Computer Science and Artificial Intelligence Laboratory

The Computer Science and Artificial Intelligence Laboratory (CSAIL) studies computing and its application to finding solutions to many of the most challenging problems of our lives, our work, and our world. CSAIL research is focused on developing the architectures and infrastructures of tomorrow's information technology, and on creating innovations that will yield long-term improvements in how people live and work. Laboratory members conduct research in almost all aspects of computer science, including artificial intelligence, the theory of computation, systems, machine learning, and computer graphics, as well as exploring revolutionary new computational methods for advancing health care, manufacturing, energy, and human productivity.

With approximately 50 research groups working on hundreds of diverse projects, CSAIL researchers focus on finding innovative ways to make systems and machines operate faster and better—and more safely, more easily, and more efficiently for the benefit of humanity. CSAIL projects fall into three areas of inquiry:

- Artificial intelligence: We seek to understand and develop both living and artificial systems capable of intelligent reasoning, perception, and behavior.
- Systems: We seek to discover new principles, models, metrics, and tools for both hardware- and software-based computer systems.
- Theory: We seek to understand the mathematics of computation and its wide-ranging, real-world consequences.

CSAIL has a long history of technological innovations that have affected how people interact and do business. CSAIL is known as the incubator of some of the greatest technological advances of the past 30 years, including the internet, personal computing, mobile computing, open-source software, microprocessors, robotic surgery, and social networking.

CSAIL's current research addresses some of the grand challenges of the 21st century, including developing personalized learning, securing cyberspace, advancing health informatics, reverse-engineering the human brain, enhancing virtual reality, developing tools for scientific discovery, improving urban infrastructure, and ensuring the health of the environment. Computing is central to solving these challenges and CSAIL contributes to making computing more capable by addressing fundamental algorithmic and systems questions at the core of computing. Key CSAIL initiatives currently under way include tackling the challenges of big data, developing new models for wireless and mobile systems, securing computers and the cloud against cyber attacks, rethinking the field of artificial intelligence, and developing the next generation of robots. Advanced software-based medical instrumentation and medical informatics systems to aid clinical decision making are being investigated. Advancements in biological research are also under way, including developments in the field of computational biology and the application of machine learning to the interpretation of complete genomes and the understanding of gene regulation.

CSAIL research is sponsored by a large number of diverse sources, from US government contracts to the private sector. US government sponsors include the Air Force Research Laboratory and the Air Force Office of Scientific Research, the Army Research Office, the Defense Advanced Research Project Agency (DARPA), the Department of Defense Research and Engineering, the US Department of Education, the US Department of Energy, the Intelligence Advanced Research Projects Activity, the National Institutes of Health, the National Institute of Justice, the National Science Foundation, the Navy (including the Office of Naval Research and Naval Air Systems Command), and the Space and Naval Warfare Systems Center. US and international industrial sponsors include Boeing, BMW of North America, Ford, Foxconn, General Electric, Intel Corporation, Lockheed Martin Advanced Technology Laboratories, Microelectronics Advanced Research Corporation, Nissan Motor Company, Nippon Telegraph and Telephone Corporation, Nokia, Northrop Grumman Corporation, Qatar Computing Research Institute, Quanta Computer Inc., Shell, Siemens, and State Farm Insurance. Other organizations sponsoring research include Aarhus University, Battelle Memorial Institute, Delta Electronics Foundation, DSO National Laboratories, Epoch Foundation, Industrial Technology Research Institute, Nanyang Technological University, and the Singapore–MIT Alliance.

Research Projects

CSAIL has many single- and multi-investigator projects as well as a number of virtual centers and large-scale projects. Large-scale projects and collaborations include the following:

Qmulus Project

In 2005, CSAIL started a 10-year, $45.5 million research collaboration called T-Party with Quanta Computer, Inc.. The project was renamed Qmulus in 2010 to reflect the shift in CSAIL's research emphasis toward cloud computing. We describe below the progress made during fiscal year 2014 in the four broad areas of research funded by Qmulus:

- Cloud technologies**:** The general emphasis of Qmulus research is in hardware and software security. We are building a secure processor that can be used to maintain privacy in big-data applications, using oblivious random access memory so that memory access patterns do not leak private information (professor Srinivas Devadas). In another hardware-related project, we are constructing a next-generation prototype database machine called BlueDBM, a distributed flash-based storage system with hardware acceleration (professor Arvind Mithal). In software, our research focuses on three aspects: cloud security, multi-core applications, and cloud infrastructure (professors Frans Kaashoek, Robert Morris, and Nikolai Zeldovich). Finally, we focused on a number of database-related applications that might be a good target for the BlueDBM system (professor Sam Madden).
- Multimedia**:** We continue to develop and improve human-language technologies for natural human–computer interactions and scientific discovery, focusing on health- and wellness-related applications such as logging food intake, interactively exploring an online database of drug side effects, and analyzing research literature to shed light on the relationship

between wellness and environmental toxins such as glyphosate in genetically modified foods (Dr. James Glass and Dr. Stephanie Seneff). We continue to develop low-cost hardware implementation for speech recognition (professor Anantha Chandrakarsen). Finally, we continue to improve our multi-view autostereoscopic systems, focusing on view-zone transitions, view expansion, and the creation of custom parallax-barrier masks for ultra-high-definition displays (professor Wojciech Matusik).

- Health care: We improved our method of using computational photography and motion magnification in medical applications by introducing a phase-based video processing technique, and by introducing a real-time improvement using a Riesz pyramid extension (professors Fredo Durand and William Freeman). In cloud-based medicine and health care, we developed new tools for visualizing large data sets to support interactive knowledge discovery; developed novel techniques to analyze cardiovascular signals relative to heart rate; analyzed large data sets for risk stratification; and developed techniques for enhanced presentation of videos (professor John Guttag).
- Education technology: With respect to flexible content creation and viewing, we continue to improve the Pentimento video software user and web interfaces, and we started to use this tool in two MIT courses (professor Fredo Durand). We continue to use crowd computing techniques for education, emphasizing the improvement of video lectures and tutorials, as well as the visualizing and clustering of student program submissions (professor Jason Miller). Using human-language technologies, we developed tools to link various contents for a given course, and to link similar courses on the same Massive Open Online Course (MOOC) platform (professor Victor Zue). We continue to develop tools to implement the "see one, do one, teach one" pedagogy for online teaching of Course 6.004x (professor Lucas Ward and Dr. Chris Terman). Finally, we investigated how to standardize MOOC student online learning behavior data with the goal of schematizing observations collected from different platforms (Dr. Una-May O'Reilly).

Qatar Computing Research Institute

In 2012, CSAIL started a seven-year, $35 million research collaboration with the Qatar Computing Research Institute to collaborate on a wide range of research topics in computer science. We describe below progress made during FY2014 in five broad areas of research:

- Cyber security: The objective is to develop new techniques that can remove many of the vulnerabilities that attackers exploit and that can predict and intercept new (zero-day) attacks that exploit previously unknown vulnerabilities. These techniques include a secure programming language (professor Solar Lezama), homomorphic and functional encryption (professors Shafi Goldwasser and Vinod Vaikuntanathan), architecture for secure computation on encrypted data (Professor Devadas), and model-based intrusion prediction (Dr. Howard Shrobe).
- Arabic speech and language development: A number of subprojects are ongoing in human-language technologies, including speech-recognition error reduction based on diacritized lexical representations (Dr. Glass), improved dialect

recognition using deep neural networks (Dr. Glass), development of a sampling-based parser that can easily handle arbitrary global features (professor Regina Barzilay), and improved semantic parsing for understanding by means of re-ranking models (Drs. Glass and Seneff).

- Advanced analytics and visualization in sports: Projects in this area include video magnification in the presence of large confounding motion (professors Durand and Freeman), backward-compatible masks for autostereoscopic screens (Professor Matusik), and the use of machine-learning techniques to build a program that automatically classifies basketball plays (Professor Guttag).
- Data management for social computing: There are two research thrusts in this effort. The Humanitarian Technologies team collaborated with the International Community of the Red Cross on its "Restoring Family Links" application, and developed a new website and tutorials for the framework (Dr. Kagal). The Social Web Architecture team continued software development and standards activities with the World Wide Web Consortium (W3C) Linked Data Platform Working Group (professor Tim Berners-Lee).
- Data Integration: We continue to extend our data integration framework with additional capabilities and to test its effectiveness. Specifically, we have expanded into a text component and a transformation component (professors Michael Stonebraker and Sam Madden).

Foxconn/Basic Research in Interactive Computing

The Basic Research in Interactive Computing project is a six-year, $5.7 million research collaboration sponsored by Hon Hai/Foxconn Technology Group. This research collaboration completed phase two and began phase three in January 2014. The project is currently funding six major research areas in several sectors of computer science, ranging from networking, human–computer interactions, and computer graphics and vision, to theory. The research is predicated on the belief that computers and information devices are fast becoming interactive; they interact with other computers, with their environments, and, above all, with humans. Each form of interaction adds a new dimension to the challenge of modeling and understanding the behavior of computer systems as well as the task of building and using these systems.

During the current year, the collaboration supported five principal investigators in the following research areas:

- Advanced hand-tracking and gesture-based interaction (professor Randall Davis)
- Reducing the energy consumption of cellular wireless interfaces (professor Hari Balakrishnan)
- Bringing software-defined networks to wireless local area networks (professor Dina Katabi)
- Multicore software technology for video-conferencing systems (professor Charles Leiserson)
- iDiary: activity modeling using audio, video, and GPS streams from phones (professor Daniela Rus)

A total of eight Foxconn engineers visited CSAIL during this academic year to facilitate technology transfer and to receive training in advanced research and development in computer science. Each engineer spent six months at CSAIL. In December 2013, video presentations summarizing the research results for the four projects were made via video conferencing from CSAIL to Foxconn in Taiwan.

The MIT Angstrom Project: Universal Technologies for Exascale Computing

The CSAIL-led Project Angstrom team was one of four teams selected by DARPA for funding under the Ubiquitous High-Performance Computing Program in 2010. The team was charged with the task of rethinking computing and creating a fundamentally new computing architecture to meet the challenges of computing in the 2020 timeframe. The CSAIL team is the only university-led team of the four; it consists predominantly of university researchers, the majority of whom are MIT faculty. The Project Angstrom team is strongly interdisciplinary, involving faculty from MIT's CSAIL, the Microsystems Technology Laboratories (MTL), and the Research Laboratory of Electronics (RLE), as well as faculty from the University of Maryland Department of Electrical and Computer Engineering (ECE). Project Angstrom's goal is to create the fundamental technologies necessary for future extreme-scale computers. Extreme-scale computers face several major challenges, the most difficult four being energy efficiency, scalability, programmability, and dependability. To address these challenges, the Project Angstrom team reexamined every layer of system design and interfaces, including circuits, hardware architecture, operating systems, runtime software systems, compilers, programming languages, and applications. Project Angstrom has pioneered self-aware computing, adding self-awareness to each of these layers to improve energy efficiency and dependability while retaining programmability and scaling to thousands of processors.

In academic year 2011–2012, the team completed an overall design of a 1,000-core Angstrom processor. In academic year 2013–2014, the team built several prototype processors that exploit many of the Angstrom Project design innovations.

- **Energy-aware processor:** The team has built a prototype of a single "tile" of Angstrom that has energy-aware features. The tile consists of a processor, an adaptive cache, and energy-monitoring circuits.
- **Execution migration machine**: The Angstrom team has prototyped hardware-level instruction-granularity thread migration in a 110-core chip multiprocessor. Implemented in 45 nanometer (nm) application-specific integrated circuit technology, the chip occupies 100 square millimeters. The fabricated chips were received in July 2013 and were tested in January 2014.
- **OMNI: In-network coherence processor—**A 11-mm-by-13-mm 36-core chip prototype, fabricated with IBM 45 nm silicon on insulator technology, carries a practical in-network coherence architecture and design. Fabricated chips were received in July 2013 and were tested in January 2014.The OMNI architecture contains distributed in-network global ordering and an adapted coherence protocol closely designed with the network, connected to 36 Freescale Power Architecture cores and Cadence on-chip memory controllers. In addition, the natural optimizer compiler is compliant with standard bus protocols, allowing system-on-chip buses to be interchanged easily.

Intel Science and Technology Center in Big Data

MIT principal investigators Samuel Madden and Michael Stonebraker continue to head the Intel Science and Technology Center in Big Data, which is based at CSAIL. Professors Madden and Stonebraker are leading a team of 20 researchers (from Brown University, MIT, Portland State University, University of Tennessee, University of California, Santa Barbara, and University of Washington) in a project to build new software tools to process and manage massive amounts of data. Specifically, the center is focused on new data management systems and new computer architectures that, together, can help users process data that exceeds the scale, rate, or sophistication of data processing that existing systems provide. The center is developing new technologies in a number of areas, including data-intensive scalable computing, machine learning, computer architecture and domain sciences (genomics, medicine, oceanography, imaging, and remote sensing).

DARPA Robotics Challenge

Announced in March 2012, the DARPA Robotics Challenge (DRC) is inspired by the Fukushima nuclear disaster and other situations in which human-like mobility and dexterity are required, but that are too hazardous for the physical presence of humans. (Other examples include the Deepwater Horizon oil rig blowout and the wildfires in Yarnell, Arizona.) Through the competition, DARPA aims to accelerate the development of remotely operated humanoid robots that can walk and climb over complex terrain within disaster zones (rubble, stairs, ladders), and perform useful tasks that require handling objects (levers, pumps, valves) and using tools (wrenches, drills, saws)—even climbing into and driving ordinary vehicles.

CSAIL formed a team in spring 2012 to compete in the Virtual Robotics Competition (VRC) in June 2013. This was a simulation-based challenge that required the teams to demonstrate perception, planning and control, user interface, and networking capabilities to command the virtual robot through three tasks in the “cloud.” Team MIT was built on a substantive new collaboration between professor Seth Teller’s research group, focusing on perception and human-robot interaction, and professor Russ Tedrake’s research group, focusing on real-time planning and control; neither group had ever worked on humanoid robots before. There were hundreds of initial applicants, but only 26 teams from eight countries qualified to compete in the VRC. Team MIT finished third, alongside teams from top industrial laboratories and teams that had been leaders in the field of humanoid robotics for years. Team MIT was also chosen as one of the teams to receive a humanoid robot built by Boston Dynamics as government-furnished equipment to use in the robot competition. The robot arrived at MIT at the end of August 2013.

Less than four months after receiving the robot, Team MIT traveled to Homestead, Florida, for the DRC trials competition. This was an extremely high-pressure, high-stakes competition, with 16 of the most advanced robots in the world facing off in a live competition, with only the top eight teams eligible to earn continued support from DARPA and entry into the DRC finals. Teams had to compete in eight challenging events, ranging from having the robot turn valves and open doors to guiding it through negotiating debris, climbing a ladder, and driving a car. Each team had only one chance at each of the challenges.

Again, Team MIT competed against the top robotics teams in the world with a student-dominated team, finishing in fourth place and advancing into the final round. The MIT team achieved this with a level of sophistication and autonomy that went far beyond what was required for the relatively simple trials and will compete in the final competition, now scheduled for June 4–5, 2015.

Team MIT has taken a long-term basic research approach to the problem presented in the DRC competition, developing fundamental new algorithms in planning, control, estimation, and human–robot interaction. Even though it is a competition, the team has have decided to publish its research results continuously, and has even decided to publish much of its software as open source software, with the priorities of advancing the field and bringing recognition for the students involved.

World Wide Web Consortium

The World Wide Web Consortium (W3C) was founded at MIT in 1994 by the inventor of the web, Tim Berners-Lee. W3C is responsible for developing and maintaining the standards that make the web work and for ensuring the long-term growth of the web. Three hundred seventy-five member organizations, including most of the world's leading technology companies, are working to enhance the capabilities of web documents and create an open web platform for application development. This platform would be available across a wide range of devices, enabling collaboration and sharing of data and information.

Recent Focus

In recent years, a great many factors (people, devices, bandwidth, and policy decisions) have extended the reach of the web in society. Video, social networking tools, user-generated content, location-based services, and web access from mobile devices are transforming many industries, including mobile communications, television and publishing, automotive and entertainment industries, games and gaming, and advertising. This transformation has led to greater demands on W3C and other organizations to build robust technology that meets society's needs in areas such as privacy, security, accessibility, and multilingual content.

W3C standards define an open web platform for application development that has the unprecedented potential to enable developers to build rich interactive experiences, powered by vast data stores, that are available on any device. Although the boundaries of the platform continue to evolve, industry leaders speak nearly in unison about HTML5 as the cornerstone for this platform. But the full strength of the platform relies on many more technologies that W3C and its partners are creating, including cascading style sheets, scalable vector graphics, web open font format, real-time communications, the Semantic Web stack, extensible markup language, and a variety of application programming interfaces (APIs). The platform continues to grow, and the W3C community, in turn, is growing to meet the demand.

The demand is also driving W3C to expand its agenda and the size of its community. W3C launched Community and Business Groups in 2011. Less than three years later, more than 4,000 people participated. By making it easier for people to participate, W3C increased the relevance and quality of its work and brought more innovators to the table for pre-standards and standards track work.

The Open Web Platform

The open web platform, a cornerstone of the W3C standardization effort, is a suite of technologies that is transforming business practices, creating new business models, and allowing greater innovation on the web while reducing product lifecycle costs compared with other systems.

W3C is designing royalty-free technologies that:

- Provide a rich interface feature set, including style, interaction, and media,
- Enrich applications through APIs for device capabilities and user data,
- Integrate data and services (mashups, integration of existing databases and services),
- Run on any device (computer, telephone, television, consumer electronics, automobile, etc.) and support interaction through a variety of input and output modes,
- Meet network and communications demands (cross-origin resource sharing, real-time communications),
- Satisfy performance and distribution requirements, enable rapid development and deployment, and facilitate maintenance, and
- Address diverse social requirements for privacy, security, multilingual content, and accessibility.

Centers and Initiatives

CSAIL Alliance Program

The CSAIL Alliance Program (CAP) was approved by the CSAIL Executive Committee in September 2013 as a broader extension of CSAIL's Industry Affiliates Program (IAP). CAP is a membership-based program that serves as a portal for industry, governmental organizations, and other institutions to engage with CSAIL. CAP's mission is to provide a proactive and comprehensive approach to developing strong connections with all CSAIL has to offer. CAP has 44 members representing a variety of industry sectors and regions, including North America, Asia, South America, and Europe. Key benefits of the program include:

- Keeping abreast of the latest research at CSAIL,
- Identifying opportunities to foster and explore new research collaborations, and
- Engaging with CSAIL students for recruiting.

Members of CAP are given access to new technologies and ideas as they move from laboratory to marketplace. Throughout the year, CAP hosts events, seminars, meetings, and activities to help connect member companies with the laboratory. To facilitate collaboration and communication with industry partners, CAP maintains a members-only website, produces research briefings, and publishes a CSAIL Student Profile Book.

CAP's flagship event is the annual meeting, held in May each year. The event is a two-day conference that showcases CSAIL's research and students as well as serves as a forum for members to network, connect with, and learn from each other. More than 100 people from 35 member companies attended the 2014 CAP annual meeting. All activities are designed to help members connect with CSAIL researchers and to create value for their organization.

bigdata@csail

The goal of the MIT Big Data Initiative at CSAIL (bigdata@csail), a multiyear effort launched in May 2012, is to identify and develop new technologies that are needed to solve the next-generation data challenges. These coming challenges will require the ability to scale well beyond what today's computing platforms, algorithms, and methods can provide. The Big Data Initiative is meant to enable people to leverage big data with tools and platforms that are reusable, scalable, and easy to deploy across multiple application domains.

Our approach includes two important aspects. First, we work closely with industry and government to provide real-world applications of the tools and platforms the initiative develops and to drive impact. Promoting in-depth interactions between academic researchers, industry, and government is a key goal. Second, we believe that the solution to the issues raised by big data will be fundamentally multidisciplinary. Our team includes faculty and researchers across many related technology areas, including algorithms, architecture, data management, machine learning, privacy and security, user interfaces, and visualization, as well as domain experts in finance, medicine, smart infrastructure, education, and science. The initiative focuses on four broad research themes: computational platforms; scalable algorithms; privacy and security; and big-data applications.

The Big Data Initiative is supported by a group of sponsoring companies, including American International Group, Alior Bank, British Telecommunications, EMC Corporation, Facebook, Huawei Technologies Co., Intel, Microsoft, Quanta, Samsung, SAP AG, Shell, and Thomson Reuters.

In November 2013, the MIT Big Data Initiative was invited to participate in a White House event, "Data to Knowledge to Action: Building New Partnerships," which highlighted new collaborations in the field of big data that were expected to have significant effects. During this event, we announced two new activities:

- A series of data challenges designed to spur innovation in how people use data to solve problems and to make decisions, and
- A new Big Data and Privacy Working Group that will bring together leaders from industry, government, and academia to address the role of technology in protecting and managing privacy.

The first MIT Big Data Challenge ran from November 2013 to February 2014, co-sponsored by Transportation@MIT, in partnership with the City of Boston. This first challenge focused on transportation in downtown Boston. The challenge made available multiple data sets, including transportation data from more than 2.3 million taxi rides as well as data from local events and social media, along with weather records. The goals were to predict demand for taxis in downtown Boston and to create visualizations that provide new ways to understand public transportation patterns in the city (http://bigdata.csail.mit.edu/challenge).

The White House Office of Science and Technology Policy and MIT cohosted a public workshop called "Big Data Privacy: Advancing the State of the Art in Technology and Practice" on March 3, 2014. The event was part of a series of workshops on big data and privacy organized by the MIT Big Data Initiative and the MIT Information Policy Project. The workshop was also the first in a series of events being held across the country in response to President Obama's call for a review of privacy issues in the context of increased digital information and the computing power to process it (http://web.mit.edu/bigdata-priv/).

During spring 2014, the Big Data Initiative also hosted a workshop at MIT on key challenges in "Big Data Analytics: Challenges in Big Data for Data Mining, Machine Learning and Statistics."

MIT Center for Wireless Networks and Mobile Computing

Wireless@MIT is an interdisciplinary center working to develop the next generation of wireless networks and mobile devices. Headquartered at CSAIL, the center creates a focal point for wireless research at MIT and addresses some of the most important challenges facing the wireless and mobile computing fields. In particular, the center is innovating in four important areas:

- Improving the use of the wireless spectrum,
- Better systems and networking for emerging mobile applications (including video/collaboration, transportation, and healthcare),
- Security and privacy, and
- Low-power design.

The center's work includes contributions from more than 50 MIT faculty, research staff, and graduate students. A key feature of the center is that it encourages collaborations between MIT faculty in various departments and laboratories. In particular, the center has faculty members from the Department of Electrical Engineering and Computer Science (EECS), the Department of Mechanical Engineering, and the Sloan School of Management. Its members also come from many laboratories including CSAIL, MTL, RLE, and the Laboratory for Information and Decision Systems.

The center has eight members—Amazon.com, Cisco, Intel, Mediatek, Microsoft, ST Microelectronics, Telefonica, and Google. The center has been frequently featured in the news because of its research outcomes.

Rethinking AI

Rethinking AI is a CSAIL research initiative that is exploring a number of unconventional approaches to the fundamental goal of developing a computational model of human-like intelligence. The projects in this research initiative come at the problem from a number of directions.

Work in learning to abstract starts from the premise that abstraction remains a key obstacle to learning. To learn abstractions from observations, we must have rich (probabilistic) representations that support the creation of templates and that enable efficient assessment of the predictive value of such templates, i.e., the ability to quickly recognize when and where templates might apply. Two key insights form stepping stones toward these goals: 1) templates must be designed in a way that naturally enables step-wise generalization, and 2) we must develop methods to quickly recognize and track hierarchically organized templates.

Our work has focused primarily on developing and implementing general ways of performing semantic image compression by creating and using templates as building blocks. These building blocks are organized hierarchically in a lattice; they represent salient features and their relations as they are present in the images, from simple edges to high-level objects. In particular, we have developed learning algorithms for optimizing how such lattices are created in response to images. The goal of this is to focus computation on uncovering objects and their parts rather than on evaluating low-level features. We learn not only what the basic building blocks are but also how to use them in interpreting images, i.e., in deciding where to look and what to look for. Our work uses hand-written Hindi letters to demonstrate what the overall approach can accomplish and how such images are interpreted quickly. We believe our framework is particularly well suited to learning abstractions across different modalities and we aim to demonstrate cross-modal transfer in the context of learning to perform visually guided actions.

Work on algorithmic encoding seeks to understand how to tailor and guarantee computational results with large numbers of unreliable, weakly communicating components. We want to adaptively coax a set of noisy, uncertain components collectively to achieve the task at hand. This is a longstanding problem with major implications across a wide range of fields, including AI, fault-tolerant computing, error-correcting codes, bioengineering, and neuroscience. Unlike previous work in fault tolerance, however, we do not conceive of the task as guaranteeing that every operation will succeed (or that recovery is possible when one does not). Instead, we assume that errors are commonplace and heterogeneous, and seek to redefine how algorithms should be written in this context.

We have previously answered fundamental questions about the relation between errors and outcomes; for example, we have derived a theoretical understanding of distributions over outcomes induced from errors. These results highlight the ways in which many local modifications of computations influence global outcomes. Turning this around gives us the means to use noisy distributed systems for modeling probability distributions over global arrangements. Our recent work has focused on answering what computations can and cannot be realized robustly. In particular, we are quantifying the relation between robustness and approximability and that relation's uses in large-scale distributed computation. Broadly speaking, computations that are robust to errors are also provably easy to approximate.

Work in probabilistic programming integrates Bayesian probability with universal computation. Instead of modeling distributions over a set of random variables, probabilistic programs model distributions over the execution histories of programs, including programs that analyze, transform, and write other programs. Probabilistic programs, as the first computationally universal systems for probabilistic modeling, have capacities for abstraction and recursion that can support reflective reasoning, natural language semantics, and learning model structure as probabilistic inference. They can also naturally express core commonsense knowledge about physical objects, intentional agents, and their causal interactions, using software simulators with stochastic dynamics and hidden parameters. These architectures have led to successful new models of people's intuitive physical and psychological reasoning. They have also begun to yield new AI systems whose unusual capabilities showcase the power of probabilistic inference on classic AI problems that had been largely left out of the statistical AI revolution of the last two decades. One engineering result from our work in the past year was showing that probabilistic programming enables solving problems from computer vision, information retrieval, statistics, and robotics using 1/100th the number of lines of code required by standard baselines.

The biggest recent surprise has been in demonstrating a system that infers three-dimensional (3-D) models of the human body from single images, even when the limbs or torso are occluded. The system uses a probabilistic computer-aided design program to generate a hypothetical 3-D form and then applies physical transformations to pose it in a scene. It interprets images by finding probable 3-D models and transformations that, when rendered and filtered, are compatible with the image. This is one of the first explicit inverse graphics approaches to compete successfully with mainstream computer vision techniques. It points the way towards perceptual systems that might have some of the robustness and richness of biological vision.

We released BayesDB, a Bayesian database that lets users query the probable implications of data tables without requiring them to design probabilistic models or implement statistical inference algorithms. Queries are written in Bayesian Query Language (BQL), a probabilistic programming language designed for automatic machine learning, much like Structured Query Language (SQL). BQL commands solve problems such as detecting predictive relationships between columns, inferring missing values, synthesizing hypothetical subpopulations, and finding similar or anomalous rows, columns, or values. Crucially, queries need not be rewritten when the models change, just as SQL queries need not be rewritten when indexes or memory layouts change; this opens the door for automatic model refinement and a broad class of new applications of statistical inference. We are integrating BayesDB with more flexible probabilistic programming languages and beginning to explore applications in public health, cancer biology, and information extraction.

Our work in parsing the minimalist program is motivated by the flood of data now available in very different modalities. This data is widely used to train programs through machine-learning algorithms, including web-based text (e.g., Google's 1 trillion n-gram corpus, used to bootstrap state-of-the-art machine translation engines), audio recordings collected from thousands of hours of recorded telephone conversations (used to train speech recognition models), and social networking text (used to provide personalization services and build recommendation engines). Much of the emphasis in the language research community has been focused on extracting surface regularities from this data. But surface regularities are insufficient, both empirically and as a research agenda.

Our work attempts to build a bridge between big-data surface regularities and models of deep constraints on representation and structure, thereby taking natural language understanding to a new level unattainable by either approach alone. These deep constraints include constraints on constituency, argument structure, subject-verb agreement, and relative hierarchical ordering between syntactic elements. We are developing ideas and algorithms that enable taking advantage of both kinds of information, an approach that we believe will lead to substantially more powerful natural language understanding systems.

Recent work has focused on demonstrating that our minimalist program ideas are language independent and capable of supporting practical programs in diverse languages. In particular, we initiated the development of a Russian-language parser in parallel with continued work on our English-language parser.

entrepreneurship@CSAIL

During the past year, CSAIL continued its entrepreneurship@CSAIL initiative to help transition CSAIL research to start-up based commercialization as well as enabling CSAIL students to pursue start-up ideas of their own design. Following an initial successful offering in spring 2013, the subject 6.S078 Entrepreneurship Project was taught in spring 2014 through EECS, with 14 registered students and the participation of two local venture capital firms. Eight different teams participated in the subject, and self-selected teams made presentations to a panel of venture experts at the end of the term.

Research Highlights

In addition to the large-scale collaborative projects and other center research, many individual and multi-investigator projects are under way. A sampling of this work appears below.

Abstractions for Multi-Material 3D Printing

Multi-material 3D printing has the potential to revolutionize the manufacturing industry, not only by reducing the required time and cost of fabrication, but by also allowing the fabrication of complex materials that would be difficult or impossible to manufacture with other technologies. Such a technology can be used for a variety of applications, from the biomedical to the electromechanical fields, and therefore has the potential to affect the general population, academic research, and a number of industrial markets.

In this area we have been developing abstractions that are necessary to scale the complexity of three-dimensional printed models. We have introduced OpenFab, a direct specification pipeline for multi-material fabrication that is inspired by the programmable pipelines used for film and real-time rendering. The pipeline supports procedural evaluation of geometric detail and material composition, using shader-like fablets that allow models to be specified easily and efficiently. We also describe a streaming architecture for OpenFab; only a small fraction of the final volume is stored in memory and output is fed to the 3-D printer with little start-up delay.

As an alternative to specifying material composition directly, it is possible to specify an object by defining its functional goal (e.g., specific color, stiffness, or refractive index). To address this problem, we have developed Spec2Fab, a general, computationally efficient process for translating functional requirements to fabricable 3-D prints. Our solution relies on two data structures: a reducer tree that efficiently parameterizes the space of material assignments and a tuner network that describes the optimization process used to compute material arrangement. We provide an application programming interface for specifying the desired object and for defining parameters for the reducer tree and tuner network. We illustrate the utility of this framework by implementing several fabrication algorithms as well as demonstrating the manufactured results.

This research is under the direction of associate professor Wojciech Matusik, in collaboration with PhD students Kiril Vidimče, Desai Chen, Jonathan Ragan-Kelley, and Szu-Po Wang, and Drs. David Levin, Piotr Didyk, and Pitchaya Sitthi-Amorn.

Cryptography

Functional encryption is a new paradigm for public-key encryption that enables fine-grained control of access to encrypted data. It provides, for instance, the ability to release secret keys associated with a keyword that can decrypt only those documents that contain the keyword. More generally, functional encryption allows the owner of a master secret key to release restricted secret keys that reveal a specific function of encrypted data. This stands in stark contrast to traditional encryption, where access to the encrypted data is all or nothing.

Professor Goldwasser invented several extensions of the functional encryption model and existing schemes. She proposed new functional encryption schemes that can evaluate functions over encrypted data in a time that is proportional to the function evaluation time over unencrypted data. This is in contrast to previous schemes, where the cost was proportional to the worst-case running time of the function on unencrypted data.

Professor Goldwasser also introduced multi-input functional encryption, where the function to be computed over encrypted data takes multiple cipher-texts as input. This provides a general tool for computing on encrypting data that allows for mining aggregate information from several different data sources (rather than from a single source, as in single input functional encryption). Examples of applications include running structured queries over multiple encrypted values stored in a database and noninteractive, differentially private data release.

This research resulted from a collaboration between professor Shafi Goldwasser, assistant professor Vinod Vaikuntanathan, associate professor Nickolai Zeldovich, PhD student Raluca Popa, and Dr. Yael Tauman Kalai with Microsoft Research.

Complexity Theory

Computational problems do not exist in isolation; rather, multiple and correlated instances of the same problem arise naturally in the real world. For example, in biology, the problem of learning the mapping from the DNA (genotype) to the observable characteristics (phenotype) of an organism draws on data from multiple specimens with highly correlated DNA that provide multiple correlated inputs. In coding theory, the objective is to recover from as many possible errors as possible. To this end, one may well take advantage of the fact that a typical hard drive contains error-correcting encodings of multiple files that are within small edit-distance from each other.

In this project, we set out to investigate what can be gained computationally from access to correlated instances rather than to a single instance of a problem. We were especially interested in settings where significant computational gain can be made in solving a single primary instance by also having access to auxiliary instances that are correlated with the primary instance via the solution space. We focused on constraint satisfaction problems (CSPs), a very expressive class of computational problems that is well studied in terms of approximation algorithms and nondeterministic polynomial-time (NP) hardness, and also in terms of average case hardness and usage for cryptography. We consider a variety of naturally occurring worst-case and average-case CSPs, and show how the availability of a small number of auxiliary instances generated through a natural generating process radically changes the complexity of solving the primary instance, from intractable to expected polynomial time.

This research resulted from a collaborated effort by professor Shafi Goldwasser, Dr. Abhishek Jain, professor Jonathan Katz, and assistant professor Elain Shi from the University of Maryland, professor Amit Sahai from UCLA, assistant professor Hong-Sheng Zhou from Virginia Commonwealth University, Dr. S. Dov Gordon from Applied Communication Sciences, Dr. Vipul Goyal from Microsoft Research, and Dr. Feng-Hao Liu from the University of Maryland.

The MOOCDB Project: Scalable Technology for Online Education

The proliferation of MOOCs and their platforms continues. MOOCs provide an opportunity to study a new kind of student learning behavior. Although they don't allow visual observation of students, it is possible to capture very detailed behavioral data by recording clickstreams, assessments, and student submissions. This behavioral data has enormous potential to inform our knowledge of how students learn and how best to teach them.

For every MOOC platform, the behavioral data is different at the lowest level. It is also voluminous and event driven, meaning different students' behaviors are intertwined. Because this poses a challenge to large-scale learner analysis, we have developed an open-standard data schema called MOOCdb that efficiently organizes the data using student and interaction mode axes. MOOCdb tables are general enough to normalize data from any platform, permitting MOOCs taught by different universities on different platforms to be compared. We validated the schema first by comparing two courses, one taught by Stanford University on the Coursera platform and the other taught by MIT on the edX platform. Stanford University colleagues then joined forces with us and we developed automating software that we have since used to normalize 50 courses' data from two different platforms and three universities, including MIT. This standardization allows shareable analysis, data representations, and predictive models across different courses, platforms, and universities, an ambitious and necessary cornerstone for understanding student behavior at a global scale.

Using normalized data, we have also developed a means of investigating why students "stop out" of MOOCs—that is, stop submitting solutions to problems. We predictively model, with machine-learning algorithms, whether a student will stop out or continue the course. This approach highlights the importance of analyzing MOOC student behavior features. We engaged the help of a broad selection of people so that simple, complex, and abstract reasons could be obtained. To collect them, we developed a systemized way of soliciting and using crowd feature proposals.

The stop-out study analyzed the first MITx MOOC: 6.002X, offered in fall 2012. It had 154,763 registrants, of whom 69,221 looked at the first problem set; 26,349 earned at least one point; and, after completing 15 weeks of study, 7,157 earned a certificate for successful completion. By designing a scalable, distributed, modular, reusable framework that steps from data organization to modeling in an iterative and agile manner, we were able to set up 91 different stop-out prediction modeling experiments to predict stop-out rates in a range of up to 13 weeks ahead, with one week's to all previous weeks' history. To solve the experiments, we employed our cloud scale platform offering multiple machine-learning algorithms and parameter optimizations. It allowed us to build 10,000 different models for these prediction problems. We found that stop-out prediction is a tractable problem. Our models achieved a receiver operating characteristic area under the curve as high as 0.95 (generally 0.88) when predicting one week in advance. Even with more difficult prediction problems, such as predicting stop-out rates at the end of the course with only one week's data, the models attained areas under the curve of 0.7. This suggests that early predictors of stop-out rates exist.

For each of the 91 experiments, we examined the importance of 27 features used by each model. We perturbed the data more than 200 times, deriving 70,000 models, to discern highly predictive features robustly. Overall, features that incorporate student problem submission engagement are the most predictive of stop out. In general, relational features, such as a student's percentile standing or laboratory grade over time, are more predictive than simple features, such as count of submissions. Features involving inter-student collaboration, such as the class forum and Wiki, can be useful in stop-out prediction. It is likely that the quality and content of a student's questions or knowledge are more important than strict collaboration frequency. We found, in particular, that the length of forum posts is predictive, but that the number of posts and number of forum responses is not. We also considered the unobservable states a student may have been in, with the hypothesis that they express a student's engagement or interest and therefore drive the behavioral observations. Using a stacked modeling approach that incorporated the probabilities of hidden states as second-level features, we obtained some of our best models.

A question educators have always wanted to answer is: how do students solve problems? Using the data from MITx's 6.002x spring 2012 course as an example, we have developed a large-scale mixed automated and manual process that allows this question to be addressed for any MOOC. The basis of the process is the reorganization of MOOCdb source data into relevant and retrieval-efficient abstractions called student resource trajectories and answer type transition matrices. The process's analysis methods use these abstractions to offer insight into a problem's level of difficulty and student behavior around a problem type, such as homework. The methods support the development of multiple models of students' problem-solving processes in terms of resource use patterns by using hidden Markov modeling.

This research is part of the MOOCDB Project and is a collaboration of the Department of Mechanical Engineering's students Colin Taylor and Fang Han, PhD student Franck Dernoncourt, Dr. Kalyan Veeramachaneni, and principal research scientist Una-May O'Reilly.

Joint Modeling of Imaging and Genetics

The search for genetic variants that increase the risk of a particular disorder is an important focus in medical research. Genetic studies of disease typically examine genetic markers and their correlations with the incidence of a disease to identify locations in the genome associated with the disorder of interest. Heterogeneity in the effects of genetic variation on the disease processes presents a significant obstacle for robust detection of associations between the genetic code and the disease phenotype.

Our research takes advantage of anatomical information revealed via noninvasive medical imaging to improve the sensitivity of genetic studies and to uncover associations of genetic and imaging markers with the disease of interest. Imaging provides a rich quantitative characterization of the disease and of the inherent heterogeneity in the population cohort. Our computational analysis methods extract associations among genetic markers, image-based measures, and clinical indicators jointly. This is in contrast to a common approach of isolating a few image-based features affected by the disease and then identifying the relevant genetic markers that explain the observed image variations. Our algorithms simultaneously estimate genetic and image-based patterns that together predict the incidence of disease.

Our experiments in the context of Alzheimer's disease have demonstrated significant advantages of exploiting image-based measures of brain tissue degeneration to discover genetic locations associated with the disease. Our current efforts aim to refine the model and to apply it in collaboration with our clinical colleagues who study genetic underpinnings of ischemic stroke and of chronic obstructive pulmonary disease. Although clinically they are very different, both disorders exhibit hereditary aspects and high variability of symptoms. Our research provides tools for analyzing such disorders and for ultimately improving diagnosis and treatment.

This research is a collaboration of graduate students Adrian Dalca and Ramesh Sridharan, Dr. Kayhan Batmanghelich and professor Polina Golland, and their clinical colleagues Dr. Natalia Rost at the Massachusetts General Hospital and Dr. Michael Cho at Brigham and Women's Hospital.

Sound Input Filter Generation for Integer Overflow Errors

Many computer security vulnerabilities involve programming errors. These errors typically involve overlooked cases that rarely (if ever) occur in the standard inputs that the program was developed to process. But an attacker may be able to find a carefully crafted input that targets the overlooked case and enables the attacker to take control of the program.

Integer overflow errors are a particularly insidious source of security vulnerabilities. These errors occur when the program attempts to compute a large number that does not fit in the memory that the computer has set aside to represent the computed value. In this case the computer produces a value that is different from the value that the programmer expected. In practice, attackers have been able to successfully exploit these errors. In a standard scenario, an integer overflow error causes the program to allocate too little memory to hold input data that the program will subsequently read in and process. The input data overflows the allocated memory to overwrite adjacent memory containing data critical to the program's execution, enabling the attacker to inject code that runs in place of the code that the programmer originally wrote.

We developed a system, Sound Input Filter Generation for Integer Overflow Errors (SIFT), that analyzes a program before it runs to find sensitive locations at which potential integer overflow vulnerabilities may occur. Taking all possible program executions into account, SIFT automatically generates an input filter that examines inputs before they are presented to the program. If the filter accepts the input, SIFT provides a guarantee that the input will not exploit the analyzed integer overflow vulnerabilities. If the filter rejects the input, the input is discarded before it reaches the protected program.

Determining whether an input will trigger an integer overflow is, like most program analysis problems that people wish to solve, undecidable. SIFT evades this undecidability result by being conservative—the generated filters may conservatively discard inputs that the program could have successfully processed without error. One critical empirical question is the practical impact of this conservative approach—does SIFT generate filters that discard too many benign inputs?

We evaluated SIFT by collecting a set of benchmark programs. All of these programs are designed to process inputs (such as image or video files) downloaded from the Internet. All of these programs also contain integer overflow errors that remote attackers can exploit (by crafting files that trigger the errors, then placing the files on the Internet for users to download). SIFT identified 58 sensitive locations at which integer overflow errors may enable an attacker to exploit the program. SIFT was able to automatically generate filters for 56 of these 58 sensitive locations.

We tested the automatically generated SIFT filters for these benchmark programs on more than 62,000 benign inputs downloaded from the internet. Recognizing that none of these inputs would trigger integer overflows at the protected sensitive locations in the protected programs, the filters correctly accepted all of these benign inputs. We also tested the filters on inputs that were specifically designed to exploit the known integer overflow errors. The filters successfully rejected all of these inputs. For all the benchmark programs, the average time required to filter each downloaded input was less than 16 milliseconds.

This research is a collaboration of PhD students Fan Long and Deokhwan Kim, Dr. Stelios Sidiroglou-Douskos, and professor Martin Rinard.

Orienteering for Robots

Man-made environments exhibit an organization and structure in which objects are usually aligned to a small number of dominant orientations. Scientists in the fields of computer vision and robotics have exploited the phenomenon in a variety of navigation and reasoning tasks. Earlier incarnations of this idea were referred to as the "Manhattan world assumption," alluding to the borough of the same name and incorporating the notion that a single coordinate axis dominates the arrangement of structures. The phenomenon is observed at many scales, from the bookshelves in a library to buildings in a large city. However, the organizational complexity often varies across environments and multiple orientations are observed. In our research we developed a probabilistic model for describing the global organization of man-made scenes by the tendency of objects to align to a small, but unknown, number of dominant orientations.

Various sensors, including the recently available RGB-D cameras (low-cost cameras that combine color imagery with depth information) and light detection and ranging (LiDAR) technology, facilitate such modeling; they allow one to measure surface normals efficiently at scales ranging from that of a small room to an entire city.

Commonly used probabilistic models rely on the observed data exhibiting properties associated with Euclidean geometry—that is, the "distance" between observed measurements is consistent with our usual notion of the distance between points in the 3-D world. The challenge of using such approaches to describe orientations is twofold. Surface normals are more naturally represented as points on the unit sphere. The usual notion of Euclidean distance does not apply, and one must take into account the curved geometry of the sphere. Additionally, the orientations of coordinate axes, the so-called Manhattan frames, are naturally represented by the set of rotations in 3-D space. This set also possesses a somewhat complicated non-Euclidean geometry.

Our approach relies on a branch of mathematics known as differential geometry and specifically on a special class known as Lie Groups. Lie Groups (named in honor of the 19th century Norwegian mathematician, Sophus Lie) are a special set of mathematical operators with geometric properties that, while non-Euclidean in the global sense, exhibit local Euclidean properties. The analogy is that if one were to navigate between two points on the surface of the earth, for points that are close to each other (say, Cambridge to Boston), one can largely ignore curvature when measuring distances; however, for points that are far (say, Cambridge to Sydney, Australia), the global surface properties are critical. By exploiting this property of the set of rotations combined with the intrinsic geometry of the sphere, we were able to develop a mathematically principled probabilistic model and computationally efficient inference algorithm for analyzing man-made environments. Such models are useful from a scene-understanding perspective in computer vision and from a navigational perspective in robotics.

This research is a collaboration of PhD student Julian Straub, Drs. Oren Freifeld and Guy Rosman, professor John Leonard, and senior research scientist John Fisher.

Laboratory-Sponsored Activities

CSAIL Outreach

Demo Day: This year CSAIL hosted its first annual Demo Day in conjunction with National Robotics Week, inviting local students to see demonstrations of a wide array of robots and related research. In attendance were students from the John O'Bryant School of Mathematics and Science in Roxbury and TechBoston Academy in Dorchester, as well as Boston Mayor Martin (Marty) Walsh, who applauded CSAIL for giving students the chance "to gain invaluable skills for jobs in a well-paying growing industry."

Middle East Education through Technology: CSAIL has been a long-time supporter of the Middle East Education through Technology (MEET) program, an innovative educational initiative aimed at creating a common professional language between young Israelis and Palestinians. MEET enables its participants to acquire advanced technological and leadership tools while also empowering them to create positive social change within their communities. Many MIT students volunteer to teach MEET summer courses at the Hebrew University in Jerusalem. CSAIL continues to provide financial support for the program.

Dertouzos Distinguished Lecture Series

Three speakers gave presentations during the 2013-2014 Dertouzos Distinguished Lecture Series. They were:

- Jitendra Malik, University of California, Berkeley, "The Three R's of Computer Vision: Recognition, Reconstruction, and Reorganization"
- Silvio Micali, CSAIL, "Proof, Secrets, and Computation"
- Shafi Goldwasser, CSAIL , "The Cryptographic Lens"

MAC50 Anniversary Symposium: Computing the Future: Celebrating 50 Years of Computer Science & Artificial Intelligence

On May 28–29, 2014, CSAIL hosted a one-and-a-half-day symposium marking the 50th anniversary of Project Multiple Access Computer and Machine Aided Cognition (MAC), along with CSAIL's 10th anniversary and its key role in the computer science revolution. All the symposium speakers were alumni who represented five decades of computer science research at MIT.

Since the founding of Project MAC on July 1, 1963, computer scientists at MIT have been solving important problems and directly affecting the quality of life in America. Project MAC was launched with a grant from the Defense Advanced Research Projects Agency

(DARPA) and subsequently was funded by both DARPA and the National Science Foundation (NSF). Project MAC led to the foundation of an official academic computer science curriculum at MIT and marked the beginning of an enormously productive era of computing research at MIT. The project also led to the founding of MIT's Laboratory for Computer Science and Artificial Intelligence Laboratory, which merged in 2003 to become CSAIL.

Project MAC was founded with the goal of developing a computing system that would allow individuals to have access to computational power much as we are able to have access to electricity for our homes. The result was time-sharing and a new paradigm of interactive computing, which laid the foundation for many of today's basic design concepts for software systems. These ideas contributed to the later development of the personal computer and personal computing. Today, the CSAIL team continues to make key advances in all aspects of computing. We are a confederation of passions, and each member of our team has a passion to improve our society and our lives through computing. From reverse-engineering the brain to making robots fly like birds, the creative and visionary spirit of Project MAC withstands the test of time. This is evident in the CSAIL community's continued ability to be at the forefront of the future of computing.

Program committee members included Jack Costanza, Adam Conner-Simons, Fernando Corbato, Randall Davis, David Gifford, Silvio Micali, Victoria Palay, Ronald Rivest, Daniela Rus, Howard Shrobe, Stephen Ward and Patrick Winston.
Symposium sponsors were the School of Engineering, the Singapore-MIT Alliance for Research and Technology Center, Foxconn, Google, Microsoft, the NSF, Qatar Computing Research Institute, Quanta Computer, and VMware.

A total of 414 people registered for the event. They comprised MIT faculty (50), MIT staff (56), MIT students (62) and alumni (66), and attendees from industry (57) as well as other interested attendees (73), sponsors, (26), and speakers (24).

Organizational Changes

Professor Daniela Rus has continued as director of CSAIL. The director's duties include developing and implementing strategies designed to keep CSAIL growing and evolving, fund raising, determining laboratory policies, and examining promotion cases.

CSAIL's leadership team includes two associate directors and the executive cabinet. These leaders assist the director with her duties and are appointed by the laboratory's director. Professor Randall Davis continued as associate director. In October 2013, principal research scientist Howard Shrobe became the second associate director for the 2013-2014 term.

The CSAIL executive cabinet meets twice a month to review and advise the director on policy, processes, and activities within the laboratory. Members named to the executive cabinet included Saman Amarsinghe, Regina Barzilay, Jack Costanza, Randall Davis, Srini Devadas, John Fisher, David Gifford, Daniel Jackson, Martin Rinard, Ronitt Rubinfeld, Karen Shirer, Howard Shrobe, Nickolai Zeldovich, and Victor Zue.

Victor Zue continued as director of international relations, managing the engagements and oversight of various important CSAIL international contracts and international contract negotations. Jack Costanza continued as the assistant director for infrastructure, overseeing information technology infrastructure and user support, building operations, and communications. Karen Shirer continued as assistant director for administration, overseeing finance and human resources.

Srini Devadas remained the space czar, overseeing the space committee and managing the allocation of space within CSAIL. The space committee implements improvements to the facilities that will increase the quality of the working environment for the laboratory's faculty, staff, and students. The space committee also includes the two assistant directors, Jack Costanza and Karen Shirer, and administrative assistant, Sonya Kovacic.

Lori Glover, managing director, launched an updated and renamed CSAIL Alliance Program (CAP). The program was previously called the CSAIL IAP.

Awards and Honors

Our faculty and staff won many awards this year, including the following:

Edward Adelson: Massachusetts Academy of Sciences, Fellow; IEEE Computer Society, Helmholtz Test of Time Award

Anant Agarwal: Association for Computing Machinery (ACM) Special Interest Group on Computer Architecture (SIGARCH), inclusion of "Anatomy of a Message in the Alewife Multiprocessor" in the volume published as *25 Years of International Conference on Supercomputing*

Hari Balakrishnan: Indian Institute of Technology, Madras, Distinguished Alumnus Award

David Clark: Internet Society, Internet Hall of Fame inductee

Konstantinos Daskalakis: ACM Conference on Electronic Commerce, Best Paper Award; Special Interest Group on Electronic Commerce, Best Student Paper Award

Erik Demaine: EECS, Steven and Renee Finn Innovation Fellow

Srini Devadas: Design Automation Conference (DAC), Top 10 Cited Author; DAC, Best Paper Hat Trick Award; ACM Special Interest Group on Programming Languages, Architectural Support for Programming Languages and Operating Systems, Most Influential Paper Award; ACM Computer and Communication Security (CCS) Conference, CCS Best Student Paper; ACM SIGARCH, inclusion of "AEGIS: Architecture for Tamper-Evident and Tamper-Resistant Processing" and "Analytical Cache Models with Application to Cache Partitioning"in the volume published as *25 Years of International Conference on Supercomputing*; IEEE Computer Society, Technical Achievement Award

DARPA Robotics Challenge Team: qualified as part of the Virtual Robotics Challenge; qualified for DRC final competition

Distributed Robotics Laboratory: IEEE and African Robotics Network, first place in "Ultra-Affordable Robot" design challenge

Joel Emer: *IEEE Micro*, "Top Picks in Computer Architecture"; ACM *Computing Reviews*, "Notable Computing Books and Articles"

William Freeman: IEEE Automatic Face and Gesture Recognition Conference, Test of Time Award; IEEE International Conference on Computer Vision, Test of Time Award

Jim Glass: Fellow, IEEE

Lori Glover: American Society of Engineering Education, College–Industry Partnership Division, Best Session Award

Shafi Goldwasser: Fellow, Massachusetts Academy of Sciences

Polina Golland: Sloan School of Management, Jamieson Prize for Excellence in Teaching; Electrical and Computer Engineering Department Heads Association, 2014 Diversity Award

D. Fox Harrell: Fellowship, Center for Advanced Study in the Behavioral Sciences at Stanford

Frans Kaashoek: ACM Special Interest Group on Operating Systems (SIGOPS), Hall of Fame Award; ACM Symposium on Operating Systems Principles (SOSP), Best Paper Award; EECS, Jamieson Teaching Award; MIT, Earll M. Murman Award

Dina Katabi: Fellow, MacArthur Foundation; Fellow, ACM; EECS, Viterbi Professor

Tom Leighton: Fellow, Massachusetts Academy of Sciences

Charles Leiserson: Fellow, American Association for the Advancement of Science; IEEE Computer Society, Taylor L. Booth Education Award; ACM, Paris Kanellakis Theory and Practice Award

John Leonard: Fellow, IEEE; International Conference on Robotics and Automation (ICRA), Best Student Paper; Associate Head of Research, Department of Mechanical Engineering

Barbara Liskov: Fellow, Massachusetts Academy of Sciences

Tomás Lozano-Pérez: MacVicar Faculty Fellowship

Robert Morris: ACM SIGOPS, Hall of Fame Award; ACM SOSP, Best Paper Award

Una-May O'Reilly: EvoStar, EvoStar Award for Outstanding Achievements in Evolutionary Computation in Europe

Aude Oliva: Guggenheim Fellow, John Simon Guggenheim Memorial Foundation

Martin Rinard, ACM International Conference on Object Oriented Programming Systems, Languages, and Applications, Best Paper Award

Daniela Rus, EECS, Andrew and Erna Viterbi Professor

Julie Shah: American Institute of Aeronautics and Astronautics, Best Intelligent Systems Paper Award (2012)

Nir Shavit: Fellow, ACM

Russ Tedrake: International Conference on Robotics and Automation, Best Paper Award; International Conference on Hybrid Systems: Computation and Control, Best Paper Award

Seth Teller: MIT IEEE/ACM Student Club, Best Undergraduate Academic Advisor

Theory of Computation Team: ACM-SIAM Symposium on Discrete Algorithms, Best Paper

Vinod Vaikuntanathan: Sloan Foundation, Research Fellowship

Nickolai Zeldovich: ACM SOSP, Best Paper Award

Key Statistics for Academic Year 2014

Faculty: 107 (13% women)

Research staff: 33 (33% women)

Administration, technical, and support staff: 88 (66% women)

Postdoctoral researchers: 74 (12% women)

Visitors: 104 (14% women)

Paid Undergraduate Research Opportunities Program Participants: 163 (37% women)

Master of engineering students: 43 (21% women)

Graduate students: 377 (20% women)

Daniela Rus
Director, Computer Science and Artificial Intelligence Laboratory