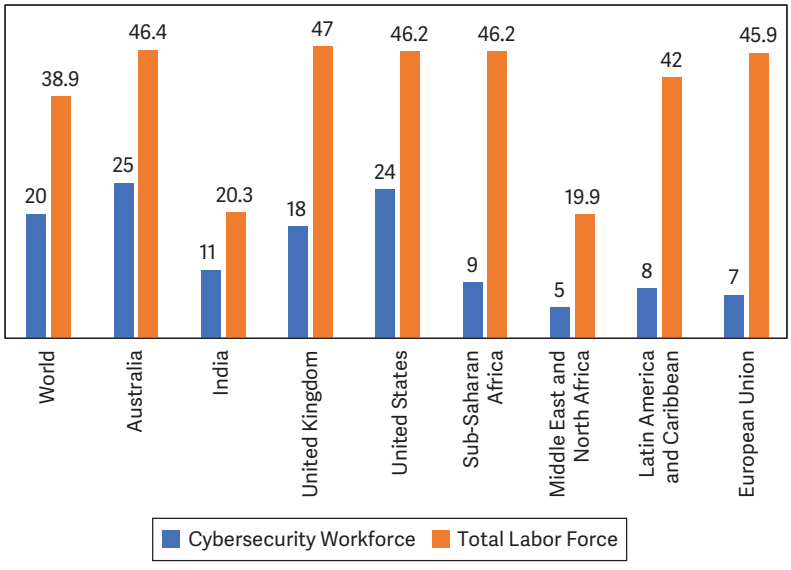


FIGURE 1. The percentage of females in the cybersecurity and total labor forces. (Sources for the cybersecurity labor force: world;[3] Australia;[4] India (forecast for 2025);[5] United Kingdom;[6] United States;[7] Sub-Saharan Africa;[8] Latin America, Caribbean, Middle East, and North Africa;[9] and European Union.[10] Source for the total labor force: World Bank.[11])

A Kaspersky Lab survey of women younger than 16 in Europe, Israel, and the United States found that 78% of the respondents had never considered a career in cybersecurity. In addition, 42% considered it important to have a gender role model in their career, and about half preferred to work in an environment that had an equal male–female balance. Cybersecurity professionals also have an image problem. For instance, one-third of the respondents thought cybersecurity professionals were "geeks," and one-quarter viewed them as "nerds." A related challenge concerns negative connotations

TABLE 1. The major causes of gender asymmetry in cybersecurity.

Cause	Explanation	Example
Regulatory and legal	Defective legal and regulatory systems	Women lack legal rights in some countries, for example, facing requirements to have their husband's or a guardian's permission to work outside the home and travel.
Cultural differences in gender roles	Hindered participation by women in formal labor markets in some cultures	In Israel, Arab–Israeli women are more likely than Jewish women to stay home to care for their children.
Societal view of gender	Mistaken societal belief that cybersecurity and technology jobs are only for men	There is low representation of women in science, technology, engineering, and mathematics fields worldwide.
Stereotypes and bias in organizational decision making and practices	Mistaken impressions among potential employees that men are more appropriate for the jobs	There is an undue emphasis on technical skills and a lack of gender-neutral language in job ads.

of terms such as *hacker* that are often associated with cybersecurity roles. Due to that, two-thirds of the respondents reported that cybersecurity jobs did not appeal to them. The respondents expressed a desire to pursue careers they were more passionate about.[18] A related outcome of this bias is that women are generally not presented with opportunities in IT fields. In a survey of women pursuing careers outside IT, 69% indicated that the main reason they did not pursue jobs in the field was because they were unaware of them.[30]

A KASPERSKY LAB SURVEY OF WOMEN YOUNGER THAN 16 IN EUROPE, ISRAEL, AND THE UNITED STATES FOUND THAT 78% OF THE RESPONDENTS HAD NEVER CONSIDERED A CAREER IN CYBERSECURITY.

Finally, stereotypes and bias in organizational decision making and practices hinder women's entry into technology jobs in general and cybersecurity-related roles in particular. For instance, the industry mistakenly gives potential employees the impression that only technical skills matter in cybersecurity,[20] which can give women the impression that the field is overly specialized and even boring. Organizations often fail to try to recruit women to work in cybersecurity. According to a survey conducted by IT security company Tessian, only about half of the respondents said their organizations were doing enough to recruit women for cybersecurity roles.[21] Gender bias in job ads further discourages women from applying. Online cybersecurity job postings often lack gender-neutral language.[22]

WOMEN'S INCREASED PARTICIPATION: STRONG SECURITY AND GOOD BUSINESS

Boosting women's involvement in cybersecurity makes both security and business sense. The cybersecurity field is facing a huge skills shortage. The gap between demand and supply in this field is predicted to reach 1.8 million workers worldwide in 2022.[10] Boosting women's participation is one way to close this gap. More importantly, women cybersecurity professionals bring important benefits that translate into strong cybersecurity. For instance, female leaders in this area tend to prioritize some key areas that males often overlook. This is partly due to their backgrounds. About 44% of women in cybersecurity have degrees in business and social sciences, compared to 30% of men.[9]

Female cybersecurity professionals put a higher priority on internal training and education in security and risk management. Women are also stronger advocates for online training, which is a flexible, low-cost way of increasing employees' awareness of security issues. Females are also adept at selecting partner organizations to develop secure software.[23] They tend to pay more attention to partner organizations' qualifications and personnel, and they assess partners' ability to meet contractual obligations. They also prefer partners that are willing to perform independent security tests.

Increasing women's participation in cybersecurity is a business issue as well as a gender concern. According to Boston Consulting Group, by 2028, women will control 75% of discretionary consumer spending worldwide.[24] Security considerations such as encryption, fraud detection, and biometrics are becoming important in consumers' buying decisions.[25] Product designs require a tradeoff between cybersecurity and usability. Women cybersecurity professionals can make better-informed decisions about such compromises for products that are targeted at female customers.

Security issues associated with major technologies and platforms, such as the Internet of Things (IoT) and social media, disproportionately affect women. For instance, smart home technologies have been highly ineffective at preventing domestic abusers from harassing and harming their former partners and child predators from gaining access to children.[26] It is reported that major U.S. technology companies—Amazon, Facebook, Apple, and Google—fill less than one-third of their leadership roles with women, and the proportion is as low as 19% at Microsoft.[27]

Apps have been widely available in the App Store (iOS), Play Store (Android), and other repositories that pose a risk to a women's safety. Some have made use of location data for stalking people in real time. For instance, the iOS app Girls Around Me, which was developed by the Russian company I-Free, leveraged data from Foursquare to scan for and detect women

checking into a user's neighborhood. The user could identify a woman he would like to talk to, connect with her through Facebook, see her full name and profile photos, and send her a message. The woman would have no idea that someone was "snooping" on her. As of March 2012, the app had been downloaded more than 70,000 times.[28] It is argued that by increasing women's involvement in decision making regarding privacy and security issues, it is possible to make IoT devices more secure and reduce predators' ability to target children and share abusive images on social media.[27]

ATTRACTING WOMEN TO CYBERSECURITY

Attracting more women to cybersecurity requires governments, nonprofit organizations, professional and trade associations, and the private sector to work together. Public–private partnership projects could help solve the problem in the long run. Parents and primary school teachers are among the most important people that can play a role in creating young girls' interest in cybersecurity and technology in general. Surveys have found that girls' interest gradually fades as they get older. For instance, a study conducted by the nonprofit trade association that issues professional certifications for the IT industry, the Computing Technology Industry Association, found that 27% of middle-school girls consider a career in technology, but the proportion reduces to 18% by the time they reach high school.[29]

This does not mean that high school is too late to develop girls' interest and engagement in cybersecurity careers. Indeed, some notable cybersecurity initiatives have targeted high-school girls. One example is Israel's Shift community, previously known as the CyberGirlz program (https://rashi.org.il/en/programs/shift-community/), which is jointly financed by the country's Defense Ministry, the Rashi Foundation, and Start-Up Nation Central. It identifies high-school girls with aptitude, desire, and natural curiosity to learn IT and helps them develop those skills. The girls participate in hackathons and training programs and get advice, guidance, and support from female mentors. Some of the mentors are from elite technology units of the country's military. The participants learn hacking skills, network analysis, and the Python programming language. They also practice simulating cyberattacks to find potential vulnerabilities. By 2018, about 2,000 girls had participated in the CyberGirlz Club and the CyberGirlz Community.

In 2017, cybersecurity firm Palo Alto Networks teamed up with the Girl Scouts of the United States of America to develop cybersecurity badges.[30] The goal is to foster cybersecurity knowledge and develop interest in the profession. The curriculum includes the basics of computer networks, cyberattacks, and online safety.[30] Professional associations can also foster interest in cybersecurity and help women develop relevant knowledge. For example, the nonprofit European private foundation Women4Cyber (https://women4cyber.eu/) was established, in 2019, to "promote, encourage, and support" women's participation in cybersecurity. By July 2021, Women4Cyber had approved national chapters in 10 European countries, and seven of the groups were fully operational.[31] Likewise, Women in Cybersecurity of Spain has started a mentoring program that supports female cybersecurity professionals early in their careers.[21]

ATTRACTING MORE WOMEN TO CYBERSECURITY REQUIRES GOVERNMENTS, NONPROFIT ORGANIZATIONS, PROFESSIONAL AND TRADE ASSOCIATIONS, AND THE PRIVATE SECTOR TO WORK TOGETHER

Some industry groups have collaborated with big companies. In 2018, Microsoft India and the Data Security Council of India launched the CyberShikshaa program to create a pool of skilled female cybersecurity professionals.[3] Some technology companies have launched programs to foster women's interest in and confidence to pursue Internet security careers. One example is IBM Security's Women in Security Excelling program, formed in 2015.[32]

At the organizational level, attracting more women to the cybersecurity field requires a range of efforts. Cybersecurity job ads should be written so that female professionals feel welcome to apply. Recruitment efforts should focus on academic institutions with high

female enrollments. Corporations should ensure that female employees see cybersecurity as a good option for internal career changes. And governments should work with the private sector and academic institutions to get young girls interested in cybersecurity.

Increasing women's participation in cybersecurity is good for women, good for business, and good for society. In the absence of appropriate measures by the private sector and policymakers, the gender disparity can lead to a vicious circle. This is because women are less likely to be attracted to a field dominated by males, and the failure to attract women can result in the further dominance of men. This, in turn, makes it even more difficult to attract women. The government and private sector should collaborate to try to create a more positive image of cybersecurity professionals. It is thus important to encourage girls and women to pursue STEM courses and degrees in K–12 and colleges. Women cybersecurity professionals should also be provided mentorships and support at all job levels.

REFERENCES

1. "14 Percent of Fortune 500 chief information security officers are women," *Cybercrime Magazine,* 2020. https://cybersecurityventures.com/14-percent-of-fortune-500-chief-information-security-officers-are-women/#:~:text=Cybersecurity%2520Ventures%2520tallied%2520the%2520female
2. B. Blum and S. Ben-Hur, "High-tech hopes—Addressing Israel's engineering shortage," *The Jerusalem Post*, Aug. 16, 2020. https://www.jpost.com/jerusalem-report/high-tech-hopes-addressing-israels-engineering-shortage-637325
3. S. Morgan, "Women represent 20 percent of the global cybersecurity workforce in 2019," *Cybercrime Magazine*, Mar. 28, 2019. https://cybersecurityventures.com/women-in-cybersecurity/
4. "Cybersecurity talent study," McAfee, 2018. https://www.mcafee.com/enterprise/en-au/assets/reports/rp-cybersecurity-talent-study.pdf
5. "Cybersecurity jobs: Why so few women?" *The Times of India*, 2018. https://timesofindia.indiatimes.com/business/india-business/cybersecurity-jobs-why-so-few-women/articleshow/86567069.cms
6. "BeecherMadden find that women now make up 18% of the cyber security industry," Beecher Madden, 2018. https://beechermadden.com/women-18-of-the-cyber-security-industry/
7. "Diversity, equity, and inclusion in cybersecurity," The Aspen Institute, Washington, DC, USA, Sept. 2021. https://www.aspeninstitute.org/wp-content/uploads/2021/09/Diversity-Equity-and-Inclusion-in-Cybersecurity_9.921.pdf
8. W. R. Poster, "Cybersecurity needs women," *Nature,* vol. 555, no. 7698, pp. 577–580, 2018, doi: 10.1038/d41586-018-03327-w.
9. "Deloitte initiative to encourage women in cybersecurity expands to Middle East," Consultancy, 2018. https://www.consultancy-me.com/news/1164/deloitte-initiative-to-encourage-women-in-cybersecurity-expanded-to-middle-east
10. "Women4Cyber registry—Database of European women in cybersecurity," European Commission, Brussels, Belgium, Jul. 7, 2020. [Online]. Available: https://digital-strategy.ec.europa.eu/en/news/women4cyber-registry-database-european-women-cybersecurity
11. "Labor force, female (% of total labor force)," World Bank, Washington, DC, USA, 2018. [Online]. Available: https://data.worldbank.org/indicator/SL.TLF.TOTL.FE.ZS?end=2019&start=1990&view=chart
12. "Laws still restrict women's economic opportunities despite progress, study finds," World Bank, Washington, DC, USA, 2019. [Online]. Available: https://www.worldbank.org/en/news/press-release/2021/02/23/laws-still-restrict-womens-economic-opportunities-despite-progress-study-finds
13. S. Thomson, "18 countries where women need their husband's permission to work," World Economic Forum, Geneva, Switzerland, 2015. [Online]. Available: https://www.weforum.org/agenda/2015/11/18-countries-where-women-need-their-husbands-permission-to-get-a-job/
14. "Legal inequalities prevent women from working," Council on Foreign Relations, New York, NY, USA, 2020. [Online]. Available: https://www.cfr.org/legal-barriers/barriers/
15. M. Lidman, "Why barely 1 in 5 Israeli Arab women works, and how high-tech can fix that," *The Times of Israel*, 2018. https://www.timesofisrael.com/why-barely-1-in-5-israeli-arab-women-works-and-how-high-tech-can-fix-that/
16. D. Peacock and A. Irons, "Gender inequality in

cybersecurity: Exploring the gender gap in opportunities and progression," *Int. J. Gender, Sci. Technol.*, vol. 9, no. 1, pp. 25–44, 2017.
17. "By age 6, gender stereotypes can affect girls' choices," NSF, 2019. https://www.nsf.gov/news/news_summ.jsp?cntn_id=190924&WT.mc_id=USNSF_51&WT.mc_ev=click
18. "Statistics: K-12 education," National Girls Collaborative Project, 2016. https://ngcproject.org/statistics
19. "PaloAlto Networks partners with US Girl Scouts on security skills," *Computer Weekly*, 2018. https://www.computerweekly.com/news/450420822/PaloAlto-Networks-partners-with-US-Girl-Scouts-on-security-skills
20. R. Maurer, "Why aren't women working in cybersecurity?" SHRM, Alexandria, VA, USA, Jan. 10, 2017. [Online]. Available: https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/women-working-cybersecurity-gender-gap.aspx
21. "Economic impact and perceptions around the cybersecurity gender gap," Help Net Security, 2020. https://www.helpnetsecurity.com/2020/03/12/cybersecurity-gender-gap/
22. R. Tarun, "Gender diversity in cybersecurity matters to the business," CSO, 2019. https://www.csoonline.com/article/3490417/gender-diversity-in-cybersecurity-matters-to-the-business.html
23. Agents of Change, "Women in the information security profession," Frost & Sullivan, San Antonio, TX, USA, 2019. [Online]. Available: https://1c7fab3im83f5gqiow2qqs2k-wpengine.netdna-ssl.com/wp-content/uploads/2019/03/Women-in-the-Information-Security-Profession-GISWS-Subreport.pdf
24. "Womenomics: Economic growth and the female consumer," The Diversity Council, 2019. https://www.thediversitycouncil.com/womenomics-economic-growth-and-the-female-consumer/
25. L. Lackie, "High-profile data breaches affecting consumer trust in big brands," ITProPortal, 2016. https://www.itproportal.com/2016/05/11/high-profile-data-breaches-affecting-consumer-trust-in-big-brands/
26. N. Bowles, "Thermostats. Locks and lights: Digital tools of domestic abuse," *NY Times*, Jun. 23, 2018. https://www.nytimes.com/2018/06/23/technology/smart-home-devices-domestic-abuse.html
27. S. Acevedo, "The cyberwar needs more women on the front lines," *Wired,* 2019. https://www.wired.com/story/opinion-the-cyber-war-needs-more-women-on-the-front-lines/
28. S. Austin and A. Dowell, ""Girls around me" developer defends app after Foursquare dismissal," *The Wall Street Journal*, Mar. 31, 2012. https://www.wsj.com/articles/BL-DGB-24194
29. K. Hustad, "Girls' interest in technology drops with age, according to new CompTIA study," *Biz Journals*, Sep. 20, 2016. https://www.bizjournals.com/chicago/inno/stories/news/2016/09/20/high-school-girls-interest-in-tech-comptia-study.html
30. C. Sottile and J. Ling Kent, "Girl Scouts fight cybercrime with new cybersecurity badge," *NBC News*, Mar. 4, 2018. https://www.nbcnews.com/tech/tech-news/girl-scouts-fight-cybercrime-new-cybersecurity-badge-n852971
31. M. Saskia Brugman, "The Women4Cyber Foundation," Global Cyber Alliance, Jul. 22, 2021. https://www.globalcyberalliance.org/the-women4cyber-foundation/
32. G. Huang, "Why women in tech should consider a career in cybersecurity," *Forbes*, 2019. https://www.forbes.com/sites/georgenehuang/2016/10/04/why-women-in-tech-should-consider-a-career-in-cybersecurity/?sh=4462a6283e6f

NIR KSHETRI is the "Computing's Economics" column editor of *Computer* and a professor in the Bryan School of Business and Economics, University of North Carolina at Greensboro, Greensboro, North Carolina, 27412, USA. Contact him at nbkshetr@uncg.edu.

MAYA CHHETRI is a professor in the Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, North Carolina, 27412, USA. Contact her at m_chhetr@uncg.edu.

EDITORS: Kyle Johnsen, kjohnsen@uga.edu
Christian Sandor, chris.sandor@gmail.com
Mark Billinghurst, mark.billinghurst@auckland.ac.nz

This article originally appeared in
IEEE Computer Graphics and Applications
vol. 41, no. 6, 2021

DEPARTMENT: SPATIAL INTERFACES

DiVRsify: Break the Cycle and Develop VR for Everyone

Tabitha C. Peck, *Davidson College, Davidson, NC, 28035, USA*
Kyla A. McMullen, *University of Florida, Gainesville, FL, 32611, USA*
John Quarles, *The University of Texas at San Antonio, San Antonio, TX, 78249, USA*

Virtual reality technology is biased. It excludes approximately 95% of the world's population by being primarily designed for male, western, educated, industrial, rich, and democratic populations. This bias may be due to the lack of diversity in virtual reality researchers, research participants, developers, and end users, fueling a noninclusive research, development, and usability cycle. The objective of this article is to highlight the minimal virtual reality research involving understudied populations with respect to dimensions of diversity, such as gender, race, culture, ethnicity, age, disability, and neurodivergence. Specifically, we highlight numerous differences in virtual reality usability between underrepresented groups compared to commonly studied populations. These differences illustrate the lack of generalizability of prior virtual reality research. Lastly, we present a call to action with the aim that, over time, will break the cycle and enable virtual reality for everyone.

Virtual reality (VR) researchers argue that the diverse array of VR applications has the potential to change the world. A subset of these applications includes driving and flight simulators, surgical training, exposure therapy, physical therapy, empathy exercises, and perspective taking. Although these applications are intended to be useful for everyone, regardless of their age, gender, race, culture, ethnicity, class, ability, neurodiversity, etc., the majority of VR development is focused on a minority of the population. This narrow focus limits what we are calling the VR research, development, and usability cycle. This cycle excludes the majority of the population from being involved in the use, research, and development of VR applications and hardware.

0272-1716 © 2021 IEEE
Digital Object Identifier 10.1109/MCG.2021.3113455
Date of current version 10 December 2021.

The VR research, development, and usability cycle (Figure 1) begins with the majority of researchers being Male, White, Educated, Industrialized, Rich, and Democratic (M-WEIRD).[9] The WEIRD population is less than 6% of the world's population. Assuming that men are roughly half of this population, the M-WEIRD population is representative of less than 3% of the world's population. At their inception, VR research questions are primarily created by M-WEIRD researchers and evaluated on M-WEIRD participants.[19] Subsequently, the knowledge gained from this limited subsection of the population is broadly applied 1) in industry by M-WEIRD developers and then 2) influences future research by M-WEIRD researchers. The hardware and software developed by industry professionals, based on knowledge gained and investigated by M-WEIRD researchers and participants is then used by M-WEIRD users. This cycle often excludes other VR users who do not identify as M-WEIRD and may limit the broad effectiveness of VR research and applications.

Research and design decisions may disproportionately dissuade non-M-WEIRD people from effectively

using VR, due to VR research and development primarily focusing on the M-WEIRD population. For example, modern head-mounted displays (HMDs) used for perceiving VR do not accommodate women or children. The interpupillary distance (IPD) accommodated in modern HMDs supports 81% of men's IPDs yet only 61% of women's and even fewer children's.[5] This mismatch in IPD can cause discomfort and eyestrain as well as in depth perception errors. This discrepancy may also explain why the Oculus Rift was found to be more nauseating for women than men.[13] In addition, the physical design of HMDs for VR is often unaccommodating of the hair and headdresses of marginalized racial and ethnic groups.

Because recreational VR has primarily been designed for M-WEIRD users, they are more likely to have positive experiences with VR. These experiences may influence some M-WEIRD users to become VR researchers and developers. The authors have witnessed this anecdotally in the students who elect to enter the field of computer science due to their interest in video games and/or VR. Further, having systems designed or advertised for M-WEIRD users may propagate the lack-of-fit model[8] and build a barrier to entry for underserved populations using VR. Regardless of how or why people decide to become developers and researchers, it is clear that the people in this cycle are majority M-WEIRD. This exclusionary practice continues the cycle where VR will be developed for and designed by a nondiverse group of people. This cycle may inadvertently exclude, limit usability, or dissuade potential VR users that do not identify as M-WEIRD from using VR or becoming researchers or developers.

The VR research, development, and usability cycle is of course simplified for illustration purposes and does not take into account additional external factors. These other factors may affect who uses and develops VR hardware and software, and include societal and family pressures, economic barriers, stereotype threats, exclusionary messaging, etc.

Immediately creating a more diverse group of researchers and developers may not be possible; however, building a diverse group of researchers and developers is an important future goal. Further, asking a more diverse group of users to quickly adopt VR may not initially result in wide acceptance, especially if applications and hardware are not designed to support these users. Nevertheless, the VR research community does have the ability to enforce participant population diversity to include a wider range of participants and to ask research questions that consider the needs of all potential users. Having researchers and reviewers commit to the simple act of diversifying participant populations may have profound impacts on research results and break the current, noninclusive VR research, development, and usability cycle so that future VR is designed to be more inclusive.

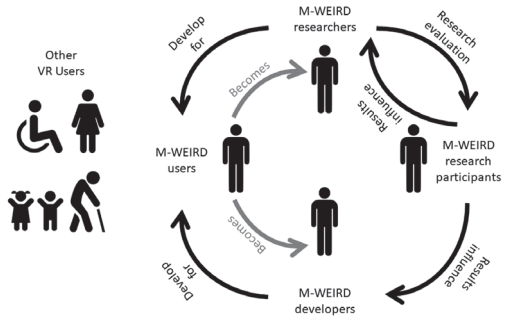


FIGURE 1. Noninclusive VR research, development, and usability cycle.

The goal of this article is to highlight the importance of inclusivity when creating VR hardware, software, and applications. There is a need to include more diverse populations in the VR research, development, and usability cycle. This work introduces the negative impact of design bias and presents research identifying the differences between different demographic groups within VR research. Finally, a call to action is presented for the VR research community to improve our current practices, with the goal of making VR more accessible for everyone.

DESIGN BIAS

Bias is defined as an unequal weighting in favor of one group compared to another. *Design bias* is "the development and dissemination of hardware and software whose characteristics systematically do not meet the needs of a subset of target users."[19] Design bias can reduce the generalizability of research study results and can negatively impact the usability and accessibility of designs. To mitigate bias, VR researchers and developers should include feedback from more diverse populations when designing VR hardware and software. One way to include diverse feedback is through *user-centered design*, a design technique used to engage and involve critical stakeholders in the co-design of the technologies they use. Ensuring a diverse participatory design population can help to reduce design bias. This involvement can occur at any stage of the design process and includes methods such as focus groups, prototype development, and developing storyboards. Historically, design bias has

resulted in negative consequences, as seen in the following cautionary example.

The automotive industry historically excludes female passengers when evaluating vehicle safety. As recently as 2010, automotive safety was primarily tested using crash-test dummies that represented an average-sized, adult male body. Because female passengers were not represented during testing, in the same car accident women had a 47% higher chance of injury compared to men.[3] Even as of today, there are no crash-test dummies based on female bodies, rather they are scaled-down male dummies, which still increases risk of injury for female passengers.

Design bias has also affected many underrepresented groups, such as women, racial and ethnic minorities, children, the elderly, transgender people, people with physical disabilities, and people with cognitive disabilities. These types of biases can appear in physical hardware or software contexts and have been observed in a variety of scenarios from facial recognition to video games.

It is generally agreed that bias can be divided into three categories: preexisting bias, technical bias, and emergent bias.[6] *Preexisting bias* occurs when the designer (knowingly or unknowingly) infuses their own inherit biases into the product they are creating. For example, this can occur in VR, during visuomotor synchrony, when the virtual world is designed such that all users must embody a white virtual avatar. *Technical bias* occurs when the system design limits the usability. This can occur when designing an interface for only right-handed users, or designing an HMD that does not support all user's IPDs. *Emergent bias* occurs over time and as people use the system. It can occur during development or after product release. Consider the case of virtual characters learning verbal interactions driven by machine learning (ML). Using biased training data, virtual characters have learned sexist and racist behaviors.[14]

PARTICIPANT UNDERREPRESENTATION

VR research often requires data collected from human participants, similar to many other research communities. Many of these research communities, including ML, human–computer interaction, and the medical industry have called for using more diverse and representative participant populations. Some of these calls to action have stemmed from research highlighting the problems of underrepresentation in participant samples in psychology research.

Similar to VR, most research in psychology is based on participants who are WEIRD.[9] Henrich *et al.*[9] highlighted several instances where non-WEIRD study participants yield very different results compared to WEIRD participants, even in studies on aspects of human perception that were once thought to be universal. For example, consider the Mueller-Lyer Illusion, which is a perceptual illusion where two lines are judged to be different lengths based upon the orientation of the arrows at the ends. Results of prior studies have shown this to be a very strong illusion in WEIRD societies, but many non-WEIRD societies are not fooled by this illusion. This suggests that since human visual perception varies greatly over such a simple visual illusion, then visual perception differences are likely to be found in VR as well.

Similar concerns have been raised about using nonrepresentative samples in ML datasets. It is widely agreed that when training an ML algorithm, choosing appropriate and representative training data is critical. If the training data are biased, then the ML algorithm predictions will also be biased. The majority of the ML literature tends to focus on sample selection that only includes observable data, meaning that the data that are chosen for training only include the data that could be directly observed. Oftentimes, the predicted outcomes could be influenced by unobservable or uncollected data. For example, suppose someone wanted to create an algorithm to predict Ph.D. student retention using data from past Ph.D. graduates. We would only observe the outcome of a student finishing a Ph.D. program if the student decided to stay in the Ph.D. program. The observed outcomes are only a consequence of a human decision-maker. There are unobservable factors (such as family dynamics, advisor relationship, financial burden, stereotype threat, department climate, and experienced microaggressions) that influence the outcome. This scenario makes it challenging to create a predictive model since the outcomes observed do not represent a random sampling of all Ph.D. students who entered the program (not just the ones who received their Ph.D.). Oftentimes, in ML, the mechanism that determines selection has no impact on the outcome being conditional on the observed attributes. Standard ML classifiers assume that data are drawn independently and normally distributed; however, the selection of examples is often biased, thus leading to biased inferences. For example, training sets consisting of a majority of light-skinned and male faces are speculated to be the reason why face-recognition systems are most accurate at identifying light-skinned male faces, and are less accurate at identifying female faces and dark-

skinned faces.[4] Using ML within VR is becoming more common, and this trend further highlights the importance of using diverse and representative datasets for training in all fields.

Finally, Peck *et al.*[19] demonstrated that female participants were significantly underrepresented in VR research and that this underrepresentation biased research findings. The change in simulator sickness after VR exposure was systematically proportional to the number of women who participated in experiments such that experiments with more women had a smaller increase in simulator sickness. If simulator sickness results were systematically related to gender, it can be assumed that other unknown measures may also be affected. This further demonstrates the importance of using diverse participant populations when performing VR research.

DIVERSITY DIMENSIONS

Even though numerous research communities have called for including more diverse participant samples, it is unclear which diversity dimensions are important to consider. The WEIRD dimensions of race/ethnicity, education, industrialization level, socioeconomic status, and government type are only a small subset of the dimensions of human diversity. Moreover, this subset of dimensions is not regularly reported in research papers, which often report little more than age and binary-gender. These additional diversity dimensions include, but are not limited to age, culture, gender, sex, mental abilities, physical abilities, sexual orientation, appearance and body, class, geographic location, language and accent, migration biographies, parental status, relationship status, and religion.

To evaluate the inclusion of diversity dimensions in human–computer interaction research, Himmelsbach *et al.*[10] investigated how many papers in the *ACM Conference on Human Factors in Computing Systems* (CHI) included diversity dimensions and counted how many dimensions were included in reports of participant demographics. They found that there was a significant increase in the number of diversity dimensions reported per paper between 2001 and 2011. However, there was no significant difference between 2011 and 2016. The average number of dimensions reported in 2016 was approximately 3 out of 14 considered dimensions. Thus, it seems that human diversity has been largely ignored in the CHI community, which overlaps significantly with VR research.

Determining the diversity dimensions that most influence usability and design remains an open research question. Collecting and reporting diversity dimensions will provide readers with a better understanding of the participant populations and may provide insight into demographic groups that should be evaluated in further detail.

PARTICIPANT POPULATION DIFFERENCES

In this section, we present numerous studies that identified response differences based on diversity dimensions. These differences could be caused by numerous factors including biological, behavioral, situational, or cultural differences. Although we present experiments identifying differences based on diversity dimensions, we do not hypothesize on what caused these identified differences.

The authors acknowledge that there are differences within and between the diversity dimension groups identified in this section. As such, we recognize that there are unique effects due to intersectionality that cannot be captured by studying one dimension in isolation. Our hope is that by bringing awareness to the necessity of research along various identity dimensions, future research can explore intersectional experiences in VR.

Gender

Numerous gender differences have been reported in VR research; however, it is likely that many more differences have never been explored. For example, studies specifically looking for gender differences have identified them in perceptual threshold studies.

Additional differences have been found between genders in response to simulator sickness. However, results are mixed and suggest that women are more likely to get sick compared to men, have comparable levels of simulator sickness to men, or based on a five-year meta-analysis may be less likely to get simulator sickness compared to men.[19] This discrepancy suggests that gender differences in simulator sickness is still an open question, where additional factors such as environment type, participant age, or participant VR experience may need to be considered to fully understand response differences within VR experiences.

Additional unexpected gender differences have been identified in VR scenarios. Women and men used hand-held trackers differently, experienced different levels of spatial immersion, and women outperformed men on spatial performance tasks. Women and men experienced embodiment in a self-avatar differently such that women were less likely to accept avatar hands that were different from their own appearance. Demonstrating preexisting bias, this effect was further

seen in a lowered embodiment score for women compared to men when an average male-sized hand was used in an experiment.[17]

Even though differences have been detected, more often than not, gender differences are rarely tested. When differences are tested, studies may only be sufficiently powered to detect large effects while gender differences are likely to be small effects.[19] The sample size in many studies is small ($n \leq 40$) and disproportionately male, further complicating the investigation of potential differences.

Race, Culture, and Ethnicity

Additional diversity dimensions that are seldom evaluated and often conflated include race, culture, and ethnicity. As noted by Henrich *et al.*,[9] people who did not identify as WEIRD perceived a perceptual illusion differently than WEIRD people. This work suggests that differences may exist due to race, culture, and ethnicity.

For example, in VR research, Almog *et al.*[1] conducted a study including Israeli, Arab, and non-Arab men and women. Participants rode in a virtual airplane. The plane took off, flew in nice weather, flew in stormy weather, and then landed. Results suggest that Arab women were significantly less likely than non-Arab women to look out the window of the airplane. This behavior was significantly correlated with their self-reported sense of presence.

Additionally, Olson[15] performed a qualitative analysis of how participants' experience growing up with racial and ethnic socialization (RES) affected decisions in a VR game including racial discrimination themes. The game, Passage Home VR, was an interactive narrative where the player is accused of plagiarism in an educational setting. The body language of the avatar determined the events in the narrative. The author found that participants' previous experiences with RES informed their decisions in VR.

The limited work in this area may be due to different interpretations of race, culture, and ethnicity worldwide as well as the challenge of creating a universal questionnaire for collecting this information. Peck *et al.*[19] proposed a questionnaire and recommended allowing participants to self-identify. Additionally, many research locations may not be racially, ethnically, or culturally diverse, thus limiting the option of investigating differences along these dimensions. Even though there are obvious challenges for investigating race, culture, and ethnicity, it is still possible to collect these participant data and to report them in papers.

Age

Though VR may be created with an adult end-user in mind, attention should be paid to VR applications for populations of any age. For example, VR has positively impacted children and older adults by improving physical activity. VR games increase physical activity in children and research has demonstrated that older adults who use VR for exercise increase their mobility and decrease their likelihood of falling. Conversely, older populations may also be more reluctant to use VR for a myriad of reasons such as user expectations, demographic segmentation in marketing, and lack of familiarity. Future efforts should aim to mitigate these issues to promote the inclusion of older adults.

Participant age can also affect presence and embodiment in VR. For example, Mcglynn *et al.*[11] investigated the influence of age on participants' sense of presence. Participants played a VR game called *Diner Duo*, in which they had to make and serve virtual hamburgers to virtual customers in a virtual diner. Results suggest that spatial presence was not significantly different across ages but older participants were less likely to notice breaks in presence.

When considering embodiment, Serino *et al.*[23] investigated the influence of age in body size perception in VR. Participants experienced an embodiment illusion with an avatar that had smaller proportions than the participants. One task for the participants was to estimate the width of their hips both before and after VR exposure. The 19–25-year-old participants increasingly underestimated hip width post-VR exposure, whereas the 26–55-year-old participants' width estimations were not significantly affected by VR. Further, Peck and Gonzalez-Franco[17] identified that participants over 30 had a lower sense of embodiment in a self-avatar compared to participants under 30.

Children are a protected population and although VR may be beneficial in numerous applications including increasing physical activity, pain distraction, or education care must be taken when developing and designing for this group. Children are physically smaller than adults and current HMDs are not designed for children and do not support their smaller IPDs. Further, children are actively in the process of cognitive development and may respond in unanticipated ways when put into virtual environments.

For example, Segovia *et al.*[21] investigated children's acquisition of false memories as an effect of VR exposure. A false memory in VR implies believing that what happened in the VE happened instead in the real world. They compared several different conditions that could impact memory—idle, mental imagery, VR

with another child's avatar, and VR with a self-avatar. Results suggested that preschool children were equally likely to acquire false memories in any condition. However, elementary school children acquired more false memories in the mental imagery and VR self-avatar conditions than the idle condition. This finding suggests that VR may have different effects on users' memory depending on age.

Disability

VR is not accessible to many people with disabilities. Unfortunately, these populations are rarely consulted during the noninclusive VR research, development, and usability cycle, which creates a major barrier to accessibility. For example, some people with balance impairments may not be able to safely stand up in many standing-based VR experiences. This limitation may prevent users with balance impairments from engaging in all parts of the experience. Oftentimes, disabled people are only considered in VR research for the sole purpose of rehabilitation or correction applications. VR designers never initiate research to study recreational uses of VR for disabled people and instead treat them as a deviation from "normal," which needs correction.[24]

There have been various studies that have investigated the unique experiences and responses that persons with disabilities have in VR. For example, numerous studies support the claim that people with multiple sclerosis (MS) have a different experience of presence than people without MS.[7] Moreover, persons with MS may respond to latency differently than persons without disabilities.[20]

Additional studies have focused on the experience of cybersickness on people with MS as compared to people without MS, specifically looking at physiological and brain wave responses. Many of the baseline differences persisted in VR. However, in some cases, VR induced completely opposite changes between the participants with MS and participants without MS. Differences were found near the parietal lobe—sensation, perception, integration of sensory input and decision-making, emotional behavior, and visual reception—and the frontal lobe, which mainly deals with motor functions.[2] These results suggest that VR may need to be made more adaptive and accessible to the needs of persons with disabilities.

Neurodiversity

VR is widely accepted as a training tool for learning various skills; however, this may not be an acceptable intervention for neurodiverse people. In fact, in some instances, VR has been harmful in practice when applied to autism intervention. Williams and Gilbert[25] conducted a survey of wearable technologies applied to autism intervention. Their work found that 90% of interventions focused on "normalizing" autistic people, viewing their traits as deficits. Only 10% of the technologies surveyed addressed user needs for sensory regulation, emotional regulation, communication, or executive function. It is critical to ensure that VR applications for neurodiverse people do not follow this pattern.

In the VR domain, Self *et al.*[22] created a VR learning environment for children with autism spectrum disorder to learn fire safety skills. All of the children in the study learned and demonstrated an understanding of fire safety skills in the virtual environment. However, these results did not always generalize to the real world. For example, when a fire alarm was triggered in the real world, several students with ASD still needed verbal prompting to exit. Moreover, aspects that can affect learning in VR, such as interfaces, may not be inclusive to neurodivergent people. Mei *et al.*[12] investigated how children with autism spectrum disorder performed basic 3-D interaction tasks, such as rotation and translation, compared to typically developed children. Participants had two tasks: 1) rotating a virtual object to match the pose of a target object, and 2) translating a virtual object to the same relative position as a target object. Results suggested that on tasks requiring attention to precision (e.g., translation along the *z*-axis) and tasks in which the interface was more ergonomically limiting (e.g., rotation tasks when trying to align the controller with the virtual object), the autism spectrum disorder group demonstrated significantly longer completion time and error prone performance than the typically developed group. These findings suggest that additional research should seek to understand how neurodiverse populations learn and apply skills, such that VR can adapt to these specific needs.

CALL TO ACTION

The previous sections presented examples of significant differences in how participants respond to VR based on different diversity dimensions. In some of the above examples, the initial research goal was not to investigate differences in diversity dimensions. For example, the difference in the subjective sense of embodiment by age was identified because of a diverse participant sample even though this was not the intended goal of the study.[17,18] This example highlights the importance of diversifying sample populations when designing for general populations instead

of relying primarily on convenience sampling of M-WEIRD college-aged participants. When working with untested populations, researchers may find unanticipated results that lead to new lines of research and generate more generalizable results across diverse populations.

We argue the VR research community is ethically obligated to develop inclusive VR hardware and applications. To support this, we propose several actions that every VR researcher can take to improve the generalizability of their research through engaging with more diverse participant populations and researchers:

Place greater emphasis on population diversity. When reviewing papers, consider if the participant population is representative of the intended population for the proposed research. If the studied population lacks appropriate diversity (gender, age, race, etc.), provide constructive criticism and require the paper to highlight this limitation and accurately quantify the results according to the lack of participant diversity. Conference and journal review criteria should include evaluation of participant diversity and place more weight on generalizability of results such that papers with greater participant diversity, or research including understudied populations are more likely to be published.

Actively recruit participants from outside your university. Recruiting primarily from university students is a common method of conducting studies in psychology, HCI, and VR, because it is convenient. However, the diversity of university students is limited to a narrow range of ages and educational backgrounds. Thus, sampling only university students may not be sufficient for generalizability. Therefore, most other populations can be considered underrepresented in the scope of VR research.

A first step in the right direction would be to recruit from the local population. For example, recruiting through online social media platforms, like Reddit, working with local businesses to hang flyers in store windows, or collaborating with local affinity groups. The people recruited through these means should be paid for their contribution, including any financial costs incurred from traveling to the laboratory to conduct the experiment (i.e., gas, parking, compensation for long travel time). If possible, you should try to bring the experiment to these outside people, rather than requiring them to travel to you.

Re-evaluate your evaluation methods. Differences may be observed between various populations in VR research; however, it is critical to understand if these differences are characteristic of the population, or if some aspect of the methodology is affecting performance. In addition to collecting quantitative experimental data, qualitative data analyzing the participants' experiences should be taken. This information will be critical in making the experiment design more inclusive. A first step in this direction would be to include an exit survey at the conclusion of each experiment, asking open ended questions about the user's experience, how they perceived the virtual world, and any factors that may have helped or hindered their performance.

Use inclusive imagery in recruitment materials, environmental scenes, and publication images. Imagery provides a subtle cue as to who is "welcome" in a space. For example, a recruitment flyer with an image of a white man may subtly indicate that women and non-white people do not belong. Additionally, hypersexualized images of women are nonprofessional and propagate hostile unwelcoming environments. Make sure the imagery is inclusive during recruitment and publication and use it to encourage the widest variety of people to participate. Show recruitment flyers to a diverse group during the design process; listen and act if someone finds the materials to be offensive or noninclusive. When creating materials, be aware of and avoid stereotypes.

Develop a relationship with affinity groups. Affinity groups are groups of people that meet and have interests. Examples include women in the workplace, working parents, lesbian, gay, bisexual, transgender, queer or questioning, etc. (LGBTQ+) affinities. A positive and collaborative relationship can increase the diversity of the project's research participants or provide diverse insights and perspective on research projects. When working with affinity groups, it is critical that you inform the group of your intentions, for example, to increase recruitment diversity or in an advisory role. Be mindful that the relationship should be mutually beneficial. You could commit to recruiting research students within this group, giving research presentations, or applying for funding to support their advisory role.

Collaborate with researchers from diverse geographic populations. While recruiting from the local general population is a step in the right direction toward increasing participant diversity, it is not sufficient. The better method of increasing participant diversity is to also recruit from beyond the local population. The easiest way to do this is to collaborate with researchers outside your institution or outside your country. Compatibility of hardware and software makes this research more challenging; however, the development of commercialized VR hardware over the past decade makes collaboration between labs more

feasible. Moreover, commercial VR systems are possible to ship for sharing with distant collaborators and participants.

Replicate experiments and include underrepresented populations. The replication crisis[16] highlights the importance of replicating previous studies and valuing replication work. Replicating previous experiments and including diverse participant populations will determine if the experiment is replaceable, supports generalizability of results, or highlights usability differences between groups. Further, when evaluating papers, place a higher value on replication work that includes or investigates previously underserved populations.

Collect and report participant diversity data. When creating your demographic survey, collect additional diversity dimension data and report these data in the participant section of your research papers. This should also be reflected in the discussion and limitations sections, either as a strength or weakness of the work. Adding information about the participant population informs other researchers about the generalizability of the data. It may also highlight interesting and unexpected differences between groups and guide further research. Administer the demographic survey after the experiment so as not to induce unexpected stereotype threats and always present inclusive choices in questions. For example, do not restrict gender to a binary response.

Actively and consistently consider the perspective of someone much different than yourself. In the design process, encourage the research team to ponder, how would someone of a different *race*, *educational background*, *gender*, *age*, *ability*, *ethnicity*, *literacy level*, *culture*, *class*, *and language* experience this system? Would there be any barriers to their engagement? How can we mitigate those challenges? When in doubt, consult with someone from that demographic. Additionally, include people from a wide variety of demographics in an advisory capacity for VR projects to help ponder and answer these questions. If your research and development team does not have the necessary diversity, hire experts from those populations to help inform the trajectory of the research.

Diversify the people in charge. Having diverse reviewers, program committees, keynote speakers, and awardees signifies that diversity is an important contribution worth acknowledging and rewarding since different perspectives enhance research quality. When diversifying committees, be aware that underrepresented populations are often asked to complete higher shares of advisory work. Respect and acknowledge their time and contributions. Do not recruit non-WEIRD personnel solely to increase a diversity quota, but to instead listen, learn, and respect their differing opinions. Further, call people out if someone questions the accomplishments of someone based on their identity, and instead highlight their well-deserved accomplishment and the barriers they may have overcome along the way.

Continually engage in professional development focused on diversity. Caring about diversity is the first step. To better understand, recognize, and respect the importance of diversity requires professional development. Commit to educate yourself in this nuanced and intricate science and treat it with the same respect you would a new academic area. This professional development is an important part of developing as researchers to improve and develop VR for a diverse and inclusive society. Pick up a book, join a reading group, listen to a podcast, or attend a lecture. Search for available resources in your area and utilize them to make yourself a better researcher.

Regardless of your experience level or background, take the first step and commit to any one of the above-mentioned actions to help break the noninclusive VR research, development, and usability cycle. If everyone takes one small step, we can start to dismantle the system and create VR experiences that are useful for and inclusive of everyone.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Rua Williams and Dr. Jessica J. Good for graciously providing their thoughts and insight on this manuscript. This material is based upon work supported by the National Science Foundation under Grant 1942146. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. I. Almog, H. S. Wallach, and M. P. Safir, "Ethnicity and sense of presence in a virtual environment: Arab women—A case in point," in *Proc. Virtual Rehabil. Int. Conf.*, 2009, pp. 78–82.
2. I. M. Arafat, S. M. S. Ferdous, and J. Quarles, "Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2018, pp. 1–120.
3. D. Bose, M. S.-G., and J. R. Crandall, "Vulnerability of female drivers involved in motor vehicle crashes: An analysis of US population at risk," *Amer. J. Public Health*, vol. 101, no. 12, pp. 2368–2373, 2011.

4. J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proc. Conf. Fairness, Accountability Transparency*, 2018, pp. 77–91.
5. N. A. Dodgson, "Variation and extrema of human interpupillary distance," *Proc. SPIE*, vol. 5291, pp. 36–46, 2004.
6. B. Friedman and H. Nissenbaum, "Bias in computer systems," *ACM Trans. Inf. Syst.*, vol. 14, no. 3, pp. 330–347, 1996.
7. R. Guo, G. Samaraweera, and J. Quarles, "The effects of avatars on presence in virtual environments for persons with mobility impairments," in *Proc. 24th Int. Conf. Artif. Reality Telexistence 19th Eurographics Symp. Virtual Environ.*, 2014, pp. 1–8.
8. M. E. Heilman, "Sex bias in work settings: The lack of fit model," *Res. Org. Behav.*, vol. 5, pp. 269–298, 1983.
9. J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world?," *Behav. Brain Sci.*, vol. 33, no. 2/3, pp. 61–83, 2010.
10. J. Himmelsbach, S. Schwarz, C. Gerdenitsch, B. Wais-Zechmann, J. Bobeth, and M. Tscheligi, "Do we care about diversity in human computer interaction: A comprehensive content analysis on diversity dimensions in research," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–16.
11. S. A. McGlynn, "Investigating age-related differences in spatial presence formation and maintenance in virtual reality," Ph.D. dissertation, School Psychol., Georgia Inst. Technol., Atlanta, GA, USA, 2019.
12. C. Mei, L. Mason, and J. Quarles, "Usability issues with 3D user interfaces for adolescents with high functioning autism," in *Proc. 16th Int. ACM SIGACCESS Conf. Comput. Accessibility*, 2014, pp. 99–106.
13. J. Munafo, M. Diedrick, and T. A. Stoffregen, "The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects," *Exp. Brain Res.*, vol. 235, no. 3, pp. 889–901, 2017.
14. G. Neff, "Talking to bots: Symbiotic agency and the case of Tay," *Int. J. Commun.*, vol. 10, 2016, Art. no. 17.
15. D. M. Olson, "Exploring the role of racial and ethnic socialization in virtual reality (VR) narratives," Master Sci. Dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., 2019.
16. H. Pashler and E.-J. Wagenmakers, "Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?," *Perspectives Neuropsychol. Sci.*, vol. 7, no. 6, pp. 528–530, 2012.
17. T. C. Peck and M. Gonzalez-Franco, "Avatar embodiment. A standardized questionnaire," *Front. Virtual Reality*, vol. 1, p. 44, 2021. doi: 103389frvir2020575943.
18. T. C. Peck, J. J. Good, and K. Seitz, "Evidence of racial bias using immersive virtual reality: Analysis of head and hand motions during shooting decisions," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 5, pp. 2502–2512, May 2021.
19. T. C. Peck, L. E. Sockol, and S. M. Hancock, "Mind the gap: The underrepresentation of female participants and authors in virtual reality research," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 1945–1954, May 2020.
20. G. Samaraweera, R. Guo, and J. Quarles, "Head tracking latency in virtual environments revisited: Do users with multiple sclerosis notice latency less?,"*IEEE Trans. Viz. Comput. Graphics*, vol. 22, no. 5, pp. 1630–1636, May 2016. doi: 101109TVCG20152443783.
21. K. Y. Segovia and J. N. Bailenson, "Virtually true: Children's acquisition of false memories in virtual reality," *Media Psychol.*, vol. 12, no. 4, pp. 371–393, 2009.
22. T. Self, R. R. Scudder, G. Weheba, and D. Crumrine, "A virtual approach to teaching safety skills to children with autism spectrum disorder," *Topics Lang. Disord.*, vol. 27, no. 3, pp. 242–253, 2007.
23. S. Serino *et al.*, "The role of age on multisensory bodily experience: An experimental study with a virtual reality full-body illusion," *Cyberpsychol., Behav., Social Netw.*, vol. 21, no. 5, pp. 304–310, 2018.
24. K. Spiel, "The bodies of TEI—Investigating norms and assumptions in the design of embodied interaction," in *Proc. 15th Int. Conf. Tangible, Embedded, Embodied Interact.*, 2021, pp. 1–19.
25. R. M. Williams and J. E. Gilbert, "Perseverations of the academy: A survey of wearable technologies applied to autism intervention," *Int. J. Hum.- Comput. Stud.*, vol. 143, 2020, Art. no. 102485, doi: 101016jijhcs2020102485.

TABITHA C. PECK is currently an Associate Professor of mathematics and computer science at Davidson College, Davidson, NC, USA. Her research interests include developing and testing usable virtual reality systems, and the investigation of the psychological implications of embodiment in self-avatars with the goal of using avatars to reduce and mitigate bias. She has been both the Journal and Conference Paper Program Chair for IEEE VR, and the Journal Paper Science and Technology Chair for IEEE ISMAR. She is a Review Editor for *Frontiers in Virtual Reality* and an Associate Editor for *Presence*. She received the Ph.D. degree in computer science from The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, in 2010. She was the recipient of numerous honorable mentions and nominations for best paper awards at IEEE VR, and received the NSF CAREER Award in 2020. She is an IEEE Senior Member. She is the corresponding author of this article. Contact her at tapeck@davidson.edu.

KYLA A. MCMULLEN is currently an Assistant Professor of computer and information science and engineering at the University of Florida, Gainesville, FL, USA. She directs the SoundPAD Lab, which focuses on the Perception, Application, and Development of 3D audio technologies for virtual and augmented reality. Her research aims to elucidate the human and computational factors that researchers must consider in the design and use of 3D audio systems. She was the General Chair for the International Community on Auditory Display (ICAD) Conference and the National Society of Blacks in Computing (NSBC). She received the bachelor's degree in computer science from the University of Maryland, Baltimore County, Catonsville, MD, USA in 2005 and the Ph.D. degree in computer science and engineering from the University of Michigan, Ann Arbor, MI USA in 2012. She was the recipient of the NSF CAREER Award in 2019. Contact her at drkyla@ufl.edu.

JOHN QUARLES is currently an Associate Professor of computer science at The University of Texas at San Antonio, San Antonio, TX, USA, and the chief technology officer of MedCognition, Inc., San Antonio, TX, USA. He has published numerous works in top conferences such as IEEE VR and top journals, such as IEEE Transactions on Visualization and Computer Graphics. His research interests include virtual, mixed, and augmented reality, serious games, and 3-D user interfaces, with focus on the accessibility of these technologies for persons with disabilities. He has been awarded significant funding from the National Institutes of Health, the Department of Defense, and the National Science Foundation. He received the bachelor's degree in computer science from The University of Texas at Austin, Austin, TX, USA, in 2004, and the Ph.D. degree in computer engineering from the University of Florida, Gainesville, FL, USA, in 2009. He was the recipient of the prestigious NSF CAREER Award in 2014. He was diagnosed with Multiple Sclerosis in 2004, resulting a variety of disabilities that have inhibited his use of virtual reality. He has the unique experience of being both a VR Researcher and an end user with disabilities, serving to inform his chosen research focus. Contact him at John.Quarles@utsa.edu.

Contact department editor Kyle Johnsen at kjohnsen@uga.edu, department editor Christian Sandor at chris.sandor@gmail.com, or department editor Mark Billinghurst at mark.billinghurst@auckland.ac.nz.

DEPARTMENT: EXPERT OPINION

Early History of Texas Instrument's Digital Signal Processor

This article originally appeared in IEEE micro vol. 41, no. 6, 2021

Wanda Gass, *retired, Texas Instuments, Dallas, TX, 75243, USA*

In the 1970s, Texas Instruments was an early player in the microprocessor industry. Its initial success as the microprocessor inside the consumer market products, calculators and digital watches, led to the development of two addition microprocessor families. The 4-bit microcomputer found success in the digital games for children (Simon Says).[1] But the single-chip 16-bit microprocessor, TMS9900, designed for the home computer market and the floating-point microprocessor, TMS9000, struggled to gain traction.[2] Therefore, in the late 1970s, several new architecture teams were created to target more specialized applications.

The architecture of the first-generation digital signal processor (DSP), TMS320C10,[3] started with the successful 4-bit microcomputer which contained the processor, on-chip data and program memory, and a few input/output pins to interact with the end user. This product used a Harvard Architecture where program memory used a customized on-chip read only memory (ROM), which included a fixed program that was in a separate memory space from the data memory. Each microcomputer was customized for the end-product application.

Initially, TMS32010 (C1X) was to have only on-chip program ROM with a fixed program for each application. The initial plan was to include an on-chip analog-to-digital (A/D) converter so that analog signals could be converted to 16-bit data prior to processing. The data path was 16 bits wide, and the accumulator was 32 bits wide, so that it could hold the results of multiplying two 16-bit numbers. A shift function was used to select which of the 16 bits were stored in data memory.

But several fundamental changes were made to the architecture during the development phase. First, the A/D converter was moved off-chip and a digital-to-analog converter was added.[4] The analog chip had a custom interface to transfer data to and from the DSP. Second, the shift and add instruction which required eight cycles to compute a multiplication was replaced with a hardware multiplier so that a multiplication could be done in a single cycle. Third, the I/O pins were redesigned to enable program memory to be fetched from off-chip RAM if the program memory address was in the upper half of the address space. The initial product was first sold in 1983.

One of the early applications for C1X was for the Julie[5] doll in 1987, which was the first commercially available consumer product capable of speech recognition.[6] The initial high-volume applications for

AUTHOR WANDA GASS WITH THE JULIE DOLL

The Julie doll is still working and Wanda uses it in her STEM outreach activities.

0272-1732 © 2021 IEEE
Digital Object Identifier 10.1109/MM.2021.3113541
Date of current version 19 November 2021.

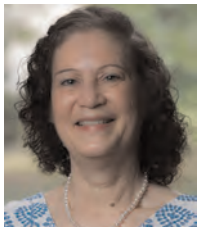
TMS320C25 in 1987 was a real-time microcontroller for positioning the head of a hard disk drive.[7] TMS320C54x was paired with an ARM in 1990 microcontroller to do signal processing for the early digital cellphones.[8]

The next commercially successful DSP architecture was C55X which targeted low-power applications such as cellphones and hearing aids in the late 1990s. The third commercially successful DSP architecture was the high-performance C6X, which was for base stations that handled multiple phone calls simultaneously in the early 2000s.

ONE OF THE EARLY APPLICATIONS FOR C1X WAS FOR THE JULIE DOLL IN 1987, WHICH WAS THE FIRST COMMERCIALLY AVAILABLE CONSUMER PRODUCT CAPABLE OF SPEECH RECOGNITION.

REFERENCES

1. H. Cragon, "The elements of single chip microcomputer architecture," *Computer*, vol. 13, no. 10, pp. 27–41, Oct. 1980, doi: 10.1109/MC.1980.1653373.
2. W. C. Rhines, "The Texas instruments 99/4: World's first 16-bit home computer," *IEEE Spectr.*, Jun. 22, 2017. Accessed: Oct. 20, 2021. [Online]. Available: https://spectrum.ieee.org/the-texas-instruments-994-worlds-first-16bit-computer
3. S. Magar, E. Caudel, and A. Leigh, "A microcomputer with digital signal processing capability," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 1982, pp. 32–33, doi: 10.1109/ISSCC.1982.1156364.
4. E. Caudel, R. Hester, and K.-S. Tan, "A chip set for audio frequency digital signal processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1982, pp. 1065–1068, doi: 10.1109/ICASSP.1982.1171583.
5. Operating Instructions for the Julie Doll. Accessed: Oct. 20, 2021. [Online]. Available: http://www.robotsandcomputers.com/robots/manuals/Julie.pdf
6. P. Rajasekaran and G. Doddington, "Microcomputer implementable low cost speaker-independent word recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1983, pp. 753–756, doi: 10.1109/ICASSP.1983.1172082.
7. Texas Instruments TMS320C25 Reference Guide. Accessed: Oct. 20, 2021. [Online]. Available: https://www.ti.com/lit/ds/symlink/tms320c25.pdf?ts=1626383904990
8. TMS320C54x Reference Guide. [Online]. Available: https://www.ti.com/lit/ug/spru131g/spru131g.pdf



WANDA GASS is currently the President of Design Connect Create, Dallas, TX, USA. She joined Texas Instruments as a member of the eight people team who designed TMS32010. In 1999, she was elected TI Fellow and was elevated to IEEE Fellow in 2007. She retired from TI in 2012 and founded a nonprofit, Design Connect Create, to inspire girls to pursue careers in STEM. She still owns a working Julie doll that she uses in her outreach activity demos. Gass received an M.S. degree in biomedical engineering from Duke University, Durham, NC, USA, in 1980. Contact her at gass@designconnectcreate.org.

EDITOR: Collette Perold, colette.perold@colorado.edu

DEPARTMENT: THINK PIECE

This article originally appeared in
IEEE Annals of the History of Computing
vol. 44, no. 1, 2022

On Logistical Histories of Computing

Matthew Hockenberry, *Fordham University, New York, NY, 10458, USA*

The history of computing is not a history of making *computers*. This is not to say it is not invested in stories of systems, accounts of components, programming languages, and application software—or in the hackers and hobbyists who made them meaningful. On the contrary, accounts of making *the* computer are endlessly fascinating. It is when we begin to talk about more than one that interest tends to falter. When the wonder of invention falls away, what is left is the minutia of manufacture—another unit on the line, another requisition, another stop on the great global supply chain. But it is precisely in this long *logistical* history—the one concerned with assembly revisions and inventory management, part prices, and labor costs—that we find the entangled origins of the messy machines people actually use.

Attempts at this history tend to be surprisingly constrained, confined to small groups or almost singular entities. Too often, it is the *reception* of the computer that is made social, not its production. But computers, in all their plural possibility, are a product only possible in the era of supply chain capitalism [1]. In exploiting legal and economic peripheries, differences in gender, race, and class, they have not just grown into patterns of labor exploitation and environmental abuse, they emerged from them. From this perspective, for example, women have always been critical to the history of computing machines, and not only in their formative roles as human calculators or early programmers. In manufacturing, female labor—and the deliberate exploitation of the liminal economic spaces woman often occupied—was central to enabling the production of the personal computer at a global scale.

In the mass manufacture of the Apple II, to follow a single thread, we leave behind the apocryphal origin of the two Steves in the garage for an already sprawling supply chain stretching from Silicon Valley to Singapore. When printed circuit boards arrived in those early days, they came to Stevens Creek Boulevard, where the nascent firm's administration occupied half of the office, and the "factory" the other. But getting them populated and into the plastic cases that marked the computer's transition from hobbyist kit to a consumer product required a low-cost labor force: a group of (so-called) "housewives" spread throughout the valley, stuffing circuit boards to be delivered daily by station wagon.

At the very least, this pattern of labor connects the computer to a longer history of domestic manufacture—to the kind of "putting-out system" once employed for shoemakers and seamstresses. But if you consider that the talk of "houses" meant cramped apartments, and that "wives" was a patronizing descriptor for working women, many newly arrived from Southeast Asia or Mexico and paid by the piece, you find troubling connections to the history of sweatshop labor in even the earliest accounts of our machine's mass-manufacture [2].

Apple, like many other computer companies, no longer has a factory. The last time they spoke about running one was before the launch of the iMac. Online ordering, Steve Jobs explained, meant a "different kind of store," and it necessitated a "different kind of factory" [3]. But while the build-to-order Power Macs he introduced in 1997 suggested a turn toward increasing variety, the longer lasting outcome was to simplify Apple's offerings. Computers were now split into parts rather than products—and fewer finished goods and standardized components coming "off the shelf" allowed companies to make their inventories "leaner." The Power Macs might have been put in for production at the company's own facilities, but the scale of the iMac's success a year later cemented the shift to a more outsourced assembly. With a software system capable of connecting the company to an interchangeable network of parts suppliers, manufacturers, and resellers, it was not difficult to switch out a few firms here and there. In two years, the factory was so "different" that Apple did not even run it. Foxconn did. And stories of sweatshops at home could be exchanged

1058-6180 © 2022 IEEE
Digital Object Identifier 10.1109/MAHC.2022.3151988
Date of current version 18 March 2022.

for more distant accounts of grueling labor regimes, untold environmental degradation, and workers protesting by threat of suicide.

It is now an almost universal truth that computers are "made in China," but that label masks the countless countries traveled on the way. The question of where computers are made, and how—at scale—is less present in our history than it should be. While there is no simple account of something made with hundreds of parts, in hundreds of places, in almost every instance these connections reveal how complex decisions about making "computers" shaped the future of "the computer." As recent scholarship increasingly turns to more detailed examinations of these logistical histories, [4] we begin to consider accounts of how, for example: the use of off-the-shelf components for systems like the IBM 5150 enabled the rapid manufacture of PC compatibles and clones; how production constraints for floppy disks and audio cassettes served to distribute different kinds of software across North America, Europe, and East Asia; how centralization of hard disk manufacture in places like Thailand accelerated solid state adoption after natural disaster disrupted the supply chain; or how production capability for plastic cases made the Apple II seem like a serious product, while the availability of translucent plastics made the iMac a playful one. We might consider how software has been developed to simulate what was once just the other side of the office, providing tools for increasingly more distant acts of assembly. Or we might begin to grapple with the industrial processing of what is left inside all these discarded boxes. After all, it is not only *the* computer that has a history, but each and every one coming off the logistical line.

REFERENCES

[1] A. Tsing, "Supply chains and the human condition," *Rethinking Marxism*, vol. 21, no. 2, pp. 148–176, 2009.

[2] See, for example, "Switching supplies grow in the bellies of computers," *Electron. Bus.*, vol. 9, pp. 120–126, Jun. 1983, and M. Moritz, *The Little Kingdom: The Private Story of Apple Computer*. New York, NY, USA: William Morrow, 1984.

[3] S. Jobs, *Apple Media Presentation*, Nov. 10, 1997.

[4] M. Hockenberry, N. Starosielski, and S. Zieger, Eds. *Assembly Codes: The Logistics of Media*. Durham, NC, USA: Duke University Press, 2021.

MATTHEW HOCKENBERRY is an Assistant Professor of Media Industries at Fordham University in New York, NY, USA. He is a historian and theorist whose work examines the media of global production, and editor (with Nicole Starosielski and Susan Zieger) of *Assembly Codes: The Logistics of Media* (Duke University Press, 2021). His current project traces how media forms shaped the emergence of new methods for logistical production and distribution in the nineteenth and twentieth centuries. Contact him at mhockenberry@fordham.edu.

Get Published in the New *IEEE Open Journal of the Computer Society*

Submit a paper to the new IEEE Open Journal of the Computer Society covering computing and informational technology.

Your research will benefit from the IEEE marketing launch and 5 million unique monthly users of the IEEE *Xplore*® Digital Library. Plus, this journal is fully open and compliant with funder mandates, including Plan S.

Submit your paper today!
Visit www.computer.org/oj to learn more.

IEEE COMPUTER SOCIETY | IEEE

IEEE

COMPUTER ARCHITECTURE LETTERS

***IEEE Computer Architecture Letters* is a forum for fast publication of new, high-quality ideas in the form of short, critically refereed technical papers.** Submissions are accepted on a continuing basis and letters will be published shortly after acceptance in IEEE Xplore and in the Computer Society Digital Library.

Submissions are welcomed on any topic in computer architecture, especially:

- Microprocessor and multiprocessor systems
- Microarchitecture and ILP processors
- Workload characterization
- Performance evaluation and simulation techniques
- Interactions with compilers and operating systems
- Interconnection network architectures
- Memory and cache systems
- Power and thermal issues at the architectural level
- I/O architectures and techniques
- Independent validation of previously published results
- Analysis of unsuccessful techniques
- Domain-specific processor architecture (embedded, graphics, network)
- High-availability architectures
- Reconfigurable computer architectures

www.computer.org/cal

Join the IEEE Computer Society for subscription discounts today!

www.computer.org/product/journals/cal

IEEE COMPUTER SOCIETY

IEEE

IEEE

SECURITY & PRIVACY

IEEE Security & Privacy **is a bimonthly magazine communicating advances in security, privacy, and dependability** in a way that is useful to a broad section of the professional community.

The magazine provides articles with both a practical and research bent by the top thinkers in the field of security and privacy, along with case studies, surveys, tutorials, columns, and in-depth interviews. Topics include:

- Internet, software, hardware, and systems security
- Legal and ethical issues and privacy concerns
- Privacy-enhancing technologies
- Data analytics for security and privacy
- Usable security
- Integrated security design methods
- Security of critical infrastructures
- Pedagogical and curricular issues in security education
- Security issues in wireless and mobile networks
- Real-world cryptography
- Emerging technologies, operational resilience, and edge computing
- Cybercrime and forensics, and much more

www.computer.org/security

Join the IEEE Computer Society for subscription discounts today!

www.computer.org/product/magazines/security-and-privacy

IEEE COMPUTER SOCIETY

IEEE

Keep up with the latest IEEE Computer Society publications and activities wherever you are.

Follow us:

@ComputerSociety

facebook.com/IEEEComputerSociety

IEEE Computer Society

youtube.com/ieeecomputersociety

instagram.com/ieee_computer_society

IEEE COMPUTER SOCIETY

IEEE

IEEE TRANSACTIONS ON

COMPUTERS

Call for Papers: *IEEE Transactions on Computers*

Publish your work in the IEEE Computer Society's flagship journal, *IEEE Transactions on Computers* (*TC*). *TC* is a monthly publication with a wide distribution to researchers, industry professionals, and educators in the computing field.

TC seeks original research contributions on areas of current computing interest, including the following topics:

- Computer architecture
- Software systems
- Mobile and embedded systems
- Security and reliability
- Machine learning
- Quantum computing

All accepted manuscripts are automatically considered for the monthly featured paper and annual Best Paper Award.

Learn about calls for papers and submission details at
www.computer.org/tc.

IEEE COMPUTER SOCIETY

IEEE

Computer

Seeks 2025–2026 Editor in Chief

APPLICATION DEADLINE: 1 MARCH 2024

Computer, the flagship publication of the IEEE Computer Society, publishes peer-reviewed articles written for and by computer researchers and practitioners representing the full spectrum of computing and information technology, from hardware to software and from emerging research to new applications. *Computer* seeks to deliver useful information for all computing professionals and students, including computer scientists, engineers, and practitioners of all levels.

Computer's Editor in Chief must be an IEEE member in good standing, who has strong familiarity with the monthly magazine, including its special issues, columns, and departments. The successful candidate will have experience attracting and developing a diverse team of talented and highly respected individuals to serve key editorial board roles and contribute excellent content on a timely and reliable basis.

SUBMIT YOUR APPLICATION TODAY!

For complete details on the application process, please visit:
computer.org/press-room/seeking-2025-editors-in-chief

IEEE COMPUTER SOCIETY

IEEE

Conference Calendar

IEEE Computer Society conferences are valuable forums for learning on broad and dynamically shifting topics from within the computing profession. With over 200 conferences featuring leading experts and thought leaders, we have an event that is right for you. Questions? Contact conferences@computer.org.

NOVEMBER

1 November

- CIC (IEEE Int'l Conf. on Collaboration and Internet Computing), Atlanta, USA
- CogMI (IEEE Int'l Conf. on Cognitive Machine Intelligence), Atlanta, USA
- DTTIS (IEEE Int'l Conf. on Design, Test, and Technology of Integrated Systems), Gammarth, Tunisia
- TPS (IEEE Int'l Conf. on Trust, Privacy and Security in Intelligent Systems and Applications), Atlanta, USA
- TrustCom (IEEE Int'l Conf. on Trust, Security, and Privacy in Computing and Communications), Exeter, UK

2 November

- CyberC (Int'l Conf. on Cyber-Enabled Distributed Computing and Knowledge Discovery), Suzhou, China

4 November

- ICEBE (IEEE Int'l Conf. on E-Business Eng.), Sydney, Australia

6 November

- ASONAM (IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining), Marrakesh, Morocco
- FOCS (IEEE Symposium on Foundations of Computer Science), Santa Cruz, CA, USA

12 November

- SC (Int'l Conf. for High-Performance Computing, Networking, Storage, and Analysis), Denver, USA

24 November

- T4E (Int'l Conf. on Technology for Education), Mumbai, India

DECEMBER

1 December

- ICDM (IEEE Int'l Conf. on Data Mining), Shanghai, China

4 December

- CloudCom (IEEE Int'l Conf. on Cloud Computing Technology and Science), Napoli, Italy
- CSDE (IEEE Asia-Pacific Conf. on Computer Science and Data Eng.), Nadi, Fiji
- ICA (IEEE Int'l Conf. on Agents), Kyoto, Japan
- UCC (IEEE/ACM Int'l Conf. on Utility and Cloud Computing), Taormina, Italy

5 December

- BIBM (IEEE Int'l Conf. on Bioinformatics and Biomedicine), Istanbul, Turkey
- RTSS (IEEE Real-Time Systems Symposium), Taipei, Taiwan

11 December

- ICAMLDL (Int'l Conf. on Advanced Machine Learning and Deep Learning), Raipur, India
- IRC (IEEE Int'l Conf. on Robotic Computing), Laguna Hills, CA, USA
- ISM (IEEE Int'l Symposium on Multimedia), Laguna Hills, USA

14 December

- BCD (IEEE/ACIS Int'l Conf. on Big Data, Cloud Computing, and Data Science Eng.), Ho Chi Minh City, Vietnam

15 December

- BigData (IEEE Int'l Conf. on Big Data), Sorrento, Italy

18 December

- HiPC (IEEE Int'l Conf. on High-Performance Computing, Data, and Analytics), Goa, India
- iSES (IEEE Int'l Symposium on Smart Electronic Systems), Ahmedabad, India

2024

JANUARY

3 January

- WACV (IEEE/CVF Winter Conf. on Applications of Computer Vision), Waikoloa, USA

17 January
- AIXVR (IEEE Int'l Conf. on Artificial Intelligence eXtended and Virtual Reality), Los Angeles, USA
- ICOIN (Int'l Conf. on Information Networking), Ho Chi Minh City, Vietnam

27 January
- ASSIC (Int'l Conf. on Advancements in Smart, Secure and Intelligent Computing), Bhubaneswar, India

FEBRUARY

1 February
- BICE (Black Issues in Computing Education Symposium), Santo Domingo, Dominican Republic

5 February
- AIMHC (IEEE Int'l Conf. on Artificial Intelligence for Medicine, Health and Care), Laguna Hills, USA
- CDKE (IEEE Int'l Conf. on Conversational Data & Knowledge Eng.), Laguna Hills, CA, USA
- ICSC (IEEE Int'l Conf. on Semantic Computing), Laguna Hills, USA

19 February
- ICNC (Int'l Conf. on Computing, Networking and Communications), Big Island, Hawaii, USA

MARCH

2 March
- CGO (IEEE/ACM Int'l Symposium on Code Generation and Optimization), Edinburgh, UK
- HPCA (IEEE Int'l Symposium on High-Performance Computer Architecture), Edinburgh, UK

11 March
- PerCom (IEEE Int'l Conf. on Pervasive Computing and Communications), Biarritz, France

12 March
- SANER (IEEE Int'l Conf. on Software Analysis, Evolution and Reengineering), Rovaniemi, Finland

16 March
- VR (IEEE Conf. on Virtual Reality and 3D User Interfaces), Orlando, USA

17 March
- SSIAI (IEEE Southwest Symposium on Image Analysis and Interpretation), Santa Fe, New Mexico, USA

APRIL

10 April
- SaTML (IEEE Conf. on Secure and Trustworthy Machine Learning), Toronto, Canada

16 April
- ICDE (IEEE Int'l Conf. on Data Eng.), Utrecht, The Netherlands

22 April
- VTS (IEEE VLSI Test Symposium), Tempe, Arizona, USA

23 April
- PacificVIS (IEEE Pacific Visualization Symposium), Tokyo, Japan

29 April
- DCOSS-IoT (Int'l Conf. on Distributed Computing in Smart Systems and the Internet of Things), Abu Dhabi, United Arab Emirates

MAY

21 May
- ISORC (IEEE Int'l Symposium on Real-Time Distributed Computing), Tunis, Tunisia

28 May
- ISMVL (IEEE Int'l Symposium on Multiple-Valued Logic), Brno, Czech Republic

IEEE COMPUTER SOCIETY CAREER CENTER

Evolving Career Opportunities

Explore new options—upload your resume today

careers.computer.org

Changes in the marketplace shift demands for vital skills and talent. The **IEEE Computer Society Career Center** is a valuable resource tool to keep job seekers up to date on the dynamic career opportunities offered by employers.

Take advantage of these special resources for job seekers:

- JOB ALERTS
- TEMPLATES
- WEBINARS
- CAREER ADVICE
- RESUMES VIEWED BY TOP EMPLOYERS

No matter what your career level, the IEEE Computer Society Career Center keeps you connected to workplace trends and exciting career prospects.

IEEE COMPUTER SOCIETY

IEEE