COMPUTING

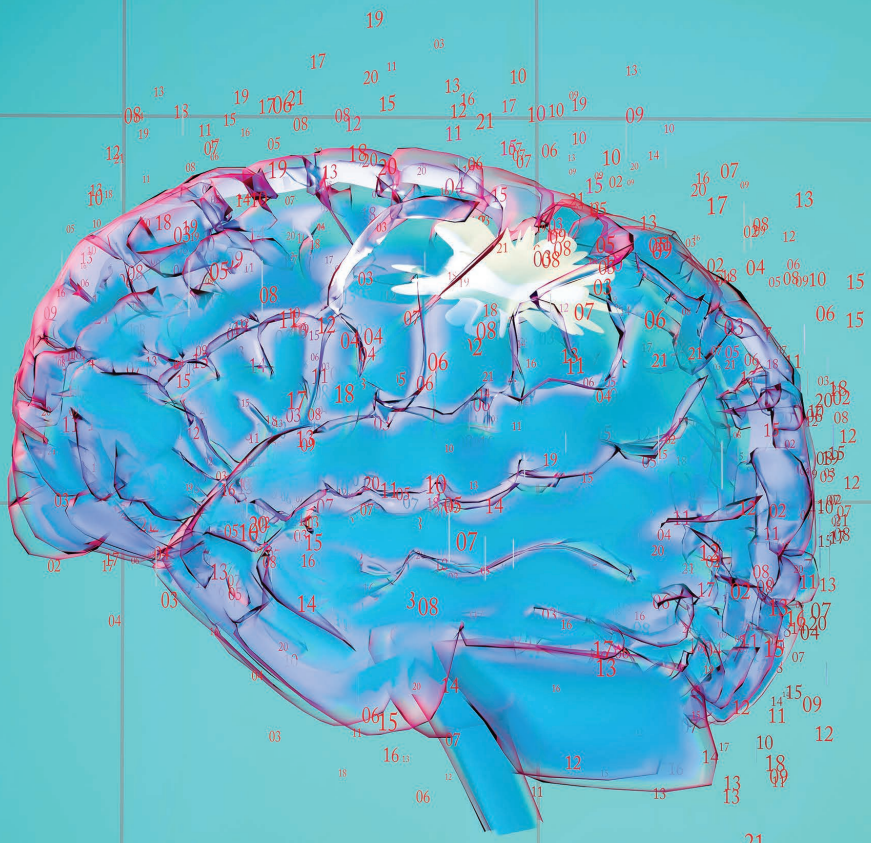
edge

> Virtual and Augmented Reality

> Machine Learning

> Security and Privacy

> High-Performance Computing



MARCH 2020

www.computer.org

IEEE

IEEE COMPUTER SOCIETY

Keep Your Career Options Open

Upload Your Resume Today!

Whether your enjoy your current position or you are ready for change, the **IEEE Computer Society Jobs Board** is a valuable resource tool.

Take advantage of these special resources for job seekers:

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

No matter your career level, the IEEE Computer Society Jobs Board keeps you connected to workplace trends and exciting new career prospects.

www.computer.org/jobs

IEEE COMPUTER SOCIETY

COMPUTING edge

IEEE COMPUTER SOCIETY computer.org • +1 714 821 8380

Printed with inks containing soy and/or vegetable oils

IEEE

IEEE COMPUTER SOCIETY

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

Computer
Jeff Voas, *NIST*

Computing in Science & Engineering
Lorena A. Barba (Interim), *George Washington University*

IEEE Annals of the History of Computing
Gerardo Con Diaz, *University of California, Davis*

IEEE Computer Graphics and Applications
Torsten Möller, *Universität Wien*

IEEE Intelligent Systems
V.S. Subrahmanian, *Dartmouth College*

IEEE Internet Computing
George Pallis, *University of Cyprus*

IEEE Micro
Lizy Kurian John, *University of Texas at Austin*

IEEE MultiMedia
Shu-Ching Chen, *Florida International University*

IEEE Pervasive Computing
Marc Langheinrich, *Università della Svizzera italiana*

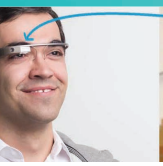
IEEE Security & Privacy
David Nicol, *University of Illinois at Urbana-Champaign*

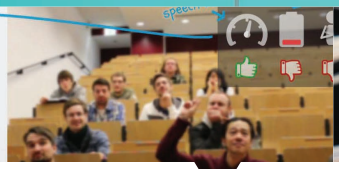
IEEE Software
Ipek Ozkaya, *Software Engineering Institute*

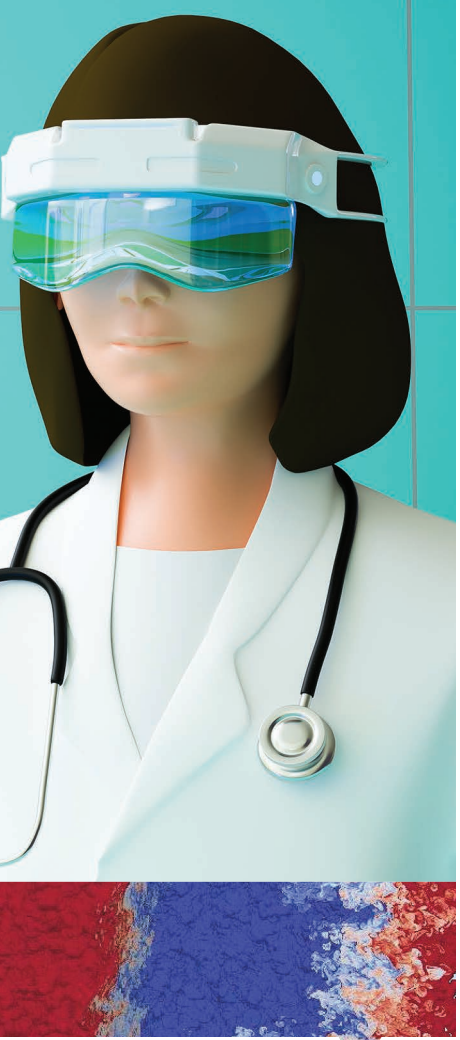
IT Professional
Irena Bojanova, *NIST*

MARCH 2020 • VOLUME 6, NUMBER 3

COMPUTING edge







38

In Situ Visualization for Computational Science

Virtual and Augmented Reality

Machine Learning

Security and Privacy

High-Performance Computing

Departments

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

Artificial Intelligence of Things in Sports Science: Weight Training as an Example

The interdisciplinary nature of sports science introduces challenges such as multifaceted data collection, accuracy in knowledge formation, and equipment usability. Artificial intelligence of things (AIoT) technology presents a feasible solution adaptable to different sports. Taking weight training as an example, the authors of this article from the November 2019 issue of *Computer* apply AIoT technology to these challenges.

Computing in Science & Engineering

Package Wavemulcor: Wavelet Multiple Regression and Correlation in R

The R wavemulcor package provides routines for the implementation of wavelet multiple regression and correlation methods. Its main functionality is the estimation of single sets of global wavelet multiple correlations, cross-correlations, and wavelet local multiple correlations over time from a multivariate time series. Read more in the November/December 2019 issue of *Computing in Science & Engineering*.

IEEE Annals of the History of Computing

Technical Fixes for Legal Uncertainty in the 1980s Software Cracking Scene

During the 1980s, many software developers and enthusiasts reacted to software's unclear status within the US copyright law by pursuing technical, rather than legislative, fixes. This article from the October–December 2019 issue of *IEEE Annals of the History of Computing* applies forensic methods to recover records of conflict between commercial software developers and the enthusiasts who sought

to "crack" commercial programs' copy protection.

IEEE Computer Graphics and Applications

A Provenance Task Abstraction Framework

Visual analytics tools integrate provenance recording to externalize analytic processes or user insights. Provenance can be captured on varying levels of detail, and activities can be characterized from different granularities. However, current approaches do not support inferring activities that can only be characterized across multiple levels of provenance. The authors of this article from the November/December 2019 issue of *IEEE Computer Graphics and Applications* propose a task abstraction framework that consists of a three-stage approach, composed of 1) initializing a provenance task hierarchy, 2) parsing the provenance hierarchy by using an abstraction mapping mechanism, and 3) leveraging the task hierarchy in an analytical tool. Furthermore, the authors identify implications to accommodate iterative refinement, context, variability, and uncertainty during all stages of the framework. They describe a use case that exemplifies the abstraction framework, demonstrating how context can influence the provenance hierarchy to support analysis. The article concludes with an agenda, raising and discussing challenges that need to be considered for successfully implementing such a framework.

IEEE Intelligent Systems

Large Basic Cone and Sparse Subspace Constrained Nonnegative Matrix Factorization with Kullback–Leibler Divergence for Data Representation

In this article from the July/August 2019 issue of *IEEE Intelligent Systems*, a new constrained NMF model with Kullback-Leibler (KL) divergence is developed for data representation. It is called large basic cone and sparse representation-constrained nonnegative matrix factorization with Kullback-Leibler divergence (conespaNMF_KL). It achieves sparseness from a large simplicial cone constraint on the base and sparse regularize on the extracted features.

IEEE Internet Computing

Effectively Linking Persons on Cameras and Mobile Devices on Networks

Nowadays, ubiquitous surveillance cameras and wireless networks facilitate understanding of online and offline human activities, and attention has been paid to analyzing people's activities by associating visual data with network data. It is critical to link mobile devices on networks to their owners in surveillance videos. However, existing works mainly rely on location-related information, while it is invalid to distinguish persons with similar movement behaviors due to the error-prone visual localization of person and wireless localization of devices. To solve this problem, in this article from the July/August 2019 issue of *IEEE Internet Computing*, the authors propose a framework called PD-Link that uses just one camera and one AP to infer links from mobile devices to their owners. To achieve its goal, PD-Link identifies a feature, person-device interactions (PDI), as the linking cue. The authors evaluate the performance of PD-Link in realistic scenarios, and the experimental results demonstrate the efficacy and effectiveness of the method.

IEEE Micro

A Logic-on-Memory Processor-System Design with Monolithic 3D Technology

In recent years, the size of transistors has been scaled down to a few nanometers, and further shrinking will eventually reach the atomic scale. Monolithic three-dimensional (M3D) ICs use the third dimension for placement and routing, which helps reduce footprint and improve power and performance of circuits without relying on technology shrinking. This article from the November/December 2019 issue of *IEEE Micro* explores the benefits of M3D ICs using OpenPiton, a scalable open-source Reduced Instruction Set Computer (RISC)-V-based multicore SoC. With a logic-on-memory 3D integration scheme, the authors analyze the power and performance benefits of two OpenPiton single-tile systems with smaller and larger memory architectures. The logic-on-memory M3D design shows a 36.8-percent performance improvement compared to the corresponding tile

design in 2D. In addition, at isoperformance, M3D shows a 13.5-percent total power savings.

IEEE MultiMedia

Smart Media Transport: A Burgeoning Intelligent System for Next-Generation Multimedia Convergence Service over Heterogeneous Networks in China

Convergence service (CS), which integrates broadband and broadcast services, is considered promising in bridging the gap between scarce bandwidth and user demands for real-time, immersive, and super high-quality service in a ubiquitous environment. However, many urgent issues remain to be tackled in media transport systems (MTS) to support CS. In this article from the July–September 2019 issue of *IEEE MultiMedia*, the challenges MTS would entail over a heterogeneous network to diverse users are first analyzed. The architecture and functionalities of the smart media transport (SMT) system are then advocated and reviewed. Several key technologies in the SMT system—including a hierarchical media-aware data model for content encapsulation, a context-aware data protection mechanism for reliable transmission, and media synchronization control based on network condition estimation—are discussed. Field trial results on campus for certain CS are provided to validate the system design and illustrate the performance of the SMT system.

IEEE Pervasive Computing

A System for Privacy-Preserving Access Accountability in Critical Environments

In this article from the April–June 2019 issue of *IEEE Pervasive Computing*, the authors present a BeagleBone-based system to increase the security and safety level of critical environments by tracking people movements. Their solution is privacy-aware, as it is possible to know who accessed a zone at a given time, with an uncertainty degree of accountability.

IEEE Security & Privacy

Did App Privacy Improve after the GDPR?

In this article from the November/December 2019 issue of *IEEE Security & Privacy*, the authors present an analysis of app behavior before and after the regulatory change in data protection in Europe. The data shows that app privacy has moderately improved after the implementation of the General Data Protection Regulation (GDPR).

IEEE Software

Critical Factors for Open Source Advancement in the US Department of Defense

Leveraging open source components in US Department of Defense (DoD) software systems remains challenging and is often met with resistance. This article from the November/December 2019 issue of *IEEE Software* describes several factors that will increase the likelihood of successfully deploying open source in DoD projects.

IT Professional

Isogeny-Based Cryptography: A Promising Post-Quantum Technique

Recent advances in quantum computing have made existing public key cryptosystems vulnerable to enormous security threats. Therefore, numerous efforts have been exploring post-quantum cryptographic techniques to address the emergence of quantum computing. The authors of this article from the November/December 2019 issue of *IT Professional* focus on one promising post-quantum cryptography known as "isogeny-based cryptography," first by reviewing some concepts of the elliptic curve isogeny, then by presenting three major methods for constructing isogeny-based difficult problems. The authors present some existing isogeny-based cryptographic primitives, such as signature and key exchange, and analyze their advantages and disadvantages. Finally, they discuss a few major challenges of isogeny research that they hope will attract more attention to isogeny-based cryptography.

EDITOR'S NOTE

Virtual and Augmented Reality in Healthcare

Virtual reality (VR) and augmented reality (AR) are entering the healthcare sphere. From training surgeons to motivating patients during physical rehabilitation, VR and AR have the potential to help improve the treatment of diverse medical conditions. VR has already been used to treat pain, anxiety, phobias, and post-traumatic stress disorder by providing distraction, relaxing environments, and safe exposure scenarios. Two articles in this issue of *ComputingEdge* describe VR and AR systems that aim to help people with certain disabilities improve their social interactions.

"A Virtual-Reality Training Application for Adults with Asperger's Syndrome," from *IEEE Computer Graphics and Applications*, presents a VR serious game that allows users to practice presentation and speaking skills and receive constructive feedback. *IEEE Pervasive Computing*'s "Designing Systems to Augment Social Interactions" discusses AR systems that use Google Glass to help people with autism and Parkinson's disease adjust the rate, volume, and pitch of their speech for better communication.

Another technology that is having an impact on healthcare is machine learning (ML). *IEEE Micro*'s "Earning Stripes in Medical Machine Learning" explains how ML algorithms can help diagnose various medical conditions through image analysis. *Computer*'s "Detecting and Alleviating Stress with SoDA" describes an ML-based system that analyzes sensor data from wearables to determine whether the user is stressed, and then suggests stress-reduction techniques (such as meditation) if needed.

For ML to benefit society, we need to make sure that ML software implementations are secure. In *Computer*'s "Security Engineering for Machine Learning," the authors categorize security threats to ML systems and emphasize the need for an architectural risk analysis of ML. The author of *IEEE Security & Privacy*'s "Unknowable Unknowns" posits that the opaque nature of ML decision-making can lead to security and privacy failings.

The final two articles in this *ComputingEdge* issue discuss tools for high-performance computing (HPC). "Software Stack in a Snapshot," from *Computing in Science & Engineering*, presents a containerization service that allows HPC users to easily migrate their software stack to the supercomputer and to deploy their preferred software version. "*In Situ* Visualization for Computational Science," from *IEEE Computer Graphics and Applications*, argues that *in situ* processing can overcome limitations caused by supercomputer I/O constraints but that more research is needed to refine the approach.

Department

A Virtual Reality Training Application for Adults With Asperger's Syndrome

Diego Rojo
Centro Universitario de Tecnología y Arte Digital

Jesús Mayor
Centro Universitario de Tecnología y Arte Digital

José Jesús García Rueda
Centro Universitario de Tecnología y Arte Digital

Laura Raya
Centro Universitario de Tecnología y Arte Digital

Editor: Mike Potel
potel@wildcrest.com

***Abstract*—Asperger's syndrome is a disorder that involves a qualitative impairment in social interactions. While most treatments are aimed at children or adolescents, in this paper we present the development of a virtual reality training application in which adults with Asperger's syndrome can train in an autonomous and controlled way how to present in public.**

Presentation and communication skills are increasingly important in professional and academic environments, e.g., making a public presentation of a student's final dissertation is a requirement in order to graduate in the European higher education area. Acquiring these skills is difficult and requires specialized training, as shown by the many courses aimed at helping people tackle their difficulties with presenting in public or even overcoming phobia of public speaking. For those with Asperger's syndrome (AS), being able to make public presentations becomes even a greater challenge and can be a determining factor in their academic or career progress.

The use of virtual reality (VR) as a tool for rehabilitation of different types of phobias and disorders has been studied since the 1990s, proving its effectiveness in overcoming different phobias such as acrophobia or arachnophobia. However, the use of technology to improve the social skills of Asperger's syndrome has mainly focused on research and development of three-dimensional (3-D) virtual environments,[1,2] but without using VR head-mounted displays (HMDs).

The report by Kothgassner *et al.*[3] supports the hypothesis that the physiological reaction to speaking in front of an audience is similar, whether in a real-life environment or in a virtual reality environment using an HMD. This suggests

Digital Object Identifier 10.1109/MCG.2018.2884272
Date of current version 22 March 2019.

 Published by the IEEE Computer Society

that the use of VR is a good approach to overcome a phobia of public speaking.

To this end, we have developed CicerOn VR: Virtual Speech Coach, a project of the Research Chair in Accessible Technologies created by Indra Sistemas S.A., Fundación Universia, and Centro Universitario de Tecnologıa y Arte Digital. The project's objective is to support people with Asperger's syndrome in the process of overcoming their difficulties when public speaking, by designing and developing an immersive virtual reality serious game. In this environment, users can train in a gradual and controlled way how to interact with others and present in public.

ASPERGER'S SYNDROME

Asperger's syndrome is an Autism Spectrum Disorder characterized by qualitative impairment in social interaction. Individuals have no clinically significant general delay in language, but show restricted, repetitive, and stereotyped patterns of behavior, interests, and activities.[4] All this leads to a deterioration of the individual's social, academic, and work activity.

Most studies of treatments to improve symptoms of AS are aimed at children and adolescents because, according to scientific evidence, an intensive early intervention (between 0 and 6 years) can modify the poor prognosis generally associated with children with AS.[5,6] It should be noted that most of these treatments do not make use of digital technologies, although in recent years interest in technology has increased due to promising results from relevant research projects.[7]

Despite this, Fuentes-Biggi *et al.*[6] emphasize the need to continue supporting people with AS into their adult years, following a continuing education plan. The development of CicerOn is focused precisely on serving as a support tool during this adult stage, helping college students and adults with AS to improve their communication skills.

APPLICATION DESIGN AND DEVELOPMENT

CicerOn is a serious game that allows the user to talk to different avatars in different virtual environments. It is an experience that supports a virtual reality exposure therapy in a user-friendly way due to its gamified design. The different levels of the game provide a hierarchy of increasing exposure to the phobia of public speaking, encouraging motivation to continue the training. An automatic speech recognition system evaluates the speaking of the users, showing feedback on their tasks. This allows for a gradual improvement of their speaking ability.

The application is aimed at users with Asperger's syndrome aged 15 and over. People can use it without supervision, at any time and without having to go to a specialized center. The strong feeling of presence when using VR with an HMD allows users to more easily transfer what they learn within the virtual world to the real world.

Hardware

We evaluated different alternative virtual reality platforms to determine the most appropriate one for our purposes. Since the aim of the project was to expose the user to a systematic desensitization of public speaking, the preferred head-mounted display should have a microphone and a pair of headphones. In the end, we decided to use mobile virtual reality platforms, as they allow us to reach the largest number of users thanks to their lower price and ease of use.

Although we have implemented a multiplatform application, user testing has been done with Samsung Gear VR from Oculus (see Figure 1). In addition, the users have a controller that allows them to move and interact with the virtual environment in an intuitive way, using the joystick for navigation and the buttons for interaction with the objects.

Application Design

CicerOn is a serious game whose goal is to make users speak aloud before different audiences.

During the design process of CicerOn, we carried out several meetings with both psychologists and people with AS. The first meetings were aimed at eliciting the main features to be included when designing a virtual reality application for the target users. Regarding interests and motivations, these include puzzle and riddles, which in the end became the main mechanisms implemented in CicerOn. Other particularities to be taken into account are the typical AS user's

Figure 1. Samsung Gear VR head-mounted display and a Bluetooth controller used for testing.

need for detailed instructions, lack of empathy, inability to understand irony and sarcasm, and a literal understanding of messages and propositions. These characteristics vary greatly from one individual to another, but they are widely present in this group.

This application is structured in six levels. Each level is associated with a higher level of exposure to the audience, implementing an incremental exposure hierarchy. From an initial level of exposure to phobia (public speaking on a stage but without spectators), the complexity of the phobia exposure increases level after level. The increasing rate of difficulty, in terms of exposure to the phobic stimulus, has been designed by psychologists specializing in patients with Asperger.

CicerOn's storyline is based on Egyptian mythology. Figuring out riddles and puzzles, users travel through different countries to find a lost object as if they were explorers. To complete each level, users must read aloud a final text. These texts contribute to the narrative, unveiling some key elements of the game's story.

The mechanics and structure of the game are designed to be repetitive, in order to foster the creation of a routine in players. Sudden changes of scenarios without preview explication are avoided. Routines are made explicit to the players by means of checklists and transition videos. The length of these videos and their appearance are always the same.

Users' main objective in each level is to find all the fragments of a ripped postcard (see Figure 2), hidden all over the level. Each fragment contains a riddle that provides the user with a clue for finding the next one so that the user can finally complete the postcard. The last fragment will take the user near the final reading area, such as a stage or a lectern. Once all the fragments have been found, the complete postcard offers users a longer text to read. The user must read the text of the postcard aloud. The final text contains information on how the game story progresses and what to do in the next level. This text is divided into several blocks or paragraphs. After each block, users will receive the first feedback on their performance. We group

Figure 2. Fragments that compose ripped postcards containing individual clues which when reassembled lead the user to a final text to be read. The aesthetics of the postcards are closely related to the specific city where the action of the game is taking place.

Figure 3. Images of the comfortable base camp for the users with Asperger's syndrome to rest after completing each individual level and see the progress of the game.

the performance scores into three categories (improvable, quite good, and fantastic). Negative messages have been removed from our application and visual feedback using icons complements quantitative information. Once all the blocks have been read, an average score is displayed. To facilitate the reading flow, our application underlines the text while the user is reading.

The automatic speech recognition system included in our application evaluates the user's final reading aloud. It scores the reading according to the volume and correct pronunciation of the words. Several reports show feedback on the user's progress. If players pass a minimum score, the system will move on to the next level. We designed an additional stage (a lounge) with a familiar and comfortable look to be used as a base camp. It is the user's safe place in the virtual world to go when they feel overwhelmed. With this additional environment, we want to prevent the user from removing the VR device under conditions of over-stress due to excessive exposure to the phobic stimulus. In addition, each time users complete a level, our application takes them to this base camp, where they can also check their progress (see Figure 3).

The different levels that make up the application are detailed below. Figure 4 shows some of the scenarios associated with these levels.

Figure 4. Images of scenarios used in the application: a school auditorium in London (upper left and upper right) and a lecture hall in the Louvre museum in Paris (lower left and lower right).

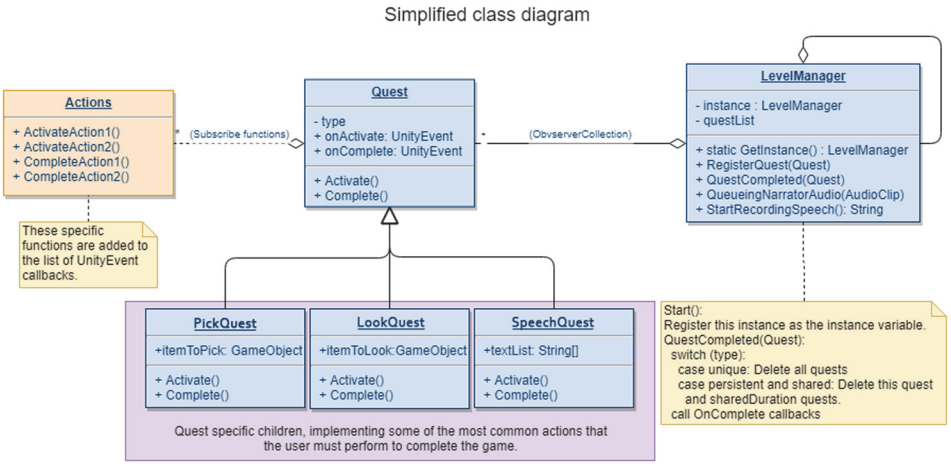


Figure 5. Object-oriented design approach facilitates the creation and reuse of multiple treatment scenarios.

- **Level 0.** Takes players to the base camp in Madrid, where the story is introduced, and the basic navigation and interaction mechanics are explained.
- **Level 1**: Located in a school auditorium in London. The initial final text to be read aloud by users explains the mythological background of the game and the goal to be achieved. The place is empty and there is no audience. A new character is introduced, who will continuously guide the user through the training.
- **Level 2.** A transition video carries the user to Paris. Once all the postcard's fragments have been found, users must read the final text in front of a nonhuman cartoon character.
- **Level 3.** Again, a transition video takes the user back to Madrid. In this case, users have to read the final text in front of three people who are not paying attention to them.
- **Level 4.** Users travel to Alexandria this time. As in the other levels, users have to discover the place where each fragment of the postcard is hidden. At each level, the number of pieces that make up the postcard is different (see Figure 2). In this case, the player makes the final reading before a group of 10 people, who are visibly, but not constantly, paying attention to the user.
- **Level 5**: The final level takes place in the city of Bubastis, where players find the lost object and the game comes to an end. Up to now, the software has increased the difficulty of the training process by modifying the size of the audience and their attitude towards users. However, the highest level of exposure to the phobic stimuli for people with Asperger's syndrome reported by psychologists is reading before a single person, who is paying all their attention to the users. In this level, users become the entire focus of attention of the spectator.

Technical Design

CicerOn is developed with Unity 2017 (unity3d.com), which provides us with a solid game engine foundation. Maya 2017 (www.autodesk.com/products/maya) is used for modeling and animating virtual environments.

Due to the repetitive narrative-action loop of the game, we have generated a system of quests that allows level designers to easily create new levels. This system is implemented by a state machine that will control the execution flow of each scene within the application. The most significant design pattern under this structure is a Singleton programming design pattern, which provides a static referencing of the script using a class called *LevelManager*. To facilitate development, there are several subclasses of class *Quest* that use an Observer design pattern, so the state machine builds itself inside the *LevelManager*. In Figure 5, we provide a class diagram that explains the basic architecture of our application.

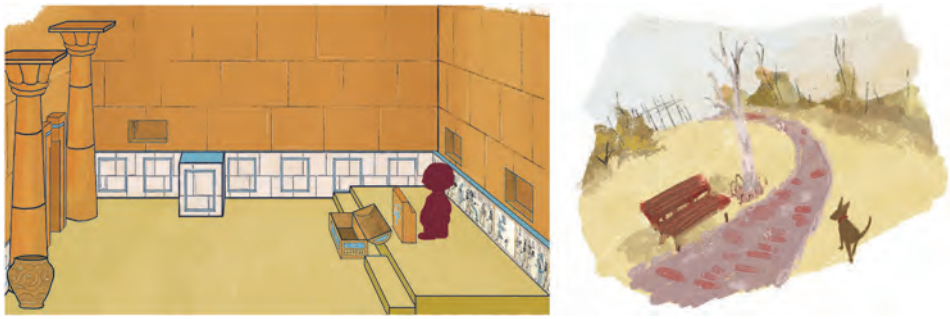


Figure 6. Sketches validated as acceptable to AS patients before construction of the 3-D models.

The developer or level designer, once they include the *Quest* script on the scene, can easily add functionality thanks to the callbacks *ActivateQuest* and *CompleteQuest*, which manage the actions that take place in any given *Quest*. This way, using the Unity editor, designers can easily add new behaviors. When several quests are available at a time, the *LevelManager* is responsible for the activation of quests according to the properties of the latter (e.g., persistence, uniqueness, and shared duration).

The *LevelManager* class also manages narrative resources and speech recognition, which can only be used sequentially. Thus, when a *Quest* asks for the playing of a piece of audio, it is queued. And if users perform an action that requires speech recognition, our easy-to-access Singleton will handle this request.

In the process of integrating an automatic speech recognition system, we encountered a major issue. At first, we tried to use an off-line recognizer called Pocket Sphinx (cmusphinx.github.io), to avoid the need for an Internet connection. Unfortunately, the quality of recognition was not good enough and caused confusion among users. The lowest results were obtained when used by people with high-frequency voices, such as women.

Due to these problems, we decided to use another Android-compatible recognizer called Wit.ai (wit.ai), which provides free speech recognition through its online services. This recognizer provided us with recognition accurate enough for the needs of the application.

Finally, good performance in VR applications is essential to prevent cybersickness in users. A requirement of at least 60 frames per seconds (fps) was determined to be necessary to obtain a smooth motion within the virtual world. Optimizing the number of polygons and the texturing of 3-D models has also been a challenge for CicerOn. For example, we have reduced to 29 856 the number of polygons of level 1, with only 29 draw calls. We achieve a realistic and detailed appearance (see Figures 3 and 4) with a continuous performance above 100 fps due to the different optimization techniques and shaders used.

Perceptual Adapted Design

Adapting functional and visual design to students with AS is essential for better use of the experience. But most of the existing applications have a children's appearance and content, which leads to rejection by adult users with AS. Our application has a cartoon or caricatured appearance but not infantilized look. Avoiding saturated and bright colors, strident shapes or overwhelming environments have been some of the guidelines followed by artists. The sketches and models used in CicerOn have been validated both by people with AS and psychologists (see Figure 6).

When designing game interactions, it is necessary to consider the possible repetitive motor mannerisms of the users. Simple interactions have been designed to improve usability and reduce the learning curve. In order to increase the users' reading-aloud time, in the first design of the application, the interactions with the objects in the game were carried out through voice. According to the first tests conducted with real users, the gameplay was perceived as boring, lowering the users' motivation, what led to a redesign of the application. This new design provides a more dynamic and intuitive interaction through a gamepad.

Figure 7. VR headsets (Samsung Gear VR) enable users to move about freely and practice realistic movements in different scenarios.

Regarding level design, the increase of difficulty during the gameplay is defined by the number and nature of spectators, ranging from none to many. Some variability has to be taken into account here, as most people with AS perceive one-on-one reading as more stressful than reading before several people. Therefore, we have defined the one-to-one reading activity as the last challenge in the game.

EVALUATION

An iterative design evaluation process has been at the heart of CicerOn's development from the very beginning. Following a "with them and for them" philosophy, some members of the development team themselves have Asperger's syndrome. They have contributed to the testing of the experience, looking for performance improvements, locating bugs and proposing adaptations in functionality, script, and appearance to make it more adapted to people with Asperger's syndrome.

In addition, a qualitative evaluation has been conducted by psychologists, pedagogues, teachers, adults with Asperger's syndrome, and their families in each development cycle (see Figure 7). As a result, modifications in the narrative, in the number of levels or in their temporal arrangement and other extra functionalities have been made.

During the design and implementation process, we worked hand-in-hand with different associations specialized in Asperger's syndrome. Among them are the PAUTA Association (Psychopedagogy of Autism and Associated Disorders) and the Asperger Madrid Association. They have also tested the application, giving us very positive and encouraging feedback regarding both the software and its potential to support phobia treatment therapies in people with Asperger's syndrome. The use of VR lowers the entry barriers for users to begin the therapy. The selected VR device (see Figure 1) is easy to handle and is not heavy or uncomfortable. The application's playful design motivated patients and caregivers to continue with the treatment. Patients claimed that, without noticing, they found themselves speaking from a stage.

After the development of CicerOn, Indra Sistemas S.A. and Fundación Universia are confident in the power of virtual reality applications to help overcome other phobias of people with Autism Spectrum Disorder. A video of our system in use is available at youtu.be/zFOkOkoOtUM.

REFERENCES

1. J. A. Ehrlich and J. R. Miller, "A virtual environment for teaching social skills: AViSSS," *IEEE Comput. Graph. Appl.*, vol. 29, no. 4, pp. 10–16, Jul./Aug. 2009.
2. M. R. Kandalaft, N. Didehbani, D. C. Krawczyk, T. T. Allen, and S. B. Chapman, "Virtual reality social cognition training for young adults with high-functioning autism," *J. Autism Dev. Disorders*, vol. 43, no. 1, pp. 34–44, 2013.

3. O. D. Kothgassner *et al.*, "Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment," *Comput. Hum. Behav.*, vol. 62, pp. 124–135, 2016.
4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, (4th ed., text rev.). Washington, DC, USA: American Psychological Assoc., 2000.
5. F. Mulas, G. Ros-Cervera, M. G. Millá, M. C. Etchepareborda, L. Abad, and M. Téllez de Meneses, "Modelos de intervención en niños con autismo," *Rev. Neurol.*, vol. 50, no. 3, pp. 77–84, 2010.
6. J. Fuentes-Biggi *et al.*, "Guía de buena práctica para el tratamiento de los trastornos del espectro autista," *Rev. Neurol.*, vol. 43, no. 7, pp. 425–38, 2006.
7. A. L. Wainer and B. R. Ingersoll, "The use of innovative computer technology for teaching social communication to individuals with autism spectrum disorders," *Res. Autism Spectrum Disorders*, vol. 5, no. 1, pp. 96–107, 2011.

Diego Rojo is a Full-Time Lecturer with Centro Universitario de Tecnología y Arte Digital (U-tad), Madrid, Spain. His research interests include data visualization, human–computer interaction, and machine learning. He received the Master's degree in computer graphics and simulation from U-tad. Contact him at diego.rojo@acm.org.

Jesús Mayor is currently working toward the Ph.D. degree in computer science at Universidad Rey Juan Carlos, Madrid, Spain, with an interest in virtual reality, user experience, computer graphics, and acoustics simulation. He is a Full-Time Lecturer with Centro Universitario de Tecnología y Arte Digital (U-tad), Madrid, Spain. Mayor received the Master's degree in computer graphics and simulation at U-tad. Contact him at j.mayor.m90@gmail.com (www.jesusmayor.com).

José Jesús García Rueda is currently working toward the Ph.D. degree in telematics at Universidad Politécnica de Madrid, Madrid, Spain, with an interest in virtual worlds, hypermedia, and educational technology. He is a Full-Time Lecturer and Researcher with Centro Universitario de Tecnología y Arte Digital, Madrid, Spain. Contact him at jose.rueda@u.tad.com.

Laura Raya is currently working toward the Ph.D. degree in computer science at Universidad Rey Juan Carlos, Madrid, Spain, with an interest in virtual reality, visualization, gamification, and education. She is a Full-Time Associate Professor with Centro Universitario de Tecnología y Arte Digital, Madrid, Spain. Contact her at laura.raya@u-tad.com (gmrv.es/∼lraya).

■ Contact department editor Mike Potel at potel@wildcrest.com.

This article originally appeared in
IEEE Computer Graphics and Applications,
vol. 39, no. 2, 2019.

Designing Systems to Augment Social Interactions

Ionut Damian
University of Augsburg

Elisabeth André
University of Augsburg

Editor:
Albrecht Schmidt
albrecht.schmidt@ifi.lmu.de

Thanks to the rapid advancement of mobile and wearable technologies over the past decade, a new shift in social skills training lies ahead that will allow users to continuously monitor and improve their social behavior during actual social interactions. At the same time, despite the demonstrated effectiveness of social augmentation, researchers must address several limitations.

Social behavior lies at the very core of being human. We engage in social interactions multiple times every day. For example, we speak with friends, buy products from a salesperson, or hold conversations with colleagues. Yet, some types of social interactions—speaking in public, being interviewed, participating in a negotiation, and so on—often seem to be governed by a special set of rules that many of us struggle with. However, these kinds of interactions are the ones in which the outcomes are the most crucial. For example, a job interview can decide a person's employment status, or a presentation in school might have a large impact on a student's final grade. The problem is amplified by the fact that in such critical situations stress can make our bodies go into "auto pilot" mode, rendering our cognitive minds oblivious to our body's use of gestures, postures, or even speech.

Imagine a computer system able to monitor our behavior during social interactions and give us subtle feedback on how to improve. For example, when speaking in public, such a system could help us maintain eye contact with the audience or control our speaking rate by informing us if we are too slow and boring, or are too fast and unintelligible. The same system could also help us make a better impression during job interviews by correcting our body posture and use of gestures. Moreover, persons who suffer from various disabilities, such as autism or Parkinson's disease, could use the system to better regulate their behavior, thus avoiding misunderstandings and generally increasing their functional independence. Such a system would effectively "augment" the user's social skills.

 Published by the IEEE Computer Society

These social augmentation systems can be seen as an extension of Douglas Engelbart's framework for augmenting human intellect1 and the personal augmentation concepts of Cassandra Xia and Pattie Maes.[2] Whereas they focused on problem solving, memory, decision making, motivation and mood, we focus on a different yet equally important domain of human intellect: social behavior.

DESIGN REQUIREMENTS

Since social augmentation systems are meant to be used during real social interactions, a careful design of the system is paramount. To get a grip on this issue, we devised a set of requirements for the design of social augmentation systems. These have been informed by both literature and empirical research.[3]

To support future research in social augmentation, we developed the open source SSJ framework (hcm-lab.de/ssj). It offers the ability to quickly create mobile social augmentation systems using off-the-shelf Android devices and Bluetooth-connected sensors.

First, the user should be able to correctly perceive and process the information delivered by the social augmentation. From a psychological point of view, the augmentation represents a secondary task for the user whereas the social interaction is the primary task. Considering this, the first requirement can be formulated as follows: The social augmentation must be able to momentarily draw enough attention from the primary task to allow information from a social augmentation task to be perceived and processed.

However, it is crucial that the augmentation not draw too much attention lest it distract the user and disrupt the social interaction. According to distributive attention models,[4–5] tasks can be carried out in parallel without quality degradation as long as enough processing resources are available. Thus, to reduce the amount of distraction, the social augmentation needs to be economical with its demand of resources.

Yet, if the social augmentation is to guide the user to a more desirable behavioral state, it must first be able to generally elicit a change in behavior. Thus, the provided information must be understandable, sufficiently detailed, and relevant to both the user and the moment in which it is delivered. Informing a user who, say, is making a presentation that he or she talked too loudly five minutes ago is not only irrelevant but also confusing. This leads us to the third requirement: The provided information must be appropriate for facilitating the intended change in behavior.

So far, the social augmentation can trigger a change in user behavior without much disturbance. What is still missing is the relation to the social interaction mentioned earlier. Specifically, it is important that the augmentation does not just trigger any change in behavior, but one that contributes to the goals of the user in the interaction. For example, in a job interview, the augmentation should help the user make a better impression and thus increase his or her chances of employment.

The social aspect of the augmentation means that the physical form and aspect of the system is also critical. More specifically, the augmentation system should not hinder verbal or nonverbal communication within the social interaction, nor should it break social conventions or otherwise disrupt that interaction. For example, the use of head-mounted displays (HMDs) might prevent perception of the user's gaze signals, interfering with one critical communication channel. Moreover, the augmentation must be mindful of its impact not only on the user, but also on the persons with whom the user is interacting.

Finally, privacy and transparency concerns also need to be addressed. Throughout history, many promising technologies have encountered resistance over such concerns. When the first truly mobile camera, the Kodak box camera, appeared in 1888, it was heavily criticized and even forbidden in certain public places. Thus, the delicate handling of privacy issues is crucial. One approach is to ensure users understand that social augmentation will entail the loss of some privacy. According to Jason Hong,[6] users will accept this if the perceived value of the system

matches or exceeds that of the lost privacy. It is thus paramount that the social augmentation respects the privacy of both users and bystanders.

BEHAVIORAL FEEDBACK LOOP

In previous work,[7,3] we introduced the behavioral feedback loop as the driving force behind social augmentation. In simple terms, a feedback loop occurs when the output of a system is repeatedly and continuously fed back to the system as input, thus forming a closed loop. From the point of view of social augmentation, feedback loops are particularly interesting due to their self-regulating nature. Thus, the goal of generating self-awareness discussed earlier can be directly translated to a feedback loop structure. The user's behavior (output) is recorded and fed back to the user (input) continuously, generating awareness of one's own behavior. For this, physical artifacts, such as miniaturized sensors or lightweight displays, are used to both perceive the user's behavior and deliver real-time feedback. Now, through intelligent and goal-oriented manipulation of the feedback loop, the user's behavior can be steered toward a more beneficial state for the social interaction. Figure 1 illustrates the resulting two-step pipeline of a behavioral feedback loop: The user's behavior is first analyzed in real time and then, based on its quality, feedback is automatically generated and delivered to the user.

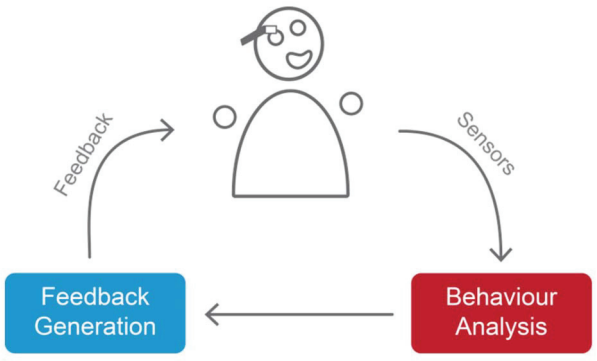


Figure 1.The behavioral feedback loop. The user's behavior is first analyzed in real time and then, based on its quality, feedback is automatically generated and delivered to the user.

EXAMPLES

Social augmentation systems can be deployed in different scenarios to help users improve the outcomes of their interactions or overcome certain disabilities. Here we provide three concrete use cases for social augmentation.

Augmenting Public Speaking

Public speaking is a distinct type of social interaction that requires speakers to deliver an informative message to their audience while entertaining and inspiring enthusiasm at the same time. This makes speaking in public a particularly stressful experience. The Logue[8] social augmentation system, shown in Figure 2, attempts to relieve some of this pressure by delivering direct and objective feedback on the quality of one's speech rate, body energy and openness, as well as providing instructions on how to improve it. To achieve this, social-signal processing techniques are employed to analyze the speaker's performance using data from a microphone and a depth camera. Based on this analysis, feedback is generated and delivered to the user unobtrusively in real time using an HMD.

Figure 2. The Logue social augmentation system delivers live visual feedback to a person speaking in public via a head-mounted display.

Augmenting Group Discussions

Unlike speaking in public, where conversation is mostly one sided and exchanges between speaker and audience are limited, traditional face-to-face interactions are more susceptible to disturbances as they contain complex exchanges of verbal and nonverbal messages that are governed by a delicate set of unwritten rules. We designed an augmentation system that attempts to overcome this problem by providing feedback using different modalities. The system helps the user control their speaking time during group discussions by providing auditory, tactile, or visual feedback in real time.[7]

Augmenting the Speech of Adults with Disabilities

One common dysfunction associated with autism is atypical prosody.[9] It impacts the rate, loudness, and pitch of the speaker's voice, potentially leading to misunderstandings during social interactions. Louanne Boyd and her team[10] developed a social augmentation system that targets adults with autism. Using Google Glass, their system provides visual feedback whenever the user exhibits an atypical vocal pitch or loudness.

A similar approach was proposed by Roisin McNaney and her colleagues[11] for helping people with Parkinson's, who often "have an impaired perception of how loud they are speaking," better regulate the loudness of their voice. Their LApp system uses Google Glass to continuously monitor the loudness of the user's voice and gives visual feedback whenever it drops below a predefined threshold.

CONCLUSION

Social skills training has evolved over the past few decades from using manual and analog forms of knowledge transfer to intelligent virtual simulation environments, which can automatically react and adapt to the learner. Now, thanks to the rapid advancement of mobile and wearable technologies over the past decade, a new shift in social skills training lies ahead that will allow users to continuously monitor and improve their social behavior during actual social interactions.

Despite the demonstrated effectiveness of social augmentation, researchers must address several limitations. As with all types of augmentation, there is a risk of the user learning to rely on the system too much. This could make the user dependent on the system and thus unable to act without it. Similarly, the user might focus too much on "pleasing" the system and ignore the real world—for example, cues from interlocutors. This could result in abnormal behavior because social augmentation is meant to help users improve their current social behavior, not be a substitute for it. Nevertheless, intelligent feedback design and advanced behavior analysis routines could be used proactively to minimize such risks.

REFERENCES

1. D.C. Engelbart, *Augmenting Human Intellect: A Conceptual Framework*, government report AFOSR-3223, Stanford Research Institute, 1962; www.dougengelbart.org/pubs/papers/scanned/Doug_Engelbart-AugmentingHumanIntellect.pdf.
2. C. Xia and P. Maes, "The Design of Artifacts for Augmenting Intellect," *Proc. 4th Augmented Human Int'l Conf.* (AH), 2013, pp. 154–161.
3. I. Damian, *Social Augmentation Using Behavioural Feedback Loops*, thesis, Universität Augsburg, 2017; opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/37750.
4. D. Navon, "Resources—A Theoretical Soup Stone?," *Psychological Rev.*, vol. 91, no. 2, 1984, pp. 216–234.
5. C.D. Wickens, "Multiple Resources and Performance Prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, 2002, pp. 159–177.
6. J. Hong, "Considering Privacy Issues in the Context of Google Glass," *ACM Comm.*, w3c_recommendation, vol. 56, no. 11, 2013, pp. 10–11.
7. I. Damian, T. Baur, and E Andre, "Measuring the Impact of Behavioural Feedback Loops on Social Interactions," *Proc. 18th ACM Int'l Conf. Multimodal Interaction* (ICMI), 2016, pp. 201–208.
8. I. Damian et al., "Augmenting Social Interactions: Realtime Behavioural Feedback using Social Signal Processing Techniques," *Proc. 33rd Ann. ACM Conf. Human Factors in Computing Systems* (CHI), 2015, pp. 565–574.
9. L. Kanner, "Autistic Disturbances of Affective Contact," *Acta Paedopsychiatrica*, vol. 35, no. 4, 1968, pp. 100–136.
10. L.E. Boyd et al., "SayWAT: Augmenting Face-to-Face Conversations for Adults with Autism," *Proc. 2016 CHI Conf. Human Factors in Computing Systems* (CHI), 2016, pp. 4872–4883.
11. R. McNaney et al., "LApp: A Speech Loudness Application for People with Parkinson's on Google Glass," *Proc. 33rd Ann. ACM Conf. Human Factors in Computing Systems* (CHI), 2015, pp. 497–500.

ABOUT THE AUTHORS

Ionut Damian is a postdoctoral researcher at the Institute of Computer Science, University of Augsburg. Contact him at damian@hcm-lab.de.

Elisabeth André is a full professor of computer science and chair of Human-Centered Multimedia in the Faculty of Applied Informatics at the Institute of Computer Science, University of Augsburg. Contact her at andre@hcm-lab.de.

This article originally appeared in
IEEE Pervasive Computing, *vol. 17, no. 1, 2018.*

IEEE COMPUTER SOCIETY

PURPOSE: The IEEE Computer Society is the world's largest association of computing professionals and is the leading provider of technical information in the field.

MEMBERSHIP: Members receive the monthly magazine *Computer*, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.

COMPUTER SOCIETY WEBSITE: www.computer.org

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 and 18 journals. Refer to membership application or request information as noted above.

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

Conferences/Education: The society holds about 200 conferences each year and sponsors many educational activities, including computing science accreditation.

Certifications: The society offers three software developer credentials. For more information, visit www.computer.org/certification.

BOARD OF GOVERNORS MEETING

28 – 29 May: McLean, Virginia

EXECUTIVE COMMITTEE

President: Leila De Floriani
President-Elect: Forrest Shull
Past President: Cecilia Metra
First VP: Riccardo Mariani; **Second VP:** Sy-Yen Kuo
Secretary: Dimitrios Serpanos; **Treasurer:** David Lomet
VP, Membership & Geographic Activities: Yervant Zorian
VP, Professional & Educational Activities: Sy-Yen Kuo
VP, Publications: Fabrizio Lombardi
VP, Standards Activities: Riccardo Mariani
VP, Technical & Conference Activities: William D. Gropp
2019–2020 IEEE Division VIII Director: Elizabeth L. Burd
2020-2021 IEEE Division V Director: Thomas M. Conte
2020 IEEE Division VIII Director-Elect: Christina M. Schober

BOARD OF GOVERNORS

Term Expiring 2020: Andy T. Chen, John D. Johnson, Sy-Yen Kuo, David Lomet, Dimitrios Serpanos, Hayato Yamana
Term Expiring 2021: M. Brian Blake, Fred Douglis, Carlos E. Jimenez-Gomez, Ramalatha Marimuthu, Erik Jan Marinissen, Kunio Uchiyama
Term Expiring 2022: Nils Aschenbruck, Ernesto Cuadros-Vargas, David S. Ebert, William Gropp, Grace Lewis, Stefano Zanero

EXECUTIVE STAFF

Executive Director: Melissa A. Russell
Director, Governance & Associate Executive Director: Anne Marie Kelly
Director, Finance & Accounting: Sunny Hwang
Director, Information Technology & Services: Sumit Kacker
Director, Marketing & Sales: Michelle Tubb
Director, Membership Development: Eric Berkowitz

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: +1 800 678 4333; Fax: +1 714 821 4641;
Email: help@computer.org

IEEE BOARD OF DIRECTORS

President: Toshio Fukuda
President-Elect: Susan K. "Kathy" Land
Past President: José M.F. Moura
Secretary: Kathleen A. Kramer
Treasurer: Joseph V. Lillie
Director & President, IEEE-USA: Jim Conrad
Director & President, Standards Association: Robert S. Fish
Director & VP, Educational Activities: Stephen Phillips
Director & VP, Membership & Geographic Activities: Kukjin Chun
Director & VP, Publication Services & Products: Tapan Sarkar
Director & VP, Technical Activities: Kazuhiro Kosuge

IEEE

revised 14 January 2020

Column: Micro Economics

Earning Stripes in Medical Machine Learning

Shane Greenstein
Harvard Business School

Today, we are living through one of those heady situations in which scientific, technical, and commercial frontiers all simultaneously advance in a grand interrelated dance. Advances in computer technology in the last decade opened up the potential for big gains in applications of neural networks aimed at recognizing and diagnosing visual images. Many startups and established firms are making decisions about how to develop and deploy such software, and what products to develop next.

Advances in computer technology in the last decade opened up the potential for big gains in applications of neural networks aimed at recognizing and diagnosing visual images. Many startups and established firms are making decisions about how to develop and deploy such software, and what products to develop next.

While the science continues to expand into some headline popping territory, and the technologist find new prototypes for mind-altering forecasting, what has happened within markets? What has found its way into practical applications with commercial appeal?

We can start with a simple observation: If real people are spending real money on an application of machine learning, then it must have created new economic value. However, it can be difficult to provide an overview of a broad moving target.

Today's column goes inside the actions of one firm's experience as a way to provide insight into the broad trends. It will focus on the factors shaping the creation of value from applying machine learning in medical imaging. The focus will be on a firm called Zebra Medical Vision, an Israeli startup in the business of using machine learning to improve X-rays. Its experience illustrates many of the same technology trends and strategic issues faced by other machine learning startups, particularly those in healthcare markets.[1]

SETTING

You probably have not heard of Zebra, even though it is doing well by the norms of its field. It is a startup competing with other startups and major technology companies. Zebra's product is a series of software tools that use machine learning algorithms to read radiology scans for signs of different medical conditions.

Zebra is best known in the industry for charging $1 for each scan. The founders hoped the simple pricing would make it easier for Zebra's algorithm to be used more widely. The $1-per-scan has now become focal for all firms. Of course, they have accomplished more than merely set a pricing standard, and that is what will receive attention.

Digital Object Identifier 10.1109/MM.2019.2932278
Date of current version 10 September 2019.

For some years now, Zebra has faced a situation common to many startups as they expand. Zebra has access to a cutting-edge technology, but they have had to decide how to apply it to a promising commercial applications. More concretely, Zebra has had to simultaneously find training data, hire in competitive labor markets, and negotiate FDA approvals and other government regulations.

Zebra is located in the entrepreneurial culture of Herzliya, a suburb of Tel Aviv. For its goals, that location has both advantage and disadvantage.

Here is the upside. The area benefits from nearby universities and a local start-up culture, as well as research offices for large technology companies including Apple and Microsoft.

Israel's size is also an advantage, believe it or not. Its population is smaller than Sweden and bigger than Denmark. That made it easier to make a deal for data. Zebra has access to anonymized healthcare data through Israel's taxpayer-funded government healthcare system. It is easier to make that sort of deal in a smaller country, so Zebra got an earlier start than most other firms.

Specifically, there are four Israeli HMOs, and Zebra made an arrangement to use one HMO's data. Most people in Israel remain in the same HMO throughout their life. Because all health data are digitized and each patient is tracked using an anonymized patient identification number, health information is organized in such a way that it can provide accessible, anonymized medical data to inquiring parties. After coming to terms, Zebra gained access to tens of thousands of X-ray and CT scans.

What is the disadvantage of that location? For one, the tight labor market for talent. In its current position, Zebra has found itself competing with Google, Apple, Microsoft, and startups to hire new machine learning PhDs, who are looking for competitive salaries and interesting, cutting-edge work. To be sure, it has found some. Founded in December 2013, Zebra employed approximately 40 people in 2018, including its three founders, and it has been growing since then.

The second disadvantage is distance from big markets, which are in the United States, Europe, and India. The founders spend significant time travelling to meet with their partners in hospitals across the globe.

Perhaps most interesting, location does not put Zebra at a technical disadvantage. It uses recent breakthroughs in data storage and computing power by employing Google's Tensor Flow. That makes data-heavy analysis possible anywhere in the world with access to major internet lines and the technical talent to develop the software. In that sense, Zebra occupies the same technical playing field as everyone else.

TECHNOLOGY

What precisely does Zebra's algorithm do? It relies on images from X-rays and CT scans, which are both radiation-based images that create a view of the internal parts of a body, with shades of gray depicting muscles and fat, and white showing bones and metal. X-rays are the single images from a scan, and CT scans are a series of X-ray scans that together make up a three-dimensional image.

Since the 2000s, these images have been increasingly digitized, which allowed for faster, more efficient, and cheaper methods of analysis. The digital copies allowed researchers to use collections of the scans more easily for research purposes, especially for machine learning purposes. Zebra builds on this digital foundation and the research into how to use it for algorithm development.

To date, Zebra's algorithms use large amounts of data within outlined parameters, i.e., in a framework of *supervised* learning. Accordingly, each data set addresses one application at a time—i.e., spinal fractures, brain bleeds, lung congestion, and so on. Data are analyzed by a computer program for recognizable patterns, which the program could then recognize in other data sets, or refine when the original set has more data added in.

In Zebra's case, a series of images have been preanalyzed, with the sections of the images that showed if there was a potential problem marked and annotated for the computer to learn from. All scanned X-rays for training data must be read by multiple doctors in order to give the software reliable data from which to learn. Scans that are

performed with out-of-date equipment, or by technicians who are not trained on the equipment, can lead to difficult-to-analyze results.

That said, X-rays, while a valuable medical tool, are still often unclear or misread. Hence, radiology is not a perfect science. Related, no machine learning algorithm will be perfect.

So what does "great but imperfect" software do? In common parlance, great software reduces "false positives" and "false negatives." A great test can find nearly all positive diagnoses of a condition. A great test also correctly identify when a person does not have a condition.

A key aspect of Zebra's strategy, unlike many other healthcare machine learning developers, is to work closely with healthcare providers to develop its software. Zebra positions its software as a tool that can help radiologists, not replace them.

Notably, a key aspect of Zebra's strategy, unlike many other healthcare machine learning developers, is to work closely with healthcare providers to develop its software. Zebra positions its software as a tool that can help radiologists, not replace them. The software for detecting bone fractures, for example, came from conversations with radiologists who explained that a tool could help them screen the elderly with high risk. More broadly, they have learned from that experience that automated diagnosis creates value when the software is a tool for screening, or an instrument to aid decision making triage in urgent care. At some point, it may also become a useful tool for "second opinions" or "finding clues humans missed, and developing new concerns to explore."

Zebra still faces the same dilemmas of this business as everyone else: whether Zebra should focus its development on products for the developing world (at first China and India) or focus on the United States and Europe. In China and India, Zebra's diagnostic software can help with medical care in areas that have X-ray and CT scanners, but often not enough trained providers to accurately read the results. These countries also have shorter timelines for government approval, which makes it easier for Zebra to go to market. Conversely, in the United States and Europe, Zebra's software will be positioned differently and potentially will be used on every appropriate scan in a busy hospital, allowing Zebra to charge more overall and gather more data to improve its accuracy.

After exploring the potential of all options, to date Zebra has tended toward commercializing in Europe and the United States, enduring the regulatory scrutiny of the European Commission and the Federal Drug Administration. Several of its products have already made it through the FDA approval process. By the norms of the field that makes Zebra a leader in commercial applications.

COMMERCIALIZATION

Commercialization is the act of translating technical knowledge into valuable products and services. Zebra has focused most of its energies around commercialization, namely, product development, and the development of efficient processes for developing algorithms for specific diagnoses.

Today they are deploying those processes in specific medical settings. The central issue for management is "which diagnoses should gain their priority and why?

Think of Zebra's experience as a good barometer of progress in commercialization of machine learning for medical applications. Despite all the excitement, the conclusion is irrefutable: While plenty of value could be created, we are just at the beginning of deployment into mainstream medical practice.

REFERENCE

1. S. Greenstein and S. Gulick, *Zebra Medical Vision*. Harvard Business School Case, Nov. 2018, Art. no. 619053.

Shane Greenstein is a professor at the Harvard Business School. Contact him at sgreenstein@hbs.edu.

This article originally appeared in IEEE Micro, *vol. 39, no. 5, 2019.*

ADVERTISER INFORMATION

Advertising Coordinator

Debbie Sims
Email: dsims@computer.org
Phone: +1 714-816-2138 | Fax: +1 714-821-4010

Advertising Sales Contacts

Mid-Atlantic US:
Dawn Scoda
Email: dscoda@computer.org
Phone: +1 732-772-0160
Cell: +1 732-685-6068 | Fax: +1 732-772-0164

Southwest US, California:
Mike Hughes
Email: mikehughes@computer.org
Cell: +1 805-208-5882

Northeast, Europe, the Middle East and Africa:
David Schissler
Email: d.schissler@computer.org
Phone: +1 508-394-4026

Central US, Northwest US, Southeast US, Asia/Pacific:
Eric Kincaid
Email: e.kincaid@computer.org
Phone: +1 214-553-8513 | Fax: +1 888-886-8599
Cell: +1 214-673-3742

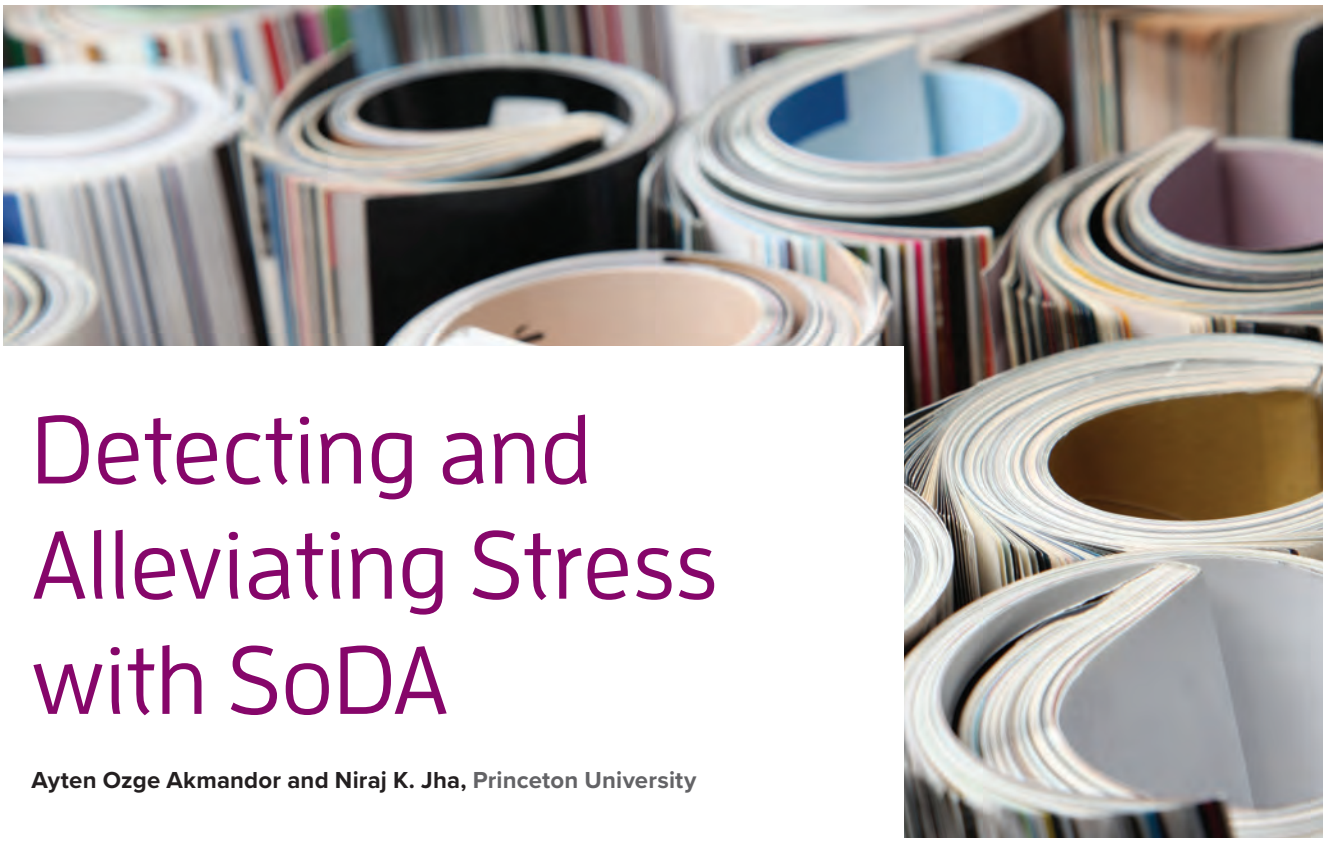
Midwest US:
Dave Jones
Email: djones@computer.org
Phone: +1 708-442-5633 Fax: +1 888-886-8599
Cell: +1 708-624-9901

Jobs Board (West Coast and Asia), Classified Line Ads

Heather Bounadies
Email: hbuonadies@computer.org
Phone: +1 623-233-6575

Jobs Board (East Coast and Europe), SE Radio Podcast

Marie Thompson
Email: marie.thompson@computer.org
Phone: +1 714-813-5094



Detecting and Alleviating Stress with SoDA

Ayten Ozge Akmandor and Niraj K. Jha, Princeton University

This installment of Computer's *series highlighting the work published in IEEE Computer Society journals comes from* IEEE Transactions on Multi-Scale Computing Systems.

How wonderful would it be if our body-worn accessories, such as wearable medical sensors (WMSs), were able to track our health in a round-the-clock fashion? Breakthroughs in the field of machine learning and rapid advancements in sensor technologies hold great promise to actualize this dream through automation of decision-making processes.

For example, in "Keep the Stress Away with SoDA: Stress Detection and Alleviation System" (*IEEE Trans. Multi-Scale Computing Systems*, vol. 3, no. 4, 2017, pp. 269–282), Ayten Ozge Akmandor and Niraj K. Jha use machine learning and WMSs to implement a continuous, user-friendly stress coach.

As shown in Figure 1, Akmandor and Jha's proposed system—called SoDA—collects physiological signals through five WMSs: electrocardiogram, galvanic skin response, respiration rate, blood pressure, and blood oximeter. SoDA processes the data by eliminating unwanted noise, outliers, and interpersonal variations. Then, SoDA extracts features for each section of the processed data and inputs them into previously trained machine-learning models. These models determine whether the user is stressed, and activate the stress alleviation protocol if needed.

The stress alleviation protocol starts with the most effective stress therapy (determined during the training stage) and tracks the physiological signals. If the physiological signals indicate that the user is benefiting from the therapy, SoDA proceeds with it for a predefined time period; otherwise, it switches to the next therapy and repeats the steps. SoDA continuously collects, processes, and analyzes the data. Because it is awake 24/7 or as long as the

 Published by the IEEE Computer Society

user wears the sensors, SoDA has the potential to take action at the onset of a stressful situation and help the user circumvent its negative effects.

Depending on the user's expectations, current condition, and available resources, SoDA offers two modes: *individualized* and *generalized*. The individualized mode utilizes only the corresponding user's data to build the machine-learning model. Because the model requires the user's data, the user needs to be available for data collection for approximately two hours. In the generalized mode, the machine-learning model is trained with previously collected data from a large group of individuals. Because the individualized mode is personalized to the user, it exhibits a higher classification accuracy. The generalized mode eliminates the need for personalized data collection, at the expense of reduced classification performance. Akmandor and Jha carried out analyses on 32 individuals under both modes. They obtained a stress detection accuracy of 95.8 and 89.3 percent, respectively, for the individualized and generalized modes. Moreover, the authors also tested the effects of various stress alleviation therapies (for example, micromeditation, good news, and warm stone) and verified their efficacy by comparing them with a control case that did not include stress alleviation therapy.

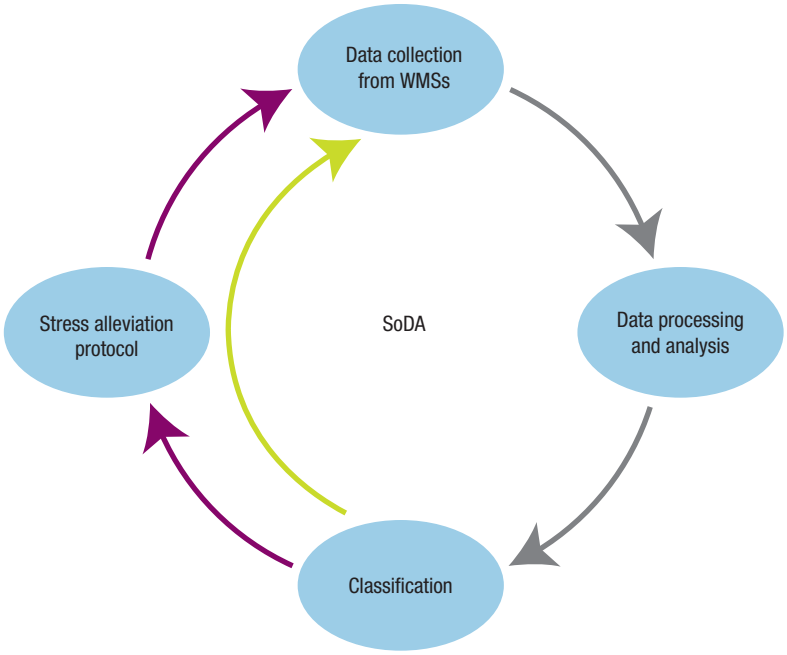


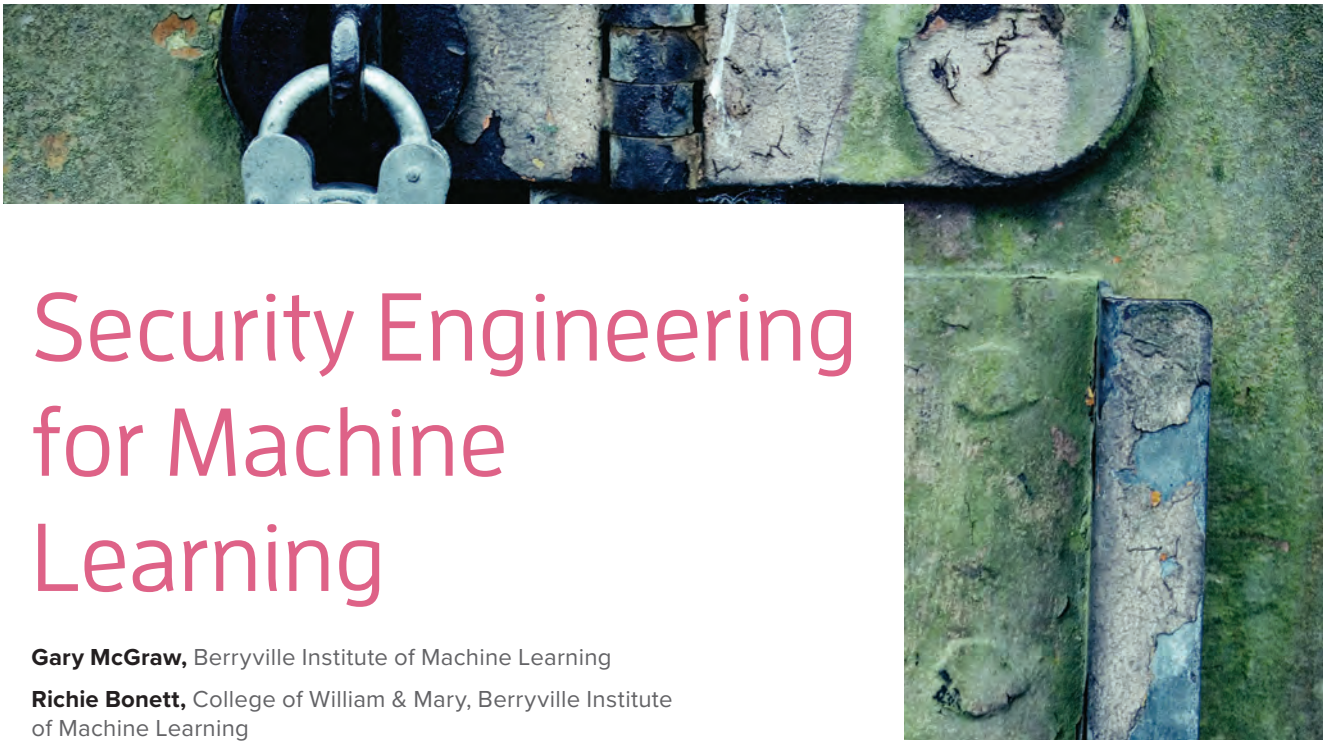
Figure 1. The SoDA stress detection and alleviation system. WMS is wearable medical sensor.

SoDA provides end-to-end stress coaching by not only detecting stress but also suggesting therapies, tracking physiological signals, and modifying the therapy accordingly when needed. Although SoDA exhibits high performance and real-time response and offers a user-friendly setup and multiple stress therapies, the authors state the need for longer experimental duration, integration with more WMSs, a larger experimental population, and an increased set of stress therapy options. SoDA's current implementation and the stated future steps open up the opportunity to tackle other serious health conditions.

AYTEN OZGE AKMANDOR is a third-year graduate student in electrical engineering at Princeton University. Contact her at akmandor@princeton.edu.

NIRAJ K. JHA is a professor of electrical engineering at Princeton University. Contact him at jha@princeton.edu.

This article originally appeared in
Computer, *vol. 51, no. 7, 2018.*



Security Engineering for Machine Learning

Gary McGraw, Berryville Institute of Machine Learning

Richie Bonett, College of William & Mary, Berryville Institute of Machine Learning

Harold Figueroa and Victor Shepardson, Ntrepid, Berryville Institute of Machine Learning

Artificial intelligence is in the midst of a popular resurgence in the guise of machine learning (ML). Neural networks and deep learning architectures have been shown empirically to solve many real-world problems. We ask what kinds of risks ML systems pose in terms of security engineering and software security.

Machine learning (ML) appears to have made impressive progress on many tasks, including image classification, machine translation, autonomous vehicle control, and playing complex games, such as chess, Go, and Atari video games. This has led to much breathless popular-press coverage of artificial intelligence and elevated deep learning to an almost magical status in the eyes of the public. ML, especially of the deep-learning sort, is not magic, however. It is simply sophisticated associative-learning technology based on algorithms developed over the past 30 years. In fact, much of the recent progress in the field can be attributed to faster CPUs and much larger data sets rather than to any particular scientific breakthrough.[1]

ML has become so popular that its application, although often poorly understood and partially motivated by hype, is exploding. In our view, this is not necessarily a good thing. We are concerned with the systematic risk invoked by adopting ML in a haphazard fashion. Our research is focused on understanding and categorizing security-engineering risks introduced by ML at the design level.

While the idea of addressing the security risk in ML is not a new one, most previous work has focused on either particular attacks against running ML systems (a kind of

Digital Object Identifier 10.1109/MC.2019.2909955
Date of publication: 30 July 2019

 Published by the IEEE Computer Society

dynamic analysis) or on operational security issues surrounding ML. Just for the record, we encourage these lines of inquiry.

Our research focuses on three threads: building a taxonomy of known attacks on ML, exploring a hypothesis of representation and ML risk, and performing an architectural risk analysis (sometimes called a threat model) of ML systems in general. We report our progress here.

A TAXONOMY OF ML ATTACKS

Attack taxonomies in security have a long history.[2] One of the motivations behind building such a taxonomy is to guide engineering tradeoffs made at the design level using real-world data about how fielded systems are attacked. For that reason, we are building a taxonomy of ML attacks.

In practice, fielded ML systems as targets run the gamut from white box, which are fully open source and trained on public data, to black box, which map inputs to outputs via an application programming interface to an unknown transformation function. Between the two extremes lie many other possibilities including ML systems based on an open-source model with proprietary hyperparameters and training data and a black-box model that leverages transfer learning from an existing white-box model.[3]

Attacks on ML systems can be divided into two types: manipulation attacks, which alter system behavior by tweaking input, training data, or the model itself, and extraction attacks, which surreptitiously discern secret information in the ML system. Additionally, attacks can be classified by which part of the system they target (input, training data, and model). This results in a taxonomy of six categories as shown in Table 1.

Input-manipulation attacks (also known as *adversarial examples* and *evasion attacks*) are by far the most common kind of ML attack discussed in the literature. The attacker creates an input to an operating ML system that reliably produces a different output than its creators intend. Successful attacks include stop-sign misclassification, spam misidentification, and broken language processing.[4]

Training-data manipulation attacks (also known as *poisoning* and *causative attacks*) are attacks on an operating model via the training process. The attacker modifies the data corpus used to train ML systems, with the intent of impairing or influencing future system behavior. For example, an attacker may publish bogus data to interfere with medical diagnoses or influence financial time-series forecasting models.[5] In the infamous case of Microsoft Research's Tay, Internet trolls successfully implemented a data-manipulation attack to turn the chatbot into a bigot.

There are few examples of model-manipulation attacks in the literature. However, one can imagine an attacker publishing a white-box model with certain latent behavior that is meant to be unwittingly adopted by third parties and later exploited by the attacker. Given the increasing adoption of transfer learning and the fact that releasing code, and even model parameters, under a permissive open-source license is common in ML, we believe this attack category deserves attention.

Input-extraction attacks (also known as *model inversion*) apply in cases where model output is public but inputs are supposed to remain secret. In this case, an attacker, given outputs, attempts to recover inputs. Attacks include inferring features of medical records from the dosage recommended by an ML model and producing a recognizable image of a face given only the classification and confidence score in a face-recognition model.[6]

Training-data extraction attacks (also called *model inversion*) involve extracting details of the data corpus that an ML model was trained on.[7] ML research focuses much of its attention on the learning model to the exclusion of attention on data, yet data are clearly known to be crucially important to a trained system's behavior. Real-world ML systems often incorporate proprietary data and data with serious privacy implications.

Model-extraction attacks target any less-than-fully white-box ML system and attempt to open the box and copy the target's behavior or parameters. Examples include theft of a proprietary model and enabling white-box attacks on what was designed to be a black-box model.[8]

Work on this taxonomy is ongoing. (In the interest of space, we have not included as many references as we would like in this section. See Berryville Institute of Machine Learning for more information: https://berryvilleiml.com/references/.)

A WORKING HYPOTHESIS ON REPRESENTATION

Our work is informed by a hypothesis about representation in ML systems that we are actively exploring. Control over input, output, and hidden representations is essential to understanding the attacks we described in the preceding section.

ML systems are conventionally evaluated on a held-out test set drawn from the same distribution as the training data. This prevents overfitting to specific examples in the training data but guarantees nothing about

TABLE 1. The six attack categories.

Input manipulation	Input extraction
Training data manipulation	Training data extraction
Model manipulation	Model extraction

generalization to a different data distribution in production. Input-manipulation attacks exploit precisely this weakness by targeting a region of input space in which system behavior is not understood. Similarly, data-manipulation attacks mold the training distribution to an attacker's intent. In an adversarial setting, we must understand ML representations over the entire potential input space, not just the training-data distribution. Representations that are unstable and corruptible can be easily (and often undetectably) tampered with. Improved representation strategies can lead directly to more secure ML systems.

Better representation approaches may also lead to more robust operation in challenging contexts well beyond avoiding adversarial dynamic activity. System robustness in the face of both limited and very noisy data can protect against catastrophic failure, especially when ML systems are applied to situations that stray beyond their training.

These ideas are not new. In our view, basic principles in representation have been discovered multiple times in multiple disciplines and published under multiple names. For example, in the numerical computation and statistical communities, phenomena such as ill-conditioning, collinearity, and outliers have long been described and are well understood. Their detrimental effects on computation and estimation are modeled through concepts such as condition numbers and statistical leverage and mitigated through techniques for regularization and outlier detection.

Our view is that an overfocus on pure learning strategies without regard to representational fluidity may be accidentally adding risk to current ML systems. We would like to take advantage of the progress that exists in various adjacent fields to explore representation issues that can improve ML systematically (mostly from a security perspective). Increased attention to representation can help in two ways: achieving more stable and efficient signal-content representations and supporting the complementary concern of modeling signal typicality.

EXAMPLE: ANOMALY DETECTION IN TRAINING DATA

Anomaly-detection ideas can be directly applied to input data based on some measure of the data's typicality during both ML training and operations. During training, such an approach can protect against anomalous input with high leverage that may poison the model. Anomaly detection in input can also be applied during operation to assess the typicality of the test input against the training data and offer a model-independent way of determining whether the ML system is likely to perform as expected. In both cases, anomaly scores that describe observed data drift can give us an indication of when we're interpolating and when we're extrapolating.

EXAMPLE: DEFENSIVE INPUT TRANSFORMATION

Input transformation can be used to defend against some kinds of ML attacks, especially in the input-manipulation category. There is often a great deal of extraneous variation (for example, nonrelevant variation with respect to the ML system tasks) found in the raw input to an ML system. As a result, the ML system is likely to include some of this extraneous information in its learned hidden representations. In some sense, the bad extra information becomes entangled with the good.

Because of this, the ML system can become susceptible to bad-extra-information-based attacks. As an example, just because an image is slightly noisy, an ML recognition system should not make a silly categorization error (for example, turtle → rifle or stop sign → speed limit sign). Well-known input-manipulation attacks do exactly this with low-level noise, relying on entanglement of the noise signal with the task-relevant signal in the distributed/learned representation being built and used by the ML system.

This is not a new phenomenon. In linear-inversion problems, such as image deblurring, a numerically rank-deficient, ill-conditioned operator cannot be inverted in the presence of noise without careful consideration of the representation implicit in the process inversion. Information from subspaces associated with small singular values must be attenuated or discarded altogether. ML systems should take advantage of this knowledge.

WILD SPECULATION

Evolved sensory systems found in nature do this kind of attenuation and discarding thing all the time [think of the bandwidth limitations in human hearing (hertz) and vision (nanometers), for example]. Raw input in biological systems is limited in a task-opportunity and risk-dependent way. The auditory and visual systems of different mammal, bird, and insect species have all evolved to reflect niche opportunities and risks (and are all divergent from each other in numerous ways; bandwidth is an easy one to observe).

In our view, the adaptations displayed by these systems are neither completely reliant on nor entirely gleaned through Hebbian learning but, rather, implemented in aspects of the anatomy and physiology of various organisms that were established through genomic evolution. As we experiment with learning systems, we should use a variety of learning algorithms, some of which may be able to achieve different kinds of search and increase robustness by introducing different types of error and nonlinearity.

TOWARD A THOROUGH ARCHITECTURAL RISK ANALYSIS OF ML

We are interested in building security into ML systems from a security-engineering perspective. This means understanding how ML systems are designed for security (including what representations they use), teasing out possible engineering tradeoffs, and making such tradeoffs explicit. We are

This article originally appeared in Computer, *vol. 52, no. 8, 2019.*

also interested in the impact of including an ML system as a component in a larger design. Our basic motivating question is how do we secure ML systems proactively while we are designing and building them?

Early work in security and privacy in ML has taken an operations-security tack focused on securing an existing ML system and maintaining its data integrity. For example, Nicolas Papernot uses Salzter and Schroeder's famous security principles to provide an operational perspective on ML security.[8] In our view, this article does not go far enough into ML design to satisfy our goals. A key objective of our work is to develop a basic architectural risk analysis (sometimes called a *threat model*) of a typical ML system.[9] Our analysis will take into account common design flaws, such as those described by the IEEE Center for Secure Design.[10]

Securing a modern ML system must involve diving into the engineering and design of the ML system itself. Our work sets out a taxonomy of known attacks against existing ML systems, describes a hypothesis of representation that may help make ML systems more secure, and hints toward a more complete architectural risk analysis of ML.

REFERENCES

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015. doi: 10.1038/nature14539.
2. G. McGraw and G. Hoglund, *Exploiting Software*. Reading, MA: Addison-Wesley, 2004.
3. B. Wang, Y. Yao, B. Viswanath, H. Zheng, and B. Y. Zhao, "With great training comes great vulnerability: Practical attacks against transfer learning," in *Proc. 27th USENIX Security Symp.*, 2018, pp. 1281–1297.
4. X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, 2019, pp. 1–20.
5. S. Alfeld, X. Zhu, and P. Barford, "Data poisoning attacks against autoregressive models," in *Proc. 30th AAAI Conf. Artificial Intelligence*, 2016.
6. M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Computer Communications Security*, 2015, pp. 1322–1333.
7. R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proc. 2017 IEEE Symp. Security Privacy*, 2017, pp. 3–18.
8. N. Papernot, "A marauder's map of security and privacy in machine learning," presented at the 11th ACM Workshop Artificial Intelligence and Security With 25th ACM Conf. Computer and Communications Security, Toronto, Canada, Oct. 19, 2018.
9. G. McGraw, *Software Security*. Reading, MA: Addison-Wesley, 2006.
10. I. Arce et al., "Avoiding the top 10 software security design flaws," IEEE Center for Secure Design, Nov. 13, 2015. [Online]. Available: https://cybersecurity.ieee.org/blog/2015/11/13/avoiding-the-top-10-security-flaws/

GARY MCGRAW is a cofounder of the Berryville Institute of Machine Learning. Contact him at gem@garymcgraw.com.

RICHIE BONETT is with the College of William & Mary and the Berryville Institute of Machine Learning. Contact him at richiebonett@gmail.com.

HAROLD FIGUEROA is with Ntrepid and the Berryville Institute of Machine Learning. Contact him at harold.figueroa@gmail.com.

VICTOR SHEPARDSON is with the Berryville Institute of Machine Learning. Contact him at victor.shepardson@gmail.com.

Digital Object Identifier 10.1109/MC.2019.2926563

Daniel E. Geer, Jr.
In-Q-Tel

Unknowable Unknowns

Kurt Gödel proved that there are problems for which it is impossible to construct an algorithm that always leads to a correct yes-or-no answer; those problems are undecidable. Alan Turing proved that the halting problem is undecidable in Turing machines. Alfred Tarski proved that truth in the standard model of a system cannot be defined within that system. Olav Lysne proved that it is not possible to verify electronic equipment procured from untrusted vendors, and that a vendor cannot build a system that supports verification by untrusting customers. Ben-David et al. proved that scenarios exist where learnability can neither be proved nor refuted. Finally, Donald Rumsfeld made commonplace the phrase *unknown unknowns*.

And so we come to artificial intelligence (AI), which is to say self-modifying algorithms, which is to say machine learning. Readers of *IEEE Security & Privacy* are well aware of the interrogatability problem "Monsieur Algorithme, why did you make this decision?" an acute concern in multiple subject matter areas, of which cybersecurity is assuredly one. Most *IEEE Security & Privacy* readers agree that all security tools are dual-use, freighting "Why did you make this decision?" with significantly more than mere curiosity or a search for optimality.

There are some who say that a self-modifying algorithm, if purposefully and skillfully constructed, can tell the "why" of its decisions, tell that why in a form we humans can appreciate, and then, perhaps, nod in knowing acceptance. One hopes that this will soon be true, but as of now, it is not.

In other words, and for the time being, black-box interrogation of AI models—similar in spirit to a statistician's sensitivity testing—has the potential to become the default method of assessing an AI model's behaviors. This default will last at least so long as no one has success in understanding a priori how a model works. It is merely stop-gap, relative to Rumsfeld's unknown unknowns: the probability of not asking enough of the right type of questions to characterize a data-driven model is as great as the probability that the model was trained on incomplete or biased data.

It is logical to presume that as AI models increase in complexity, they become more opaque. This parallels a problem we in cybersecurity know only too well: that of trying to understand the attack surface of growing and/or dynamic software installations. Unprovability thus becomes acute, including in the case of cybersecurity, where the mutation rate for offense and defense alike mean not just learning but unlearning.

In some areas other than cybersecurity, handing off the keys to AI models offers immediate, iterative improvement in tailored operations, efficiency, and safety. In the arms race that is cybersecurity, using adaptive algorithms to thwart other adaptive algorithms is so attractive as to seem necessary, and so necessary as to seem attractive.

It is logical to presume that as AI models increase in complexity, they become more opaque.

The financial services industry has already demonstrated some apparent truths worth considering, the principal of which is that we (humans) can build systems more complex than we can manage, complete with behaviors that we cannot predict. Perhaps the question is whether self-modification is, or can be made to be, a safe enough technology to implement, and if so, how does this decision vary by the realm of application?

Digital Object Identifier 10.1109/MSEC.2019.2898636
Date of publication: 2 April 2019

This article originally appeared in
IEEE Security & Privacy, *vol. 17, no. 2, 2019.*

In human society, it is natural for the occasional interrogator who asks "Why did you do that?" to demand an action reversal based on the answer to the question. In the digital policy world, Article 15, Section 1(h) of the European Union's General Data Protection Regulation reads "The data subject shall have the right to obtain from the controller (…) access to personal data and the following information: 1(h), the existence of automated decision making [and] meaningful information about the logic involved as well as the significance and the envisaged consequences of such processing for the data subject." Cybersecurity decisions will certainly encounter Article 15's requirement, and for cybersecurity services that only know what to interdict by being trained on "normal day" data, there is no real answer to Section 1(h)'s requirement as to whether there was hidden malignancy in the training data—i.e., that is an unknowable unknown.

The author suggests that an exclusive embrace of machine learning for cybersecurity is a Faustian bargain—but it's a free country. ■

Daniel E. Geer, Jr. is the chief information security officer of In-Q-Tel. Contact him at dan@geer.org.

Digital Object Identifier 10.1109/MSEC.2019.2903679

Column

Software Stack in a Snapshot

Nils Heinonen
Argonne National Laboratory

Editors: James J. Hack, jhack@ornl.gov; **Michael E. Papka**, papka@anl.gov

***Abstract*—The Argonne Leadership Computing Facility is deploying Singularity to allow HPC resources to adopt "containers"—a technology that has benefitted non-HPC resources like cloud computing servers for a few years now. For HPC users, containerization will allow them to easily migrate their software stack between resources with minimal effort. For HPC facilities themselves, containerization gives users more autonomy to deploy the version of the software that best meets their specific needs.**

Scaling code for massively parallel architectures is a common challenge the scientific community faces. When moving from a system used for development—a personal laptop, for instance, or even a university's computing cluster—to a large-scale supercomputer like those housed at the Argonne Leadership Computing Facility (ALCF), a U.S. Department of Energy Office of Science User Facility, researchers traditionally would only migrate the target application; the underlying software stack would be left behind.

To help alleviate this problem, the ALCF has deployed the service Singularity. Singularity, an open-source framework originally developed by Lawrence Berkeley National Laboratory and now supported by Sylabs, Inc., is a tool for creating and running containers (platforms designed to package code and its dependencies so as to facilitate fast and reliable switching between computing environments)—albeit one intended specifically for scientific workflows and high-performance computing (HPC) resources.

"There is a definite need for increased reproducibility and flexibility when a user is getting started here, and containers can be tremendously valuable in that regard," said Katherine Riley, the ALCF's director of science. "Supporting emerging technologies like Singularity is part of a broader strategy to provide users with services and tools that help advance science by eliminating barriers to productive use of our supercomputers."

GROWING DIVERSITY, GROWING SERVICES

The demand for such services has grown at the ALCF as a direct result of the HPC community's diversification.

Digital Object Identifier 10.1109/MCSE.2019.2900203
Date of current version 26 April 2019.

 Published by the IEEE Computer Society

When the ALCF first opened, it was catering to a smaller user base representative of the handful of domains conventionally associated with scientific computing (high energy physics and astrophysics, for example). HPC is now a principal research tool in new fields, such as genomics, which perhaps lack some of the computing culture ingrained in certain older disciplines. Moreover, researchers tackling problems in machine learning, for example, constitute a new community. This creates a strong incentive to make HPC more immediately approachable to users so as to reduce the amount of time spent preparing code and establishing migration protocols, and thus hasten the start of research work.

Singularity, to this end, promotes strong mobility of computing and reproducibility due to the framework's employment of a distributable image format. This image format incorporates the entire software stack and runtime environment of the application into a single monolithic file. Users thereby gain the ability to define, create, and maintain an application on different hosts and operating environments. Once a containerized workflow is defined, its image can be snapshotted, archived, and preserved for future use. The snapshot itself represents a boon for scientific provenance by detailing the exact conditions under which given data were generated: in theory, by providing the machine, the software stack, and the parameters, one's work can be completely reproduced. Because reproducibility is so crucial to the scientific process, this capability can be seen as one of the primary assets of container technology.

SUITED FOR SCIENCE

ALCF users have already begun to take advantage of the service. Argonne computational scientist Taylor Childers (in collaboration with a team of researchers from Brookhaven National Laboratory, Lawrence Berkeley National Laboratory, and the Large Hadron Collider's ATLAS experiment) led ASCR Leadership Computing Challenge and ALCF Data Science Program projects to improve the performance of ATLAS software and workflows on DOE supercomputers. Every year ATLAS generates petabytes of raw data, the interpretation of which requires even larger simulated datasets, making recourse to leadership-scale computing resources an attractive option.

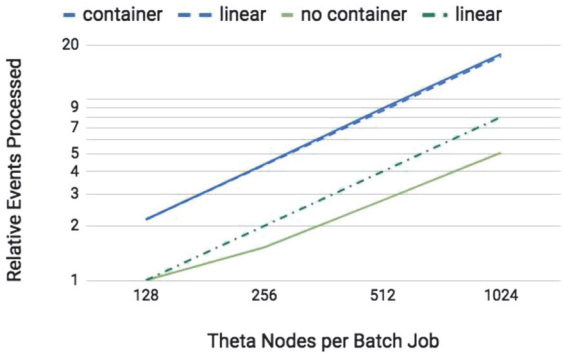


Figure 1. This plot shows the number of events simulated (solid lines) in the ATLAS detector with and without software inside a container. Linear scaling is shown (dotted lines) for reference. Image: J. Taylor Childers, Argonne National Laboratory.

The ATLAS software itself—a complex collection of algorithms with many different authors—is terabytes in size and features manifold dependencies, making manual installation a cumbersome task. Yoda, an MPI-enabled Python application the team developed to communicate between CERN and ALCF systems and ensure all nodes in the latter are supplied with work throughout execution, permitted the researchers to run the ATLAS software on Theta (the ALCF's Cray XC40 system based on the second-generation Intel Xeon Phi processor) inside a Singularity container. As depicted in Figure 1, the use of Singularity resulted in linear scaling on up to 1024 of Theta's nodes, with event processing improved by a factor of four. Containerization also effectively circumvented the software's relative "unfriendliness" toward distributed shared file systems by accelerating metadata access calls; tests performed without the ATLAS software suggested that containerization could speed up such access calls by a factor of seven.

"All told, with this setup, we were able to deliver to ATLAS 65 million proton collisions simulated on Theta using 50 million core-hours," said Childers.

While Singularity can present a tradeoff between immediacy and computational performance (because the containerized software stacks, generally speaking, are not written to exploit massively parallel architectures), the

data-intensive ATLAS project demonstrates the potential value in such a compromise for some scenarios, given the impracticality of retooling the code at its center.

Because containers afford users the ability to switch between software versions without risking incompatibility, the service has also been a mechanism to expand research and try out new computing environments. Rick Stevens, Argonne's Associate Laboratory Director for Computing, Environment, and Life Sciences (CELS), leads the Aurora Early Science Program project virtual drug response prediction. The machine learning-centric project, whose workflow is built from the CANcer Distributed Learning Environment (CANDLE) framework, enables billions of virtual drugs to be screened singly and in numerous combinations while predicting their effects on tumor cells. Their distribution made possible by Singularity containerization, CANDLE workflows are shared between a multitude of users whose interests span basic cancer research, deep learning, and exascale computing. Accordingly, different subsets of CANDLE users are concerned with experimental alterations to different components of the software stack.

"CANDLE users at health institutes, for instance, may have no need for exotic code alterations intended to harness the bleeding-edge capabilities of new systems, instead of requiring production-ready workflows primed to address realistic problems," explained Tom Brettin, strategic program manager for CELS and a co-principal investigator on the project. Meanwhile, through the support of the Exascale Computing Project, CANDLE is being prepared for exascale deployment.

Containers are a relatively new technology for HPC, and their role may well continue to grow. "I don't expect this to be a passing fad," said Riley. "It's functionality that, within five years, will likely be utilized in ways we can't even anticipate yet."

ACKNOWLEDGMENTS

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Nils Heinonen is a Writer and Editor with Argonne National Laboratory, Lemont, IL, USA. He received his bachelor's degree in mathematics and English from the University of Minnesota, Minneapolis, MN, USA. Contact him at nheinonen@anl.gov.

This article originally appeared in Computing in Science & Engineering, *vol. 21, no. 3, 2019.*

COMPSAC 2020
Madrid, Spain
July 13-17, 2020

COMPSAC is the IEEE Computer Society Signature Conference on Computers, Software and Applications. It is a major international forum for academia, industry, and government to discuss research results and advancements, emerging challenges, and future trends in computer and software technologies and applications. The theme of COMPSAC 2020 is "Driving Intelligent Transformation of the Digital World".

Staying relevant in a constantly evolving digital landscape is a challenge faced by researchers, developers, and producers in virtually every industry and area of study. Once limited to software-enabled devices, the ubiquity of digitally-enabled systems makes this challenge a universal issue. Furthermore, as relevance fuels change, many influencers will offer solutions that benefit their own priorities. Fortunately, history has shown that the building blocks of digital change are forged by those conducting foundational research and development of digital systems and human interactions. Artificial Intelligence is not new, but is much more utilized in everyday computing now that data and processing resources are more economically viable, hence widely available. The opportunity to drive the use of this powerful tool in transforming the digital world is yours. Will your results help define the path ahead, or will you relegate those decisions to those with different priorities for utilizing intelligence in digital systems? COMPSAC has been and continues to be a highly respected venue for the dissemination of key research on computer and software systems and applications, and has influenced fundamental developments in these fields for over 40 years. COMPSAC 2020 is your opportunity to add your mark to this ongoing journey, and we highly encourage your submission!

COMPSAC 2020, organized as a tightly integrated union of symposia, will focus on technical aspects of issues relevant to intelligent transformation of the digital world. The technical program will include keynote addresses, research papers, industrial case studies, fast abstracts, a doctoral symposium, poster sessions, and workshops and tutorials on emerging and important topics related to the conference theme. Highlights of the conference will include plenary and specialized panels that will address the technical challenges facing researchers and practitioners who are driving fundamental changes in intelligent systems and applications. Panels will also address cultural and societal challenges for a society whose members must continue to learn to live, work, and play in the environments the technologies produce.

Authors are invited to submit original, unpublished research work, as well as industrial practice reports. Simultaneous submission to other publication venues is not permitted. All submissions must adhere to IEEE Publishing Policies, and will be vetted through the IEEE CrossCheck portal. Further info is available at www.compsac.org. Conference authors and authors of previously published papers should visit **https://ieeecompsac.computer.org/2020/jc-cj/** to view special publishing opportunities available at COMPSAC.

Organizers
Standing Committee Chair: Sorel Reisman (California State University, USA)
Steering Committee Chair: Sheikh Iqbal Ahamed (Marquette University, USA)
General Chairs: Mohammad Zulkernine (Queen's University, Canada), Edmundo Tovar (Universidad Politécnica de Madrid, Spain), Hironori Kasahara (Waseda University, Japan)
Program Chairs in Chief: W. K. Chan (City University, Hong Kong), Bill Claycomb (Carnegie Mellon University, USA), Hiroki Takakura (National Institute of Informatics, Japan)
Workshop Chairs: Ji-Jiang Yang (Tsinghua University, USA), Yuuichi Teranishi (National Institute of Information and Communications Technology, Japan), Dave Towey (University of Nottingham Ningbo China, China), Sergio Segura (University of Seville, Spain)
Local Chairs: Sergio Martin (UNED, Spain), Manuel Castro (UNED, Spain)

Important Dates
Main conference papers due: Extended to 13 February 2020
Main conference paper notification: 3 April 2020
Journal/Conference (JC) submissions due: 10 April 2020
Journal/Conference (JC) notifications: 30 April 2020
Workshop papers due: 9 April 2020
Workshop papers notification: 1 May 2020
Camera-ready and registration due: 15 May 2020

IEEE COMPUTER SOCIETY

Photo: King Felipe III in Major Square, Madrid
Photo credit: Iria Castro - Photographer (Instagram @iriacastrophoto)

Department: Visualization Viewpoints
Editor: Theresa-Marie Rhyne, theresamarierhyne@gmail.com

In Situ Visualization for Computational Science

Hank Childs
University of Oregon

Christoph Garth
Technische Universität Kaiserslautern

Janine Bennett
Sandia National Laboratories

Bernd Hentschel
RWTH Aachen University

***Abstract*—*In situ* visualization is an increasingly important approach for computational science, as it can address limitations on leading edge high-performance computers and also can provide an increased spatio-temporal resolution. However, there are many open research issues with effective *in situ* processing. This article describes the challenges identified by a recent Dagstuhl Seminar on the topic.**

There are two processing paradigms for visualizing data: *in situ* processing, i.e., processing data as it is generated, and post hoc processing, i.e., processing data well after it is generated (see Figure 1). Research results on *in situ* visualization started appearing approximately a quarter-century ago, albeit using terms such as coprocessing[1] and runtime visualization.[2] Despite promising findings, post hoc processing has retained its role as the dominant processing paradigm for scientific visualization. Post hoc processing enjoys advantages with respect to simpler interfacing (i.e., via files instead of code), increased ability for the end-user to explore data, and fewer consequences when encountering error conditions, among others. In contrast, *in situ* processing enjoys advantages with respect to access to more spatio-temporal data, decreased time to access data, and increased computational power. That said, it is difficult to make definitive statements comparing the merits of *in situ* and post hoc processing, since each can be implemented with variations and optimizations that mitigate their respective weaknesses. For example, post hoc processing can improve an access time via multiresolution methods, while *in situ* processing can improve on fault tolerance by performing concurrent visualizations on separate system resources.

In high-performance computing, the dominance of post hoc processing has faded over the last five years, and given way to a surge of interest around *in situ* processing.[3] The primary driver

Digital Object Identifier 10.1109/MCG.2019.2936674
Date of current version 1 November 2019.

Processing paradigms for scientific visualization

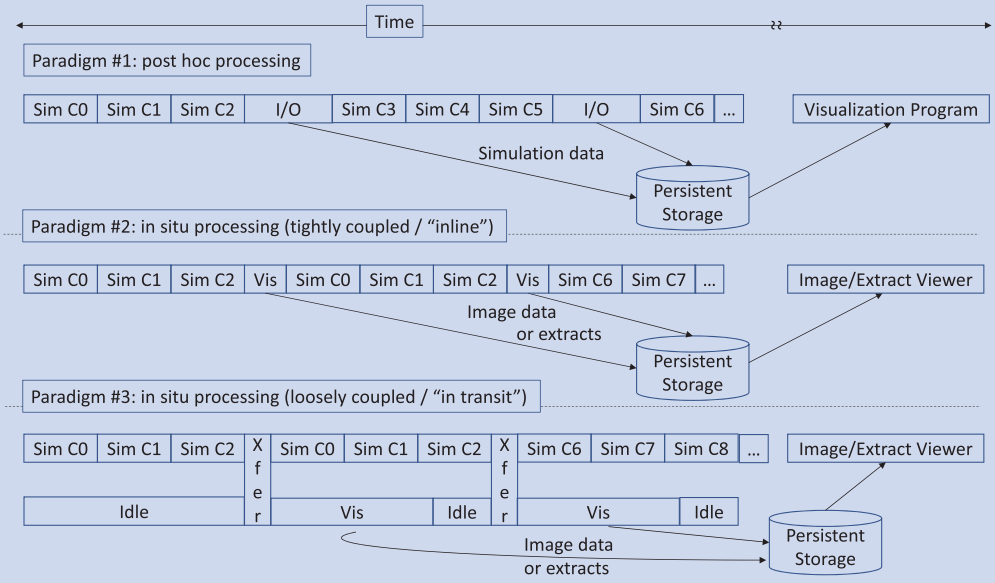


Figure 1. Contrasting post hoc and *in situ* processing. This diagram shows three processing paradigms, one for post hoc and two for *in situ*, running horizontally and divided by dashed lines. For each of the three workflows, time runs from left to right, with a break in the timeline indicating that a user views data sometime after it is stored. In each of the workflows, a simulation runs in so-called "cycles," where a simulation advances flow from one state to the next. These are denoted "Sim C1" for the first cycle, "Sim C2" for the second cycle, etc. In the top workflow, post hoc processing, the simulation stores its state to persistent storage every three cycles. In this notional example, the "I/O" rectangle is wider than any of the simulation cycle rectangles to indicate that I/O has become very expensive on supercomputers. For simulations with very quick cycle times, the I/O rectangle may be hundreds of times wider than a simulation cycle rectangle (i.e., take hundreds of times longer). Of course, I/O may also occur less frequently than every three cycles. The top workflow concludes when a user starts a separate visualization program to load the simulation data from persistent storage. This visualization program may run on the same resources as the simulation code, or on distinct resources. The middle workflow denotes a common variant of *in situ* processing that is sometimes referred to as "tightly coupled" or "inline" *in situ* visualization. In this mode, both visualization and simulation run on the same resources, in an alternating manner. The output of the visualizations are images or extracts, and these images or extracts can be viewed after (or during) the simulation via a viewer which likely runs on distinct resources. The final workflow denotes another common variant of *in situ* processing, "loosely coupled" or "in transit." With this workflow, additional resources for visualization are run concurrently to the simulation, and data is transferred from the simulation resources to the visualization resources. In this workflow, the resources for visualization are often smaller than those for simulation, making periods of idle time acceptable.

behind switching processing paradigms is to address I/O constraints on leading-edge supercomputers. On these machines, the ability to generate data is increasing much faster than the ability to store data to persistent storage; compute has gone up by a factor of approximately 100× over the last decade, with I/O performance typically only increasing by a factor of 10×. Consequently, visualization performance using the post hoc paradigm, already shown a decade ago to be limited on supercomputers by I/O performance,[4] becomes poorer and poorer with each new generation of hardware. Of course, *in situ* processing addresses this problem by avoiding the usage of disk altogether. Today, *in situ* processing is being used increasingly on supercomputers, including a

State-of-the-art *in situ* visualization

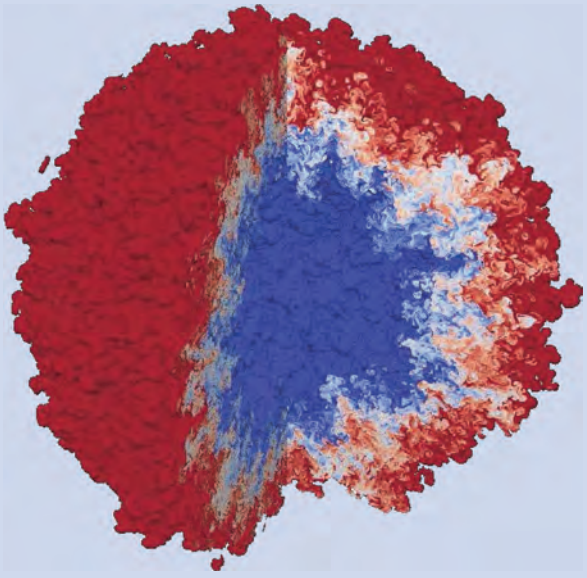


Figure 2. Example of a state-of-the-art *in situ* visualization, made by Lawrence Livermore National Laboratory on the Sierra supercomputer (courtesy Matthew Larsen and Cyrus Harrison). This image is of an idealized Inertial Confinement Fusion (ICF) simulation of a Rayleigh–Taylor instability with two fluids mixing in a spherical geometry. The image was created with the *in situ* visualization library Ascent[15], which performed its tasks using the same resources as the simulation code, in this case 4096 nodes and 16,384 GPUs. The mesh contained 98 billion hexahedrons, and rendering occurred in approximately 300 ms for each frame. By running *in situ*, all I/O costs were avoided; instead, time that could have been spent on I/O was spent making a movie flying around the mixing layer.

recent example of visualizing 98 billion cells on 16,384 GPUs (see Figure 2). Unfortunately, despite some successes, *in situ* processing is far from a solved problem, as there are multiple barriers impeding its widespread use.

The open research problems with *in situ* processing were the subject of a Dagstuhl Seminar in July 2018, titled "*In Situ* Visualization for Computational Science," which brought together practitioners from scientific visualization, computational science, and high-performance computing. The workshop participants identified ten *in situ* processing challenges that require significant research, and also "cross-cutting challenges" that do not require research. These challenges were documented in a workshop report[5] that serves as the source material for this article, although two pairs of challenges were combined for readability.

In all, the authors of this article believe that significant research is still needed to enable *in situ* processing to meet widespread production needs. We hope this article will inform our community about the challenges identified at the Dagstuhl Seminar and will also encourage additional research in this space.

DATA QUALITY AND REDUCTION

An important consideration for *in situ* processing is whether the desired visualizations are known *a priori*. When they are, the visualizations can be specified ahead of time and carried out as the data are generated. When they are not known *a priori*, *in situ* processing is more complicated, since it is not clear which visualizations to carry out. One approach is to refuse *in situ* processing, i.e., to do post hoc processing instead, although saving data less frequently. The problem with this approach is that I/O constraints may cause the data to become so sparse temporally that important phenomena may be lost (i.e., a phenomenon begins after one time slice is saved and ends before the next one) or that features cannot be tracked over time.

Another approach when lacking *a priori* knowledge, and the overarching research challenge described in this section, is to use a combination of *in situ* and post hoc processing. Specifically, data are transformed and reduced *in situ*, and the resulting reduced data are saved to disk, to be explored later in the traditional post hoc manner. The benefit of this idea is that

the reduced data could be small enough to store sufficient temporal frequency even in the face of I/O constraints.

The following topics were identified as specific research challenges for data quality and reduction.

- There are myriad of possible techniques for reducing scientific data.[6] Some techniques are general, while others are useful for very specific visualizations. Overall, it is unclear which techniques are the best match for given usage scenarios.
- This "*in situ* reduction + post hoc exploration" paradigm creates a tension between reduction and data integrity. If too much reduction is performed, then the result may be a loss of data integrity and, thus, meaningless visualizations. Similarly, if data integrity is held as paramount, then reduction may be minimal, and I/O problems will persist. The challenge, then, is to determine the acceptable levels of reduction and integrity and also to meet these levels.
- Temporal analysis is commonplace, and requires special attention, both on how to achieve better reduction-integrity tradeoffs and on how to predict which small scale features will be important and, thus, should not be reduced since they will expand over time.
- Reduction introduces error, and this error should be treated in a way that is acceptable to domain scientists, i.e., error bounds, verification, considerations for error propagation, etc.

Solving these challenges could help with both efficiency (e.g., reducing I/O costs or even reducing the need to repeat a simulation) and effectiveness (e.g., enabling scientific discoveries by ensuring that sufficient data are available for exploration).

WORKFLOWS

There are a variety of instantiations that *in situ* processing can take. In one form, the *in situ* routines are provided as a library that is compiled into a simulation code, sharing memory and compute resources. In another form, the *in situ* routines are part of their own executable which runs on distinct resources. There are also many other possible instantiations, involving multiple modules, multiple data transfers, etc.

Workflows are a mechanism for encoding each of these instantiations with data moving from one task to another. The form of these tasks can vary, from subroutines within one binary to separate executables on distinct computing resources. Overall, the workflow methodology may seem heavyweight for the simplest instantiations of *in situ* processing, but it is widely viewed as necessary as more and more modules are incorporated to solve analysis problems.

The workshop participants differentiated between workflow specification and workflow execution since they each have their own challenges. Workflow specification refers to how tasks are defined, including possible inputs and outputs of a task, dependencies between tasks, and resource requirements. Workflow specification informs workflow execution, since it defines the set of possible actions to take but is distinct since it does not concern itself with how the workflow is executed.

The following were identified as specific research challenges for workflow specification.

- How do we specify desired workflow behavior as conditions change? These changes can come from the data (need more resources to explore a phenomenon) or from the system (faults or unresponsive nodes).
- How should the results of certain operations in the workflow drive future operations in the workflow? How do we specify this?
- How can resource priorities be embedded in the specification?
- How should tradeoffs between declarative and procedural approaches affect workflow specifications?

For workflow execution, the challenge is to develop systems that realize workflow specifications, and also to execute them efficiently. Realizing a workflow is involved: forming a set of tasks, scheduling those tasks on hardware, and handling error conditions. The tasks may have

multiple granularities ranging from scheduling within a single program to scheduling across many programs. Further, the penalty for poor decisions can be high, for example, when stalls occur when waiting for resources to become available.

The following were identified as specific research challenges for workflow execution.

- What strategies balance flexibility and efficiency? How can abstractions avoid being too coarse (which prevents optimizations) or too fine (which becomes too complex to manage)?
- What should the data interfaces between modules be?
- How can workflow research outside the community be leveraged? What are the visualization community's unique requirements?

The benefits of solving this problem include improved resource utilization (via elasticity to fit resources optimally), by preventing wasted cycles when there is a fault, and potentially by getting algorithms to run on the types of hardware where they are most efficient. Other benefits include simplifying the job of workflow execution, increased user productivity, and particularly in code reuse and in collaboration. Finally, we note that the workshop discussion and this summary are particularly indebted to ideas from a U.S. Department of Energy workshop on workflows.[7]

EXASCALE SYSTEMS

Exascale computing poses multiple fundamental changes with respect to visualization. First, exascale hardware architectures will be different than our previous generation of supercomputers because innovative approaches will be needed to achieve so much computing power within energy and cost constraints. The most important change for the visualization community is the relative decrease in the I/O bandwidth (i.e., the driver for *in situ* processing). Other notable changes include the pervasive use of accelerators, multiple accelerators per node, billion-way concurrency, an increased focus on power usage, and deep memory hierarchies. Second, exascale machines will be used in a different way than our previous generation of supercomputers. One example is a shift in science uses cases, including multiphysics solvers, ensembles, and the inclusion of machine learning. This diversity of use cases motivates another section in this article: exascale simulations will produce new, nontraditional types of data.

The following were identified as specific research challenges for exascale systems.

- Ensure that our algorithms can run scalably at exascale-level concurrencies, and/or developing production workflow capabilities to enable visualization on a subset of the compute allocation.
- Leverage exascale hardware features into our algorithms when they can improve performance, such as nonvolatile random access memory (NVRAM).
- As the gap between execution rate and RAM grows, *in situ* visualization will need to adapt, both in terms of efficient execution, and in terms of minimizing effects on simulation codes.

Solving these challenges will allow exascale hardware to be used efficiently. Ultimately, the solution may align with the principles of codesign—if we better understand the behavior of visualization software, then we can not only adjust our research regarding *in situ* approaches to perform better given supercomputer constraints but also affect the design of future supercomputers.

ALGORITHMIC CHALLENGES

Some staple visualization algorithms are difficult to parallelize efficiently and/or apply to large data sets. In particular, some algorithms exhibit global access patterns that may not even be feasible due to prohibitive communication costs. Examples include particle advection (difficult to parallelize efficiently) and some topological techniques (sometimes global in nature). If we plan to run such algorithms *in situ*, then we will need to develop efficient parallel versions of these algorithms or to identify alternate techniques that are suitable for high-performance computer architectures. An example of the latter

approach would be (perhaps) using convolution-type algorithms in the place of particle tracing flow visualization.

The following were identified as specific research challenges with respect to algorithms.

- Are there any features in future high-performance computer architectures that are helpful for difficult-to-parallelize algorithms? (e.g., deep memory hierarchies.)
- Can we use alternate forms, such as compressed data or higher order elements, to improve scalability? If so, how does that affect the accuracy of the results?
- If exploration-oriented use cases mandate that some of these algorithms be interactive, then how would that change our *in situ* design?
- What are the *in situ* configurations that promote efficiency? For example, if running at lower concurrency is more efficient, then this would inform how to carry out *in situ* processing. This topic is explored more in the section on Cost Models.

Solving these challenges is important; failing to do so will limit the techniques we can deliver to stakeholders, in turn limiting the potential for insight from simulations on the next generation of high-performance computers.

EXASCALE ENABLES NEW USE CASES AND NEW TYPES OF DATA

The typical *in situ* visualization use case is a single simulation generating a high-resolution mesh. However, increased computational power is enabling new types of outputs that motivate new types of visualization. One very important, and increasingly prominent, use case is that of ensemble analysis, i.e., where simulation codes produce ensembles instead of single output. This change requires different processing paradigms, has different properties with respect to efficiently using hardware, and requires different visualization approaches to be effective. A notable example of success on this front is the Melissa project,[8] which was used to analyze an ensemble of 80,000 parallel simulations and avoided 288 TB of storage. Alternate data sources also include simulations that are run alongside experiments and that incorporate experimental data in their visualizations, such as the Xi-CAM effort.[9] Finally, this challenge is not limited to ensembles and including experimental data. Other examples include computational steering or, when mesh resolutions get sufficiently high, a multi-scale representation of data.

The following were identified as specific research challenges for new use cases/new types of data.

- What techniques will the visualization community need to incorporate to support these new types of data? In this case, it is important to note that new collaborations may be required to succeed, for example, bringing in mathematicians and statisticians.
- How should the *in situ* processing approach be adapted to include data wrangling for new types of data (ensembles and experimental)? What data models facilitate exchange and are acceptable to both visualization and simulation stakeholders?
- How should workflows and abstractions support multiple, simultaneous goals?

The challenges involved with these new usages of supercomputers are driven by the domain scientists. While some of the solutions are arguably outside visualization, our community is developing useful infrastructure to address some of these problems, and broadening these infrastructures to include more approaches will benefit all.

COST MODELS

When running *in situ*, visualization routines often have to perform their tasks within a given time budget. Failure to complete their task in the allotted time typically means that either the visualization is aborted or that the simulation stalls. This situation can be avoided by assessing feasibility *a priori*: for a given set of visualization tasks and its parameters (e.g., isosurfacing task and parameters of specific isolevels), a given data set, and given resources, then can the task be completed within T seconds? Currently, feasibility assessments are not rigorous, e.g., previous experiences, extrapolations from smaller data sets, etc. Cost models provide a more rigorous way to assess feasibility.

Cost models take an input workload (visualization tasks, parameters, data set, resources) and produce an estimate of how long it will take to complete the task. If the estimate is accurate, then these models can be used to answer feasibility questions. Unfortunately, cost models are often hard to generate, although our community has had some successes.[10–12] The following were among the specific research challenges identified for cost models.

- How can we design cost models without extensive studies sweeping many sets of parameters? Which parameters can be fixed or are not relevant?
- Can machine learning be used to solve this problem?
- What are the unique challenges specific to visualization?
- How accurate do the cost models need to be? How tolerant can they be of inaccuracies?

Solving this problem will enable more efficient use of resources. If a cost model reveals that tasks can be completed in under the time budget, then the time can be returned to the simulation or more visualization tasks can be performed.

CONVERGENCE OF HPC AND BIG DATA

Developments happening outside visualization, high-performance computing (HPC), and computational science communities appear likely to affect each of these three fields. In particular, Big Data processing models produce ideas and implementations that may inform solutions to our research challenges. Similarly, recent machine learning/deep learning developments are poised to alter approaches in many fields, with the potential to benefit *in situ* visualization as well. Further, these activities impact hardware designs, and their components may well appear on supercomputers in the near future. The overarching challenge, then, is how to leverage the breakthroughs from these fields to our own.

The following were among the specific research challenges identified for the convergence of HPC and Big Data.

- How can machine learning be used to enable *in situ* data reduction or to optimize *in situ* workflows?
- How can the knowledge extracted from Big Data and/or machine learning help improve *in situ* visualization approaches?
- What elements of programming environments for Big Data (which are accessible to new programmers) can be incorporated for *in situ* frameworks?

Overall, the benefit of embracing Big Data/machine learning would be the harnessing of industry resources to do tasks more quickly and effectively than we could do otherwise.

SOFTWARE COMPLEXITY, HETEROGENEITY, AND USER-FACING ISSUES

This research challenge focuses on making *in situ* visualization software accessible and usable to a large number of stakeholders. At the workshop, it was identified that there were three cross-cutting problems that fuel this challenge. The first is software complexity, i.e., *in situ* software is complex and hard to use and also exists in a difficult context—linking two complex software packages, each with their own constraints, data models, parallelization strategies, etc. The second is heterogeneity, i.e., heterogeneity limits adoption due to a cross product of options in hardware architecture, software tools, and usage. The third is user-facing issues, i.e., users are reluctant to adopt new technologies and in particular *in situ*, because it creates additional dependencies for their code, the dependent software they invest may have an uncertain lifetime, they inherit reliability issues from the software they adopt, etc. The following were among the specific research challenges identified for increasing *in situ* adoption.

- How can we minimize our intrusion into the simulation code?
- How can we deal with diverse hardware (i.e., performance portability) and software? That said, it was noted that the VTK-m project[13] is addressing part of the hardware heterogeneity problem, although it does not support deep memory hierarchies, networks, etc.

- How can we isolate faults so that reliability issues with *in situ* visualization software do not affect simulation codes?

Failing to make *in situ* techniques accessible to domain scientists will mean that some stake holders go forward without it, which could potentially result in lost discoveries. Further, solving the problem has a cost benefit—reduced integration times, reduced loss in computational time due to faults, etc.

PRACTICAL ISSUES FACING *IN SITU*

The workshop participants felt that some challenges were worthy of documentation, even though they were not research challenges per se. The challenges were discussed in two panel sessions: "software engineering and deployment" and "programming and funding issues/interdisciplinary/pipeline." In all cases, there was a recognition that the problems discussed were not unique to *in situ* visualization, but it was also recognized that *in situ* visualization exacerbates these problems.

For the first panel, the primary theme was on the differences in priorities between *in situ* tools researchers/developers and their target user community, specifically for the following topics.

- Requirements, in that scientists and engineers, prefer specialized tools, tailored to their environment, while *in situ* tool developers prefer general-purpose solutions that can be deployed in many simulation codes.
- Adoption, with respect to multiple issues: effective communication between communities, committing to software with an uncertain lifetime, and simulation codes increasing their dependencies on external software packages.
- Accessibility, in particular, the tension between commercial software and open source.

For the second panel, a major focus was on the significant investment needed to develop *in situ* tools, and that acquiring funding for this investment can be difficult. An important aspect of this conversation was the role of industry collaboration and the unclear division of labor between research and production.

CONCLUSION

This article follows 10 years after an article by Kwan-Liu Ma, also on the challenges for *in situ* visualization.[14] Comparing the predecessor article and the Dagstuhl Seminar shows the progress we have made over the last decade, as well as the areas where little has changed. The previous article asked several questions that are still highly relevant today: What are the optimal configurations for simulation and visualization to share the hardware? How should we reduce data and yet maintain its integrity? What are the best ways to exchange data between simulation and visualization codes? How do we run algorithms efficiently in parallel? Happily, some of the questions raised in Kwan-Liu's article have been answered in the last decade: Will simulation scientists accept part of their allocation being used for visualization? (Yes.) Can existing commercial and open-source visualization software tools be extended to support *in situ*? (In most cases, serious adaptation was needed, but that work has happened or is in progress.) Finally, some new questions have emerged: How to deal with new use cases and new types of data (experimental, ensembles, etc.)? How to incorporate workflow methodology? How to deal with the new architectural features on modern supercomputers? How can we use cost models to achieve efficiency? And how can we ride the wave of activity in Big Data and machine learning? As *in situ* processing is still in a nascent phase, it would not surprise the authors to have another article 10 years from now that has a similar breakdown, with some current problems well addressed, others unsolved, and new, unforeseen problems emerging.

ACKNOWLEDGMENTS

This article is derived from a Dagstuhl report that had many contributors. The authors feel particularly indebted to participants that led the writing of individual research challenges, specifically E.W. Bethel, P.-T. Bremer, K. Isaacs, K. Moreland, B. Muite, J. Patchett, T. Peterka, D. Pleiter, D. Pugmire, B. Raffin, A. Ribes Cortes, H.-W. Shen, R. Sisneros, and M. Srinivasan. They are also very appreciative to the other seminar participants who greatly contributed to the general discussion: A. Bauer, T. Carrard, M. Dorier,

S. Frey, N. Gauger, M. Hadwiger, C. Hansen, K. Heitmann, I. Hotz, J. Kruger, M. Larsen, P. Messmer, K. Ono, M. Parashar, V. Pascucci, N. Rober, U. Rude, F. Sadlo, G. Weber, R. Westerman, and H. Yu. Sandia National Laboratories is a multimission laboratory managed and operated by the National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-NA0003525. Some of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract number DE-ASC52-07NA27344 (LLNL-JRNL-787609). This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

REFERENCES

1. R. Haimes, "pV3: A distributed system for large-scale unsteady CFD visualization," in *Proc. 32nd Aerosp. Sci. Meeting Exhibit*, 1994, Paper 321.
2. K.-L. Ma, "Runtime volume visualization for parallel CFD," in *Parallel Computational Fluid Dynamics*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 307–314.
3. A. C. Bauer *et al.*, "In situ methods, infrastructures, and applications on high performance computing platforms," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 577–597, Jun. 2016.
4. H. Childs *et al.*, "Extreme scaling of production visualization software on diverse architectures," *IEEE Comput. Graph. Appl.*, vol. 30, no. 3, pp. 22–31, May/Jun. 2010.
5. J. C. Bennett, H. Childs, C. Garth, and B. Hentschel, "In situ visualization for computational science," *Dagstuhl Rep.*, vol. 8, no. 7, pp. 1–43, 2019.
6. S. Li, N. Marsaglia, C. Garth, J. Woodring, J. Clyne, and H. Childs, "Data reduction techniques for simulation, visualization and data analysis," *Comput. Graph. Forum*, vol. 37, no. 6, pp. 422–447, Sep. 2018.
7. E. Deelman *et al.*, "The future of scientific workflows," *Int. J. High Perform. Comput. Appl.*, vol. 32, no. 1, pp. 159–175, 2018.
8. T. Terraz *et al.*, "Melissa: Large scale in transit sensitivity analysis avoiding intermediate files," in *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.*, 2017, Paper 61.
9. R. J. Pandolfi *et al.*, "Xi-cam: A versatile interface for data visualization and analysis," *J. Synchrotron Radiat.*, vol. 25, no. 4, pp. 1261–1270, 2018.
10. M. Larsen, C. Harrison, J. Kress, D. Pugmire, J. S. Meredith, and H. Childs, "Performance modeling of in situ rendering," in *Proc. Int. Conf. High Perform. Comput. Netw. Storage Anal.*, Salt Lake City, UT, USA, Nov. 2016, pp. 276–287.
11. V. Bruder, S. Frey, and T. Ertl, "Prediction-based load balancing and resolution tuning for interactive volume raycasting," *Vis. Inform.*, vol. 1, no. 2, pp. 106–117, 2017.
12. M. Dorier *et al.*, "Adaptive performance-constrained in situ visualization of atmospheric simulations," in *Proc. IEEE Int. Conf. Cluster Comput.*, Sep. 2016, pp. 269–278.
13. K. Moreland *et al.*, "VTK-m: Accelerating the visualization toolkit for massively threaded architectures," *IEEE Comput. Graph. Appl.*, vol. 36, no. 3, pp. 48–58, May/Jun. 2016.
14. K.-L. Ma, "In situ visualization at extreme scale: Challenges and opportunities," *IEEE Comput. Graph. Appl.*, vol. 29, no. 6, pp. 14–19, Nov./Dec. 2009.
15. M. Larsen *et al.*, "The ALPINE in situ infrastructure: Ascending from the ashes of strawman," in *Proc. 3rd Workshop In Situ Infrastructures Enabling Extreme Scale Anal. Vis.*, Denver, CO, USA, Nov. 12–17, 2017, pp. 42–46.

Hank Childs is currently an Associate Professor with the Department of Computer and Information Science, University of Oregon, Eugene, OR, USA. His research interests focus on scientific visualization, high-performance computing, and the intersection of the two. He received the Ph.D. degree in computer science from the University of California, Davis, CA, USA, in 2006. Contact him at hank@uoregon.edu.

Janine Bennett is a Principal Member of the Technical Staff with Sandia National Laboratories, Livermore, CA, USA. She began her career doing research on scientific visualization and topology, and has expanded her interest into issues relating to exascale computing and computational science. She received the Ph.D. degree in computer science from the University of California, Davis, CA, USA, in 2008. Contact her at jcbenne@sandia.gov.

Christoph Garth is a Professor of computer science with Technische Universität Kaiserslautern, Kaiserslautern, Germany. His research interests include large-scale data analysis and visualization, *in situ* visualization, topology-based methods in visualization, and interdisciplinary applications of visualization.

He received the Ph.D. degree from Technische Universität Kaiserslautern in 2007 and then spent four years as a Postdoctoral Researcher with the University of California, Davis, CA, USA. Contact him at garth@cs.uni-kl.de.

Bernd Hentschel is currently a Data Scientist with d.velop AG in Gescher, Germany, a major SME in the area of enterprise content management. Previously, from 2010 to 2018, he co-led the Virtual Reality Group with RWTH Aachen University, Aachen, Germany. His research interests include the analysis of domain-specific features in large simulation data, parallel visualization algorithms, and immersive visualization. He studied computer science with RWTH Aachen University, from where he received the Dipl.-Inform. degree in 2003 and the Dr. rer. nat. degree under the supervision of Prof. Dr. T. W. Kuhlen in 2009. Contact him at hentschel@vr.rwth-aachen.de.

Contact department editor Theresa-Marie Rhyne at theresamarierhyne@gmail.com.

This article originally appeared in
IEEE Computer Graphics and Applications,
vol. 39, no. 6, 2019.

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

Submit a paper today to the premier new open access journal in computing and information 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

CALL FOR PAPERS

IEEE 21st International Conference on Information Reuse and Integration for Data Science (IEEE IRI 2020)

Las Vegas, NV, USA / 11 – 13 August 2020

The IEEE IRI conference is a recognized forum for researchers and practitioners from academia, industry, and government to present and exchange ideas that address real-world problems with real-world solutions.

This conference is now seeking excellent, novel, and contemporary papers covering all aspects of Data—including Scientific Theory and Technology-Based Applications.

The conference includes, but is not limited to, the areas listed below:

- Application—Autonomous Vehicles, Business, Education, Engineering, Healthcare, the Internet of Things, Math, Military, Multimedia, NLP, Robotics, Science, Security, Social Networking, Space, Vision, et al.
- Contemporary as well as Novel Data Mining Techniques
- Data & Knowledge Representation and Management
- Data Science & Technologies—Heuristic Acquisition
- Data Visualization
- Graph Models
- Machine Learning & AI
- Predictive Data Analysis & Intelligence
- Predictive Modeling
- Recommender Systems
- Statistical Analysis
- Theory

This year, for the first time, IRI will hold classified military sessions at the nearby Naval Information Warfare Center (NIWC) in San Diego. Presentation of classified IRI military papers and attendance at these sessions is open to those holding an active US government Secret clearance. A clearance is not required for any other IRI sessions. SECRET presentations pertaining to the following areas of interest are encouraged:

- Artificial Intelligence, Heuristics, and Explanation-Based Learning
- Machine Learning
- Computer Vision
- Hypersonic Flight
- Autonomous Vehicles
- Predictive Maintenance
- Logistics
- Silicon Compilers
- Quantum Theory and Application

www.bit.ly/iri20-cfp

Important Deadlines

Abstract submission: 15 April 2020

Full paper research/industry/application/gov't track deadline: 22 April 2020

Short paper track deadline: 22 April 2020

Poster and demo paper track deadline: 8 May 2020

Full/short paper acceptance notification: 1 June 2020

Poster/demo paper acceptance notification: 15 June 2020

Camera ready submission deadline: 20 June 2020

Author registration due: 1 July 2020

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

SUBMIT TODAY

IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING

SCOPE

The *IEEE Transactions on Sustainable Computing* (*T-SUSC*) is a peer-reviewed journal devoted to publishing high-quality papers that explore the different aspects of sustainable computing. The notion of sustainability is one of the core areas in computing today and can cover a wide range of problem domains and technologies ranging from software to hardware designs to application domains. Sustainability (e.g., energy efficiency, natural resources preservation, using multiple energy sources) is needed in computing devices and infrastructure and has grown to be a major limitation to usability and performance.

Contributions to *T-SUSC* must address sustainability problems in different computing and information processing environments and technologies, and at different levels of the computational process. These problems can be related to information processing, integration, utilization, aggregation, and generation. Solutions for these problems can call upon a wide range of algorithmic and computational frameworks, such as optimization, machine learning, dynamical systems, prediction and control, decision support systems, meta-heuristics, and game-theory to name a few.

T-SUSC covers pure research and applications within novel scope related to sustainable computing, such as computational devices, storage organization, data transfer, software and information processing, and efficient algorithmic information distribution/processing. Articles dealing with hardware/software implementations, new architectures, modeling and simulation, mathematical models and designs that target sustainable computing problems are encouraged.

SUBSCRIBE AND SUBMIT

For more information on paper submission, featured articles, calls for papers, and subscription links visit:

www.computer.org/tsusc

IEEE COMPUTER SOCIETY

IEEE COMMUNICATIONS SOCIETY

CEDA IEEE Council on Electronic Design Automation

IEEE

IEEE INTERNATIONAL SYMPOSIUM ON HARDWARE-ORIENTED SECURITY AND TRUST

4–7 May 2020 • San Jose, CA, USA • DoubleTree by Hilton

Join dedicated professionals at the IEEE International Symposium on Hardware Oriented Security and Trust (HOST) for an in-depth look into hardware-based security research and development.

Key Topics:

- Semiconductor design, test and failure analysis
- Computer architecture
- Systems security
- Cryptography and cryptanalysis
- Imaging and microscopy

Discover innovations from outside your sphere of influence at HOST. Learn about new research that is critical to your future projects. Meet face-to-face with researchers and experts for inspiration, solutions, and practical ideas you can put to use immediately.

REGISTER NOW: www.hostsymposium.org

IEEE COMPUTER SOCIETY · IEEE · ttc · IEEE TC on Security & Privacy

Conference Calendar

Questions? Contact conferences@computer.org

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.

Find a region: Africa ■ Asia ▲ Australia ◆ Europe ● North America ◗ South America ★

April

5 April

- ISPASS (Int'l Symposium on Performance Analysis of Systems and Software) ◗

9 April

- MIPR (IEEE Int'l Conf. on Multimedia Information Processing and Retrieval) ▲

13 April

- Mobile Cloud (IEEE Int'l Conf. on Mobile Cloud Computing, Services, and Eng.) ●
- SOSE (IEEE Int'l Conf. on Service-Oriented System Eng.) ●
- DAPPS (IEEE Int'l Conf. on Decentralized Applications and Infrastructures) ●
- JCC (IEEE Int'l Conf. on Joint Cloud Computing) ●
- AITest (IEEE Int'l Conf. on Artificial Intelligence Testing) ●
- BigDataService (IEEE Int'l Conf. on Big Data Computing Service and Machine Learning Applications) ●

14 April

- PacificVis (IEEE Pacific Visualization Symposium) ▲

15 April

- COOL Chips (IEEE Symposium on Low-Power and High-Speed Chips and Systems) ▲

20 April

- ICDE (IEEE Int'l Conf. on Data Eng.) ◗

May

3 May

- FCCM (IEEE Int'l Symposium on Field-Programmable Custom Computing Machines) ◗

4 May

- HOST (IEEE Int'l Symposium on Hardware Oriented Security and Trust) ◗

11 May

- CCGrid (IEEE/ACM Int'l Symposium on Cluster, Cloud, and Internet Computing) ◆
- ICFEC (IEEE Int'l Conference on Fog and Edge Computing) ◆

18 May

- SP (IEEE Symposium on Security and Privacy) ◗
- FG (IEEE Int'l Conf. on Automatic Face and Gesture Recognition) ★
- IPDPS (IEEE Int'l Parallel and Distributed Processing Symposium) ◗

20 May

- ISMVL (IEEE Int'l Symposium on Multiple-Valued Logic) ▲

23 May

- ICSE (IEEE/ACM Int'l Conf. on Software Eng.) ▲

2469-7087/20

30 May
- ISCA (ACM/IEEE Int'l Symposium on Computer Architecture) ●

June

7 June
- ARITH (IEEE Int'l Symposium on Computer Arithmetic) ◗

14 June
- CVPR (IEEE Conf. on Computer Vision and Pattern Analysis) ◗

15 June
- ICHI (IEEE Int'l Conference on Healthcare Informatics) ●

16 June
- EuroS&P (IEEE European Symposium on Security & Privacy) ●

19 June
- JCDL (ACM/IEEE Joint Conf. on Digital Libraries) ▲

22 June
- CBMS (IEEE Int'l Symposium on Computer-Based Medical Systems) ◗
- CSF (IEEE Computer Security Foundations Symposium) ◗

26 June
- SmartCloud (IEEE Int'l Conference on Smart Cloud) ▲

29 June
- DSN (IEEE/IFIP Int'l Conf. on Dependable Systems and Networks) ●

30 June
- MDM (IEEE Int'l Conf. on Mobile Data Management) ●

July

6 July
- ICME (IEEE Int'l Conf. on Multimedia and Expo) ●

8 July
- ICDCS (IEEE Int'l Conf. on Distributed Computing Systems) ▲

13 July
- COMPSAC (IEEE Computers, Software and Applications Conference) ●

August

31 August
- RE (IEEE Int'l Requirements Eng. Conf.) ●

September

21 September
- ASE (IEEE/ACM Int'l Conf. on Automated Software Eng.) ◆

28 September
- ICSME (IEEE Int'l Conf. on Software Maintenance and Evolution) ◆
- SecDev (IEEE Secure Development) ◗

October

18 October
- MODELS (ACM/IEEE Int'l Conf. on Model Driven Eng. Languages and Systems) ◗

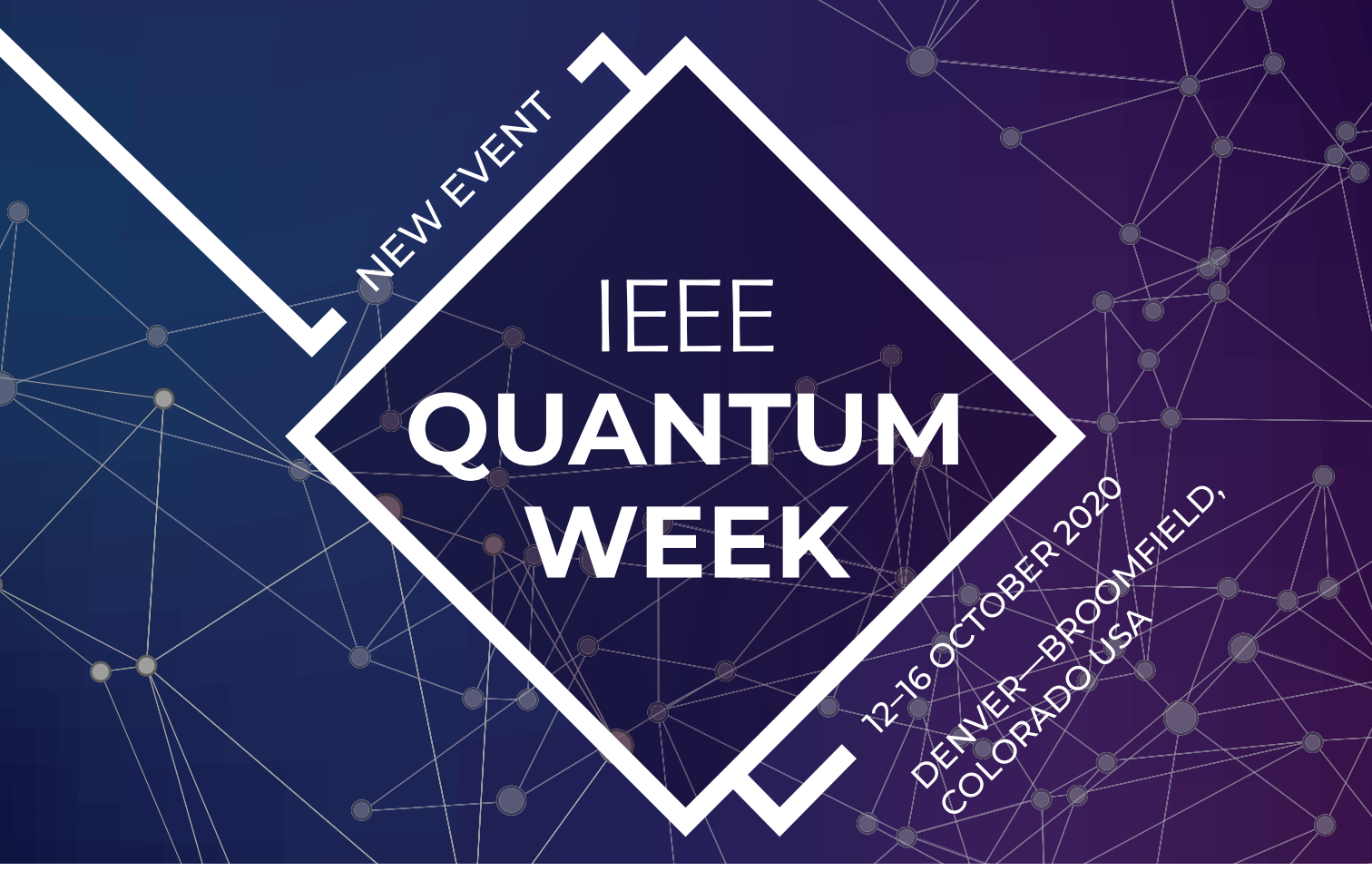
21 October
- FIE (IEEE Frontiers in Education Conf.) ●

25 October
- VIS (IEEE Visualization Conf.) ◗

November

15 November
- SC ◗

16 November
- FOCS (IEEE Symposium on Foundations of Computer Science) ◗
- LCN (2020 IEEE Conf. on Local Computer Networks) ◆



IEEE Quantum Week 2020 Is Open for Submissions

Participation opportunities are available for the inaugural IEEE International Conference on Quantum Computing and Engineering (QCE 2020) to be held 12–16 October 2020, in Denver—Broomfield, CO.

IEEE Quantum Week aims to be a leading venue for presenting high-quality original research, ground-breaking innovations, and compelling insights in quantum computing, engineering, and technologies.

Authors are invited to submit proposals for technical papers, posters, tutorials, workshops, and panels. Submission schedules are available at qce.quantum.ieee.org/important-dates.

IEEE Quantum Week includes the following technical paper tracks:

- Quantum Communications, Sensing, Cryptography
- Quantum Photonics and Optics
- Quantum Computing
- Quantum Algorithms & Information
- Quantum Applications and Simulating Nature
- Quantum Engineering
- Quantum Benchmarks & Measurements
- Quantum Education

Papers accepted by IEEE QCE will be submitted to the IEEE Xplore Digital Library. The best papers will be invited to the journals *IEEE Transactions on Quantum Engineering* (*TQE*) and *ACM Transactions on Quantum Computing* (*TQC*).

Submission instructions and details:
qce.quantum.ieee.org/callforcontributions

IEEE QUANTUM · IEEE COMPUTER SOCIETY · IEEE ComSoc IEEE Communications Society · IEEE CSC Council on Superconductivity · IEEE Photonics Society · IEEE