INNOVATION

with an impact

ON SOCIETY,

EDUCATION,

SCIENCE *and*

TECHNOLOGY

Information Sciences Institute

2016

ANNUAL REPORT

USC Viterbi

School of Engineering

SEPTEMBER 2017

Information Sciences Institute

is a world leader in research and development

of advanced information processing, computing

and communications technologies.

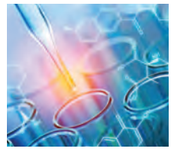
CALIFORNIA
4676 Admiralty Way #1001
Marina del Rey, CA 90292
310.822.1511

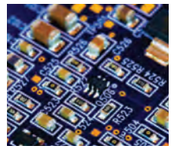
MASSACHUSETTS
890 Winter Street
Waltham, MA 02451
781-622-9790

VIRGINIA
3811 Fairfax Drive #200
Arlington, VA 22203
703.243.9422

Report Preparation Janice Wheeler (Editor); Yigal Arens, Prem Natarajan | **Design** Russo Design, Los Angeles | **Printing** Clear Image Printing

Information Sciences Institute





ISI researchers today drive revolutionary advances in a wide variety of areas ranging from machine learning and artificial intelligence to cybersecurity, from novel electronics to high-performance computing architectures and quantum computing, from heterogeneous computing environments to data science and scientific workflows, and from health informatics to forecasting of societal and cyber events. The age we live in is the most exciting time to be a researcher in the information and computing sciences. It is also a tremendously exciting time to be at ISI.

In the spring of 2017, ISI established a new research center in suburban Boston—the first geographical expansion of its operations in more than three decades. The establishment of this new center is a testament to the vitality and aspirations of the ISI community, as it seeks to further expand the scope and impact of its scientific and technological operations. During 2016 and early 2017, we witnessed a record number of new researchers joining the ISI family. In an environment of unprecedented and intense competition for technical talent, it is particularly satisfying that all three ISI locations—California, Virginia, and Massachusetts—have attracted exciting new talent from academia and from industry.

The times in which we live are, in substantial part, defined by the ascendancy of Artificial Intelligence (AI), which is poised to affect every aspect of human endeavor, from shopping to entertainment to manufacturing. Some even refer to it as the fourth industrial revolution. For more than four decades, ISI has been recognized as a world-class center of AI research, known for broad-ranging and enduring contributions to several areas including cognitive architectures, natural language processing and understanding, forecasting, computer vision, behavioral analytics, computational social science, scientific workflows, digital government, and biomedical informatics. This past year, Yolanda Gil, a research director at ISI and a research professor of computer science, was elected to serve as president of the Association for the Advancement of Artificial Intelligence (AAAI) for a two-year term beginning in 2018. AAAI is the world's premier professional society in AI, and Dr. Gil's election to the influential position is testimony to the leadership that ISI's researchers continue to exhibit in the professional communities to which they belong.

> "THE BEST WAY TO PREDICT THE FUTURE IS TO INVENT IT."
> — Alan Kay

As ISI prepares to celebrate its 45th anniversary, it is an appropriate time to reflect on the many and varied contributions that ISIers have made over the years to their chosen fields of endeavor, and to remember some of the key personalities in the Institute's history. *Keith Uncapher* founded the institute in 1972, *Herb Schorr,* as the second director of ISI, contributed to its transformation into a vibrant center of competitively funded research, *Jon Postel* led the request for comments (RFC) series in the early years of the Internet, *Paul Mockapetris* led the design and implementation of the Domain Name System (DNS), and *Danny Cohen* led the design of the network voice protocol (the direct precursor of voice-over-IP) and established the MOSIS fabless foundry in the early 1980s. I begin my fifth year as the executive director of ISI, drawing inspiration from its unique history of accomplishments and from the exciting prospects for its future.

Premkumar Natarajan
Michael Keston Executive Director, Information Sciences Institute
Vice Dean of Engineering, Viterbi School
Research Professor of Computer Science
University of Southern California

Information Sciences Institute's main offices in Marina del Rey are located on the coast of Southern California, with a view of the Pacific Ocean and the Santa Monica Mountains (13-story building to the right).

In Waltham, Massachusetts, offices for Information Sciences Institute are located close to major universities and numerous high-tech companies.

In Arlington, Virginia, offices for Information Sciences Institute are located in close proximity to Washington D.C. and Northern Virginia's technology centers.

ISI established the position of **Research Director** to recognize our leaders who head a sizable scientific research group, and develop, plan and direct the research and funding strategy for that group while also making critically valuable contributions to the current and future success of the Institute. During 2016, we had one new appointment to the position of ISI Research Director.

SHRI NARAYANAN

Shri is a Research Director at ISI and holds joint appointments at USC as a professor of electrical engineering and computer science (Viterbi School of Engineering), professor of linguistics, psychology, and neuroscience (Dornsife College of Letters, Arts and Sciences), and professor of pediatrics (Keck School of Medicine). He is also director of USC's Signal Analysis and Interpretation Lab and the Ming Hsieh Institute. His areas of research include signals and systems modeling, with an interdisciplinary emphasis on speech, audio, language, multimodal and biomedical problems and applications—all with direct societal relevance. Shri is a Fellow of the National Academy of Inventors, Acoustical Society of America, Institute of Electrical and Electronics Engineers, International Speech Communication Association, and American Association for the Advancement of Science.

ISI established the position of **Research Team Leader** to recognize researchers who not only lead a sizable scientific research group, but who are also instrumental in seeking out and recruiting other scientists and researchers to ensure ISI's technical integrity and its reputation as a leader in 21st-century technology. During 2016, we had one new appointment to the position of ISI Research Team Leader.

SCOTT MILLER

Scott Miller is a Research Team Leader at ISI with three decades of experience leading teams in disciplines such as information extraction, machine translation, forecasting, and information retrieval. Prior to joining ISI, Scott was a senior technical director with the Speech and Language Department at BBN, where he led Intelligence Community research efforts in information extraction and served as principal investigator on DARPA and IARPA efforts. Previous roles include founding Translingual Technologies, a start-up that created novel machine translation technologies, and serving as chief scientist at Basis Technology Corp. where he led the development of machine-learning methods for syntax-based machine translation and multilingual text processing.

The **Research Lead** position was established by ISI to acknowledge researchers who continuously make noteworthy contributions to their field of research and to the technological advancement of Information Sciences Institute. We are pleased to announce the 2016 appointment of five new Research Leads.

ELIZABETH BOSCHEE

Elizabeth is a Research Lead at ISI whose primary research areas include information extraction and language understanding. Prior to joining ISI, she was a lead scientist at BBN where she led numerous research efforts sponsored by DARPA, IARPA, ONR, and the IC. Her most recent work has focused on creating systems to automatically compile knowledge bases about persons, organizations, and events by reading the news or social media; these systems have been deployed to the DOD/IC community and been made broadly available to the research community by IARPA.

EMILIO FERRARA

Emilio is a Research Lead at ISI and a research assistant professor in USC's Viterbi School of Engineering. His field of research falls at the intersection of network science, data science, machine learning, and computational social science. Before joining ISI in 2015, Emilio was an assistant professor at Indiana University's School of Information and Computing and a research scientist at Indiana University's Network Science Institute. He is currently a member of ISI's Intelligent Systems Division, and is also the first member of ISI to be selected as a researcher-in-residence under the Michael Keston Researcher-in-Residence program.

IVAN SANCHEZ ESQUEDA

Ivan has been with ISI since 2011 and is a Research Lead in the Computational Systems and Technology Division. He is an active researcher in the field of nanoelectronics, having published over 30 papers in this field — including a Best Paper Award on the modeling of multi-gate field-effect transistors. Ivan's expertise lies in the physics of nanoelectronic materials and devices and the impact of defects in the operation, performance, variation, and reliability of devices and integrated circuits. Areas of research encompass computational nanoelectronics, electronic transport and scattering physics, low-dimensional semiconductors, quantum transport and quantum confinement effects, and physics-based simulations of semiconductor devices.

ANDREW SCHMIDT

Andrew has been a member of the ISI community since 2011, and is a Research Lead working with the Reconfigurable Computing Group in ISI's Computational Systems and Technology Division. Andrew's research expertise lies in reconfigurable computing and FPGA, computer architecture, high-performance computing, and embedded systems with a focus on trust and security. His work extends into many areas of research—including autonomous on-board processing for space-based systems with NASA. Andrew has coauthored a textbook in the area of embedded systems and reconfigurable computing.

FEDERICO SPEDALIERI

Federico joined ISI in 2010 after several years as a post-doctoral scholar at California's Jet Propulsion Laboratory and UC Los Angeles. He is a Research Lead at ISI and a research assistant professor in USC's Ming-Hsieh Department of Electrical Engineering. Federico has been a part of the USC-Lockheed Martin Quantum Computing Center since its inception, and he is currently bridging the gap between the quantum annealing paradigm and its applications to machine learning, natural language processing, scheduling and planning, and many other research areas. At ISI Federico is a member of the Computational Systems and Technology Division.

ISI RESEARCH TEAM IN WALTHAM, MASSACHUSETTS

Information Sciences Institute in Waltham, Massachusetts, hosts a dedicated and highly regarded team of computer science researchers. Back row: Joel Barry, Elizabeth Boschee, Huaigu Cao, and Xujun Peng. Front row: Jonathan Habif and Scott Miller.

The Institute Achievement Award is presented to members of the ISI community whose accomplishments and contributions are innovative, exemplary, and greatly add to the reputation and technological growth of ISI. This year awards were presented to three outstanding researchers at ISI's 2016 annual holiday celebration. Congratulations to the following recipients:

PEDRO SZEKELY

2016 Institute Achievement Award for addressing challenging knowledge graph research problems and translating advances into real-world applications.

Pedro has been directing the Domain-Specific Insight Graphs (DIG) project, exhibiting outstanding leadership and management skills. His vision and courage in tackling challenging problems at scale, his ability to manage large, heterogeneous research teams, and his commitment to the success of the program are models for ISI researchers.

EWA DEELMAN

2015 Institute Achievement Award for technical contributions and leadership in the field of scientific workflow systems for high-performance computing.

An innovator in the area of scientific workflow management, Ewa introduced the concept of decoupling the workflow description from the workflow execution, which enabled scientists to describe workflows in an abstract, resource-independent way that a planner could map to available distributed resources.

JOHN HEIDEMANN

2015 Institute Achievement Award for pioneering work in mapping and understanding the Internet.

John's pioneering work on Internet census (a way of counting live machines on the Internet through probing) has helped establish active probing as a viable, reliable source of host and network liveness, and a signal of Internet outages. John was the first to perform regular, wide sweeps of the Internet to detect live hosts, something that had previously been thought of as a non-viable method for measuring the Internet population.

CHIEF INFORMATION OFFICER **DEREK LAZZARO**

ISI's Computing and Information Services Division (CIS) supports the Institute's research mission by providing cutting-edge IT infrastructure and services.

Lead by recently hired Chief Information Officer Derek Lazzaro, the CIS team delivers a diverse catalog of IT services that enable collaborative research, innovation, and effective cybersecurity.

The CIS has a significant mission scope, with responsibility for nearly 2,000 computers, multiple petabytes of data, several data centers, and three geographically diverse sites.

Within the last year, the CIS team launched a major infrastructure refresh project. With the recent upgrades, ISI researchers now benefit from dense 100-gigabit data center networking, 100-gigabit next-generation firewalls, and even 100-gigabit WAN connectivity to the USC campus and the Internet. Other key systems have also been upgraded and modernized, including the virtual computing cluster, the ISI identity management system, and the ISI website.

A major focus for the CIS team is cybersecurity. Working collaboratively with researchers, the CIS team helps create resilient plans to protect sensitive data, comply with regulations, and defeat the advanced threats that target research organizations.

CHIEF FINANCIAL OFFICER **JAMES WHALEN**

Information Sciences Institute is a highly dynamic environment that competes vigorously for research funding—primarily from a number of federal and state agencies, but also from foundations, institutions, and industry. Federal funding sources range from the academic, such as the National Science Foundation, to those with applied research programs, such as the National Institutes of Health, to those with focused research applications, such as agencies within the US Intelligence Community.

ISI's Administrative Services Division plays a significant role in creating a competitive advantage and assisting the ISI research community in understanding and fulfilling all administrative requirements surrounding both proposals and awards. While it is necessary to provide exemplary and innovative research to meet our customer's expectations, it is also crucial to understand the administrative requirements of funding agencies and to help research teams meet those in a timely and efficient manner that generates fully documented compliance. The division's organizational structure and staffing, and its related business processes and management tools, are designed to support these administrative objectives.

Our **PreAward Team** supports both simple one-task proposals and very large proposals involving dozens of tasks, multiple subcontractors and time-phased summaries. The team also assists in USC's negotiations with diverse research sponsors. The **Finance and Accounting Group** provides assistance with manpower planning for research teams—both small and large. This is correlated with monthly forecasts to facilitate optimal application of resources within funding constraints and research objectives.

Overseeing all aspects of Administrative Services is Chief Financial Officer James Whalen. Following a successful 19-year career in the aerospace industry, James joined ISI in 1998. He is dedicated to sound financial management and to applying tested management techniques, such as contingency theory, to optimize interactions among staff, business processes and suites of software applications.

GREPSEC Workshop organizers and attendees. ISI's Terry Benzel, a key workshop facilitator, can be seen in row 2, second from left.

ISI AND THE GREPSEC WORKSHOPS

Information Sciences Institute has enjoyed a four-year relationship with the NSF-supported GREPSEC Workshop. The Institute has been involved since the workshop's inception in 2013 and during its subsequent development—all the while guiding its specific goals of encouraging and supporting more women and underserved groups who have an interest in computer security, and exposing them to the wide spectrum of areas that need attention. As one of ISI's top research scientists puts it: "We all benefit from a diverse set of experts, as with any field. There is quantifiable strength in diversification."

Computer security is a wide field, encompassing security for networks, operating systems, and software systems; language-based security; security for forensics and privacy; and the legal and policy issues inherent in all computer security systems. GREPSEC covers these varied topics, drawing on expertise from industry and academia, and the workshops have been structured to show the wide range of problems within the field—as well as some of the attempted solutions. The ratio of speakers to students is deliberately high since we want to encourage one-on-one mentoring. GREPSEC has been successful in bringing both students and faculty together, and there is strong evidence that these workshops are effective in encouraging more female and underrepresented graduate students to attend top research conferences.

Tameem Albash, Federico Spedalieri and Itay Hen
ISI's Publications Recognition Award | honors the achievements of ISI's most prolific authors. This quantum computing group published a total of 11 articles.

Ewa Deelman
2015 Institute Achievement Award | *for technical contributions and leadership in the field of scientific workflow systems for high-performance computing.*

Emilio Ferrara
Junior Scientific Award | from the Complex Systems Society *for the design and application of novel network-science models, algorithms, and tools to study phenomena occurring in large, dynamical techno-social systems.*

Yolanda Gil
President-Elect | Association for the Advancement of Artificial Intelligence (AAAI), an association with more than 4,000 members, June 2016.

John Heidemann
2015 Institute Achievement Award | *for pioneering work in mapping and understanding the Internet.*

Kevin Knight, Marjan Ghazvininejad, and Xing Shi
Computer-Generated Poetry | winning entry in contest designed to test whether a computer poet can pass as a human poet. Dartmouth College, April 2016.

Kuan Liu, Xing Shi, Anoop Kumar, Linhong Zhu, and Prem Natarajan
2016 ACM RecSys Challenge | *5th place winning team (out of over 100 competing teams).*

Daniel Moyer (advisor Greg Ver Steeg)
Young Scientist Award | for *"A Continuous Model of Cortical Connectivity"* which constructs a new model for brain connectivity that can assist neuroscientists in diagnosing brain diseases and treatment effects (Medical Image Computing and Computer Assisted Intervention Society Conference, Athens, 2016).

Pedro Szekely
2016 Institute Achievement Award | *for addressing challenging knowledge graph research problems and translating advances into real-world applications.*

Joseph Touch
Outstanding Service Award | IEEE Communications Society, Internet Technical Committee.

Linhong Zhu, Majid Ghasemi-Gol, Pedro Szekely, Aram Galstyan, and Craig Knoblock
Best Research Paper Award | for "Unsupervised Entity Resolution on Multi-type Graphs," 2016 International Semantic Web Conference (ISWC).

David Barnhart
Guest Speaker | Advanced Manufacturing Partners for Southern California Exhibit at Space Technology Expo.

Terry Benzel
Co-organizer | NSF-supported GREPSEC workshops whose mission is to encourage and support more women and underrepresented groups interested in computer security, and to introduce them to the wide spectrum of areas that need attention.

Gully Burns
Research Topic Editor | Frontiers in Neuroscience—"Discovery Informatics in Neuroscience—Extracting & Shaping Actionable Knowledge from Texts and Data."

Active Reviewer | Frontiers System

Working Group Leader | FORCE11 "Force Protocols."

Stephen Crago
Chair, Steering Committee | IEEE Transactions on Big Data.

Steering/Organizing Committee | Fault Tolerance Spaceborne Computing Employing New Technologies.

Program Committee | Area Chair for Systems Software and Hardware—IEEE Cloud.

Program Committee | International Conference on Edge Computing.

Senior Member | IEEE

Ewa Deelman
Associate Editor | IEEE Transactions on Parallel and Distributed Systems.

Workshop Chair | SC'16 Technical Program, Salt Lake City, November 2016.

Posters Chair | Cluster 2016, Taipei Taiwan, September 2016.

Senior Member | IEEE

Matthew French
Senior Member | IEEE

General Chair & Program Committee Member
2016 Int'l Symposium on Field Programmable Custom Computing Machines.

Yolanda Gil
General Chair | International Semantic Web Conference (ISWC) 2016.

Chair | ACM's Special Interest Group in Artificial Intelligence (SIGAI).

Advisory Board | Semantic Web Science Association (SWSA).

Co-Chair | Technology and Architecture Committee of the NSF EarthCube Program.

Founding Member | Leadership Council of the NSF EarthCube program.

Founding Editorial Board Member | Journal Data Science: Methods, Infrastructure, & Applications.

Advisory Board Member | Computer Science Section, MethodsX Journal.

Founding Editorial Board Member | AI Matters (ACM SIGAI Quarterly Newsletter).

Editorial Boards | Artificial Intelligence Journal; Journal of Web Semantics; Applied Ontology Journal.

Kevin Knight
General Chair | Conference of the North American Association for Computational Linguistics (NAACL).

Kristina Lerman
Editorial Boards | Nature Scientific Reports; ACM Transactions on Internet Technology.

Organizer | Social Informatics 2016; IPAM Analytics Workshop; ICWSM Tutorials Chair.

Andrew Schmidt
Co-chair | High Performance Computing Tract—Int'l Conference on Reconfigurable Computing and FPGAs.

Technical Program Committee | Int'l Symposium on Field Programmable Custom Computing Machines.

Reviewer | ACM Transactions on Reconfigurable Technology and Systems (TRETS).

Joseph Touch
Tutorial | "Root Causes of Communication Latency," IEEE CCNC 2016.

Chair | OSA Nonlinear Optics Technical Group.

Organizer | DOE Networking 2025 organizing committee.

Team Leader | IANA Ports Expert Review.

Guest Editor | Special issue on Optical Computing, *Nanophotonics*

Panelist | IEEE Rebooting Computing Workshop.

Board of Advisors/Steering Group | Asia Future Internet Forum (AsiaFI).

Committee Member | Internet Society IRTF Applied Network Research Prize.

Member | IETF Transport Area Review Triage (ART).

Traveling Lecturer | Selected by the Optical Society.

Gabriel Weisz
Technical Program Committees | Int'l Conference on Reconfigurable Computing and FPGAs; Intl Conference on Field Programmable Logic and Applications.

ADMINISTRATIVE STAFF	
Lida Dimitropoulou	Project Manager
Derek Lazzaro	Chief Information Officer, Computing & Information Services
Lechelle Revilla	Accountant I
Crisca Robles	Accountant I

SCIENTISTS AND ENGINEERS	
Sunny Boonsook	System Admin / DevOps Engineer II
Elizabeth Boschee	Senior Computer Scientist / Research Lead
Martin Brooks	System Admin / DevOps Engineer IV
Huaigu Cao	Computer Scientist
Joshua Chudy	Programmer Analyst II
Kaushik Datta	Computer Scientist
Wesley Hardaker	Senior Computer Scientist
Amir Hashmani	System Admin / DevOps Engineer IV
Christophe Hauser	Computer Scientist
Vivek Kale	Computer Scientist
Mayank Kejriwal	Computer Scientist
Victor Mendez, Jr.	Network Admin/DevOps Engineer IV
Scott Miller	Supervising Computer Scientist / Research Team Leader
Joshua Monson	Computer Scientist
Heidi Morgan	Senior Computer Scientist
Shri Narayanan	Research Director
Scott Novotney	Computer Scientist
Xujun Peng	Research Programmer II
Zheng Tang	Research Programmer I
Hongsuda Tangmunarunkit	Computer Scientist
Joshua Zusman	VLSI Design Engineer

POSTDOCTORAL SCHOLARS	
Peter Fennell	Understanding information diffusion in online social networks
Daniel Garijo	Provenance, scientific workflows, linked data, semantic web, artificial intelligence
K.S.M. Tozammel Hossain	Data mining and algorithmic techniques in computational biology and social networks

DOCTORAL STUDENT	RESEARCH AREA	ADVISOR
Nazanin Alipoufard	Computational behavior	Kristina Lerman
Zoe Gonzalez Izquierdo	Quantum annealing architectures	Federico Spedalieri, Itay Hen
Roelof Groenewald	Classical optimization techniques	Federico Spedalieri, Itay Hen
David Kale	Machine learning in high-impact domains	Greg Ver Steeg
Brendan Kennedy	Machine learning and forecasting	Prem Natarajan
Hannes Leipold	Quantum computing	Federico Spedalieri
Zekun Li	Computer vision	Prem Natarajan
Jeremy Liu	Machine learning and quantum computing	Michael Orosz
Daniel Moyer	Machine learning, connectomics, network theory	Greg Ver Steeg
Aqib Nasar	Network measurement	John Heidemann, Ethan Katz-Bassett
Sivaramakishman Ramanathan	SDN and security	Jelena Mirkovic
Stephen Rawls	Computer vision and optical character recognition	Prem Natarajan
Yuan Shi	Machine learning to adapt to sensor failures	Craig Knoblock
Rajat Tandon	Network security	Jelena Mirkovic
Nazgol Tavabi	Cyber attack forecasting	Kristina Lerman
Geoffrey Tran	Fault tolerance in cloud-based analytics	Stephen Crago
Binh Vu	Mining online sources to predict cyber attacks	Craig Knoblock
Xin-Zeng Wu	Network structure and dynamics	Kristina Lerman
Kun Yue	Audio signal processing	Young Cho
Xiaofan Zhang	Classical optimization techniques	Federico Spedalieri, Itay Hen

PH.D. GRADUATES	ADVISOR	RESEARCH	CURRENTLY AT
Farshad Kooti	Kristina Lerman	Predicting and modeling human behavioral changes using digital traces	Facebook Data Science Group
David Lee	Jeffrey Draper	Modeling the reliability of highly scaled field-programmable gate arrays in ionizing radiation	Sandia National Labs
Paul Wood	Alefiya Hussain, Saurabh Bagchi	Improving the resilience of cyber-physical systems under strategic adversaries	Research Scientist, Purdue University

M.S. GRADUATES	ADVISOR	RESEARCH	CURRENTLY AT
Nada Aldarrab	Kevin Knight	Decipherment of historical manuscripts	Univ. Southern California

NEW MASTER'S AND UNDERGRAD STUDENTS

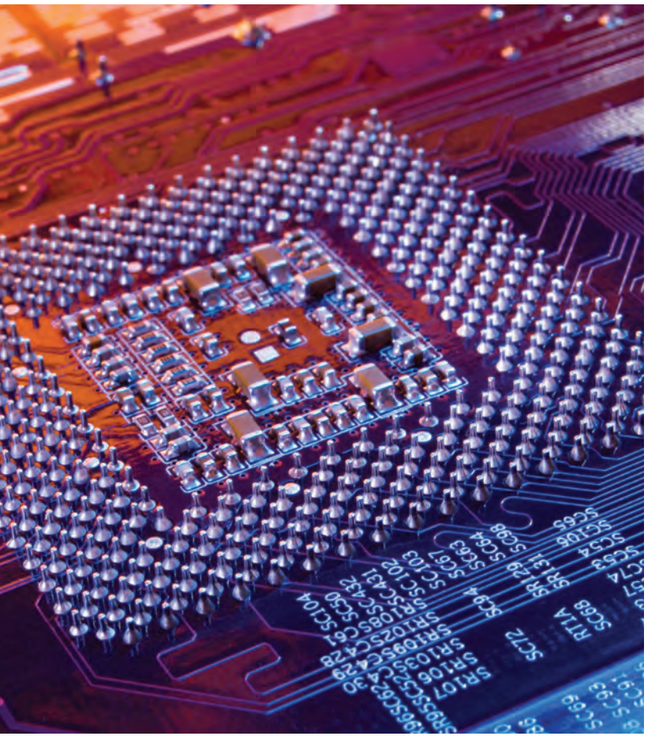
STUDENT	ADVISOR	STUDENT	ADVISOR
Tushar Agarwal	Kristina Lerman	Joel Mathew	Prem Natarajan
Nada Aldarrab	Kevin Knight	Dipa Maulik	Craig Knoblock
Suresh Alse	Craig Knoblock	Neil Mehta	Itay Hen
Sumedh Baikady	Young Cho	Saurabh Mishra	Yolanda Gil
Vaishnavi Bharadwaj	Kevin Knight	Hirak Modi	Young Cho
David Brenn-Cogen	Craig Knoblock	Palak Modi	Craig Knoblock
Xiaotian Chen	Young Cho	Ujwal Mysore	Young Cho
Yinyi Chen	Craig Knoblock	Sarvotham Pai	Kevin Knight
Nilay Chheda	Craig Knoblock	Abhinav Palia	Jelena Mirkovic
Shreya Chowdhury	Joseph Touch	Weiwu Pang	Young Cho
Alyssa Deng*	Yolanda Gil	Jay Patel	Prem Natarajan
Yi Ding	Craig Knoblock	Aniket Pednekar	Craig Knoblock
Matthew Dinh	Young Cho	Jay Priyadarshi	Kevin Knight
Ismail Enchikalathil	Prem Natarajan	Wayne Quadros	Joseph Touch
Schuyler Fried	Itay Hen	Maya Ram	Craig Knoblock
Nikunj Gala	Craig Knoblock	Jaydeep Ramani	Jelena Mirkovic
Jiachang Ge	Young Cho	Guarangi Raul	Craig Knoblock
Ameya Hanamsagar	Jelena Mirkovic	John Runburg	Itay Hen
Rebecca Hao	Craig Knoblock	Senthil Sampath	Prem Natarajan
Eric Heiden	Craig Knoblock	Yash Singh	Prem Natarajan
Yao Guang Hoh	Young Cho	Yongchao Shang	Young Cho
Priyambada Jain	Kristina Lerman	Naiqing Song	Kevin Knight
Rashit Jain	Craig Knoblock	Nikhila Sreekanth	Craig Knoblock
Suraj Jayakumar	Craig Knoblock	Archit Srivastava	Kristina Lerman
Jason Jewik*	Yolanda Gil	Karishma Surekha	Craig Knoblock
Yibo Ji	Young Cho	Zheng Tang	Craig Knoblock
Aditya Joshi	Prem Natarajan	Vincentius Utanto	Young Cho
Gaurav Kashyap	Young Cho	Catherine Wang*	Yolanda Gil
Geoffrey Knopf	Alefiya Hussain	Xunzhi Wang	Young Cho
Ravi Raju Krishna	Craig Knoblock	Yubo Wang	Young Cho
Yeshavanta Kubusada	Prem Natarajan	Jordan Woods	Young Cho
Sanjana Lakkadi	Kristina Lerman	Hao Wu	Craig Knoblock & Michel Sika
Vishnukarthik Lakshmanan	Gully Burns	Le Xiao	Jelena Mirkovic
Stephen Lee	Michel Sika & Stephen Crago	Yao Xiao	Young Cho
Yiye Lin	Prem Natarajan	Jung-Jung Yeh	Kevin Knight
Surabhi Lodha	Craig Knoblock	Brian Zhang	Itay Hen
Jeff Marshal	Itay Hen	Barret Zoph	Kevin Knight
Joe Mathai	Prem Natarajan	Sizhuo Zou	Yolanda Gil

** Denotes high school student*

VISITING SCHOLARS AND RESEARCHERS

SCHOLARS AND RESEARCHERS	ISI HOSTS
Laura Alessandretti, Director's Intern, City University London *Applied mathematics*	John Paul Walters John Wroclawski
Alessandro Bessi, Postdoc, IMT Institute for Advanced Studies *Cyber attack forecasting*	Kristina Lerman
Keith Burghardt, Postdoc, University of Maryland *Behavioral data science*	Kristina Lerman
Gabriel Durkin, Visiting Scholar, UC Berkeley *Quantum metrology*	Federico Spedalieri Itay Hen
Peter Fennell, James S. McDonell Foundation Post-Doc, United Kingdom *Information diffusion*	Kristina Lerman
Rosa Filgueira, Visiting Scholar, University of Edinburgh, United Kingdom *Data-intensive computing and workflow systems*	Ewa Deelman
Luciano Gallegos Marin, Visiting Scholar, Brazil *Sentiment analysis, urban analytics*	Kristina Lerman
Floran Geigi, PhD, Visiting Scholar, Graz University, Austria *Information integration*	Kristina Lerman
Guilherme Pereira Gribeler, Florida Institute of Technology *Undergraduate student (summer)*	Ewa Deelman
Samrat Jha, University of Maryland, College Park *Undergraduate student (summer)*	Ewa Deelman
Vishnukarthik Lakshmanan, Visiting Scholar, University of Southern California *Web programming; TechKnAcq interface*	Gully Burns
Allon Percus, Visiting Scholar, Claremont Graduate University *Network science*	Kristina Lerman
Victoria Stodden, Associate Professor, University of Illinois Urbana Champaign *Reproducibility in computational science*	Yolanda Gil

THE SECURE AND ROBUST ELECTRONICS CENTER (SURE) | DIRECTOR MATTHEW FRENCH



Global production trends have introduced significant vulnerabilities into electronics hardware. Integrated circuits that once were created by a single company, possibly under a single roof, are now produced by "fabless foundries" with few or no facilities of their own. Instead, state-of-the-art production relies on a complex, international design and manufacturing process. The same development flow and novel physics that are pushing chips to new levels also expose new security risks.

Designs may be generated with tools from multiple software vendors. Source code may be written by different companies, possibly in different countries. Key technical elements may be licensed from third parties. Foundries and their customers rarely know with absolute certainty whose hands a chip has passed through, whether a latent issue has occurred, and how the chip will perform in all circumstances.

In addition, modern nanoscale fabrication is adding complexity to reliability and resiliency issues. The performance of individual transistors has become increasingly variable, making overall system reliability challenging. Shrinking voltage margins increase susceptibility to errors that historically were seen only in space environments. Additionally, wear-out and aging effects appear sooner in the lifetime of integrated circuits and with more variability.

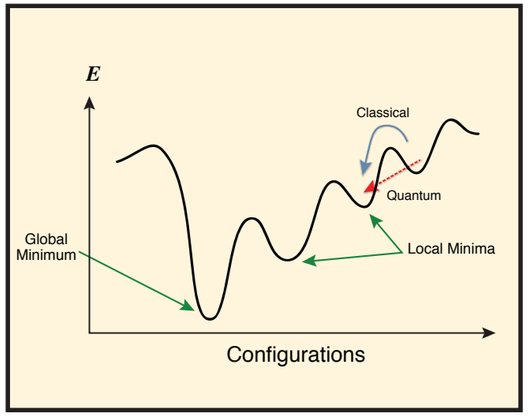
The issues could have profound consequences for everything from misfires or failures of major U.S. infrastructures, both physical and virtual, to individual smartphone performance.

USC's Information Sciences Institute established the *Secure and Robust Electronics Center (SURE)* to address these issues. SURE seeks to investigate and address state-of-the-art manufacturing complexities that have significantly compromised four key aspects of chip production: trust, security, resilience and reliability. SURE researchers have over a decade of experience in this emerging field, with supporters that include DARPA, IARPA, NASA, Army Research Office, Air Force Research Laboratory and Defense Threat Reduction Agency. As experienced industry collaborators, we aim to accelerate the scale, pace and impact of hardware robustness and security technology development; that which were once military and aerospace concerns, are now rapidly expanding into the commercial sector.

The SURE Center combines the research thrusts of ISI with MOSIS's fabless foundry capabilities, enabling SURE to address full system device design and production, including front-end design and FPGA programming, microarchitecture, integrated circuits, CAD, and fabrication.

USC-LOCKHEED QUANTUM COMPUTATION CENTER
DIRECTORS DANIEL LIDAR, ROBERT LUCAS, AND FEDERICO SPEDALIERI

Faculty, researchers and students are performing basic and applied research into quantum computing and are collaborating with researchers around the world. The *USC-Lockheed Quantum Computation Center* (QCC) houses a D-Wave 2X quantum annealing system, manufactured by D-Wave Systems, Inc. QCC was the first organization outside of D-Wave to house and operate its own system, and it has conducted research on three different generations of the processor. Currently, there are only two other operational systems in the US—one installed at NASA Ames Research Center and operated jointly by NASA and Google, and another at Los Alamos National Laboratory. Operating quantum computing systems is demanding: the systems need to be kept at near absolute temperature and electromagnetically shielded to protect the fragile quantum states from degradation by external noise.



USC-Lockeed Quantum Annealing

THE POSTEL CENTER FOR EXPERIMENTAL NETWORKING | DIRECTOR JOE TOUCH

The *Postel Center for Experimental Networking* is an endowed facility focusing on network research and service to the Internet community. In its early years, the Center focused on Internet history and supporting network research tools and resources through visiting scholars. Recently, we explored Internet policy matters and issues underlying our understanding of the Internet architecture and its protocols. We developed reference documents for the Internet community to clarify architectural concepts (tunnels, middleboxes, GRE fragmentation, transport port use) and to extend legacy protocols for new uses (both TCP and UDP option space extensions).

The Postel Center is named in memory of **Dr. Jon Postel**—a brilliant and dedicated scientist who made many key contributions to the formative days of the ARPANET, including protocol design and verification, multimedia computing and communications, electronic commerce, the domain name system, and many other specific Internet protocols. Jon was widely known for the influence he exerted on the management of the Internet, recognizing early on that packet-switching research would need organization and a modicum of discipline if it were to realize its full potential as a universal communication medium. Jon created the activities that eventually grew into the RFC Editor, which issues and controls the many documents that specify how Internet computers interoperate, and also initiated (circa 1981) the Internet Assigned Numbers Authority (IANA), the central coordination function for the global Internet.

Currently we are exploring new missions and future directions, and moving away from support for shared resources that are increasingly available via both public and commercial enterprises. This past year we refocused the Center to expand on ISI's rich history in developing and managing Internet resources. This includes developing materials to educate the Internet community on "best practices" for network conservation, e.g., in the documentation of protocols and management of shared resources, such as number and name spaces. Additionally, we are establishing new collaborations that will help us explore and document ISI's role in Internet history.

ISI CENTER FOR COMPUTER SYSTEMS SECURITY | DIRECTOR CLIFFORD NEUMAN

The ISI *Center for Computer Systems Security* conducts research and provides education in the crucial disciplines of computer, network, and application security. Among the current research activities, Center staff are developing new architectures for isolation in networked systems, and they are studying resilience to cyber-attacks in the critical infrastructure systems of the power grid and oil and gas extraction.

The work on critical infrastructure demonstrates how cyber-attacks affect the operational resilience of the infrastructure, impeding the delivery of power to consumers or oil and gas to refineries. It also identifies remediation strategies among candidate actions. Center researchers are exploring cross-infrastructural dependencies to identify how attacks in one area impact the resilience of other infrastructures.

In addition to its research activities, Center researchers are frequently called upon by the media to explain events involving privacy, cyber security, and cyber crime. In the past year, Center Director Clifford Neuman was quoted in more than 50 publications and appeared more than 15 times in television and national network news segments and radio programs, where topics ranged from security of the Internet of Things, WikiLeaks' disclosure of hacking techniques, breaches of the Dallas emergency warning notification system, security of our election systems, and hacks that occurred during the 2012 campaigns.

ISI's Center for Computer Systems Security was targeted in 2011 as a DHS- and NSA-designated Center of Academic Excellence in information assurance research. Center staff lead several education programs in computer security and have redesigned the master's of science degree in security engineering to apply the existing focus on high assurance systems to the Internet and cloud-focused computing environments now in demand by consumers and security. A new cross-disciplinary class was also introduced that focuses on privacy and the practical legal and international considerations of computer security in today's and future systems.

LEIDOS AND INFORMATION SCIENCES INSTITUTE COLLABORATION

Leidos, Inc., a national security, health, and engineering research firm headquartered in Arlington, VA, and Information Sciences Institute are collaborating to provide scientific, engineering, systems integration and technical services to the U. S. Defense and Intelligence communities and other civil agencies, and to selected commercial markets. Leidos and researchers at Information Sciences Institute are currently working together on the following significant research effort.

Evaluation of Mapping NLP-Based Algorithms into Reconfigurable Computing Kernels (EMBARK)

Human-computer interaction has evolved significantly with the introduction of digital assistants such as Siri, Cortana, and Alexa opening the door to what next-generation computing platforms will be capable of achieving. At the heart of these technologies is natural language processing (NLP) which allows a computer to extract meaning from natural language input instead of structured computer code, such as C++ or Java. The complex algorithms developed to enable such capabilities as information extraction, information retrieval, and sentiment analysis are pushing the envelope of existing computing platforms. CPU- and GPU-based systems have typically been leveraged for their ease of programmability; however, these systems still have performance and power limitations. EMBARK has evaluated how reconfigurable computing, using field programmable gate arrays (FPGAs), can accelerate not only the computation, but improve the development productivity by leveraging NLP community toolkits to systematically provision FPGA parallel resources, and support dynamic reconfiguration and run-time management. NLP algorithms and neural nets were evaluated, matching computation problems to the FPGA resources that will most effectively boost that software's performance and productivity. EMBARK achieved improvements in terms of the framework, front-end analysis, system assembly and run time. Results included improvements in clock speeds and programmability by a factor of up to 10X, loosening of resource constraints, and better run-time management. The research team now hopes to integrate EMBARK with user applications and get the results into the hands of actual developers.

NORTHROP GRUMMAN CYBERSECURITY RESEARCH CONSORTIUM

Northrop Grumman's Cybersecurity Research Consortium (NGCRC), founded in 2009, is a groundbreaking partnership of industry and academia formed to advance research, facilitate collaboration among the nation's top scientists, and accelerate solutions to counter the fast-changing threats from cyberspace. The consortium addresses some of the world's leading cyber problems, including attribution in cyberspace, supply chain risk, and securing critical infrastructure networks. Members of the consortium coordinate research projects, share information and best practices, develop curricula, write joint case studies, and provide numerous learning opportunities and applications for students and the defense community overall.

USC's Viterbi School of Engineering joined the Northrop Grumman Cybersecurity Research Consortium in 2013, with Information Sciences Institute serving as the lead organization for establishing and advancing the partnership and expanding the consortium's breadth of investigation into the most pressing cyber threats to the economy and national security. ISI brought with it a strong reputation for leadership in big data, cybersecurity, computer science, and informatics. The other three partners in this research consortium are also leading cybersecurity research universities—Carnegie Mellon University, Massachusetts Institute of Technology, and Purdue University.

Through NGCRC, ISI researchers are developing a capability for retroactive cross-site sharing of cybersecurity datasets, effectively replaying the relevant portions of data after a determination is made regarding which details are most relevant, and while still complying with policies on data sharing. In other work with NGCRC, ISI researchers are collaborating with Viterbi's Energy Institute and the Center for Energy Informatics to deploy tools to assess operational resilience in critical infrastructure (the smart grid and oil and gas distribution), and develop game-theoretic models to mitigate the impact of attacks on these systems.

Collaborative research aimed at securing critical infrastructure networks.

COMPUTATIONAL SYSTEMS AND TECHNOLOGY | DIRECTOR STEPHEN CRAGO

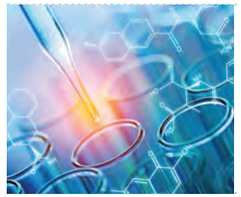
A world leader in computing technologies—from basic to applied research—the CS&T Division conducts research on quantum computing, advanced medical devices, onboard processing for space missions, radiation mitigation, smart grids, cloud computing, and integrated circuits.

INTELLIGENT SYSTEMS | DIRECTOR YIGAL ARENS

One of the world's largest artificial intelligence groups, the IS Division focuses on natural language processing, knowledge technologies, information integration, robotics, machine intelligence, biomedical knowledge engineering and data integration, and multimedia analysis.

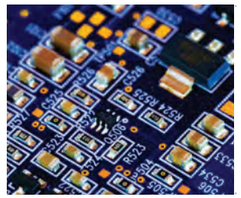
INTERNET AND NETWORKED SYSTEMS | DIRECTOR JOHN WROCLAWSKI

The I&NS Division focuses on Internet measurements, cyber-physical systems like the smart grid, cybersecurity challenges facing the Internet, methodologies and tools for large-scale cyber and cyber-physical networked systems, clean-break initiatives, and the Domain Name System.



INFORMATICS SYSTEMS RESEARCH | DIRECTOR CARL KESSELMAN

This division has a broad research agenda focusing on creating new types of sociotechnical systems that enable and accelerate discovery in domains of high societal impact. It specializes in highly collaborative user-driven research in which work is evaluated in the context of operational, high-impact science.



MOSIS | DIRECTOR WES HANSFORD

MOSIS is an efficient, affordable way to prototype and volume-produce multi-project wafers (MPW) and related services that drive IC innovation. From design spec interpretation through mask generation, device fabrication and onto assembly, MOSIS is a trusted expert interface to the semiconductor ecosystem.

DIRECTOR **STEPHEN CRAGO**

ISI is one of the world's leaders in computing technologies, ranging from basic to applied research. Our Computational Systems and Technology (CS&T) Division is focused on:

BIG DATA

DECISION SYSTEMS

HETEROGENEOUS CLOUD AND EMBEDDED COMPUTING

MICROARCHITECTURE, INTEGRATED CIRCUITS, AND ADVANCED ELECTRONICS

QUANTUM COMPUTING

RECONFIGURABLE COMPUTING AND WIRELESS NETWORKS

SCIENCE AUTOMATION TECHNOLOGIES

SYSTEM SOFTWARE AND COMPILERS

TRUSTED ELECTRONICS AND COMPUTING

From Theoretical to Hands-On

Our current initiatives include theoretical adiabatic quantum computing through the USC-Lockheed Martin Quantum Computation Center and hardware security through ISI's Secure and Robust Electronics Center. CS&T projects include system software for heterogeneous clouds and hardware-software design of unique chips and field-programmable gate arrays. We're also exploring algorithms and data structures to help understand and control large-scale complex systems such as oil field geometry, megacity emergency response and legacy software security.

CS&T teams are creating wireless networking technologies for battlefields and other difficult environments, along with social media platforms for people who lack trustworthy Internet access. We're advancing our scientific automation tools—which enable researchers to focus on conducting science, not managing data—already used by astronomers, physicists and earthquake specialists, including the LIGO team that was the first to detect the gravitational waves that Einstein predicted.

Intellectual Leadership, Diverse Expertise

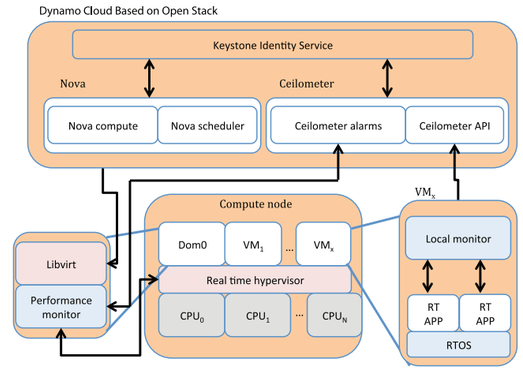
We provide intellectual leadership within the broader research community through participation in workshops like the White House OSTP National Strategic Computing Initiative Workshop, the IEEE Rebooting Computing Initiative, and trusted electronics workshops run by the SURE Center.

Our 40 computing technology researchers, research programmers and graduate students represent a wide range of disciplines, including electrical engineering, computer science, physics, and math. Led by Stephen P. Crago, CS&T researchers are based in Arlington, Virginia and in Marina del Rey, California.

Dynamo: A Framework for Dynamic Real-Time Workloads in a Reliable Cloud

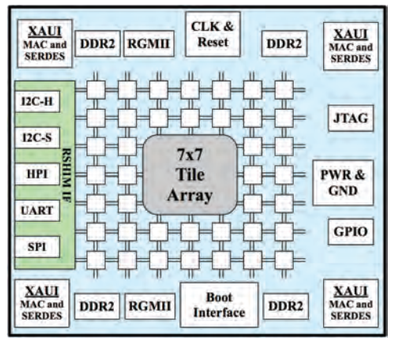
Today's clouds are capable of high performance at tremendous scale. Unsurprisingly, the cloud represents a high-value target for cyber attackers interested in compromising cloud services and degrading critical infrastructure. The current cloud model pushes the responsibility of reliability onto the end-user, effectively forcing users into existing platform models in order to provide a degree of resilience. While this model is generally effective for existing IT-oriented applications (e.g., databases, directory services, general-purpose computation, etc.), it is poorly suited for mission-critical applications composed of heterogeneous architectures, real-time requirements, and critically, a spectrum of reliability requirements.

The *Dynamo* team is developing tools, APIs, and cloud services to enhance the reliability of real-time and heterogeneous workloads in a real-time cloud. Using a dynamic resource management framework, Dynamo enables run-time resource reallocation to enhance the performance and reliability of real-time applications. Dynamo enables applications to leverage a spectrum of resiliency strategies, allowing adaptive performance and reliability trade-offs to occur dynamically and automatically.



The Dynamo cloud and real-time resource management framework

Dynamo is a component of the cloud computing group at ISI who are finding ways to integrate heterogeneous computing into dynamic public and private secure cloud platforms for data centers, high-performance computing, and embedded computing. ISI is one of the first organizations to develop support for heterogeneous computing in the OpenStack cloud environment, and is now developing support for real-time applications and the deployment of such technology to new application domains.



The Maestro Processor

Maestro: A Radiation Hardened Space Processor

As Earth-orbiting satellites generate increasingly large datasets, new approaches to data ingest and processing must be considered. One approach is to increase the satellite-to-Earth downlink capacity, and process these datasets on the ground. A second approach, however, would allow the satellite to perform processing and/or preprocessing on board. This would reduce downlink requirements, and may improve the timeliness of the data through autonomous satellite operation.

The Maestro processor is a 49-core radiation-hardened-by-design (RHBD) processor for space missions. Maestro is derived from the Tilera TILE64 processor, inheriting a full complement of high-speed interfaces, a complete Linux distribution and a development toolchain. Our team of researchers is developing system software for the Maestro processor, including enhancements to the processor's compiler. Most visibly, USC/ISI is extending the Maestro toolchain with GCC compiler support, which will enable full OpenMP support and additional language front-ends. The switch to GCC has already resulted in improved performance for many applications. In addition, both OpenMP and the additional language front-ends will dramatically improve software portability to the Maestro processor.

Radiation Mitigation Through Arithmetic Codes (RadMAC)

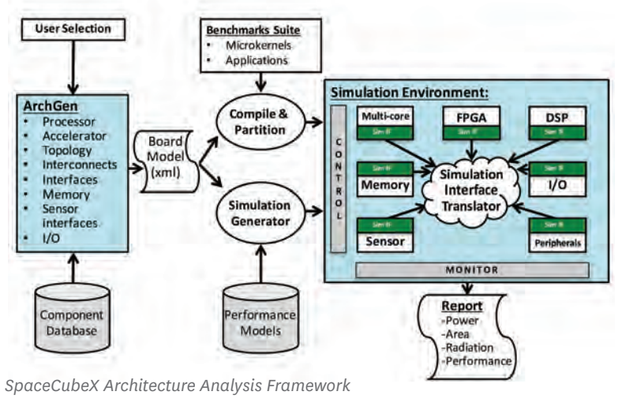
Sponsored by the Defense Threat Reduction Agency, the RadMAC project is developing low-overhead techniques to mitigate faults caused by solar particles in conventional, non-process-hardened modern electronics without the typical reduction in performance and efficiency. The project reached a significant milestone in 2016 with the demonstrated detection and correction of errors in prototype chips fabricated in a conventional 28nm Samsung process, as well as in a commercial FPGA, when exposed to radiation from a heavy ion source. With a demonstrated five-orders-of-magnitude reduction in sensitive cross-section, the approach requires only 20% of the power, 40% of the area and incurs a third of the performance penalty compared to state-of-the-art methods.

The RadMAC prototype chip

As a showcase, the program is currently developing a Low Overhead Hardened Reconfigurable Array (LOHRA) in a 16-nm TSMC FinFET technology, along with an accompanying toolset that supports high-level programming.

SpaceCubeX Project

NASA's next-generation Earth science missions—which will help scientists answer critical 21st-century questions about global climate change, air quality, ocean health, and ecosystem dynamics—are projected to increase on-board processing requirements by 100 to 1,000x over the current generation of satellite systems due to higher resolution instruments and the move to continuous observations. Hybrid computing architectures, the combination of multi-core CPU processors with specialized coprocessors such as DSPs or FPGAs, provide the high computational efficiency required in constrained size, weight, and power (SWAP) platforms. However, finding the optimal selection of devices, topology, interconnects, memories, etc., is currently a manual process and often yields mission-specific point solutions.



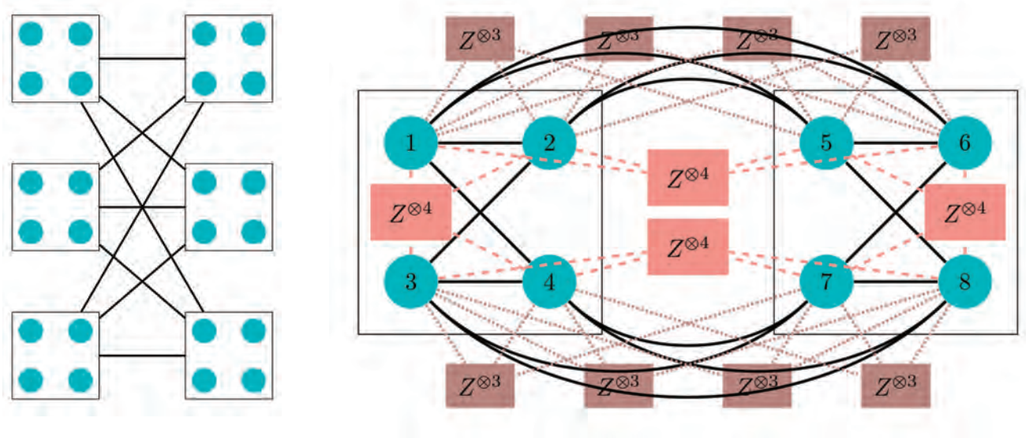
SpaceCubeX Architecture Analysis Framework

The goal of the Reconfigurable Computing Group's *SpaceCubeX* project is to develop a co-design framework which enables rapid trade space exploration of hybrid board designs, while also mapping a benchmark suite of earth science applications to the architectures and producing quantified performance results. The framework shown to the left allows a developer to describe architectures of interest using a component database, which then automatically generates the testbed of heterogeneous processing simulators or emulators. The framework includes all of the test harnessing to virtually connect different types of processor simulators and emulators and add in the performance modeling of board-level aspects such as interconnect speed, instrument interfaces, and interactions with deep memory. To date, the team — which includes USC/ISI, NASA GSFC, and NASA JPL — has performed over 300 benchmark experiments over five principal architectures, and identified improvements of more than 20,000x over the current state of practice flight hardware. These results will be used to drive the architecture for the SpaceCube3.0 on-board space processor, whose goal is to become the common processing hardware used on all NASA Earth science missions.

Quantum Annealing

Moving forward, QCC researchers are working to help design and develop the next generation of quantum annealers. Current systems have several limitations born of the engineering compromises that had to be reached in order to build a scalable device. We believe that these issues may be hindering the computational advantages of the quantum system. One of these limiting factors is the restricted connectivity of the device: each qubit interacts only with a handful of other qubits. This constrains the type of problems that can be natively implemented in the processor, and even though techniques have been developed to implement more general problems, the overhead required greatly reduces the size of the instances that can actually be solved. Hence, increasing the connectivity between qubits is expected to extend the realm of real-world problems that could be solved using quantum annealers. One of our main research focuses will be to collaborate in the design of alternative architectures that maximize the connectivity while allowing for a scalable implementation of quantum annealing. We will also study the potential advantages of exploiting interactions involving more than two qubits.

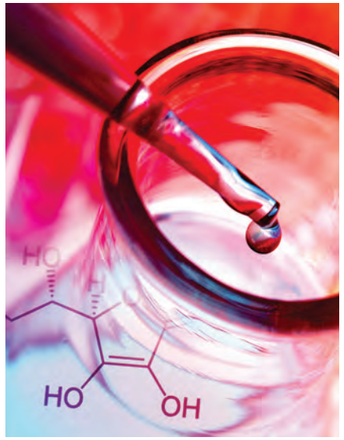
Current quantum annealing systems are designed to solve combinatorial optimization problems that are of great practical importance in many applications. This can be done using only a restricted type of quantum interaction between the qubits. However, it is still a matter of debate in the community whether these interactions are sufficient to manifest at least part of the power of quantum computing. More general quantum interactions are required in order to achieve the full power of quantum computation. It is very challenging to implement these general interactions on an actual device. Another focus of our research will be two-pronged: 1) understanding "how much" of this interaction needs to be added to currently existing architectures in order to start seeing a clear computational advantage for quantum devices, and 2) collaborating in the design and modeling of physical qubits capable of implementing these interactions.



Schematic of SCECA. Left: An encoded K_3,3. Each box with four circles represents an encoded qubit comprising four physical qubits. Right: Expanded view of a row from the left figure, showing all Z operators between a pair of encoded qubits on opposite sides of the bipartition of the K_(n,n). For each such pair, each physical qubit has four 2-body, three 3-body, and three 4-body Z terms.

USC BIOCOR

Funded by the National Institute of Standards and Technology (NIST), *Biomedical Devices and Equipment Consortium Organization to Roadmap Industry* (BIOCOR) is a collaborative initiative to encourage U.S. manufacturing of advanced medical devices. BIOCOR brings together expertise from the University of Southern California's Information Sciences Institute (lead), USC's Keck School of Medicine, Columbia University and the University of Minnesota. BIOCOR's major objectives are 1) identify and prioritize major technical and manufacturing challenges facing the nation's biomedical devices and equipment (BDE) industry, and 2) develop a technology roadmap, along with companion technology maturation and development plans, to address the challenges. During 2016, the USC BIOCOR team hosted two major workshops focused on medical device manufacturing challenges and the role that digital health will play in the industry. Key findings include the need for standards in the industry and for collaborative technology platforms, the reduction of costs in manufacturing devices in low volumes, and the streamlining of regulatory compliance processes.

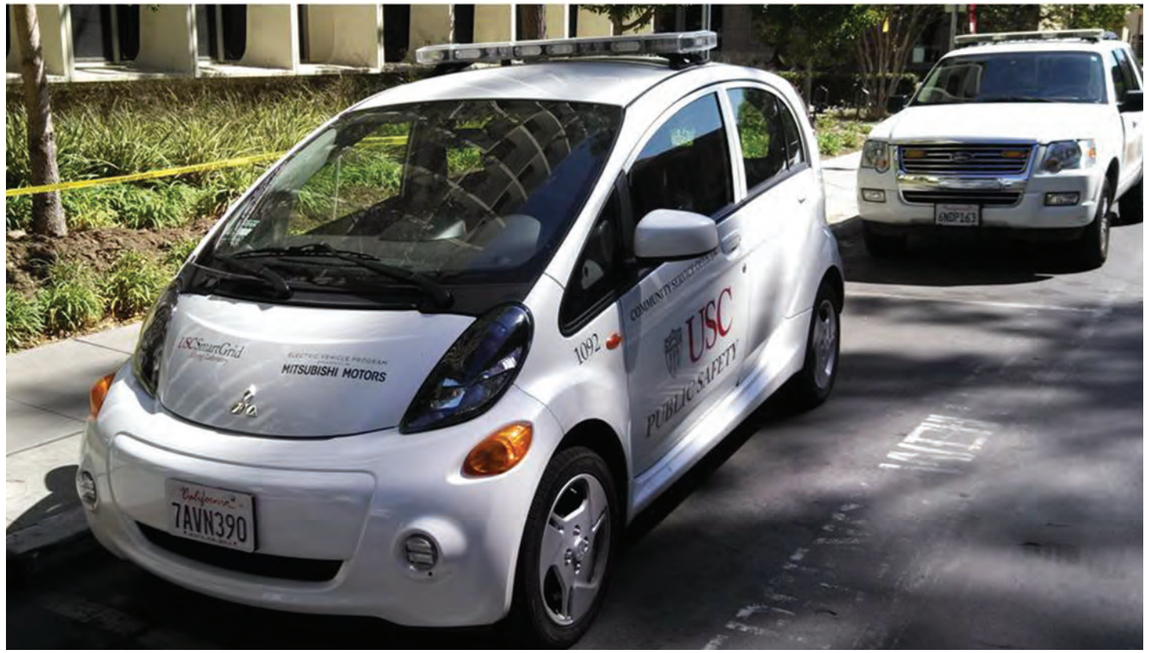


Backpack

In 2016 ISI was awarded a grant by the U.S. Department of State for *Backpack: Offline Content Distribution Through Mobile Phones*. Backpack expands upon the success of the mobile Android application Toosheh to accommodate additional modes of data transfer and to develop and deploy an Apple iOS version of the application.

USC Smart Grid Project

A major accomplishment in 2016 was the successful conclusion of the *Smart Grid Regional Demonstration Projec*t (SGRDP), funded by the U.S. Department of Energy and Los Angeles Department of Water and Power (LADWP). This multimillion-dollar, six-year effort focused on research into dynamic demand response (DR), cyber security, the impact of electric vehicles on the power grid and methods for engaging the customers to become partners with the utility. Led by the Decision Systems Group, the USC SGRDP team was composed of researchers from both ISI and the USC campus. Over the six-year period of the project, the project team published over 100 research papers and developed working technologies deployed on LADWP's 52,000-customer testbed. In addition, the Decision Systems Group led efforts to instrument USC's University Park Campus, turning the campus into the *USC Smart Grid Living Laboratory* where research into smart grid technologies continues.



On loan from Mitsubishi, this MiEV vehicle was used for the Smart Grid project. Use data for these electric vehicles was collected by the Smart Grid team for miles driven, type of activity, number of charges and duration of each charge.

Development of Quantum Monte Carlo Methods

ISI has developed a Monte Carlo algorithm designed to simulate quantum as well as classical systems at equilibrium, bridging the algorithmic gap between quantum and classical thermal simulation algorithms. The method is based on a novel decomposition of the quantum partition function that can be viewed as a series expansion about its classical part. ISI has demonstrated that the algorithm is optimally suited to tackle quantum many-body systems that exhibit a range of behaviors from fully quantum to fully classical, in contrast to many existing methods. This research also illustrates how our method allows for the unification of quantum and classical thermal parallel tempering techniques into a single algorithm, and illustrates its practical significance.

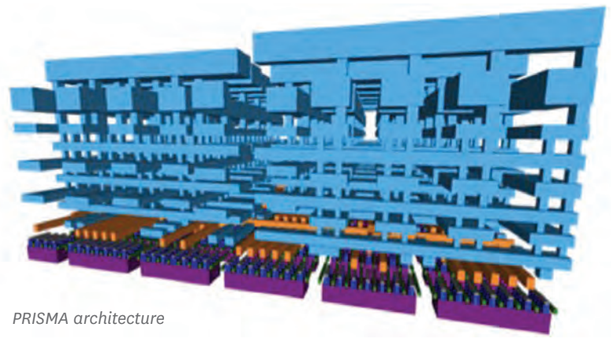
HARP platform and the USC/ISI FPGA testbed

Heterogeneous Accelerator Research Program

Over the past year the Reconfigurable Computing Group at the University of Southern California's Information Sciences Institute (USC/ISI) and the FPGA/Parallel Computing Group at USC have been working collaboratively as part of Intel's Heterogonous Accelerator Research Program (informally known as HARP). Intel developed HARP to enable research across hardware and software for performance of diverse workloads in the data center as computer architectures move beyond homogeneous parallelism to incorporate application-customized hardware. The HARP platform is a tightly integrated dual socket Intel Xeon processor and Altera StratixV FPGA which provides high bandwidth coherent memory access, pairing fast sequential general-purpose computing with a fine-grained, highly parallel co-accelerator. The research platform comes at an exciting time, where significant investments into integrating FPGAs into high-performance computing are being made, including those by Microsoft, Intel, and Amazon.

The USC/ISI team was selected as 1 of 20 universities and research institutes out of over 130 submissions worldwide to participate in the first phase of the program. Each team was awarded a HARP platform; to date, a combined fourteen graduate students, research staff, and faculty members from USC/ISI have used the shared resource across research topics including image processing, distributed big data and graph analytics, natural language processing, software-defined networking, and memory access utilization and optimization. The USC/ISI work resulted in four publications, four invited talks and demonstrations and one poster, with impressive results in natural language processing (170x speedup), deep learning (5x), and high-throughput sorting (2-5x). In December 2016 ISI was invited to present USC/ISI's research and experiences at Intel's HARP Workshop in Hillsboro, OR. USC/ISI was recently recognized by Intel as a leader in HARP research and is one of three universities selected to receive the second-generation HARP platform to continue research with applications and algorithms, operating systems, scalability, and tools. USC/ISI and HARP represent a tremendous opportunity to showcase our research expertise on advanced technology across numerous projects and to spawn research collaboration in multi-core processors and FPGAs.

PRISMA Ptychography-Based Rapid Imaging of Nano-Structures with Multilayer Assemblies

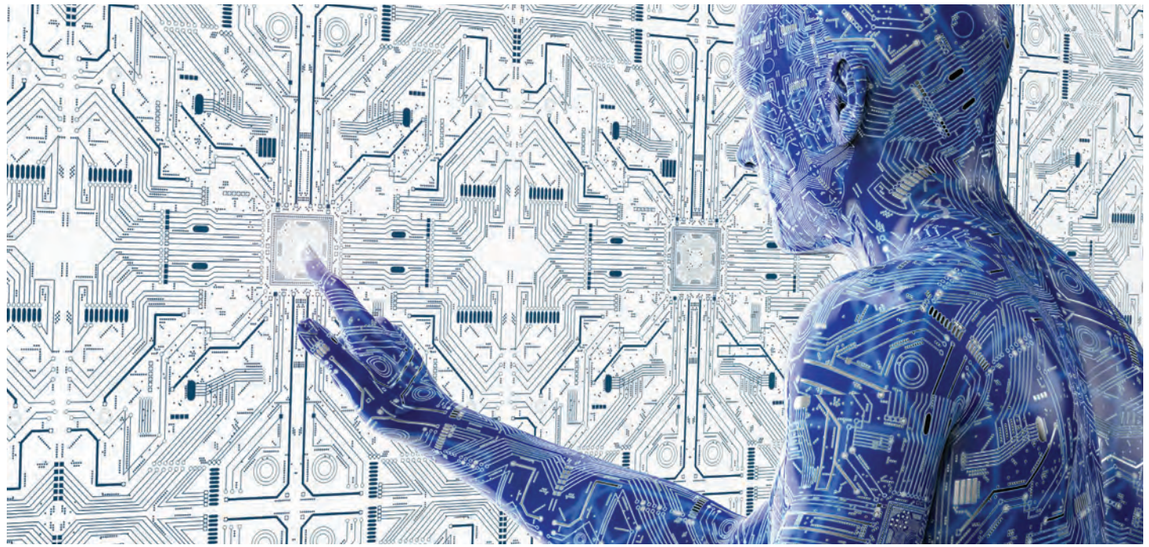


PRISMA architecture

Modern integrated circuits (ICs) are highly complex devices, where variations at the nanoscale can have dramatic impacts on performance and correctness. The ability to image these devices would revolutionize the semiconductor industry. However, imaging these devices, and the dozen or more metal layers that compose them, is extraordinarily challenging. X-ray ptychography combined with tomography represent today's most promising approach to imaging ICs nondestructively; however, the computational requirements of ptychography and tomography at the sub-10 nanometer scale are substantial. ISI's PRISMA team is developing highly scalable ptychography and tomography algorithms that integrate machine learning with high-performance accelerators to achieve the petaflop-scale performance on petabyte-scale data sizes, enabling sub-10 nanometer x-ray imaging of modern ICs in a matter of days.

DIRECTOR **YIGAL ARENS**

The Intelligent Systems Division (ISD) comprises more than 100 faculty, research staff, USC graduate students, and short- and long-term visiting researchers. Most ISD researchers hold graduate degrees in computer science or related disciplines, and many also serve as research faculty in the USC Viterbi School of Engineering—mainly in the Department of Computer Science.



ISD is one of the world's largest artificial intelligence (AI) groups. It is known especially for its work in natural language processing, machine translation, and information integration. We also explore biomedical data integration and engineering, computational behavior, adaptive robotics, social networks, and video, image and multimedia analysis. We build working prototypes and partner with industry to create commercial applications. These are ISD's primary research thrusts.

Natural language processing and machine translation, for which ISD is internationally renowned; this includes statistical machine translation, question answering, summarization, ontologies, information retrieval, poetry generation, text decipherment, and more.

Knowledge technologies, involving interactive knowledge capture, intelligent user interfaces, semantic workflows, provenance, and collaboration technologies, with a focus on scientific data analysis and discovery.

Information integration, using artificial intelligence and machine learning techniques to solve complex information integration problems, with applications to areas ranging from describing artwork to discovering exploited children.

Biomedical data integration that provides a semantically consistent view of, and efficient query access to, distributed, heterogeneous biomedical data.

Biomedical knowledge engineering, and innovative methods of building biomedical informatics systems based on cutting-edge AI techniques.

Robotics, in particular modular, self-reconfiguring robots and control methods.

Machine intelligence and data science, focusing on developing efficient algorithms to analyze data from a variety of applications areas, including computational social science, cybersecurity, and biomedicine.

Video, image and multimedia analysis, including document image processing and face recognition.

Teaching Data Science to Non-Programmers

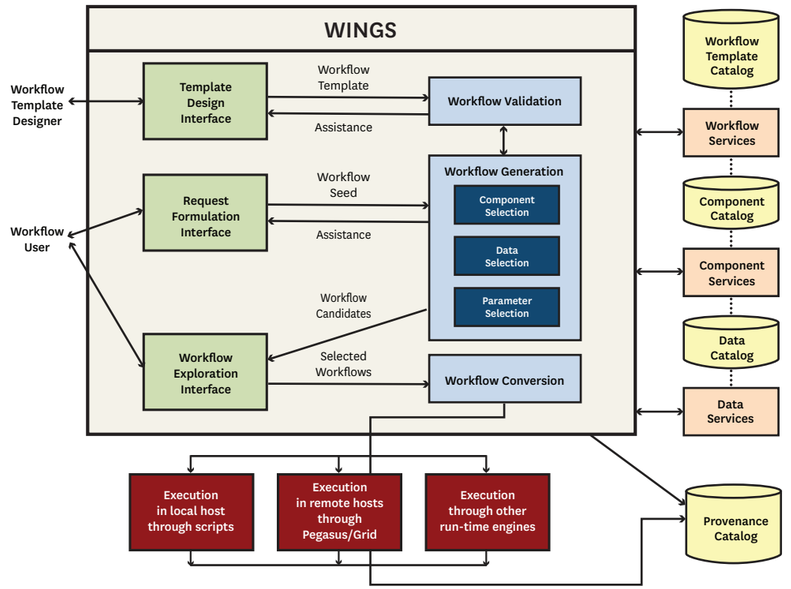
Big data analytics has emerged as a widely desirable skill in many areas. Although courses are now available on a variety of aspects of big data, there is a lack of broad and accessible courses for non-programmers that enable them to learn about big data analytics in practice. This poses severe limitations to our ability as a society to take advantage of our vast digital data resources.

With a grant from the National Science Foundation, ISI researchers have developed course materials to train non-programmers in important data science topics such as machine learning, parallel computing, visualization, ontologies and semantic metadata, and data stewardship. Through lessons and practice, students learn about analyzing text data, multimedia data, geospatial data, time series data, and network data. Analytic methods are captured in semantic workflows that ensure valid use of methods and provide an easy-to-use visual programming environment where students can practice within predefined lesson units.

Key features of the educational materials include:

- **Exposing students** to well-understood end-to-end data analysis processes that have proven successful in several challenging domains and represent the state-of-the-art.
- **Allowing students** to easily experiment with different combinations of data analysis processes, represented as workflows of computations that they can easily reconfigure.
- **Providing students** with structured lessons to analyze real-world and scientific data, posing significant challenges to the students over and above what is learned in textbooks.
- **Guiding students** to achieve target performance levels in each lesson as they experiment with different algorithms and datasets.

This course was offered at USC in the spring and fall of 2016 as part of the Informatics Program, and taught in diverse majors such as political science, education, communications, and business. The course is now a requirement for all new interdisciplinary degrees in Informatics at USC.



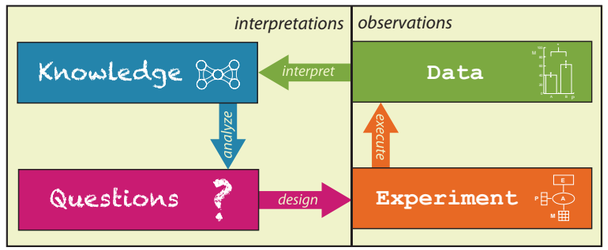
Example of a workflow run by a student to process a large document using a MapReduce strategy to parallelize the computations.

Machine Reading of Scientific Evidence

Scientific literature is a written record of the most crucial aspects of human knowledge, and it is made up of tens of millions of highly technical, structured documents. Although much progress is being made to develop AI methods to process natural language in general, we lack models that effectively process the complex language of science. In particular, *machine reading* systems typically make no distinction between concrete scientific observations and the interpretive restatement of interpretations echoed from cited documents (which are notoriously unreliable). Thus, driving machine reading systems from interpretive text that is easier to read inherently uses less reliable information as input. We follow a premise that observational assertions (i.e., statements about the results of experiments themselves) are generally more reliable than interpretive assertions (i.e., statements about the meaning of results), and that interpretations based on citations are even less reliable. Based on this premise, we use AI methods to locate and process text in the results section that describe the primary experimental results and implications of the article itself.

Deep DIG

Domain Insight Graphs are an area of great expansion within the information integration group at ISI. We seek to extend this core technology by applying the infrastructure developed by the DIG group to the more intricate, involved domains of neuroscience and molecular biology. DIG provides a scalable, linked data framework with many powerful, general capabilities. We seek to leverage these tools in biomedical informatics applications to enable data sharing and a standardized interlinked platform for scientific analysis and reasoning.



The Cycle of Scientific Investigation: This research investigates how to automate and accelerate the scientific process represented by this diagram. Different elements of the research relate to different sections of the cycle—such as representing experimental knowledge or extracting information from the literature to populate an informatics system.

Creating Reading Lists from Corpora of Pedagogical Text

In order to learn difficult subjects, students are often confronted with collections of technical documents that they must read and understand. Under IARPA seedling funding, we develop unsupervised methods designed to assist students dealing with large-scale document collections that can create reading lists for students on the fly. This project, called *TechKnAcq* (from Technical Knowledge Acquisition), processes text from the field of natural language processing and uses information theoretical methods to estimate dependency relationships between concepts in the document collection. This work is also being applied to ISI's ERUDITE project.

Knowledge Engineering from Experimental Design (KEfED)

Experimental work in science is driven by the creation and analysis of data under controlled conditions. In this project we treat experiments as generative processes and attempt to leverage their design to better understand the structure of data generated by them. This year, we have moved forward by working with members of the broader community to apply the principles we developed in the context of real scientific projects. This includes preliminary collaborations with the *SourceData platform*, the *Immunoepitope Database (IEDB)* and consortium members of the *Ontology for Biomedical Investigation (OBI)*. We are also promoting this general idea within the *FORCE11* community as a working group concerned with protocols.

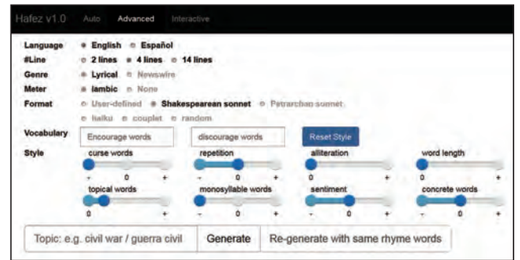
Development of natural language processing (NLP)

Development of Natural Language Processing for All the World's Languages

ISI plays a key role in the DARPA LORELEI program, which seeks to develop natural language processing capabilities for all the world's languages. We translate foreign languages into English; we find names of people, places, and organizations; and we assess reports during disaster situations in order to determine how best to allocate aid resources. In 2016 we made several key advances in *neural machine translation*, a promising new method for translating between human languages using deep learning. One advance allows us to learn translation patterns from a large-resource language pair (such as French-to-English), then transfer those patterns to help us better learn to translate a small-resource language (such as Uzbek-to-English). This helps extend the reach of machine translation to many more languages. A second advance lies in understanding what these deep, neural models are doing. We developed a kind of software "microscope" that lets us inspect what grammatical features of the input language these systems capture and how they make decisions when creating output language. As AI systems grow more complex and autonomous, we need such understanding in order to improve and control them.

Poetry Turing Test

In April 2016, Dartmouth College researchers organized the first *PoetiX Challenge*—a Turing test designed to see whether a computer poet could pass as a human. The format was exceedingly challenging: given a topic word or phrase, the computer must generate an unbounded number of distinct 14-line sonnets on the topic, written in iambic pentameter and obeying rhyme constraints. Organizers presented these poems to a judge, along with a sonnet written by a professional human poet. Our ISI team submitted the winning entry and received a $3000 prize. Their system used a novel combination of finite-state technology, word vectors, and recurrent neural networks. While it outperformed other computer systems, the poetry does not yet fool human judges, so the challenge remains. A human/machine collaborative version of this system is available at http://52.24.230.241/poem/advance.



Shall I compare thee to a summer's day?
Thou art more lovely and more temperate.
Rough winds do shake the darling buds of May,
And summer's lease hath all too short a date.

Excerpt from Sonnet 18, Wm. Shakespeare

DISK: Automated Discovery of Scientific Knowledge

In many areas of science, sensors and instruments are continuously collecting data. Yet most research projects analyze data at a particular point in time, and once articles are published they are rarely revisited to account for new data. In some cases, this makes sense since more data may only be tangentially related, and thus may not be relevant to include in a joint analysis. However, in many cases the availability of additional data may significantly affect prior results—by confirming with additional evidence or invalidating results. In addition, the new data may enable new types of analyses, leading to important revisions of prior findings or to entirely new findings. The goal of this project is to automatically and continuously analyze scientific data as it becomes available so that scientists can be alerted if their prior studies are affected or if new results are gleaned.

In collaboration with Stanford University, where pioneering work in this area was done in the 1980s, researchers at ISI are developing a new framework for automated discovery from data repositories that tests user-provided hypotheses using expert-grade data analysis strategies, and reassesses hypotheses when more data becomes available. The new framework, called DISK, represents discovery strategies as lines of inquiry that express how to test a hypothesis by finding and analyzing relevant data. DISK is being used to reproduce the discoveries in a seminal cancer multiomics paper published in *Nature*, the first proteogenomics study ever published.

DiSPARITY: Digital, Semantic and Physical Analysis of Media Integrity

In the age of Photoshop and social media, a picture may still be worth a thousand words, but not all those words are true all the time. The goal of our DiSPARITY research effort is to create techniques that can assess the truthfulness, or integrity, of an image or a video clip—thereby allowing users to determine whether and how much to rely upon the specific visual information at hand.

Sponsored by DARPA's Medifor Program, the DiSPARITY team will characterize visual and semantic signatures of manipulated digital media (images and videos) by considering pixel-level attributes, by analyzing the physics of the scene, and by investigating the semantics of the image in the context of other related images and videos. We are also characterizing the genealogy of a manipulated image given a world data set. Key challenges to achieving these goals include the wide variety of acquisition devices, the increasing sophistication of manipulation tools and techniques (including rapid advances in computer graphics technology), and the sheer volume of data to be analyzed.

Can you trust what you see?

EFFECT: Forecasting Cyber Attacks

On October 21, 2016, dozens of popular sites, including Twitter, Netflix, and Paypal, became unreachable due to a massive cyber attack directed against the infrastructure of the Internet. Attackers exploited a vulnerability in the software used by Internet of Things (IoT) devices that enabled them to commandeer vast numbers of such devices and use them to unleash a massive denial-of-service attack against the Domain Name System services provider. The overwhelmed domain name service provider stopped routing Internet traffic to its proper destinations, which resulted in widespread disruptions of normal activities.

The attack did not come out of the blue. In fact, there were signs suggesting it would happen weeks prior to the attack. If caught early enough, these signs could have allowed organizations, such as the domain name services provider, to anticipate and prepare for the attack. The first signs emerged on September 5th, in a blog post that described a vulnerability in the operating system of IoT devices, which would enable an attacker to remotely gain control of the device. On September 30th, the software kit exploiting this vulnerability was publicly released. The news of this release quickly spread online, and it took a little more than three weeks for hackers to weaponize the software to take over thousands of IoT devices and launch an attack.

The goal of the IARPA-sponsored EFFECT project is to monitor publicly available data sources to anticipate cyber attacks, and ISI is leading a team that is working to address this challenge. The team includes researchers with complementary skills in information extraction, data fusion and integration, cyber security, socio-behavioral modeling, and machine learning. It is building a system that monitors a multitude of data sources on the open and dark web, extracts indicators of the preparation and planning of cyber attacks, and analyzes these indicators to produce warnings of attacks. Members of the EFFECT team include these academic and industry partners: ISI, Arizona State University, Raytheon BBN, Hyperion Gray, Lockheed Martin ATL, and Ruhr U. Bochum.

Optical Character Recognition (OCR)

Detecting and recognizing text in images and videos is one of the newest areas of research at ISI. This area of work is often referred to as optical character recognition (OCR), and the ISI OCR research group is focused on trainable and script-independent techniques, i.e., techniques that can be easily ported to multiple domains and languages/scripts. In recent years, deep learning has become the dominant technology for a variety of pattern recognition tasks, including OCR. We are pursuing an approach that combines the power of deep learning with more traditional language models— thereby effectively exploiting both the visual/appearance information in images and videos, and the linguistic information inherent in text. Our technology has been demonstrated on English and Arabic documents, both machine printed and handwritten. The work has already resulted in several publications at top-tier conferences such as ICCV and ICDAR.

ISI also participated in a pilot competition sponsored by the US Census Bureau to assess the feasibility of automatically recognizing handwritten names in historical census documents from 1990, with the goal of eventually expanding the program to process all the millions of documents from 1990 and prior decades to aid in social sciences research. The ISI team won the pilot competition by recognizing 90% of first and last names correctly.

SIGINT-Based Anticipation of Future Events (SAFE)

In the SAFE project, USC's Information Sciences Institute collaborates with Virginia Tech, Harvard, Northeastern, and Carnegie Mellon universities to create techniques for generating probabilistic forecasts of events. These techniques will include feature extraction and prediction methods (multivariate time series models) that automatically sample SIGINT data to detect group-level behavior changes that may be a signal to anticipate and respond to events of interest. While there have been some recent successes in automated event forecasting, the SAFE project tackles several research challenges that require fundamentally new solutions. The challenges include multiple modalities of data, non-stationarity of the data distributions, and the inherently multiscale nature of the underlying processes (both spatial-temporal, and organizational)—all of which need to be accounted for in the modeling framework in order to generate accurate and timely forecasts.

DIRECTOR **JOHN WROCLAWSKI**

ISI's Internet and Networked Systems Division has provided key impetus to the advance of world-changing technologies from the earliest days of the Internet to the present.

Today's group carries out a broad program of research spanning computer networking, cybersecurity, and interaction between the cyber and physical worlds. Key areas of study include:

Research on Research. A unique program focused on developing new experimental methodologies and research tools for use in the large-scale, complex, cyber and cyber-physical networked systems space. The goal of this work is to improve the *research process itself* across these focus areas, providing users nationally and worldwide with powerful new tools for experimental cybersecurity and networked systems study.

Complex Systems Cybersecurity. A portfolio centered on large-scale, system-level challenges in this increasingly central domain. Our work focuses on cybersecurity challenges facing the Internet, the emerging "Internet of Things," and cyber-physical systems, with research that addresses topics ranging from distributed denial-of-service resilience to new security solutions specifically suited to inexpensive, mass-produced IoT components.

Cyber-Physical Systems. A focus on the smart grid and similar large-scale problem domains. Building on the key observation that energy, transportation, and similar physical-world networks share many properties with cyber networks, our research seeks to apply cyber-network principles to the creation of robust, resilient, real-world networked critical infrastructures for modern society. Our research ranges from fundamental mathematical modeling techniques to the development of novel, domain-specific experimental infrastructures and tools.

Internet Measurement. A world-recognized Internet measurement group whose work in this area seeks to improve our fundamental ability to understand the Internet's operation and evolution over time, as it continues to evolve into a core structure of modern society.

Internet-Scale Naming and the Domain Name System. A research program centered on the health and robustness of the Internet's Domain Name System, or DNS, as it faces new challenges in the modern era. Combining an aggressive research activity with in-production operation of the system's "B" root server—one of 13 in the world—our work continues to advance this fundamental and critical component of today's global Internet and tomorrow's Internet of Things.

'Clean-Break' Initiatives. Designed for areas ranging from optical networking to software systems integrity. Here, our overriding goal is to "try something different"—looking at established problems through new lenses, with the aim of radically altering the capabilities or economics of the best-known solution.

A defining element of our research philosophy is the leverage of powerful cross-couplings between research itself and the tools for performing it. Beyond pure research, we frequently design and build cutting-edge research infrastructure—as catalyst and enabler for our own work, as a transition path for our results, and as a service to the larger community.

DETER: Experimental Methodologies for a Secure Cyber Future

ISI is home to the DETER Project, an internationally recognized research effort sponsored by DHS, NSF and DARPA, whose goal is to develop next-generation methodologies, tools, and infrastructures for experimental cybersecurity research. Rather than focusing on individual results, DETER seeks to advance the state of the art in cybersecurity R&D overall, with particular application to complex, distributed, multi-organizational, networked cyber systems.

Central to this activity, the project develops and operates DETERLab, a unique systems modeling and emulation environment established to support experimental cybersecurity research. DETERLab serves as a national resource for the cybersecurity research, development, and education communities, and as an immediate technology transfer-to-practice vehicle for new research results.

To further support its objectives, the DETER project has established a rich national and international collaboration program. Key partners have used the DETER software distribution as a base for similar experimental facilities, frequently specialized for differing user communities and technical domains. Recent work has focused on collaboration with the National University of Singapore's National Cyber Lab and the Israel National Cyber Bureau's CREATE Testbed.

These collaborations are central to a principal objective of the project in 2016: catalyzing creation of an ecosystem of cyber experimentation and cyber experimenters in the future. Project work furthering this ecosystem is being realized in new methods and tools for modeling complex multi-discipline cyber scenarios, development of experimentation libraries and sharing technologies, and adoption of advanced software development and testbed generation techniques that facilitate broad use of the DETER technology itself.

To further advance this objective, in 2015 ISI and SRI International, with support from the US National Science Foundation, developed and led a national-scale activity to vision, study and codify next-generation requirements for cybersecurity experimentation of the future (CEF), with results captured in a broad-ranging report to the research community, its research sponsors, and society at large.

This year, the project moved aggressively to respond to needs identified in that report. Our work focused on the development and prototyping of radical new concepts in three broad areas: advances in experimental methodologies and techniques; new approaches to rapid and effective sharing and information synthesis; and implementation of advanced experimental infrastructure capabilities. In 2017, the project will demonstrate early results of this work to the larger cybersecurity research community, with the aim of catalyzing interest in the power of these next-generation capabilities by showing what is possible.

Beyond supporting research, creating effective, hands-on strategies and approaches for students to study modern cybersecurity concepts and techniques is a challenge of worldwide interest. The DETER project has developed and made freely available a set of specialized support capabilities and teaching exercises for use in undergraduate and graduate cybersecurity classes. These advanced capabilities give faculty and students direct access to new, project-based forms of instruction, such as classroom "capture the flag exercises" that combine effective, compelling pedagogy with broad applicability and simple administration. DETER educational support is currently being used by over 8,000 students in institutions ranging from top-tier universities to small community colleges serving underrepresented populations.

AmLight Express and Protect: Responding to the Network-Intensive Challenges of the Large Synoptic Survey Telescope (LSST)

Building the LSST optical telescope

The LSST is a large-aperture, wide-field, ground-based optical telescope under construction in northern Chile that will rapidly survey the nighttime sky in the Southern Hemisphere. The 8.4 meter telescope will take a picture of the sky every 17 seconds, producing a 6.4 gigabyte image. The LSST will take more than 800 panoramic images each night with its 3.2 billion-pixel camera, recording the entire visible southern sky twice each week. Each image must be transferred to the archive site for LSST at the National Center for Supercomputing Applications (NCSA) in Champaign, Illinois, within five seconds in order for processing to be completed in time to generate "transient alert" notifications to the worldwide astronomical community within 60 seconds. These transient alerts will identify any change in the sky observed, from moving objects such as asteroids or comets to distant exploding supernovae that may hold the key to understanding dark energy. To meet the requirements of the transient alert system, LSST must rely upon a robust network service that can provide the bandwidth needed to transport 6.4-gigabyte images within five seconds from the LSST base site in La Serena, Chile, to the archive site at NCSA for roughly 10 hours every night, 365 nights a year, over the 10-year period of the LSST survey.

The more than 200 petabytes of images and data products produced by LSST will address some of the most pressing questions about the structure and evolution of the universe, as well as questions of fundamental physics. LSST is designed based on four scientific areas: 1) understanding the nature of dark matter and dark energy; 2) cataloging our solar system, from tracking potentially hazardous asteroids to mapping remote comets in the Kuiper belt and beyond; 3) mapping the transient optical sky; and 4) understanding the formation and structure of our own Milky Way galaxy. While these scientific areas have driven the requirements for the LSST system, astronomers expect the impact of the available data to touch on all areas of the exploration of our universe, and complement investments in existing facilities such as ALMA (Atacama Large Millimeter Array), Gemini, and NOAO telescopes.

Building the AmLight Express & Protect Network

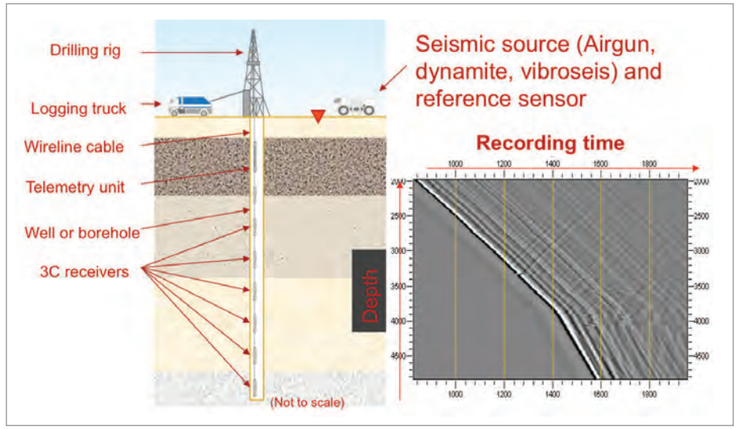
AmLight Express and Protect (AmLight-ExP) is building a network that is scalable, highly available, reliable, and will provide high-throughput and guaranteed bandwidth in response to LSST's transient alert system requirements. The AmLight-ExP team is implementing a hybrid network strategy that combines optical spectrum (Express) and leased capacity (Protect) that builds a reliable, leading-edge diverse network infrastructure for research and education.

Recursive Network Architecture

We explored the implications of our unifying Internet architecture approach in establishing equivalences, e.g., between a link and a tunnel, a router and a multipoint subnet, and a NAT as a hybrid router/host. These equivalences have begun to influence Internet standards, notably in the way in which tunnels handle fragmentation and in support for multipath routing for home routers. We established new collaborations with UCLA, which adopted our approach for their computer networks course, as well as continued ongoing collaboration with Keio University on the implication of recursive architectures on quantum networking.

NexSoft Commercialization Effort Based on Research Results of CiSoft PTG and CiSoft SEN

The Open Acceleration Systems Research Group is making substantial research progress is three areas: 1) self-sustaining cyber-physical systems, 2) biomimetic acoustic signal separation, and 3) subcomponent-level power partitioning in embedded systems. Due to the commercial relevance of this research and its results, in 2016 we expended significant effort attempting to transfer the technology to relevant industry. In collaboration with Chevron and CiSoft, we developed self-sustaining ruggedized wireless sensor network nodes that are capable of recording and transmitting high-quality audio signals from oil fields to cloud computing facilities 24/7 indefinitely. We also adapted our biomimetic audio signal separation technology to separate mixed signals collected at the pipelines into multiple independent signals. Furthermore, the research results from prior CiSoft projects have become the focus of the next-generation CiSoft effort, NexSoft. Therefore, we played a large part in technology transfer discussions with Rockwell and Schlumberger. Our group also submitted two entries to the 2016-17 Maseeh Entrepreneurship Prize Competition. One topic was a low-cost advanced hearing aid app using our biomimetic audio separation technology; the other topic concerned health and security monitoring software for automobile engine control units. Both entries reached the finals round, and the power monitoring entry took 4th place overall. In collaboration with Silicon Ribbons, we are also developing a noninvasive algorithm to optimize NVidia GPU performance based on accurate core-level power under a Small Business Technology Transfer grant from NSF.



Deployed energy-harvesting sensors for cold oil blockage detection research (CiSoft SEN)

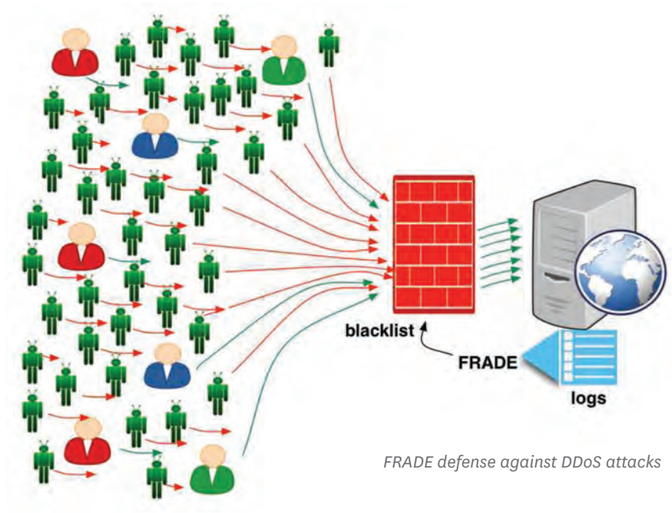
Optical Turing Machine

We completed our investigation into the feasibility and implications of digital optical computing, culminating in seven invited presentations, participation in a panel at the IEEE Rebooting Computing Workshop, an invited paper presentation at the IEEE Summer Topicals, several upcoming journal publications, and a stint as guest editor on a special issue of Optical Computing for Nanophotonics. We established new collaborations with George Washington University and BBN, in addition to ongoing collaborations with HP Enterprises and A*Star.

Fighting DDoS with FRADE and SENSS

Distributed denial-of-service (DDoS) attacks have greatly increased in frequency and severity over the past few years. In a DDoS attack, the attacker uses thousands of compromised machines to send excessive traffic to the victim server. This overloads the server and makes it unable to serve its legitimate clients. In 2016 we saw attacks against the French hosting site OVH and the DNS replication site Dyn. These attacks brought down large sites like Twitter, Reddit and Spotify, exceeded 1 TBps and were launched from hundreds of thousands of bots. An alarming aspect of DDoS is that it can affect even the largest, well-provisioned sites, and it is surprisingly easy to launch. For example, the largest attack to date (Dyn attack) was launched by a single gamer who paid $7,500 to rent 100,000 bots.

The *STEEL Lab* at ISI works on building defenses against DDoS. We have two on-going projects: FRADE and SENSS. The FRADE project, funded by the National Science Foundation, builds defenses against application-level (level 7) DDoS attacks on Web servers. In such attacks, bots send legitimate-looking Web page requests to a server to exhaust its resources. Even a modest rate of several thousand requests per second can bring down medium-sized Web servers. FRADE builds models of how human users interact with servers. These models capture the dynamics and semantics of human user interactions, and how humans process server responses. Bot behavior usually violates these models, which enables FRADE to identify bots and blacklist them. Some bots may modify their behavior to fit FRADE models. However, this lowers the bot's firepower—thus the attacker will need many more bots for a successful attack. In our tests, attacks were detected and fully filtered within seconds to minutes, and the botnet size for a successful attack was increased by three orders of magnitude. FRADE is currently being tested on DeterLab and will be released as open source in 2017. It is compatible with any Web server technology. More information about these and other projects in the STEEL Lab can be found at http://steel.isi.edu.



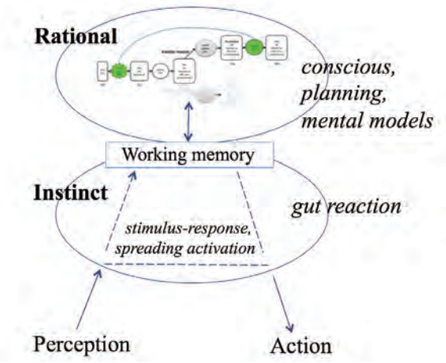
FRADE defense against DDoS attacks

SENSS projects, funded by the Department of Homeland Security, build defenses against large-volume DDoS attacks. In such attacks, the victim's link to its upstream ISP is congested with the attack traffic. The victim cannot diagnose or filter these attacks itself and must request help from upstream ISPs. Currently, there is no secure, automated way to implement this collaborative response. However, SENSS is building APIs at the victim and the ISPs that can be used for such collaboration. An ISP deploying SENSS exposes simple and generic APIs to interested networks. Attack victims use these APIs to observe and control their own traffic and routes in SENSS ISPs, and can combine them to design customized diagnosis and mitigation for many distributed attacks. We have evaluated SENSS in large-scale simulations and emulations. It performs well in sparse deployment and has a modest operating cost. It is fully deployable in today's ISP infrastructure. This excellent performance, combined with strong security and robustness, makes SENSS a game-changing solution in the fight against distributed attacks. We currently plan to transition SENSS into deployment at several regional ISPs in the summer of 2017.

Modeling Human Behavior to Improve Tools and Policies for Cybersecurity

Human behavior is a key determining factor in assessing the effectiveness of an organization's cyber defenses, including deployed hardware and software defenses as well as policies. This project aims to observe and model important aspects of human behavior in order to predict the likely consequences of changes to the security posture of an organization.

We have designed a cognitive agent architecture, DASH, that captures important aspects of human decision-making in cybersecurity, including dual-process reasoning and mental models. We built agents on DASH to model defenders, attackers and end users; this was done in collaboration with Dartmouth College, Univ. Pennsylvania, and Arizona State Univ., and with individuals whose backgrounds include psychology and the social sciences.



We recently demonstrated how increasing password strength requirements at a site can both weaken the site's security and couple it to otherwise independent sites due to prevalent coping strategies. We implemented tools to capture and reason automatically about the provenance of agent parameters based on psychological experimentation or data analysis and to explore their consequences on policy recommendations.

Enabling Experimentation on Internet-Like Systems through Flexible Models

The Internet has become a critical resource in modern society for businesses, governments and individuals. Its central role requires that new and existing technologies be well studied and experimented upon to understand their capabilities and effects. Strong models of the Internet are required for these experiments to be useful, valid and accurate. Our current research focuses on a mechanism to model the Internet and Internet-like systems.

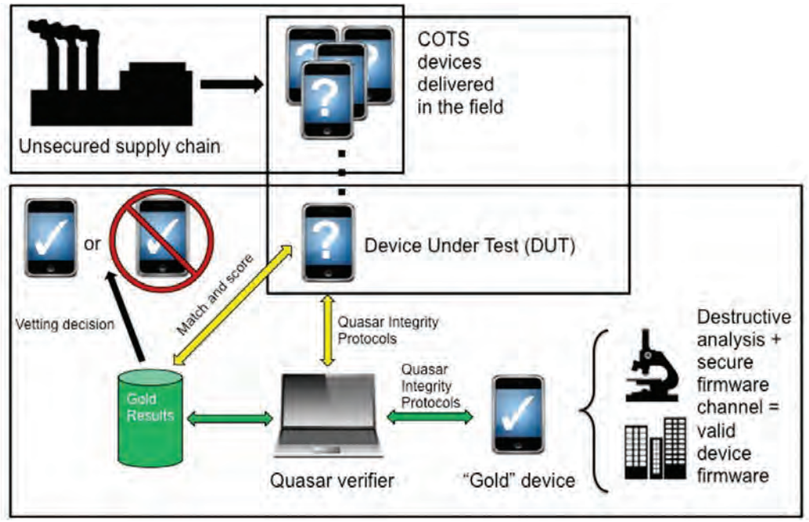
It is computationally difficult to accurately model all parts of a system as large and complex as the Internet. However, it is sufficient to model with high fidelity the parts of the Internet that are relevant to a researcher. A researcher studying routing protocols may need an extremely high-fidelity router-level model of the Internet but not require much fidelity at the link-level. Conversely, someone studying transport protocols may need detailed link-level and congestion modeling with little concern for router implementation. To this end, our models and the languages that express them are designed to provide flexibility to the researcher based on his experimental goals. We are developing small languages that allow the modeling of specific portions of Internet-like systems in high fidelity, as well as the tools to realize these models on the DeterLab Testbed. Further, the decoupling of model and realization allows domain experts to express models and non-domain experts to use them.

This work has been of great benefit to DARPA's EdgeCT program which requires high-fidelity link-level models of the DISN WAN. Our research enabled DISN domain experts to build models of the WAN through our modelling languages without requiring the details of the model's realization. Other EdgeCT partners use these models on DeterLab without a detailed understanding of the WAN being modeled or the realization of that model. This provides the massive benefit of being able to model complex Internet-like systems for experimentation without requiring the sharing of domain knowledge. Our continued research in this area will allow researchers to experiment on many Internet-like systems with fidelity tailored to their experiment without requiring domain knowledge of the modeled system. This will enable researchers to conduct experiments more quickly while also providing an improved experiment quality.

Detecting and Mitigating Malicious Software Tampering of Embedded Devices

Securing computers against attack is a wickedly difficult problem. However, Information Sciences Institute is engaged in a line of research that aims to secure systems by providing assurance that the right software—and nothing else—is installed. Especially compounding the pernicious nature of this problem is the risk of attackers compromising a computer before it is ever used! Attackers may subvert the software that comes preinstalled from the manufacturer on the increasingly powerful embedded computers found inside almost every product purchased by businesses or governments. These embedded computers, with performance scaling with Moore's law, are manufactured cheaply but often lack dedicated security features that protect software in high-end systems.

Research conducted under the QUASAR project has developed an architecture and implementation for robust vetting of preinstalled software on embedded devices such as cell phones and other consumer electronics. Our software implementation works by loading a unique set of integrity testing routines onto the device-under-test, and then engaging in an integrity protocol. Challenges are sent to the device, and only an untampered device executing precisely the QUASAR testing routines—and nothing else—is capable of furnishing the correct set of responses. An attacker, through subversion of the QUASAR testing routines, may attempt to hide the presence of tampered, malicious software, but QUASAR uses numerous strategies to defeat even the cleverest of attackers. These include randomized strategies in which challenges are interleaved with power-cycling to force malicious software to remain visible in permanent storage or else be erased from ephemeral RAM when the power is turned off. By combining several distinct techniques, the architecture eliminates reliance on any single technique or testing routine, ultimately raising the bar for the attacker and preventing tampering from going undetected. Each individual testing routine is also simpler to design and understand, providing greater assurance in its contribution to the overall vetting determination.



Detecting tampering of embedded devices

For widespread deployment, this software suite must be usable by IT staff in large business organizations and government agencies without special training or security expertise, and furthermore must not require disassembling the device or accessing internal storage. QUASAR meets these challenges by fully automating every aspect of the vetting procedure, while requiring only standard network and power connections.

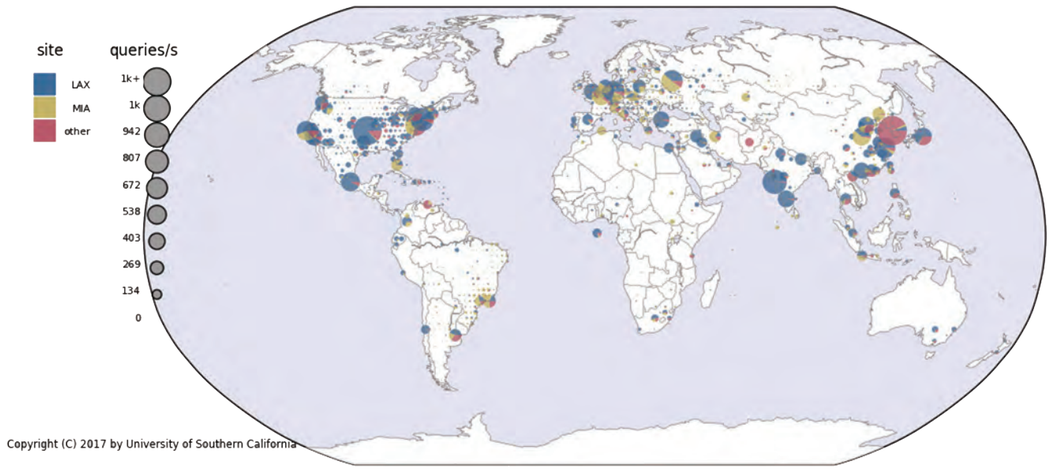
B-Root Stewardship and Growth at ISI

The Internet Domain Name System (DNS) provides the names that power websites and e-mail addresses—the dot above dot-com and dot-edu. Information Sciences Institute (ISI) maintains and manages B-Root, one of the 13 different root servers that are at the top of this system. The Internet Domain Name System was developed by ISI in the late 1980s, and we have managed and maintained a DNS Root server ever since—with a focus on both its mission of research and as a service to the Internet at large.

ISI's revitalization of B-Root took significant strides in 2016 when we increased our staff and built a second instance of the service in Miami to augment our Los Angeles site; B-Root later activated the Miami site using IP anycast. As a result, B-Root now has additional capacity, disaster resilience, and lower latency to Europe and South America. Additionally, during this period we saw a major jump in our research built around B-Root and its development as a research infrastructure for the global DNS community. The research on DNS includes a paper that analyzes anycast latency and a second paper that analyzes the 11/30/2015 denial-of-service attacks on the DNS roots. These papers were jointly authored with the University of Twente and SIDN Labs (both in the Netherlands).

We are continuing to support the global DNS research community. ISI hosted DINR, the DNS and Internet Naming Research workshop in November 2016. This workshop brought together 61 authors and attendees from academia and industry. Seven universities were represented, as well as industry and non-profit organizations, with both U.S. and international participation. Matt Larson, ICANN's vice president of research, was the keynote speaker, and 21 presentations by participants identified topics spanning research infrastructure, security and privacy, and new applications of DNS. Additionally, ISI researchers led two BOF (birds-of-a-feather) sessions on the *Future of Internet Naming* at IETF-97 in Seoul and IETF-98 in Chicago.

ISI is working with the DNS research community through these workshops and other pilot projects. The Institution has long been a leader in research infrastructure for networking to create opportunities for the research community. Our role in DNS research infrastructure is expanding—as shown by our research results, workshops, and ongoing work.



This map is from a new anycast mapping tool currently under active research at ISI. Colors depict which parts of the world are served by the B-Root's Los Angeles site (red, most of the US), or the Miami site (blue, parts of China), or a mix of both (white, Brazil, Argentina, and Eastern Europe). Circle sizes are proportional to how many IPv4 addresses are served in each two-degree square.

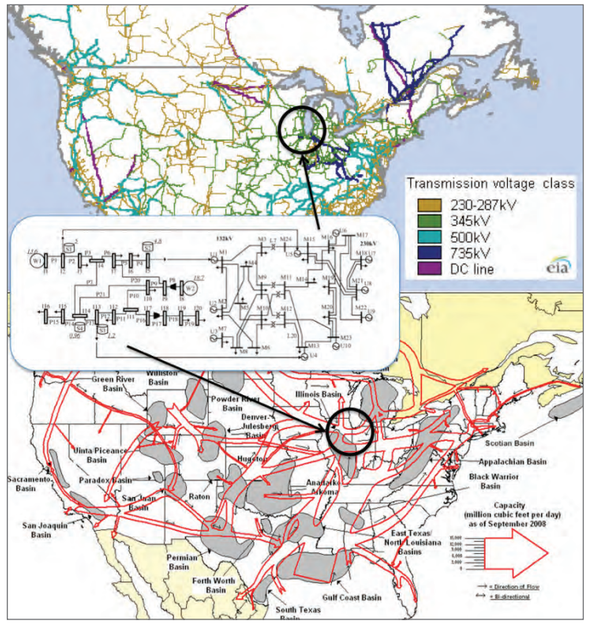
Methodologies for the Study of Next-Generation Cyber-Physical Systems

This research aims to fundamentally advance the state-of-the-art in computational and experimental methods for the study of large-scale, complex cyber-physical systems, and then to apply these methods to address critical infrastructure challenges across domains such as energy systems, smart cities, transportation, natural gas, and water supplies.

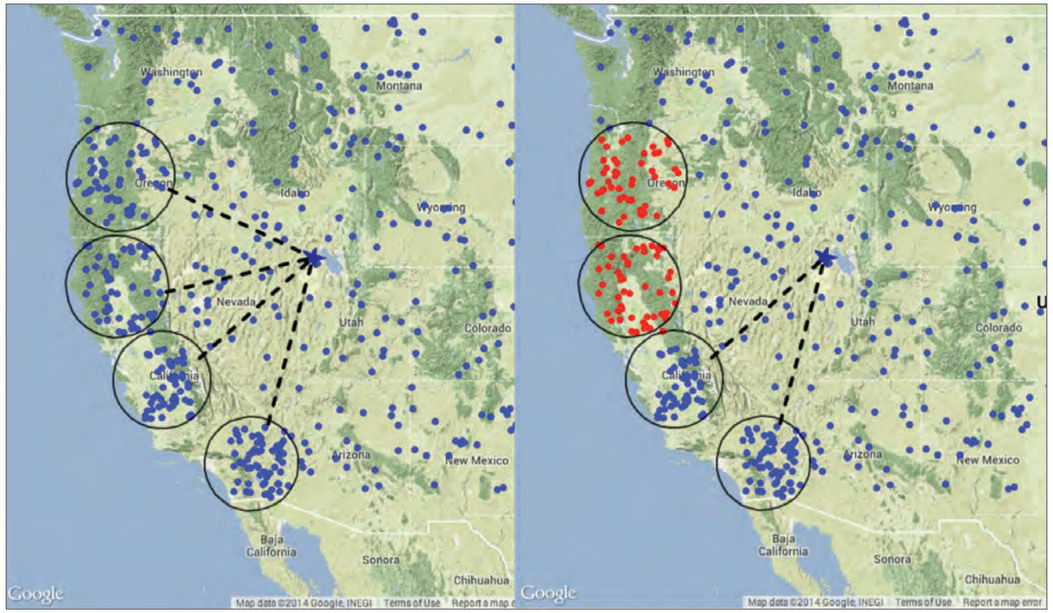
Recent technology advances in areas such as smart and ubiquitous sensors, communications, and distributed control are rapidly driving the scale and complexity of interconnected cyber-physical systems and networks to new levels. Building on strong ties to the computer networking community, which has already confronted similar issues of extreme complexity, decentralized administration, and rapid, near-exponential growth, ISI researchers are leading the development of new modeling, experimentation, and analysis strategies for the rapidly advancing world of large-scale cyber-physical systems. Our core goal is to create new methodologies and tools that will provide clear, correct, and actionable insights into the design and operation of these increasingly central components of modern society.

Combining a strong, multidisciplinary team at ISI with collaborations across multiple USC laboratories, private and public corporations, and other academic institutions, our research synthesizes ideas and methods from disciplines that include statistics, information theory, systems, theoretical and computational mathematics, optimization, economics, policy, network science, and control theory. These disciplines are relevant both for understanding complex systems and for presenting design principles and architectures that allow for system quantification and management. Our work seeks to integrate these areas: fostering new collaborations, introducing new paradigms and abstractions, and utilizing the power of experimentation to anticipate and address future challenges.

Key to our work is the observation that sophisticated, powerful tools lead directly to new research capabilities and results. We focus on two broad areas of tool construction: mathematical modeling of cyber-physical systems, and methods, frameworks, and tools that support experimental study of cyber-physical control algorithms. The past year has seen significant advances in both.



NSF RIPS Project: *Resiliency of interconnected gas and power networks in the presence of attacks. The top part of the figure is the power grid; the bottom part is the gas network. The inset illustrates how the interconnections between the two networks occur at various points in North America. This work is in collaboration with MIT.*



NSF FRESCO Project: *Evaluating the impact of DoS attacks on the stability of the power grid. The nodes under attack are shown in red. This work is in collaboration with NCSU and MIT.*

Mathematical Modeling of Cyber-Physical Systems. A significant highlight of 2016 is the development of a new dynamical systems modeling and simulation toolkit integrated directly into the very popular SciPy (Scientific Python) framework. This toolkit provides a simulation capability that is currently lacking in the research community—that is, the ability to develop, analyze and simulate systems whose structure is dynamic, but the nature of the structural changes that will take place over the course of the simulation are not known a priori. These sorts of dynamics are very common in the CPS space; e.g., consider a power system relay that is hacked and opens a transmission line breaker unexpectedly. The current state-of-the-art in simulation tools allows us to ask the question: What if a particular breaker trips in a simulation? We know a priori the nature of the structural dynamic. However if the question becomes: *How resilient is a CPS system to attack X?*—but we don't know how attack X will manifest itself in the physical realm, then our current simulation tools fall short. Our new framework, called DyPy (dypy.org), is rooted in the state-of-the-art theory of differential-algebraic equations. It will see a general release to the research community in 2017.

Integrated Experimental Frameworks. The tight coupling of physical and cyber aspects of CPS infrastructures make the design and analysis of control mechanisms in the presence of security threats particularly challenging. An obvious difficulty arises from the scale of the infrastructure. While each individual component may be relatively straightforward in its operation, the coupling between these components and systems leads to extremely complex overall system behavior. This makes the full system's response to faults or cyber attacks very difficult to understand. A more subtle challenge is to be found in the design of these systems, where the task is to balance the diverse requirements and constraints of the composite system. An optimum control strategy for one system element may not align with global requirements; this leads to compromises. While there are mature tools for the design and analysis of individual subdomains within energy cyber-physical systems, we have developed a methodological experimentation framework to explore the composite cyber-physical design space. The framework enables propagating and evaluating the uncertainty within the CPS infrastructure and capturing the complex cyber-physical interactions.

DIRECTOR **CARL KESSELMAN**

ISI's Informatics Systems Research Division pursues a broad research agenda focused on creating new types of sociotechnical systems that enable and accelerate discovery in domains of high societal impact. Launched in 2008, the division takes a holistic, systems-oriented approach, working in areas from basic network services architectures, data management abstractions, computer security, user interfaces, human factors, and domain-specific algorithms. The division specializes in highly collaborative user-driven research, in which we evaluate our work in the context of operational, high-impact domain science.

In earlier work, we developed grid computing infrastructures to support the creation and operation of "virtual organizations" as a foundation for collaboration and discovery. This work focused on understanding methods for sharing computing and storage infrastructure across distributed resource providers and collaborators. The resulting methods played a role in two Nobel prizes—e.g., all the data analysis for discovering the Higgs boson was performed on a global grid infrastructure, and the recent discovery of gravity waves took place on a data grid.

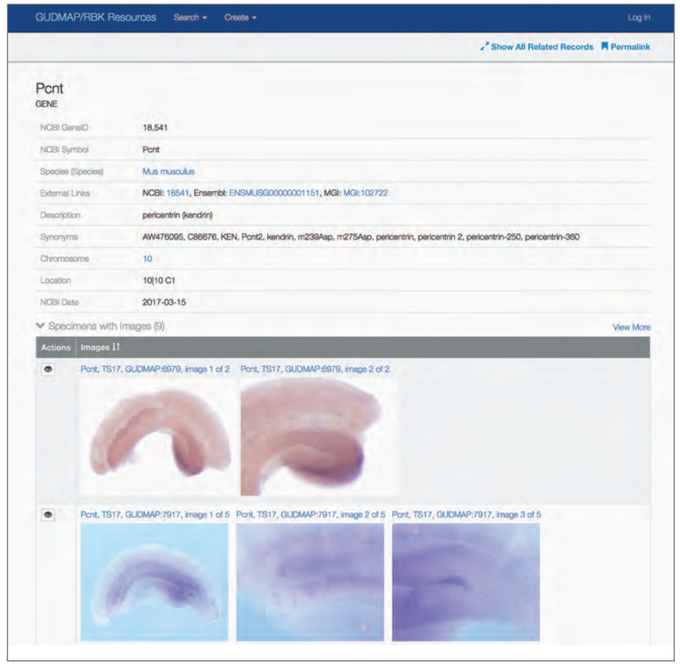
More recently, the division has focused on biomedical applications. Our current collaborations cover a broad range of applications—from basic science to clinical use cases spanning molecular biology, basic neuroscience, neuroimaging, stem cell research, and cranio-facial dysmorphia.

Our researchers work closely with ISI's highly regarded artificial intelligence, networking, and distributed systems experts, as well as with two of USC's nationally ranked Viterbi School of Engineering units: the Daniel J. Epstein Department of Industrial and Systems Engineering, and the Department of Computer Science. The division participates in many collaborative projects with the faculty in the Dornsife College, the Keck School of Medicine, and the Osterow School of Dentistry. Members of the division play a leadership role in the Michelson Center for Convergent Biosciences, including establishing the new Center for Discovery Informatics as part of the university's convergence biosciences initiative. The division also plays a central role in four international consortiums. Most recently, the Informatics Systems Research Division was choosing to house the data collection for the GenitoUrinary Development Molecular Anatomy Project—an international consortium devoted to understanding the development of the genitourinary trace in humans.

The Informatics Systems Research Division is developing innovative new approaches to accelerating scientific discovery by focusing on eliminating the complexities associated with assembling, organizing and manipulating complex, large-scale data collections. In his 1960 seminal paper "Man-Machine Symbiosis," J. C. R. Licklider observed that: "My choices of what to attempt and what not to attempt [are] determined to an embarrassingly great extent by considerations of clerical feasibility, not intellectual capability." Surprisingly, the situation has not changed significantly over the past 50 years, and with the advent of "big data," it has in many ways become worse. The goal of the ISR Division is to understand how to architect, assemble and operate complex social-technical systems that will eliminate the barriers observed by Licklider, and result in a radically faster time to new discoveries.

Deriva: This is an ecosystem of tools developed around the concept of a data-centric architecture and scientific digital asset management. In the past year, in collaboration with the G-Protein Coupled Receptor Consortium and international industry/academic collaboration, a Deriva-based platform has been in operation across the consortium and used on a daily basis by scientists as an integral part of their effort to understand the structure of a critically important class of proteins that is currently not well understood.

GUDMAP: The past year has also seen our scientific digital asset management approach being applied to a new domain of developmental biology. As part of a newly awarded research contract, the ISR Division is now responsible for all data acquisition, integration and discovery of data being generated by the NIH-funded *GenitoUnanary Molecular Anatomy Project (GUDMAP)*. In this project we are assembling detailed data associated with organ development, with the goal of understanding how these complex systems form, and why they sometimes don't develop correctly. Coupled with our existing projects in kidney reconstruction (rebuildingakidney.org) and the Dervia platform described above, we are playing a significant role in kidney-related research, and ISI is home to a unique collection of scientific data.



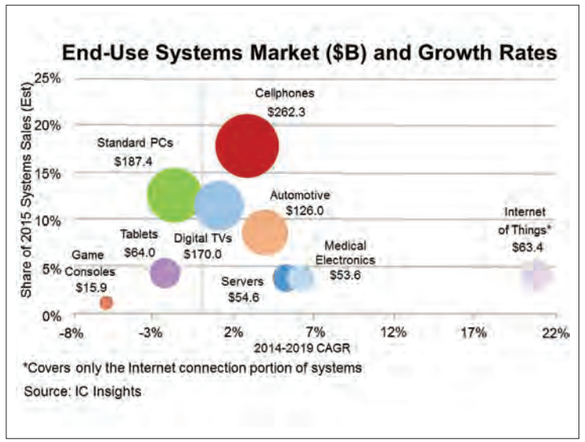
The Deriva platform is dramatically simplifying the discovery and access to complex scientific data. This figure shows that Deriva automatically organizes and presents data that captures how a specific gene (Pcnt) impacts the development of the genitourinary tract, as indicated by the colors in the microscope images. Without Deriva, scientists would have to organize, by hand, all of this data, and it would be very difficult to assemble and search.

DIRECTOR **WES HANSFORD**

For more than 35 years, custom IC designers have relied on the MOSIS service at ISI for an efficient, affordable way to prototype and volume-produce their devices. Since 1981, MOSIS has processed an average of more than seven IC designs per day.

Many turn to MOSIS for our special expertise in providing multi-project wafers (MPWs) and related services that drive IC innovation. This "shared mask" model combines designs from multiple customers or diverse designs from a single company on one mask set. It's a practical prototyping channel that allows designers to debug and perform essential design adjustments before making a substantial strategic investment. Today, with mask costs soaring, more designers than ever are using MPWs to manufacture proven devices and prototype new designs on a single wafer.

Beyond MPWs, customers are increasingly choosing MOSIS as their resource partner for volume-production. From design spec interpretation through mask generation, device fabrication and onto assembly, MOSIS is their trusted expert interface to the semiconductor ecosystem. In addition to our commercial service, MOSIS is part of the following active research programs:



DARPA CRAFT (Circuit Realization At Faster Timescales)—Developing a custom integrated circuit design flow and methodology that will: 1) sharply reduce the amount of effort required to design high-performance custom integrated circuits, 2) greatly facilitate porting of integrated circuit designs to secondary foundries and/or more advanced technology nodes, and 3) strongly increase reuse of integrated circuit elements. In support of CRAFT, MOSIS is organizing 14/16nm private MPW runs.

DARPA DAHI (Diverse Accessible Heterogeneous Integration)—Developing transistor-scale heterogeneous integration processes to intimately combine advanced compound semiconductor (CS) devices, as well as other emerging materials and devices, with high-density silicon complementary metal-oxide-semiconductor (CMOS) technology. The ultimate goal of DAHI is to establish a manufacturable, accessible foundry technology for the monolithic heterogeneous co-integration of diverse devices and complex silicon-enabled architectures on a common substrate platform. MOSIS is teamed with Northrop Grumman Corp., and is organizing private MPW runs (65/45nm).

IARPA TIC (Trusted Integrated Chips)—Developing and demonstrating split-manufacturing, with a new approach to chip fabrication where security and intellectual property protection can be assured. MOSIS is on the government team, and is organizing private MPW runs (130/65/28nm).

COMPUTATIONAL SYSTEMS AND TECHNOLOGY

A New Metric to Measure Cache Utilization for HPC Workloads
Aditya Deshpande and Jeffrey Draper
MEMSYS 2016 - *International Symposium on Memory Systems Conference*

A Study of Complex Deep Learning Networks on High Performance, Neuromorphic, and Quantum Computers
Tomas Potok, Catherine Schuman, Steven Young, **Federico Spedalieri**, **Jeremy Liu**, and **Ke-Thia Yao**
Machine Learning in HPC Environments, 2016

A Window into Occupant-driven Energy Outcomes: Leveraging Sub-metering Infrastructure to Examine Psychosocial Factors Driving Long-term Outcomes of Short-term Competition-based Energy Interventions
Kyle Konis, **Michael Orosz,** and **Nicole Sintov**
Energy and Buildings, January 2016

Accelerating Soft-Error-Rate (SER) Estimation in the Presence of Single Event Transients
Ji Li and **Jeffrey Draper**
Proc. of Design Automation Conference (DAC), 2016

Adiabaticity in Open Quantum Systems
Lorenzo Venuti, **Tameem Albash**, Daniel Lidar, and Paolo Zanardi
Phys. Rev. A 93, 2016

Analyzing Users in Parallel Computing: A User-Oriented Study
Stephan Schlagkamp, **Rafael Ferreira da Silva**, Johanna Renker, and Gerhard Rinkenauer
International Conference on High Performance Computing & Simulation, 2016

Anomaly Detection for Scientific Workflow Applications on Networked Clouds
Prathamesh Gaikwad, Anirban Mandal, Paul Ruth, **Gideon Juve**, **Dariusz Krol**, and **Ewa Deelman**
IEEE International Conference on High Performance Computing & Simulation (HPCS), 2016

Asterism: Pegasus and Dispel4py Hybrid Workflows for Data-Intensive Science
Rosa Filgueira, **Rafael Ferreira da Silva**, Amrey Krause, **Ewa Deelman**, and Malcom Atkinson
7th International Workshop on Data-Intensive Computing in the Clouds (DataCloud'16), 2016

Automated Demand-Based Vertical Elasticity for Heterogeneous Real-Time Workloads
Geoffrey Tran, **Yu-An Chen**, **Dong-In Kang**, **John Paul Walters**, and **Stephen Crago**
9th Annual IEEE International Conference on Cloud Computing, June 2016

Automating Environmental Computing Applications with Scientific Workflows
Rafael Ferreira da Silva, **Ewa Deelman**, **Rosa Filgueira**, **Karan Vahi**, **Mats Rynge**, **Rajiv Mayani**, et al.
Environmental Computing Workshop (ECW'16), 2016

Consecutive Job Submission Behavior at Mira Supercomputer
Stephan Schlagkamp, **Rafael Ferreira da Silva**, William Allcock, **Ewa Deelman**, and Uwe Schwiegelshohn
25th ACM International Symposium on High-Performance Parallel & Distributed Computing (HPDC), 2016

Driver Hamiltonians for Constrained Optimization in Quantum Annealing
Itay Hen and Marcelo Sarandy
Physics Rev. A 93, 2016

Dynamic and Fault-Tolerant Clustering for Scientific Workflows
Weiwei Chen, **Rafael Ferreira da Silva**, **Ewa Deelman**, and Thomas Fahringer
IEEE Transactions on Cloud Computing, Vol. 4, 2016

Employing High Performance Computing to Realize a Cyber Quick-Reaction Training Environment
Brian Castello, John Tran, Douglas Hire, **Robert Lucas**, and **Ke-Thia Yao**
MODSIM World, 2016

Energetic Cost of Superadiabatic Quantum Computation
Ivan Coulamy, Alan Santos, **Itay Hen** and Marcelo Sarandy
Frontiers in ICT 3, 2016

Enhancing Reproducibility for Computational Methods
Victoria Stodden, Marcia McNutt, David Balley, **Ewa Deelman**, **Yolanda Gil**, Brooks Hanson, et al.
Science, Vol. 354, 2016

Hypervisor Performance Analysis for Real-Time Workloads
Geoffrey Tran, **Yu-An Chen**, **Dong-In Kang**, **John Paul Walters**, and **Stephen Crago**
IEEE High Performance Extreme Computing Conference (HPEC), September 2016

Increasing Waiting Time Satisfaction in Parallel Job Scheduling via a Flexible MILP Approach
Stephan Schlagkamp, Matthias Hofmann, Lars Eufinger, and **Rafael Ferreira da Silva**
Intl Conference on High Performance Computing & Simulation, 2016

Joint Soft-Error-Rate (SER) Estimation for Combinational Logic and Sequential Elements
Ji Li and **Jeffrey Draper**
Proc. of the IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2016

Mean Field Analysis of Quantum Annealing Correction
Shunji Matsuura, Hidetoshi Nishimori, **Tameen Albash**, and Daniel Lidar
Physics Rev. Letters, 2016

Mitigating Single Event Upsets in Arithmetic Logic Using Residue Codes
Michel Sika, **Stephen Lee,** and **Hao Wu**
Proceedings of the Radiation Effects and Components Conference (RADECS), 2016

Nested Quantum Annealing Correction
Walter Vinci, **Tameen Albash**, and Daniel Lidar
npj Quantum Information 2, 2016

OSG-GEM: Gene Expression Matrix Construction Using the Open Science Grid
William Poehlman, **Mats Rynge**, Chris Branton, Desinghu Balamurugan, and Frank Feltus
Bioinformatics and Biology Insights, Vol. 10, 2016

Pegasus in the Cloud: Science Automation Through Workflow Technologies
Ewa Deelman, Karan Vahi, Mats Rynge, Gideon Juve, Rajiv Mayani, and **Rafael Ferreira da Silva**
IEEE Internet Computing, Vol. 20, 2016

Performance Analysis of an I/O-Intensive Workflow Executing on Google Cloud and Amazon Web Services
Hassan Nawaz, **Gideon Juve**, **Rafael Ferreira da Silva**, and **Ewa Deelman**
18th Workshop on Advances in Parallel and Distributed Computational Models (APDCM), 2016

PGen: Large-Scale Genomic Variations Analysis Workflow and Browser in SoyKB
Yang Liu, Saad Khan, Juexin Wang, **Mats Rynge**, Yuanxun Zhang, Shuai Zeng, Shiyuan Chen, et al.
BMC Bioinformatics, Vol. 17, 2016

Practical Engineering of Hard Spin-Glass Instances
Jeffrey Marshall, Victor Martin-Mayor and **Itay Hen**
Physics Rev. A 94, 2016

Quantum Annealing for Constrained Optimization
Itay Hen and **Federico Spedalieri**
Physics Rev. Applied 5, March 2016

Reducing Data Movement with Approximate Computing Techniques
Stephen Crago and Donald Yeung
IEEE Conference on Rebooting Computing, October 2016

Scaling GIS Analysis Tasks from the Desktop to the Cloud Utilizing Contemporary Distributed Computing and Data Management Approaches: A Case Study of Project-Based Learning and Cyberinfrastructure Concepts
Tyson Swetnam, Jon Pelletier, Craig Rasmussen, N. Callahan, N. Merchant, E. Lyons, **Mats Rynge**, et al.
XSEDE16 Conference on Diversity, Big Data, and Science at Scale, 2016

Science Automation in Practice: Performance Data Farming in Workflows
Dariusz Krol, Jacek Kitowski, **Rafael Ferreira da Silva**, **Gideon Juve**, **Karan Vahi**, **Mats Rynge**, and **Ewa Deelman**
21st IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2016

Simulated Quantum Annealing with Two All-to-All Connectivity Schemes
Tameem Albash, Walter Vinci, and Daniel Lidar
Physics. Rev. A 94, 2016

SpaceCubeX: Initial Simulation-level Results of Hybrid On-board Processing Architectures
Matthew French, **Andrew Schmidt**, **Gabriel Weisz**, Tom Flatley, Gary Crum, Jonathan Bobblit, et al.
NASA Earth Science Technology Forum, 2016

Surface-potential-based Compact Modeling of BTI
Ivan Sanchez-Esqueda, et al.
IEEE International Reliability Physics Symposium, 2016

System Engineering a Scalable Cyber Warfare Test Environment
Ke-Thia Yao, **John Tran**, Daniel Davis, Daniel Burns, and **Robert Lucas**
ITEA Journal, 37(1), March 2016

The Maestro Flight Experiment: A 49-Core Radiation Hardened Processor in Space
Craig Rogers, **David Barnhart**, and **Stephen Crago**
IEEE Aerospace Conference, March 2016

Toward an End-to-end Framework for Modeling, Monitoring, and Anomaly Detection for Scientific Workflows
Anirban Mandal, Paul Ruth, Ilya Baldin, **Dariusz Krol**, **Gideon Juve**, **Rajiv Mayani**, **Rafael Ferreira da Silva**,**Ewa Deelman**, et al.
Workshop on Large-Scale Parallel Processing (LSPP 2016*)*

Tunneling and Speedup in Quantum Optimization for Permutation-Symmetric Problems
Siddharth Muthukrishnan, **Tameen Albash**, and Daniel Lidar
Physics Rev. X 6, 2016

Understanding User Behavior: From HPC to HTC
Stephan Schlagkamp, **Rafael Ferreira da Silva**, **Ewa Deelman**, and Uwe Schwiegelshohn
International Conference on Computational Science (ICCS), 2016

Using Simple PID Controllers to Prevent and Mitigate Faults in Scientific Workflows
Rafael Ferreira da Silva, **Rosa Filgueira**, **Ewa Deelman**, Erola Pairo-Castineira, Ian Overton, and Malcolm Atkinson
11th Workflows in Support of Large-Scale Science (WORKS'16), 2016

What Goes On Behind Closed Doors? How College Dormitory Residents Change to Save Energy During a Competition-based Energy Reduction Intervention
Nicole Sintov, **Ellen Dux**, Agassi Tran, and **Michael Orosz**
International Journal for Sustainability in Higher Education, Vol. 17 Issue 4, 2016

Workflow Performance Profiles: Development and Analysis
Dariusz Krol, **Rafael Ferreira da Silva**, **Ewa Deelman**, and Vickie Lynch
Euro-Par 2016*: Parallel Processing Workshops*

INTELLIGENT SYSTEMS

A Comparison Between Deep Neural Nets and Kernel Acoustic Models for Speech Recognition
Zhiyun Lu, Dong Quo, Alireza Bagheri Garakani, **Kuan Liu**, Avner May, Aurélien Bellet, et al.
IEEE Int'l Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016

A Document Image Quality Assessment Using Discriminative Sparse Representations
Xujun Peng, **Huaigu Cao**, and **Prem Natarajan**
12th Annual IAPR Workshop on Document Analysis Systems (DAS), 2016

A Multi-media Approach to Cross-lingual Entity Knowledge Transfer
Di Lu, Xiaoman Pan, **Nima Pourdamghani**, Heng Ji, Shih-Fu Chang and **Kevin Knight**
Proceedings, Annual Meeting of the Association for Computational Linguistics, 2016

A Scalable Approach to Incrementally Building Knowledge Graphs
Gleb Gawriljuk, Andreas Harth, **Craig Knoblock,** and **Pedro Szekely**
TPDL 2016 - *20th International Conference on Theory and Practice of Digital Libraries*, 2016

Abstract Meaning Representations as Linked Data
Gully Burns, **Ulf Hermjakob**, and **Jose-Luis Ambite**
The Semantic Web: 15th International Semantic Web Conference, Japan, October 2016

An Automatic Approach for Building Place-Name Datasets from the Web
Ying Zhang, **Yao-Yi Chiang, Craig Knoblock, Ciang Li**, L. Du, S. Liu, and **Sanjay Singh**
Proceedings of 19th AGILE International Conference on Geographic Information Science, 2016

Assessing the Navigation Effects of Click Biases and Link Insertion on the Web
Florian Geigl, **Kristina Lerman**, Simon Walk, Marcus Strohmaier, and Denis Helic
Hypertext Conference, 2016

Automated Detection of Discourse Segment and Experimental Types from the Text of Cancer Pathway Results Sections
Gully Burns, Pradeep Dasigi, Anita de Waard, and Eduard Hovy
Database: Journal of Biological Databases & Curation (Oxford), 2016

Automated Detection of Feeding Strikes by Larval Fish Using Continuous High-Speed Digital Video: a Novel Method to Extract Quantitative Data from Fast, Sparse Kinematic Events
Eyal Shamur, Miri Zilka, **Tal Hassner**, Victor China, Alex Liberzon, and Roi Holzman
Journal of Experimental Biology (JEB), Volume 219, No.11, 2016

Automated Hypothesis Testing with Large Scientific Data Repositories
Yolanda Gil, Daniel Garijo, Varun Ratnakar, **Rajiv Mayani,** Ravali Adusumilli, Hunter Boyce, et al.
Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems (ACS), 2016

BotOrNot: A System to Evaluate Social Bots
Clayton Davis, Onur Varol, **Emilio Ferrara**, Alessandro Flammini, and Fillippo Menczer
WWW'16: Proceedings of the 25th International Conference on World Wide Web, 2016

Capturing the Interplay of Dynamics and Networks Through Parameterizations of Laplacian Operators
Xiaoran Yan, Shang-hua Teng, **Kristina Lerman**, and Rumi Ghosh
Peer Journal of Computer Science, 2016

Computationally Efficient Template-Based Face Recognition
Yue Wu, **Stephen Rawls**, **Wael AbdAlmageed**, and **Prem Natarajan**
International Conference on Patten Recognition (ICPR), 2016

Curating Central Injection Studies from the Literature Using a General Purpose Knowledge Management Strategy
Gully Burns, Alice Hernandez, and Arshad Khan
Annual Meeting for the Society for Neuroscience, San Diego, 2016

Cyber-Innovated Watershed Research at the Shale Hills Critical Zone Observatory
Xuan Yu, Chris Duffy, **Yolanda Gil,** Lorne Leonard, Gophal Bhatt, and Evan Thomas
IEEE Systems Journal, 10(3), 2016

Cycles of Scientific Investigation in Discourse
Gully Burns, Anita de Waard, Pradeep Dasigi, and Eduard Hovy
Biocreative, 2016

Dense Correspondences Across Scenes and Scales
Moria Tau and **Tal Hassner**
IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), 38(5), 2016

Detection of Promoted Social Media Campaigns
Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini
Tenth International AAAI Conference on Web and Social Media, 2016

Dimensional Data
Ekaterina Merkurjev, Andrea Bertozzi, **Xiaoran Yan**, and **Kristina Lerman**
Inverse Problems, 2016

Distribution and Inference
Jerry Hobbs and **Jonathan Gordon**
ESSLLI Workshop on Distributional Semantics and Linguistic Theory (DSALT), 2016

Do We Really Need to Collect Millions of Faces for Effective Face Recognition?
Iacopo Masi, Anh Tuan Tran, **Tal Hassner,** Jatuporn Toy Leksut and Gerard Medioni
European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016

Does String-Based Neural MT Learn Source Syntax?
Xing Shi, Inkit Padhi, and **Kevin Knight**
Proceedings, Conference on Empirical Methods in Natural Language Processing, 2016

Dynamically Generated Metadata and Replanning by Interleaving Workflow Generation and Execution
Yolanda Gil and Varun Ratnakar
Proceedings of the Tenth IEEE International Conference on Semantic Computing (ICSC), 2016

Efficient Graph-based Document Similarity (Best Research Paper Nominee)
Christian Paul, Achim Rettinger, Aditya Mogadala, **Craig Knoblock**, and **Pedro Szekely**
The Semantic Web. Latest Advances and New Domains. 13th Extended Semantic Web Conference, 2016

Emotions, Demographics and Sociability in Twitter Inactions
Kristina Lerman, Megha Arora, Luciano Gallegos, Ponnurangam Kumaraguru, and David Garica
International Conference on the Web and Social Media, 2016

Enhancing Reproducibility for Computational Methods
Victoria Stodden, Marcia McNutt, David Bailey, **Ewa Deelman**, **Yolanda Gil,** Brooks Hanson, Michael Heroux, et al.
Science, 354, 2016

Evaluating Induced CCG Parsers on Grounded Semantic Parsing
Yonatan Bisk, Siva Reddy, John Blitzer, Julia Hockenmaier, and Mark Steedman
Conference on Empirical Methods in Natural Language Processing, 2016

Evidence of Online Performance Deterioration in User Sessions on Reddit
Philipp Singer, **Emilio Ferrara**, **Farshad Kooti**, Marcus Strohmaier, and **Kristina Lerman**
PLoS ONE, 11(8), 2016

Extracting Structured Scholarly Information from the Machine Translation Literature
Eunsol Choi, **Matic Horvat**, **Jonathan May, Kevin Knight,** and **Daniel Marcu**
Proceedings, International Conference on Languages Resources and Evaluation, 2016

Face Recognition Using Deep Multi-Pose Representations
Wael AbdAlmageed, Yue Wu, Stephen Rawls, Shai Harel, **Tal Hassner**, Iacopo Masi, Jongmoo Choi, Jatuporn Lekust, Jungyeon Kim, Ram Nevatia**, Prem Natarajan**, et al.
IEEE Winter Conference on Applications of Computer Vision, 2016

From Image to Translation: Processing the Endangered Nyushu Script
Tongtao Zhang, Aritra Chowdhury, Nimit Dhulekar, Jinjing Xia, **Kevin Knight**, Heng Ji, et al.
ACM Transactions on Asian and Low-Resource Language Information Processing, 2016

Generating English from Abstract Meaning Representation
Nima Pourdamghani and **Kevin Knight**
Proceedings, International Natural Language Generation, 2016

Generating Topical Poetry
Marjan Ghazvininejad, Xing Shi, Yejin Choi, and **Kevin Knight**
Proceedings, Conference on Empirical Methods in Natural Language Processing, 2016

Geography of Emotion: Where in a City are People Happier?
Luciano Gallegos, **Kristina Lerman,** Arthur Huang, and David Garcia
WWW Workshop on Modeling Social Media (MSM), 2016

Grapheme-to-Phoneme Models for (Almost) Any Language
Aliya Deri and **Kevin Knight**
Proceedings, Annual Meeting of the Association for Computational Linguistics, 2016

How Business Clusters and Destination Choice are Connected: A Model Based on Social Media Data
Arthur Huang, **Luciano Gallegos**, and **Kristina Lerman**
Proceedings of 6th TRB Innovations in Travel Modeling Conference, 2016

How the Structure of Wikipedia Articles Influences User Navigation
Daniel Lamprecht, **Kristina Lerman**, Denis Helic, and Markus Strohmaier
New Review of Hypermedia and Multimedia, 2016

Information is Not a Virus, and Other Consequences of Human Cognitive Limits
Kristina Lerman
Future Internet, 8(2), 2016

Intelligent Systems for Geosciences
Yolanda Gil
Annual Meeting of the Geological Society of America, September 2016 *(Invited)*

LATCH: Learned Arrangements of Three Patch Codes
Gil Levi and **Tal Hassner**
IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016

Latent Space Model for Multi-Modal Social Data
Yoon-Sik Cho, **Greg Ver Steeg**, **Emilio Ferrara**, and **Aram Galstyan**
Proceedings of World Wide Web Conference (WWW), 2016

Learning Document Binarization from Data
Yue Wu, **Prem Natarajan**, **Stephen Rawls**, and **Wael AbdAlmageed**
IEEE International Conference on Image Processing (ICIP), 2016

Learning the Semantics of Structured Data Sources
Mohsen Taheriyan, **Craig Knoblock, Pedro Szekely,** and **José Luis Ambite**
Journal of Web Semantics, 2016

Leveraging Entity Linking and Related Language Projection to Improve Name Transliteration
Ying Lin, Xiaoman Pan**, Aliya Deri,** Heng Ji, and **Kevin Knight**
Proceedings, ACL 2016 *Workshop on Named Entities*, 2016

Leveraging Linked Data to Discover Semantic Relations within Data Sources
Mohsen Taheriyan, **Craig Knoblock, Pedro Szekely**, and **José Luis Ambite**
ISWC 2016 - *15th International Semantic Web Conference*, 2016

Leveraging the Contributions of the Casual Majority to Identify Appealing Web Content
Tad Hogg and **Kristina Lerman**
Proceedings, 4th AAAI Conference on Human Computation and Crowdsourcing, 2016

Maximizing Correctness with Minimal User Effort to Learn Data Transformations
Bo Wu and **Craig Knoblock**
Proceedings of the 12th International Conference on Intelligent User Interfaces, 2016

Modeling Concept Dependencies in a Scientific Corpus
Jonathan Gordon, **Linhong Zhu**, **Aram Galstyan**, **Prem Natarajan**, and **Gully Burns**
Proceedings, Annual Meeting of the Association for Computational Linguistics, Berlin, 2016

Modified Cheeger and Ratio Cut Methods Using the Ginzburg-Landau Functional for Classification of High-Dimensional Data
Ekateruba Merkurjev, Andrea Bertozzi, **Xiaoran Yan**, and **Kristina Lerman**
Inverse Problems, 2016

Multi-Source Neural Translation
Barret Zoph and **Kevin Knight**
Proceedings, Annual Meeting of the North American Association for Computational Linguistics, 2016

Name Tagging for Low-Resource Incident Languages Based on Expectation-Driven Learning
Boliang Zhang, Xiaoman Pan, Tianiu Wang, **Ashish Vaswani**, Heng Ji, **Kevin Knight** and **Daniel Marcu**
Proceedings, Annual Meeting of the North American Association for Computational Linguistics, 2016

Natural Language Communication with Robots
Yonatan Bisk, Deniz Yuret, and **Daniel Marcu**
Proceedings, Annual Meeting of the North American Association for Computational Linguistics, 2016

Network Structure and Resilience of Mafia Syndicates
Santa Agreste, Salvatore Catanese, Pasquale De Meo, **Emilio Ferrara**, and Giacomo Fiumara
Information Sciences 351, 2016

Obfuscating Gender in Social Media Writing
Siva Reddy and **Kevin Knight**
Proceedings, EMNLP Workshop on Natural Language Processing and Computational Social Science, 2016

OntoSoft: A Distributed Semantic Registry for Scientific Software
Yolanda Gil, Daniel Garijo, Saurabh Mishra, and Varun Ratnakar
Proceedings of the Twelfth IEEE Conference on eScience, 2016

OSoMe: the IUNI Observatory on Social Media
Clayton Davis, Giovanni Ciampaglia, Luca Aiello, Keychul Chung, Michael Conover, **Emilio Ferrara**, et al.
Peer Computer Science 2, 2016

Partitioning Networks with Node Attributes by Compressing Information Flow
Laura Smith, **Linhong Zhu**, **Kristina Lerman**, and Allon Percus
ACM Transactions on Knowledge Discovery Data, 11(2), 2016

Pooling Faces: Template-Based Face Recognition with Pooled Face Images
Tal Hassner, Iacopo Masi, Jungyeon Kim, Jongmoo Choi, Shai Harel, **Prem Natarajan**, et al.
IEEE CVPR Workshop on Biometrics, 2016

Portrait of an Online Shopper: Understanding and Predicting Consumer Behavior
Farshad Kooti, Kristina Lerman, Luca Maria Aiello, Mihajlo Grbovic, Nemanja Djuric, and Vladan Radosavljevic
9th ACM International Conference on Web Search and Data Mining, 2016

Pose-Aware Face Recognition in the Wild
Iacopo Masi, **Stephen Rawls**, **Prem Natarajan**, and Gerard Medioni|
IEEE Computer Vision and Pattern Recognition (CVPR), 2016

Predicting Online Extremism, Content Adopters, and Interaction Reciprocity
Emilio Ferrara, Wen-Qiang Wang, Onur Varol, Alessandro, Flammini, and **Aram Galstyan**
International Conference on Social Informatics, 2016

Proceedings of the Fifteenth International Semantic Web Conference
Eds: Paul Groth, Elena Simperl, Alasdair Gray, Marta Sabou, Freddy Lecue, Markus Krötzsch, Fabian Flöck, and **Yolanda Gil**
Springer Verlag, 2016

Relative Value of Diverse Brain MRI and Blood-Based Biomarkers for Predicting Cognitive Decline in the Elderly
Sarah Madsen, **Greg Ver Steeg**, Madelaine Daianu, Adam Mezher, Neda Jahanshad, Talia Nir, Xue Hua,
Boris Gutman, **Aram Galstyan**, et al.
SPIE Medical Imaging, 2016

Reproducibility in Computer Vision: Towards Open Publication of Image Analysis Experiments as Semantic Workflows
Ricky Sethi and **Yolanda Gil**
Proceedings of the Twelfth IEEE Conference on eScience, 2016

Scalable Query and Analysis for Social Networks: An Integrated High-Level Dataflow System with Pig and Harp
Tal-Lon Wu, Bingjing Zhang, Clayton Davis, **Emilio Ferrara**, Alessandro Flammini, Filippo Menczer, and Judy Qiu
Big Data in Complex and Social Networks, 2016

Scalable Temporal Latent Space Inference for Link Prediction in Dynamic Social Networks
Linhong Zhu, Dong Guo, Junming Yin, **Greg Ver Steeg**, and **Aram Galstyan**
IEEE Transactions on Knowledge and Data Engineering, 2016

Semantic Labeling: A Domain-Independent Approach
Minh Pham, Suresh Alse, **Craig Knoblock, and Pedro Szekely**
15th International Semantic Web Conference, 2016

SemEval-2016 Task 8: Meaning Representation Parsing
Jonathan May
Proceedings, International Workshop on Semantic Evaluation (SemEval), 2016

Simple, Fast Noise Contrastive Estimation for Large RNN Vocabularies
Barret Zoph, Ashish Vaswani, **Jonathan May,** and **Kevin Knight**
Proceedings, Annual Meeting of the North American Association for Computational Linguistics, 2016

Software and Workflow Provenance: Documenting Scientific Methods
Yolanda Gil
American Geophysical Union Fall Meeting, December 2016 *(Invited)*

Social Bots Distort the 2016 US Presidential Election Online Discussion
Allessandro Bessi and **Emilio Ferrara**
First Monday 21 (11), 2016

Social Politics: Agenda Setting and Political Communication on Social Media
Xinxin Yang, Bo-Chiuan Chen, Mrinmoy Maity, and **Emilio Ferrara**
International Conference on Social Informatics, 2016

Special Issue: Geoscience Papers of the Future
Eds: Cedric David, Christopher Duffy, **Yolanda Gil,** Scott Peckham, and Karan Venayagamoorthy
Earth and Space Science, 2016

Style in the Age of Instagram: Predicting Success within the Fashion Industry using Social Media
Jaehyuk Park, Giovanni Ciampaglia, and **Emilio Ferrara**
Proceedings of the 19th ACM Conference on Computer-Supported Communication, 2016

Supertagging with LSTMs
Ashish Vaswani, **Yonatan Bisk**, Kenji Sagae, and Ryan Musa
Proceedings, Annual Meeting of the North American Association for Computational Linguistics, 2016

Teaching Big Data Analytics Skills with Intelligent Workflow Systems
Yolanda Gil
Proceedings of the Sixth Symposium on Educational Advances in Artificial Intelligence (EAAI), 2016

Temporal Learning and Sequence Modeling for a Job Recommender System
Kuan Liu, **Xing Shi**, **Anoop Kumar**, **Linhong Zhu**, and **Prem Natarajan**
Proceedings of the ACM Recommender Systems Challenge, 2016

Texture Instance Similarity via Dense Correspondences
Tal Hassner, Gilad Saban and Lior Wolf
IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016

The CUDA LATCH Binary Descriptor: Because Sometimes Faster Means Better
Christopher Parker, Matthew Daiter, Kareem Omar, Gil Levi and **Tal Hassner**
Workshop on Local Features: State of the art, open problems and performance evaluation, at the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016

The DARPA Twitter Bot Challenge
V. S. Subrahmanian, Amos Azaria, Skyler Durst, Vadim Kagan, **Aram Galstyan, Kristina Lerman, Linhong Zhu, Emilio Ferrara**, et al.
IEEE Computer Magazine, 2016

The Information Sieve
Greg Ver Steeg and **Aram Galstyan**
International Conference on Machine Learning (ICML), 2016

The IS-GEO Research Collaboration Network: Fostering Collaborations for Intelligent Systems Research to Support Geosciences
Yolanda Gil and Suzanne Pierce
American Geophysical Union Fall Meeting, December 2016 *(Invited)*

The Geoscience Paper of the Future: Best Practices for Documenting and Sharing Research from Data to Software to Provenance
Xuan Yu, **Yolanda Gil,** Cédric David, Ibrahim Demir, Bakinam Essawy, Robinson Fulweiler, Jonathan Goodall, et al.
American Geophysical Union Fall Meeting, December 2016
Communications of the ACM 59 (7), 2016

The Importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors
Jon-Patrick Allem and **Emilio Ferrara**
Journal of Medical Internet Research 18 (8), 2016

The Majority Illusion in Social Networks
Kristina Lerman, Xiaoran Yan, and **Xin-Zeng Wu**
PLoS ONE, 11(2), 2016

The Rise of Social Bots
Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini
Communications of the ACM 59 (7), 2016

Towards a Dataset for Human Computer Communication via Grounded Language Acquisition
Yonatan Bisk, Daniel Marcu, and **William Wong**
Proceedings, AAAI Workshop on Symbolic Cognitive Systems, 2016

Toward Interpretable Topic Discovery via Anchored Correlation Explanation
Kyle Reing, **David C. Kale**, **Greg Ver Steeg**, and **Aram Galstyan**
ICML Workshop on Human Interpretability in Machine Learning, 2016

Towards the Geoscience Paper of the Future: Best Practices for Documenting and Sharing Research from Data to Software to Provenance
Yolanda Gil, Cédric David, Ibrahim Demir, Bakinam T. Essawy, Robinson Fulweiler, et al.
Earth and Space Sciences, 3(10), 2016

Transfer Learning for Low-Resource Neural Machine Translation
Barret Zoph, Deniz Yuret, **Jonathan May,** and **Kevin Knight**
Proceedings, Conference on Empirical Methods in Natural Language Processing, 2016

Twitter Session Analytics Profiling Short-Term Behavioral Changes
Farshad Kooti, Esteban Moro and **Kristina Lerman**
*Proceedings of 8th International Conference (SocInfo*2016*), Springer,* 2016

Unsupervised Entity Resolution on Multi-type Graphs
Linhong Zhu, **Majid Ghasemi-Gol**, **Pedro Szekely, Aram Galstyan**, and **Craig Knoblock**
In *ISWC* 2016 *- 15th International Semantic Web Conference*, 2016

Unsupervised Neural Hidden Markov Models
Ke Tran, Yonatan Bisk, Ashish Vaswani, **Daniel Marcu,** and **Kevin Knight**
Proceedings, EMNLP Workshop on Structured Prediction, 2016

Variational Information Maximization for Feature Selection
Shuyang Gao, **Greg Ver Steeg**, and **Aram Galstyan**
Advances In Neural Information Processing Systems, 2016

Why Neural Translations are the Right Length
Xing Shi, Kevin Knight, and Deniz Yuret
Proceedings, Conference on Empirical Methods in Natural Language Processing, 2016

INTERNET AND NETWORKED SYSTEMS

A Candidate Approach for Optical In-Network Computation
Joseph Touch, Yinwen Cao, Morteza Ziyadi, Ahmed Almainman, Ahmed Mohajerin Araiei, et al.
IEEE Summer Topicals, July 2016 (Invited Paper)

Accelerating Physical Level Sub-Component Power Simulation by Online Power Partitioning
Siddharth Bhargav and **Young Cho**
IEEE Int'l Symposium on Quality Electronic Design, 2016

All Optical Signal Level Swapping and Multi-Level Amplitude Noise Mitigation Using Three Optical Parametric Gain Regions
Yinwen Cao, Morteza Ziyadi, Youichi Akasaka, Amirhossein Mohajerin-Ariaei, Jeng-Yauan Yang, Ahmed Almaiman, **Joseph Touch**, et al.
Optics Letters, V41 N4, February 2016

Anycast Latency: How Many Sites Are Enough?
Ricardo de O. Schmidt, **John Heidemann,** and Jan Harm Kuipers
Technical Report ISI-TR-2016-708. USC/Information Sciences Institute

Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event
Giovane Moura, Ricardo de O. Schmidt, **John Heidemann**, Wouter de Vries, Moritz Müller, et al.
Proceedings of the ACM Internet Measurement Conference, November 2016 | Technical Report ISI-TR-2016-709b

Assessing Co-Locality of IP Blocks
Manaf Gharaibeh, Han Zhang, Cristos Papadopoulos, and **John Heidemann**
Proceedings of the Passive and Active Measurement Workshop, Crete, Greece, March 2016

AuntieTuna: Personalized Content-Based Phishing Detection
Calvin Ardi and **John Heidemann**
Proceedings, NDSS Workshop on Usable Security, 2016

Authorization and Access Control: ABAC
Ted Faber, Stephan Schwab, and **John Wroclawski**
Book chapter, The GENI Book, Springer Int'l Publishing, Switzerland, 2016

Beliefs about Cybersecurity Rules and Passwords: A Comparison of Two Survey Samples of Cybersecurity Professionals Versus Regular Users
Ross Koppel, **Jim Blythe**, Vijay Kothari, and Sean Smith
SOUPS Workshop on Security Fatigue, 2016

Defending Cyber-Physical Attacks on Oil Pipeline Systems: A Game-Theoretic Approach
Yatin Wadhawan and **Clifford Neuman**
International Workshop on AI for Privacy and Security (PRAISE), The Hague, Netherlands, 2016

Demontration of Automatically Phase-Locked Self-Homodyne Detection with a Low-Power Pilot Tone Based on Brillouin Amplification and Optical Frequency Combs
Yinwen Cao, Ahmed Almaiman, Morteza Ziyadi, Peicheng Liao, Amirhossein Mohajerin-Ariaei, Changjin Bao, **Joseph Touch**, et al.
Optical Fiber Communications Conference, 2016

Demonstration of Multiplexing and Transmission of QPSK-to-16QAM Channels over 100 km using Wave Mixing for Aggregation and Noise Mitigation
Ahmed Mohajerin Ariaei, Morteza Ziyadi, Ahmed Almaiman, YinWen Cao, Changjing Bao, Peicheng Liao**, Joe Touch**, et al.
CLEO 2016

DETERLab and the DETER Project
John Wroclawski, Terry Benzel, James Blythe, Ted Faber, Alefiya Hussain, Jelena Mirkovic, and **Stephan Schwab**
Book chapter, The GENI Book, Springer Int'l Publishing, Switzerland, 2016

Do You See Me Now? Sparsity in Passive Observations of Address Liveness
Jelena Mirkovic, Genevieve Bartlett, John Heidemann, Hao Shi, and **Xiyue Deng**
Technical Report ISI-TR-2016-710. USC/Information Sciences Institute

DNSSEC Roadblock Avoidance
Sarah Krishnaswamy, **Wes Hardaker**, and Olafur Gudmundsson
RFC 8027 (Best Current Practice), November 2016

Evaluating Resilience of Gas Pipeline Systems Under Cyber-Physical Attacks: A Function-Based Methodology
Yatin Wadhawan and **Clifford Neuman**
CM Workshop on Cyber-Physical Systems Security and Privacy (CPS-SPC), 2016

Experimental Demonstration of Phase-Sensitive Regeneration of a 20-40 Gb/s QPSK Channel Without Phase-Locked Loop Using Brillouin Amplification
Ahmed Almaiman, Yenwin Cao, Morteza Ziyadi, Amirhossein Mohajerin-Araei, Peicheng Liao, Changjing Bao, **Joseph Touch**, et al.
European Conference on Optical Communication, 2016

Experimental Demonstration of Phase-Sensitive Regeneration of a BPSK Channel Without Phase-Locked Loop Using Brillouin Amplification
Ahmed Almaiman, Yinwen Cao, Morteza Ziyadi, Amirhossein Mohajerin-Ariaei, **Joseph Touch**, et al.
Optics Letters, V41 N23, December 2016

Experimental Demonstration of Tunable Homodyne Detection on WDM and Dual-Polarizing PSK Channels by Automatically Locking the Channels to a Local Pump Laser Using Nonlinear Mixing
Ahmed Almaiman, Morteza Ziyadi, Ahmed Mohajerin-Ariaei, Yinwen Cao, Mohammad Chitgarha, Peicheng Liao**, Joseph Touch**, et al.
Optics Letters, V41 N12, June 2016

Experimental Investigation of Quasi-Periodic Powser Spectrums in Raman-Assisted Phase Sensitive Amplifier for 10/20/50-Gbaud QPSK and 10-Gbaud 16QAM Signals
Yinwen Cao, Fatemeh Alishahi, Akasaka Youichi, Morteza Ziyadi, Ahmed Almaiman, Amirhossein Mohajerin-Araei, **Joseph Touch**, et al.
European Conference on Optical Communication, 2016

Good Automatic Authentication Question Generation
Simon Woo, Zuyao Li, and **Jelena Mirkovic**
Proceedings of International Natural Language Generation Conference, 2016

Improving Long-term Accuracy of DNS Backscatter for Monitoring of Internet-Wide Malicious Activity
Abdul Qadeer, John Heidemann, and Kensuke Fukuda
Technical Report ISI-TR-2016-707. USC/Information Sciences Institute

Improving Recall and Security of Passphrases through Use of Mnemonics
Simon Woo and **Jelena Mirkovic**
Proceedings of Passwords Conference, 2016

Large-Scale Simulation of the Quantum Internet
Rodney Van Meter, Shigeya Suzuki, Shota Nagayama, Takahiko Satoh, Takaaki Matsuo, Amin Taherkhani, **Joseph Touch**, et al.
International Conference on Quantum Communication, Measurement and Computing, Singapore, 2016

Life-Experience Passwords (LEPs)
Simon Woo, Elsi Keiser, Ron Artstein, and **Jelena Mirkovic**
Proceedings of ACSAC, 2016

Measuring the Latency and Pervasiveness of TLS Certificate Revocation
Liang Zhu, Johanna Amann, and **John Heidemann**
Proceedings, Passive and Active Measurements Conference, Crete, Greece, March 2016

Middlebox Models Compatible with the Internet
Joseph Touch
USC/ISI Technical Report (ISI-TR-711), 2016

Reconfigurable Optical Inter-Channel Interference Mitigation for Spectrally Overlapped QPSK Signals using Nonlinear Wave Mixing in Cascaded PPLN Waveguides
YinWen Cao, Morteza Ziyadi, Ahmed Mohajerin-Ariaei, Ahmed Almaiman, Peicheng Liao, Changjing Bao, **Joseph Touch**, et al.
Optics Letters, V41 N14, July 2016

Simultaneous All-Optical Phase Noise Mitigation and Automatically Locked Homodyne Reception of an Incoming QPSK Data Signal
Ahmed Mohajerin-Ariaei, Morteza Ziyadi, Ahmed Almaiman, Yinwen Cao, Fatemah Alishahi, Mohammad Chitgarha, **Joseph Touch**, et al.
Optics Letters, V41 N20, Oct. 2016

Specification for DNS over Transport Layer Security (TLS)
Zi Hu, **Liang Zhu, John Heidemann**, Allison Mankin, Duane Wessels, and Paul Hoffman
Technical Report 7858. Internet Request For Comments

The Optical Turing Machine (abstract and poster)
Joseph Touch
IEEE Rebooting Computing Workshop, 2016

Tunable All-Optical WDM Channel Selection Using Raman Assisted Cascaded Parametric Amplification
Fatemeh Alishahi, Yenwin Cao, Amirhossein Mohajerin-Ariaei, Alireza Fallahpour, Ahmed Almainman, Changjing Bao, **Joseph Touch**, et al.
Conference on Lasers and Electro-Optics, 2016

Tunable Optical De-aggregation of a 40-Gbit/s 16-QAM Signal into Two 20-Gbit/s 4-PAM Signals Using a Coherent Frequency Comb and Nonlinear Processing
Morteza Ziyadi, Amirhossein Mohajerin-Ariaei, Yinwen Cao, Ahmed Almaiman, Ahmed Fallahpour, Changjin Bao, **Joseph Touch**, et al.
International Conference on Lasers and Electro-Optics, 2016

Tunable ROADM with Crosstalk Reduction for Overlapped 20-50 Gbaud QPSK WDM Channels using Wave Mixing
Morteza Ziyadi, Amirhossein Mohajerin-Araei, Yenwin Cao, Ahmed Almaiman, Changjing Bao, Fatemeh Alishahi**, Joseph Touch**, et al.
International Conference on Lasers and Electro-Optics, 2016

Validating Agent-Based Modeling of Human Password Behavior
Bruno Korbar, **Jim Blythe**, Ross Koppel, Vijay Kothari, and Sean Smith
AAAI Workshop on Artificial Intelligence for Cyber Security, 2016

INFORMATICS SYSTEMS RESEARCH

Accelerating Data-driven Discovery with Scientific Asset Management
Robert Schuler, Carl Kesselman, and **Karl Czajkowski**
Proceedings of 12th IEEE Intl Conference on eScience, 2016

ERMRest: An Entity-relationship Data Storage Service for Web-based, Data-oriented Collaboration
Karl Czajkowski, Carl Kesselman, Robert Schuler, and **Hongsuda Tangmunarunkit**
Proceedings of the 12th IEEE Intl Conference on eScience, 2016

I'll Take That to Go: Big Data Bags and Minimal Identifiers for Exchange of Large, Complex Datasets
Kyle Chard, Mike D'Arcy, Ben Heavner, Ian Foster, **Carl Kesselman,** et al.
Proceedings 2016 *Conference on Big Data*

Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-source and Incomplete Observations
Ivo Dinov, Ben Heavner, Ming Tang, Gustavo Glusman, Kyle Chard, **Mike Darcy,** Ravi Muddura, Judy Pa, Cathie Spino, **Carl Kesselman,** et al.
PLOS One, 2016

The FaceBase Consortium: A Comprehensive Resource for Craniofacial Researchers
James Brinkley, Shannon Harris, Greg Holmes, Joan Hooper, Ethylin Wang Jabs, Kenneth Jones, **Carl Kesselman**, et al.
Oxford University Press for The Company of Biologists Limited, 2016

PROGRAMS UNDER THE KESTON ENDOWMENT DIRECTORSHIP

In 2015 Michael and Linda Keston created their endowed directorship position with a generous donation. This was followed in early 2016 with the appointment of Dr. Prem Natarajan as the inaugural Michael Keston Executive Director of the USC Information Sciences Institute. Natarajan's goal is to use the endowment to support groundbreaking research in the areas of information processing, and computer and communications technologies. In consonance with these goals, he has established the *Michael Keston Lecture Series* which featured IARPA Director Jason Matheny as the 2016 inaugural speaker. Natarajan also established the *Michael Keston Researcher-in-Residence Program* that allows scientists to visit ISI for a period of time to develop new projects that can evolve into permanent areas of investigation and focus. In 2016 Emilio Ferrara, ISI research lead and research assistant professor, became our first researcher-in-residence. Also, under the endowment, Natarajan has awarded *Keston Research Grants* to three ISI researchers engaged in innovative technology that will ultimately benefit society.

MICHAEL KESTON INAUGURAL LECTURE: ANTICIPATORY INTELLIGENCE, 24 AUGUST 2016

From left: Michael Keston, Dean Yannis Yortsos of USC's Viterbi School of Engineering, IARPA Director Jason Matheny, and Michael Keston Executive Director Prem Natarajan

"My investment in ISI and technology is a return to my first love, which is engineering."
— Michael Keston

"Most of our important decisions are about the future and affecting a future outcome."

— Jason Matheny, Inaugural Lecturer and Director of IARPA

MICHAEL KESTON RESEARCH GRANT RECIPIENTS

JOHN HEIDEMAN
Internet & Networked Systems

The *Understanding Internet Outages* project aims to improve understanding and communicating information about global Internet reliability. After collecting data about outages, this project will create a new web portal for presenting that data.

WEI-MIN SHEN
Intelligent Systems

The *PipeFish* project will build a physical underwater robot that will travel through water pipes and perform real-time and in situ data collection with multiple on-board sensors. Collected data will be visualized and analyzed for fault detection in pipes.

GREG VER STEEG
Intelligent Systems

The *RNA-Seek* project attempts to generalize a recent success using gene expression data to choose more effective treatments for ovarian cancer. The goal is to better exploit gene expression data to guide treatments across a variety of cancer types.

JOHN O'BRIEN
EXECUTIVE VICE DEAN, USC VITERBI SCHOOL OF ENGINEERING
LOUISE DUNN PROFESSOR OF ENGINEERING

The ISI family will forever cherish the memory of John O'Brien, a gentleman scholar, humanist, and scientist who was a friend to many of us and a strong advocate for the Institute. John served as executive vice dean of the Viterbi School of Engineering for more than a decade and as ISI's interim executive director during 2012-2013. He made invaluable contributions to ISI, leading it with wisdom and strength during a period of leadership transition.

John's sense of morality, his innate wisdom, and his quick grasp of details made him an uncommonly effective advisor, whose opinions could be counted upon to serve no purpose other than the mission at hand. On a personal note, I am deeply grateful for everything that I received and learned in my four years of partnership with John.

Premkumar Natarajan
Michael Keston Executive Director, Information Sciences Institute
Vice Dean of Engineering, Viterbi School
Research Professor of Computer Science
University of Southern California

Information Sciences Institute

4676 Admiralty Way #1001; Marina del Rey, CA 90292

310-822-1511 | isi.edu | media@isi.edu

USC Viterbi
School of Engineering