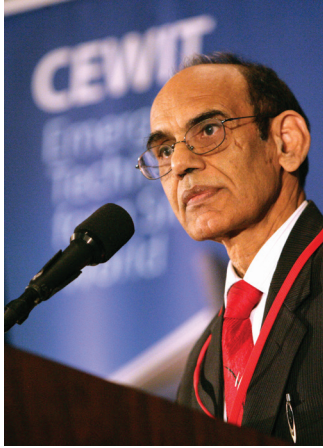# Center of Excellence
WIRELESS AND INFORMATION TECHNOLOGY
AT STONY BROOK UNIVERSITY

Selected Research Projects

Introduction

The Center of Excellence in Wireless and Information Technology (CEWIT) at Stony Brook University is a leading U.S. research institution focusing on cutting-edge research in Wireless and IT. Our focus is to conduct basic research and the commercialization of resulting technologies. We continue to shape new dimensions in the unfinished wireless and IT revolution. The extraordinary scope of the Center's activities encompasses all major IT research areas as shown on its website, cewit.org. Innovation and technology have always been the hallmark of America's global leadership and it has never been more important that we maintain our position as a prime source of technological innovations.

The information Technology world continues to change rapidly since its early years, and it will continue to advance new research insights and create new and highly profitable businesses that we have not even imagined. Artificial Intelligence (AI), Machine Learning, IoT, Data Analytics, Smart Systems, Computational Chemistry and Biology, Edge Computing, Healthcare and Biomedical Applications and other emerging technologies will span new industries. Machine Learning and AI applications will enhance the economy and computers will learn the way things work in the world. Computers will learn to perceive contexts and the ability to learn like humans. The security, connectivity, privacy, and standardization in the IoT world will be addressed to allow the extraction of insights from data. Internal data streams will be combined with data streams coming from the external world such as social media and industry for data scientists to build new algorithms to make quick and intelligent decisions. Recognizing the dominant role of Information Technology and Wireless Technology in modern medicine, CEWIT is engaged in building, prototyping and marketing of medical devices, products and technologies that support patients and clinical care providers and ultimately help improve patient care. The digital divide will be reduced, and cognitive learning will become part of smart systems and communities.

Grid computing, tying far flung supply chains, and e-commerce are still at their beginning of the creation. Data is now the raw material for the information economy, much as coal and iron ore were in the Industrial Revolution. Within the so-called "Internet of Things," sensors are being embedded in devices ranging from smart-phones, automobiles, and utility meters to assembly lines, warehouses, and hospitals to capture data in real time. Hundreds of millions of users around the globe now contribute new data, generating new knowledge and fostering collaboration on new innovations using the Internet. Fifteen out of seventeen industry sectors in the U.S. have more data stored per company than the Library of Congress.

The only way for continued economic prosperity is to reignite the basic research that has always been the economic engine for the worldwide prosperity. The U.S. and international economic priorities require that we bring together businesses, academia and the government to create the new generation technologies and solutions to solve the economic problems that we face today.

The projects described here are a modest representation of the depth and breadth of our commitment to cutting edge research in wireless and IT disciplines, and to leapfrog advances in scientific discoveries. We invite collaboration and are keen to build strategic alliances with business enterprises, academic and scientific communities and government entities.

Satya Sharma
Executive Director
The Center of Excellence in Wireless and Information Technology (CEWIT)

Table of Contents

Networks

Security

Imaging and Visualization

Table of Contents

Aging and Longevity

Internet of Things

Smart Energy and Smart Urban Systems

Algorithms and Big Data

Table of Contents

Leadership

Richard Reeder, PhD
Vice President for Research

Peter Donnelly
Associate Vice President for Technology Partnerships

Yacov Shamash, PhD
Co-Chair, CEWIT Advisory Board and Principal Investigator

Satya Sharma, PhD, MBA
Executive Director

Shmuel Einav, PhD
Director, Medical Technologies

Arie Kaufman, PhD
Chief Scientist

Sangjin Hong, PhD
Director, Globalization

Fan Ye, PhD
Director, Communication and Devices

Bin Zhang
Associate Director, Computing Services

Rong Zhao, PhD
Director, Software Systems

CEWIT Advisory Board

Russell Artzt
Co-Chair, CEWIT Advisory Board, Executive Chairman, Ringlead, Inc.

Otto Berkes
Former CTO, CA Technologies

Michael A. Bernstein, PhD
Interim President, Stony Brook University

Kamal Bherwani
Executive Chairman, Magine TV

Richard Boivie, PhD
Senior Technical Staff Member, IBM

Steven Cento
Sector Engineering Fellow, Northrop Grumman Corporation

Jennifer Costley, PhD
Director, Physical Sciences, Sustainability and Engineering, NY Academy of Sciences

Jim Harding
SVP and Chief Technology Officer, Henry Schein, Inc.

Masaaki Maeda
Strategic Advisor, Asurion, LLC

Christopher Pendergast
Chief Technology Officer, Henry Schein, Inc.

Alex Price
Group Vice President, Global Partners and Technology Ecosystem Strategy, JDA Software

Bob Sanders
Senior Vice President, Zebra Technologies

Yacov Shamash, PhD
Co-Chair, CEWIT Advisory Board, CEWIT

Satya Sharma, PhD, MBA
Executive Director, CEWIT

Kevin Tracey, MD
President, Feinstein Institute for Medical Research

Center of Excellence
WIRELESS AND INFORMATION TECHNOLOGY
AT STONY BROOK UNIVERSITY

Networks

Secure Distributed Computation and Learning Networks

Ji Liu
ji.liu@stonybrook.edu

The central goal of this project is to advance our understanding of what it takes to make various large-scale decision-making, computation, optimization and machine learning algorithms 'secure' against network and/or component failures and adversary attacks. The class of algorithms to be considered are those intended to enable the functioning of distributed networks such as sensor networks, computer networks, and multi-robot teams. In many cyber-physical applications, large volumes of heterogeneous streaming data are needed to be collected by a team of autonomous agents which then collaboratively explore a complex and cluttered environment to accomplish various types of missions. In such a distributed system, groups of agents perform a multitude of operations, including decision making, computation and estimation, optimization and learning with streaming data. To successfully and securely perform these operations in uncertain and unfriendly environments, novel concepts and methodologies are needed to craft and analyze secure algorithms capable of reliably delivering information and robustly performing desired operations. The project aims to develop a framework and computational models for teams of heterogeneous agents to reliably exploit information from various sources in uncertain and potentially unfriendly environments.

Fundamental Techniques for Incentive-aware, Efficient, and Reliable Cloudlet Management and Services

Xin Wang
x.wang@stonybrook.edu

The goal of this work is to develop some fundamental techniques that can facilitate efficient management of cloudlet resources for high cloud service performance while minimizing the system cost. These techniques are to enable efficient, reliable and low cost mobile cloud services, which will in turn support high-performance wireless applications in an energy and resource efficient manner. With this research, a cloud service provider can exploit sparse signal processing technique to remotely monitor one or multiple cloudlet clusters running in different location sites with high accuracy but low cost. To reduce the service initialization delay and operational cost, the system can cache application components, and a user can exploit the proposed rate dependent Bloom filter (RDBF) to quickly determine of some cloudlets have its required application components cached. Rather than being constrained by a limited number of options for virtual machine (VM) configurations, the system will be equipped with multi-dimensional pricing and resource negotiation capabilities for a user to flexibly negotiate multiple types of resources to maximize the benefits of both users and cloudlets while taking into account user application preference, performance requirements, and the system resource conditions.

Complementary to the research agenda, the project will carry out a broad range of education and outreach activities and facilitate technology transfer through the industrial partners and industry affiliate programs. The proposed research is part of a global effort in exploiting (mobile) cloud computing for advancing some critical fields for national economy and security, including healthcare, transportation, smart-grid, homeland security, and education. (NSF)

Spatially Oversampled Dense Multi-Beam Millimeter-Wave Communications for Exponentially Increased Energy-Efficiency

Xin Wang
x.wang@stonybrook.edu

The vast amount of spectrum available in the millimeter-wave (mmW) bands offer a path for exponential growth in data rates for wireless communications networks. In emerging systems such as fifth-generation (5G) networks, the use of mmW frequencies will potentially enable unprecedented improvements in network capacity, mobility, and spectral efficiency. However, the exploitation of mmW bands requires solutions to many technical challenges. In particular, the technology limitations present in today's implementations require new paradigms in algorithms, signal processing methods, circuit architectures, and integration methods in order for 5G wireless to become a reality. For example, there is a need for advanced channel models that let designers implement the wireless network infrastructure of the future.

There is also a need for new

algorithms, software, hardware, and electronic circuits for efficient mmW antenna array processing. This project will exploit well-known physics arising from Einstein's Special Theory of Relativity, namely the causality light-cone, to significantly improve the performance of key array signal processing components in mmW wireless base stations.

Specifically, the spatio-temporal properties of electromagnetic waves, as described by Special Relativity, are exploited in novel architectures to improve the energy efficiency, reduce the noise, and improve the linearity of array receivers. A system-wide study of spatio-temporal properties of mmW channels is combined with these architectures to design new types of mmW array receivers and optimum beam forming algorithms. (NSF)

. . . .

Precision and Robustness of Synthetic Gene Networks

Gábor Balázsi
gabor.balazsi@stonybrook.edu

The success of modern-day electronics is due to precise and robust circuit modules built from standardized and durable components. Combined with the laws of physics, electronics now allows us to explore the surface of planet Mars or visualize atomic defects in materials. Biology holds similar revolutionary capabilities, but the tools for exploiting them are still missing. This is a need that synthetic biology could address.

An engineering approach to the life sciences, synthetic biology has focused heavily on building biological device prototypes resembling electronic circuits: switches, logic gates, oscillators, counters and sensors. However, we do not know the robustness and precision for most of this quickly expanding biological tool set. By precision we mean that the biological device is reversibly fine-tunable within tolerance limits in every single cell. By robustness we mean that the biological device maintains its function over prolonged periods of time, in various physical, chemical and biological conditions.

Traditional genetic approaches to cellular control include gene deletion, knockdown and over expression. However, these methods alter protein levels drastically, without fine-tuning, and are thus imprecise. Commercial gene expression systems enable less drastic control, but still do not achieve single cell-level precision.

Optogenetics is a new technology that is successfully merging with synthetic biology to rapidly control protein levels of single cells by illumination. Yet, most newly emerging optogenetic tools are tested only transiently, without genomic integration, and without characterizing their single cell-level precision and robustness to evolution and environmental conditions. Overall, the environmental and evolutionary robustness or single cell-level precision for most synthetic biological devices have remained untested, especially in eukaryotic (e.g.,yeast and human) cells.

Future applications such as growing synthetic human organs or bio-electronic hybrid systems will require engineering precise and robust devices. Therefore, our current research directions are: (1) designing, building and experimentally characterizing synthetic gene circuits for precise cellular control; (2) experimentally and theoretically characterizing synthetic gene circuit robustness to various environmental factors and to long-term evolution; (3) understanding natural gene regulatory networks by computational modeling and synthetic biological perturbations. Our work is funded by the NIH (R35 GM122561) and by the Laufer Center for Physical and Quantitative Biology.

Networks

Discovery of Network for Applications in Supercomputer Architectures and Machine Learning

Yuefan Deng, Xiaolong Huang and Yidan Zhang
yuefan.deng@stonybrook.edu

We study the topologies of interconnection networks, which is arguably the most important component of supercomputers. A topology with small diameter (D) and minimal mean path length (MPL) can greatly reduce communication latency of the supercomputer network. Among all (N,k) regular graphs, with given graph size N and degree k, the Moore graphs and the generalized Moore graphs have the minimal diameter and minimal shortest path length, and thus win their adaptation as network topologies. We optimize MPL and discover the generalized Moore graphs that exist for certain (N,k) regular graphs. Experimental results for cubic and quartic graphs have been obtained by parallel computing and we are improving the search heuristics and will expand them to other regular graphs. Configuring and benchmarking such topologies on both real machines and simulation platform have also been conducted.
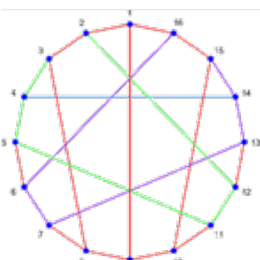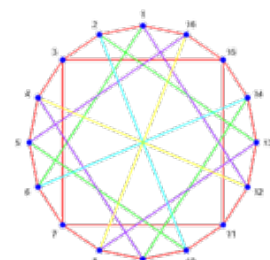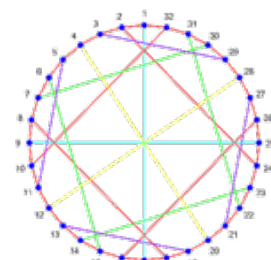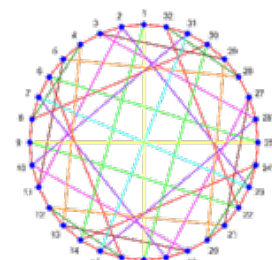
Figure 1

Multi-cast in Data Center Networks

Yuanyuan Yang
yuanyuan.yang@stonybrook.edu

Data center networks (DCNs) interconnect tens of thousands of servers to form the infrastructure for today's ubiquitous collaborative computing, which is the backbone of the clouds. Most of today's DCNs provide high bandwidth using inexpensive commodity switches. Also, many online applications (e.g., web searching) and back-end infrastructural computations (e.g., distributed file system and database) hosted by data centers require one-to-many group communication. Network-level multi-cast can greatly benefit such group communication through reducing network traffic and releasing the sender from duplicated transmission tasks, thus significantly improving application throughput and increasing network capacity. Several unique features in data centers facilitate the implementation of multi-cast. First of all virtualization of machines and networks provides plenty of flexibility. For computing tasks submitted to data centers, physical resources can be logically divided and each task can be provided a virtual network according to its scale and requirements.

We exploit the freedom of resource allocation to meet several goals: (1) to deliver satisfied user experience,i.e., the communication between members of a virtual network is nonblocking for multi-cast; First of all virtualization of machines and networks provides plenty of flexibility. For computing tasks submitted to data centers, physical resources can be logically divided and each task can be provided a virtual network according to its scale and requirements. We exploit the freedom of resource allocation to meet several goals: (1) to deliver satisfied user experience, i.e., the communication between members of a virtual network is nonblocking for multi-cast; Server redundancy is originally proposed in order to deliver high availability and fault tolerance in data centers. Nevertheless, this feature can be utilized to improve the design of data center networks and reduce network cost. For example, the number of core switches in a fat-tree data center network can be dramatically reduced if the server redundancy is taken into consideration in the overall network design. Third, the recently developed OpenFlow framework gives endless possibility to design various network control strategies based on different needs. We develop several online multi-cast scheduling algorithms to balance the traffic load in data center networks. These algorithms not only improve the utilization of network bandwidth, but also ensure that no link in the network is oversubscribed. (NSF)

(16,3) Optimal graph D=3, MPL=2.20

(16,4) Optimal graph D=3, MPL=1.75

(32,3) Optimal graph D=4, MPL=2.94

(32,4) Optimal graph D=3, MPL=2.35

Figure 1

Explore Target k-Coverage in Wireless Rechargeable Sensor Networks

Yuanyuan Yang, P. Zhou and C. Wang
yuanyuan.yang@stonybrook.edu

Energy remains a major hurdle in running computation-intensive tasks on wireless sensors. We employ a Mobile Charger (MC) to deliver wireless power to sensors via magnetic and resonant coils, which provides a promising solution to the energy problem. Instead of maintaining perpetual operation of all the sensors at the expense of high operating cost of MC, in this project, we have considered target k-coverage in Wireless Rechargeable Sensor Networks (WRSNs) by utilizing the abundant redundancy in their coverage in the network. In practice, static targets include: the radiation levels after earthquake in Fukushima nuclear reactors ; early signs of eruption of volcanoes to be detected in Indonesia; the air and water pollution of important habitats for animals. Mobile targets include: animals wandering on the field, or being attracted by food sources and moving randomly with a certain drift towards the source; diffused forest fire driven by the speed and direction of the wind. We propose a new framework, called k-coverage WRSN, where sensors are organized into clusters around each target and it is required that at least k sensors should be working in each cluster at any moment to engage in sensing tasks to cover the target. We conduct theoretical analysis on the improvement of charging capability of MC by only charging a portion of sensors. We study a distributed algorithm that can assign sensors into balanced clusters around targets.

We optimize the number of sensors being charged in each cluster while guaranteeing target k-coverage. A charging algorithm modified based on Generalized Traveling Salesman Problem (GTSP) is proposed while working sensors are picked wisely. To explore more application scenarios, we also consider mobile targets of different motion patterns such that original clusters are expanded until a re-clustering condition is met. Our extensive simulation results demonstrate significant improvements of network scalability and cost saving that MC can extend charging capability over 2-3 times with a reduction of 40% of moving cost without sacrificing the network performance. (NSF)

Figure 2

Publications: P. Zhou, C. Wang, and Y. Yang, "Leveraging target k-coverage in wireless rechargeable sensor networks", In IEEE 37th International Conference on Distributed Computing Systems (ICDCS), 2017. P. Zhou, C. Wang, and Y. Yang, "Static and Mobile Target k-Coverage in Wireless Rechargeable Sensor Networks", IEEE Transactions on Mobile Computing (TMC), 2018.

Research on Wireless Rechargeable Sensor Networks

Yuanyuan Yang
yuanyuan.yang@stonybrook.edu

In this project, we introduce the novel wireless charging technology to power wireless sensor networks. The objective is to achieve perpetual network operations as well as improve network performance. Traditional battery-powered sensor networks usually have limited lifetime that pose great challenges to meet a variety of energy-demanding applications. Energy harvesting from environmental sources can sustain network operations. However, dynamics from the energy sources may cause interruptions in network services and degrade performance greatly. The novel wireless charging technology has opened up a new dimension to replenish energy in sensor networks without wires or plugs. A charging vehicle equipped with resonant coils can move around the field to recharge nodes conveniently. Several important issues are studied in this project.
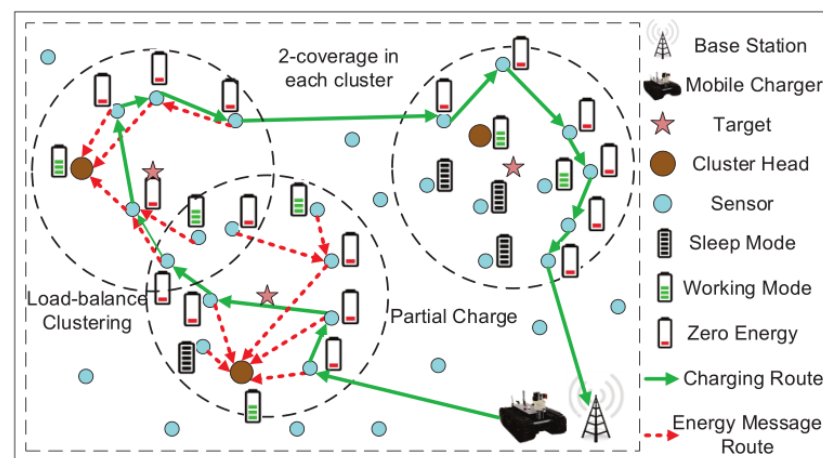
Figure 2   An example of our target k-coverage framework providing target 2-coverage for 3 targets.

Networks

The first question is how to gather real-time node's energy information in a scalable manner. To achieve this, we propose an NDN (Named Data Networking) based real time communication protocol for gathering real-time energy status that divides the network into hierarchies. We leverage concepts and mechanisms from NDN to design a set of protocols that continuously gather and deliver energy information to the vehicle to coup with a variety of recharge requests. Analytical results based on the energy neutral conditions that give rise to perpetual operation are also derived.

The second question is how to schedule vehicle(s) to achieve perpetual operations. For a single vehicle, we formulate the problem into an Orienteering problem, which is NP-hard. It aims to maximize the total energy recharged in a given time.

Since the problem is NP-hard in nature, we take reasonable approximations to simplify the problem into a Knapsack problem so we can develop polynomial-time solutions. Further, the problem on how to schedule multiple vehicles that has more scalability and robustness immediately follows. Our focus is to minimize the vehicles' total traveling cost while ensuring all nodes are functional. We formulate the problem into a Multiple Traveling Salesman Problem with Deadlines (m-TSP with Deadlines), which is also NP-hard. To accommodate energy dynamics and reduce computational overhead, we develop an online algorithm that selects the node with the minimum weighted sum of traveling time and residual lifetime. Our scheme not only improves network scalability but also guarantees the perpetual operation of networks.

The third problem is how to integrate wireless charging with traditional sensing applications such as mobile data gathering. We can combine wireless charging and data gathering utilities on a single vehicle to improve spatial-temporal efficiency. Our objective is to maximize the network utility. First, a set of sensor nodes with minimum residual energy are selected for the vehicle to perform recharge. The algorithm ensures a bounded traveling time under a given threshold. Upon traversing each node for recharging, the vehicle collects data messages in the neighborhood by multi-hop transmissions. To maximize network utility, we divide the original problem into several sub-problems to find the optimal data rates, flow routing and vehicle's stopping time. Distributed algorithms to solve these problems are proposed and convergence properties are examined. (NSF)

Novel Interconnection Networks for Data Centers

Yuanyuan Yang
yuanyuan.yang@stonybrook.edu

Driven by technology advances, massive data centers consisting of tens or even hundreds of thousands of servers have been built as infrastructures by large online service providers. Designing a cost-effective network topology for data centers that can deliver sufficient bandwidth and consistent latency performance to a large number of servers has been an important and challenging problem. A good network topology should meet the following requirements: (1) expandability, which means that expanding an existing network should not incur huge extra cost in either man power or device replacement; (2) plenty parallel paths between any pair of servers to guarantee a sufficient bandwidth and graceful degradation; (3) small network diameter so that a task can be assigned to any part of the network as required by cloud computing; (4) low cost of the interconnection structure. As the size of data centers becomes larger and larger, the cost of the interconnection structure becomes a key factor in practical applications. Many data center network topologies have been proposed recently, which can be basically divided into two categories: switch-centric networks and server-centric networks. We conduct extensive research in both categories. In this project, we mainly focus on server-centric networks and propose a novel server-centric network topology called BCube Connected Crossbars (BCCC) which can meet all the requirements described above. We also propose routing algorithms for different communication patterns in the BCCC. It is shown that the BCCC outperforms the current popular data center networks. (NSF)

Modeling and Understanding Complex Influence in Social Networks

Jie Gao
jgao@cs.stonybrook.edu

Information, beliefs, diseases, technologies, and behaviors propagate through social interactions as a contagion. Understanding of how these contagions spread is crucial in encouraging beneficial and healthy behaviors and discouraging the ones that are destructive and damaging. Rigorous, mathematical understanding of complex social contagions is not just an abstraction, but will guide applications from healthcare to

word-of-mouth advertising. The technical content of this project is inherently interdisciplinary, and its lessons will apply to related fields such as probability, economics, sociology, and statistical physics. The research efforts are integrated with the educational and outreach activities of the PIs, who have strong records of broadly disseminating cutting-edge research to high school, undergraduate, and graduate students through teaching, outreach programs, and personal mentoring.

This project will transform our understanding of social contagions by: 1) Developing a suite of technical tools to enable improved understanding of specific complex processes; 2) Determining how various parameters of cascade and social structure together impact the chances of a cascade's success or failure; and 3) Obtaining empirical evidence to both corroborate the theoretical findings, and uncover the space of realistic setting for certain parameters. Many existing models of contagion assume that increasing the number of infected (or affected) neighbors marginally decreases the chance of infection. Many contagions, such as adoption of expensive new technology, fail to have this property, but instead have more complex rules for infection. This leads to different spreading behaviors even on the same networks. Motivated by sociology research findings, this project will greatly enhance our understanding of social contagions in three aspects.

First this project will provide rigorous study of the spreading behavior of a simplified theoretical model called k-complex contagions and its interactions with structures in the underlying graph such as tie strength, unusually influential nodes, and community structures. Second,

this project presents a general model for studying cascades that is both theoretically tractable and practically motivated. The general model generalizes most previous theoretical models of complex and simple contagions and includes homophily and environmental factors on cascades. Finally, this project uses post-hoc analysis as well as real world social experiments to verify the veracity of the model and fit the parameters in different settings.

(NSF, DARPA)

....

Guaranteeing Resilience in Network Function Virtualization

Yuanyuan Yang, X. Shang and Z. Li
yuanyuan.yang@stonybrook.edu

The research objective of this project is to provide algorithmic and systematic support to the resilience of the emerging network function virtualization. Detaching network functions from their underlying hardware increases their failure rates. Connecting multiple virtualized network functions (VNFs) into a service function chain (SFC) further makes it vulnerable to any VNF failure on the chain. We have proposed a proactive local rerouting strategy to deploy backup VNFs and rerouting paths into the data center network. The strategy improves the availability of SFCs while reducing possible network congestion and latency.

We have formulated the system model into an Integer Linear Programming problem and proposed an approximation algorithm with a theoretical bound to reduce its computational complexity. Simulations have shown the significant improvement of resilience

applying local rerouting. We have also proposed a reactive supplementary rerouting strategy with an online algorithm to further improve the availability of SFCs to 100%. The online algorithm proposed is proved to have a bounded competitive ratio to the offline optimum. To further improve efficiency and eliminate the possible congestion by rerouting strategies, we have advanced our model to minimize both congestion and operating cost. For this model, we have proposed an offline algorithm with a theoretical bound and an online algorithm considering VNF migration. We currently focus on the resilience problem of SFCs in edge networks where backups and original VNFs compete for highly limited resources. Deploying backups for every SFC is no longer applicable in this situation. We will build mathematical models to describe the trade off between resilience and efficiency of SFCs in edge networks and propose algorithms to find the sweat point and improve the overall benefit.

Figure 3
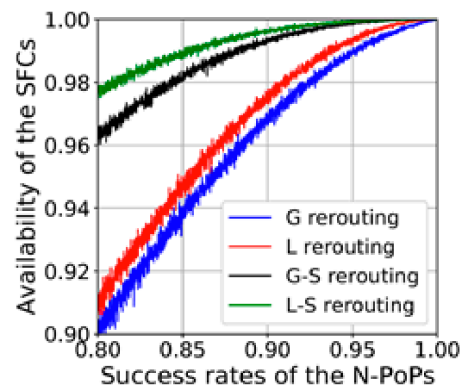
Figure 3 *The availability of SFCs when applying global rerouting (G), local rerouting (L), and the two rerouting combined with supplementary rerouting (G-S and G-S). The x-axis denotes the average success rate of nodes holding VNFs and the y-axis denotes the average availability of SFCs.*

Networks

Publications: X. Shang, Z. Li, and Y. Yang, "Partial Rerouting for High-Availability and Low-Cost Service Function Chain." 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018. X. Shang, Z. Li, and Y. Yang, "Placement of Highly Available Virtual Network Functions Through Local Rerouting." 2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). IEEE, 2018.

Dynamic Networks: Learning, Inference, and Prediction with Nonparametric Bayesian Methods

Petar Djuric
petar.djuric@stonybrook.edu

Complex networks are all around us and by nature they can be biological, social, physical and virtual. Our immune system is an intricate network of cell interactions and the brain functions as a complex network of communicating neurons.

We are part of various social circles, and the computers and mobile phones form networks comprising the World Wide Web. Networks are defined by their nodes and the links that connect the nodes. Real-world networks are dynamic in nature in that their structure evolves with time. The number of nodes is time-varying as is the number and nature of links between the nodes. The main aim of this project is to investigate nonparametric Bayesian methods for inference and prediction of dynamical networks. (NSF)

Passive RF Tag Network for Smart Spaces

Milutin Stanaćević
milutin.stanacevic@stonybrook.edu

We envision a future where every object in our living and working environment will carry one or more RF-tag type devices. Based on the backscattering tag-to-tag communication link, these RF devices will be connected in a network without the need for the central interrogating device. We devised a novel technique that enables estimation of the parameters of wireless tag-to-tag channel by a passive receiver. This empowers RF devices with the novel form of passive RF sensing that is based on the devised channel estimation. These embedded RF devices will be able to sense activities and interactions among various entities around them, both tagged and not-tagged. This enables applications such as fine grain tracking of human movements, activity and gesture recognition and human-object interactions.

These capabilities in turn will provide an ability to query and reason about the environment in order to infer a wide range of analytic information. All this will be achieved without the occupants (humans) within the environment having to carry or wear any devices (device free). (NSF)

Reconfigurable Data Processing Architecture

Sangjin Hong
sangjin.hong@stonybrook.edu

Traditional general purpose processors are especially challenged by today's energy constraints. Although the core computational structures are themselves quite efficient, most of the energy is consumed by overhead like data movement. By building a customized ASIC accelerator to match a given algorithm, one can greatly increase the performance and efficiency, but at increased cost and decreased flexibility; only applications whose accelerators are included in the chip will benefit. Coarse-grained reconfigurable architectures (CGRAs) provide the benefits of reconfigurability while eliminating many of the difficulties of FPGAs. Typically, a CGRA contains a large number of small computational elements connected with a restricted reconfigurable interconnect. Thus, a CGRA is configurable, but less configurable than an FPGA. This leads to easier mapping from an algorithm to reconfigurable resources and faster reprogramming times.

However, the parallel and tightly-connected processing elements of the CGRA can still yield large improvements in efficiency and speed over a general purpose processor. Existing approaches to CGRAs demonstrate the positive aspects of this approach, but tend to treat CGRAs as standalone systems, with aspects of control, synchronization, and real-time programmability as afterthoughts. The result is these

approaches are insufficient to provide a robust and easily usable CGRA co-processing system. To bring the efficiency gains of reconfigurable logic into the realm of general purpose processing, it is necessary for a platform to built with dynamic controllability, reconfigurability, and connectivity in mind. We propose to construct CGRA-based reconfigurable co-processors that can be tightly integrated with general-purpose processors to provide energy-efficient performance without sacrificing flexibility or ease of use. To do this we must focus on the mechanisms required to effectively control the co-processor from the point of view of the greater computing system. This will allow easy context switching and dynamic partitioning to adapt the system based on unpredictably changing workloads. The processing elements and controllers are grouped hierarchically to accelerate reconfiguration time, and to allow multiple clock domains.

Pervasive Edge Computing

Yuanyuan Yang and Y. Huang
yuanyuan.yang@stonybrook.edu

The pervasive edge computing project is focusing on studying theoretical and practical problems on edge computing over edge servers and a large number of edge devices. Cooperative caching in edge computing helps to improve the robustness and reduce the latency for data accessing. For edge devices, resources constraints and mobilities are the most crucial features. Fair caching makes heterogeneous devices cache different amounts of data based on the current remaining resources. The devices with more resources (ratio) may cache more data than those with fewer resources. This helps improving the overall performance and individual satisfactory since it prevents the overusing of certain devices. We also study the high-speed mobility scenarios and propose corresponding grouping method for cooperative caching.

Devices are organized in groups and inter-group and intra-group are dealt with differently. Mobility of devices are transferred into group membership changes and caching are reorganized mostly inside a group. This limits the affections due to mobility and improves data delivery.

Figure 4
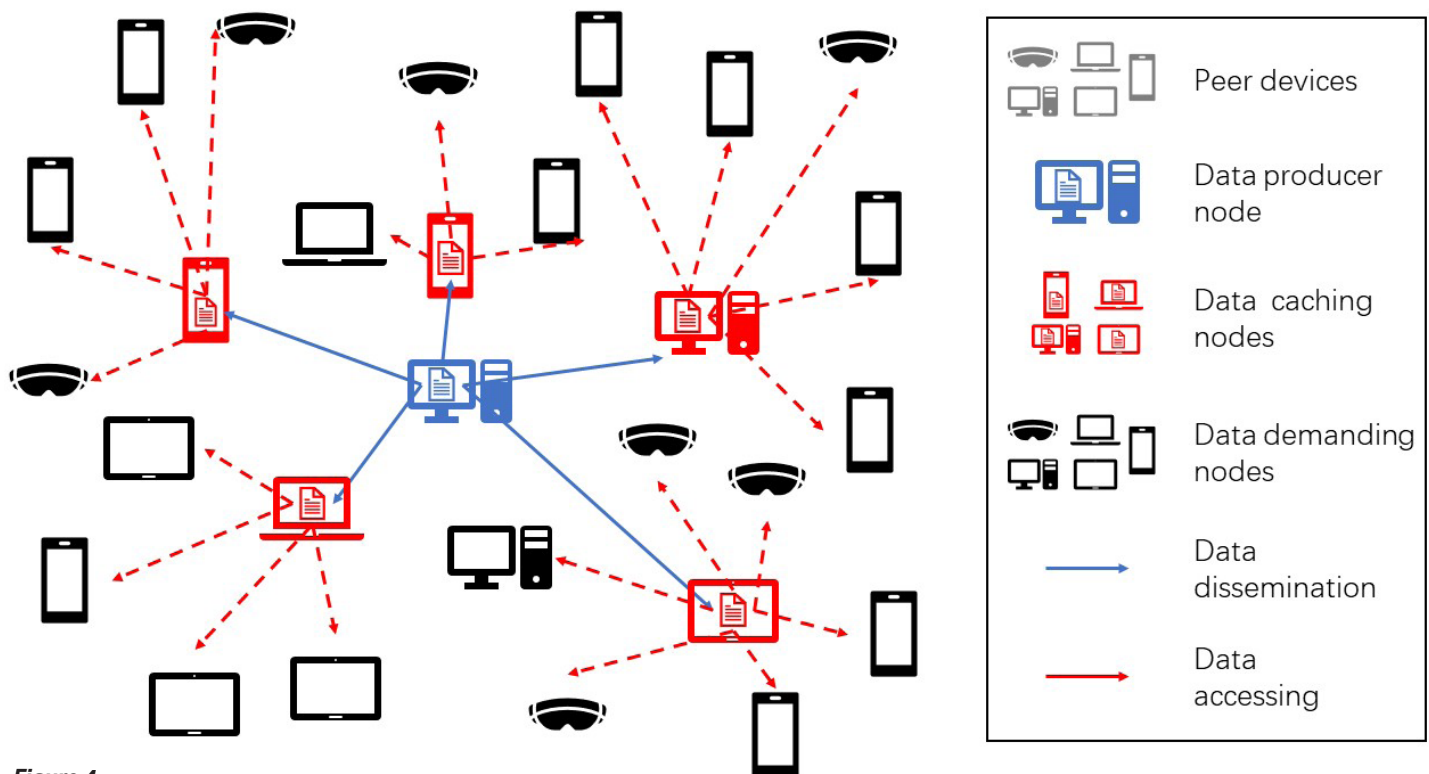
Figure 4

Security

Software Shielding Against Advanced Exploits

Michalis Polycronakis
mikepo@cs.stonybrook.edu

The exploitation of memory corruption vulnerabilities in popular software is among the leading causes of system compromise and malware infection. While there are several reasons behind this proliferation of exploitable bugs, the reliance on unsafe programming languages such as C and C++ and the complexity of modern software play a major role. The continuous discovery of previously unknown (zero-day) vulnerabilities in browsers, document viewers, and other widely used software, and the lack of effective defenses against recent exploitation techniques that leverage memory disclosure vulnerabilities, necessitate the development of additional defense mechanisms.

The main objective of this project is the design of software shielding techniques and their practical applicability to commodity software and systems. The key innovative aspects of the investigated techniques include: i) principled design that considers the strong adversarial models imposed by the latest exploitation advancements, i.e., disclosure-aided exploitation and data-only attacks, against which effective countermeasures remain an open problem; ii) novel code specialization and data protection techniques, to introduce process-level unpredictability and limit the exposure of critical data; iii) hardware-assisted implementation by leveraging recent and upcoming processor features to minimize the performance impact of the applied protections; and iv) focus on practical considerations, such as operational compatibility and non-disruptive deployment. The outcomes of this research effort are expected to improve the state of the art in defenses against advanced exploits, and achieve substantial practical impact by shielding existing vulnerable applications against exploitation, benefiting both end users and security researchers. The project also provides students the opportunity to conduct research in cybersecurity, and fosters the integration of cybersecurity into high school education through hands-on workshops for students and seminars for science teachers.

Software Specialization Against Vulnerability Exploitation

Michalis Polycronakis
mikepo@cs.stonybrook.edu

The continuous deployment of exploit mitigations in popular applications and operating systems has been making vulnerability exploitation harder. Currently deployed mitigations, however, often do not focus on the capabilities that a given bug class provides to an attacker, but instead just hinder only certain aspects of certain exploitation techniques. For instance, an arbitrary memory disclosure capability cannot only be used to bypass address space layout randomization (ASLR) and facilitate control flow hijacking, but also to leak sensitive process data- an equally severe but currently neglected exploitation threat. At the same time, modern frameworks, toolkits, and libraries greatly simplify software development while increasing productivity, but have the downside of producing bloated binary executables with a lot of code and functionality that is never exercised. As the code base of a program grows, so does the likelihood of finding bugs that can be exploited, instruction sequences that can be reused for constructing exploit code, and ways to access private or security-sensitive data. Second, the researchers are evaluating the effectiveness of these techniques against current dynamic malware analysis systems and how such defense systems might detect that queries or probes of environmental information indicate the presence of malware. Finally, the project is studying whether artificially created "wear and tear" artifacts might be injected into analysis environments to counter next-generation, environment-aware malware.

Combating Environment-Aware Malware

Michalis Polycronakis and
Nick Nikiforakis
mikepo@cs.stonybrook.edu

Tools for dynamic detection of malicious software ("malware"), such as anti-virus software, often create a protected "analysis environment" (or "sandbox") in which to test suspicious software without risk to the computer system. Malware authors have responded by developing environment-awareness techniques, to enable their malware to recognize and behave differently in a sandbox environment, thereby evading detection. Authors of defense software are endeavoring to ensure that analysis environments exhibit realistic characteristics.

This project focuses on examining "wear-and-tear" or "aging" artifacts, whose absence could enable malware to distinguish a pristine, new analysis environment from a real environment that has been seen normal use.

This project is assessing the potential for a new class of environment-aware malware that exploits usage-related artifacts that inevitably occur on real systems as a result of normal use, and which are absent in existing malware analysis environments. The project is first investigating techniques for recognizing artifacts related to past user activity and studying how malware might query or probe to acquire such information.

This project will investigate techniques for generating self-adapting software that i) reduces its attack surface by removing unneeded code and logic according to mission-specific or end-point-specific configurations and dependencies, and ii) shields itself against existing and emerging exploits by retrofitting specialized protection mechanisms that leverage the security-related primitives provided by the underlying hardware. Endpoint-specific specialization will be facilitated by a novel binary code transformation framework that relies on compiler–rewriter cooperation to enable fast and robust fine-grained code

transformation on endpoints, while achieving transparent deployment by maintaining compatibility with existing software distribution models

···

Security Policy Mining

Scott D. Stoller
stoller@cs.stonybrook.edu

The most widely used representations of access control policies are access control lists and role-based (also called group-based) policies. Over the past several years, there has been growing recognition of the advantages of attribute-based access control (ABAC), in which policies are expressed as rules stating the conditions under which a subject may perform an operation, e.g., "A user working on a project can read and request to work on a non- proprietary task whose required areas of expertise are among his/her areas of expertise." ABAC provides a high level of flexibility that supports complex, dynamic access control requirements and promotes security and information sharing. Moreover, it allows concise policies

that are easier — and hence cheaper — to understand, analyze, and maintain. Relationship-based access control (ReBAC), inspired by access control in online social networks, extends ABAC to support policies expressed in terms of relationships between subjects, resources, and other objects, as well as their attributes. The cost of manually developing an initial ABAC or ReBAC policy can be a significant barrier to adoption of these technologies by an organization. Policy mining algorithms promise to drastically reduce this cost, by partially automating the process. We are developing new algorithms for mining ABAC and ReBAC policies from access control lists (ACLs) or role-based policies, together with attribute data. Furthermore, when ACLs or role-based policies are unavailable, our algorithms can be used to mine policies from access logs; this is more challenging, because access logs contain incomplete information about authorizations. We are exploring several algorithmic approaches to this problem, including greedy algorithms, evolutionary algorithms, and machine learning (neural networks). Combining

Security

the latter two approaches enabled us to design more effective and scalable policy mining algorithms. We are also exploring techniques for efficient enforcement of ABAC and ReBAC policies in systems that do not natively support these technologies. The basic idea is to translate ABAC or ReBAC policies into a group-based policies, which are widely supported. The technical challenge is to update the translated policy efficiently as changes occur in the attribute data and the rules in the ABAC or ReBAC policy. (NSF, ONR)

Cloud Computing Security

Radu Sion
sion@cs.stonybrook.edu

As computing becomes embedded in the very fabric of our society, the exponential growth and advances in cheap, high-speed communication infrastructures allow for unprecedented levels of global information exchange and interaction. As a result, new market forces emerge that propel toward a fundamental, cost-efficient paradigm shift in the way computing is deployed and delivered: computing outsourcing. Outsourcing has the potential to minimize client-side management overheads and benefit from a service provider's global expertise consolidation and bulk pricing.

Companies such as Google, Amazon, and Microsoft are rushing to offer increasingly complex storage and computation out-sourcing services supported by globally distributed "cloud" infrastructures. Yet significant challenges lie in the path to successful large-scale adoption. In business, healthcare and government frameworks, clients are reluctant to place sensitive data under the control

of a remote, third-party provider, without practical assurances of privacy and confidentiality. Today's solutions however, do not offer such assurances, and are thus fundamentally insecure and vulnerable to illicit behavior. Existing research addresses several aspects of this problem, but advancing the state of the art to practical realms will require a fundamental leap. This project addresses these challenges by designing, implementing and analyzing practical data outsourcing protocols with strong assurances of privacy and confidentiality. It will also initiate the exploration of the cost and energy foot-prints of outsourcing mechanisms. This is essential as the main raison d'etre of out-sourcing paradigms lies in their assumed end-to-end cost savings and expertise consolidation. Yet, research has yet to explore and validate the magnitudes of these savings and their underlying assumptions. (NSF, ARO, Microsoft Research)

National Security Institute

Radu Sion
sion@cs.stonybrook.edu

Stony Brook University, the flagship state university of New York, is establishing the National Security Institute (NSI). The NSI vision and its core mission are bold: to secure our homeland by researching and developing technologies and insights for secure, trustworthy, and available communications and computing platforms. NSI's goal is to become a world leader in research, the education of professionals, security technology, business and policy, and raising awareness. NSI will span multiple disciplines and establish public-private partnerships to develop new holistic socio-technological solutions for

securing our highly-digital societies; it will engage not only in research but also in the education of professionals in defense, national and cyber-security, assurance, healthcare, and policy. A comprehensive assurance education program will be established, to train not only Stony Brook students but also the broader corporate and academic community.

NSI will leverage the team's strengths to spawn a steady stream of successful security-centric technology startups. (CEWIT)

NFS4Sec: An Extensible Security Layer for Network Storage

Radu Sion and Erez Zadok
sion@cs.stonybrook.edu,
zadok@cs.stonybrook.edu

The Network File System (NFS) is a popular method for computers to access files across networks. The latest major version of this IETF protocol, version 4, is widely accepted and includes numerous new features to improve security, performance, and usability when used over wide-area networks. However, the NFSv4's security focus is on network-wide encryption (ensuring that user data cannot be intercepted) and user authentication (ensuring that only legitimate users can access their files); it does not address end-to-end data security (i.e., persistently stored data), does not address data integrity (malicious or benign data corruptions), and more. This project extends NFSv4 with a security layer that allows one to develop multiple, composable plug-in modules to enhance the protocol's security. This layer allows for interception of protocol requests between clients and servers to

perform various useful security functions: logging access to files by users and hosts, useful for regulatory compliance reports and audits; inspecting files for malware patterns and automatically quarantining them; verifying the integrity of long-lived files against malicious changes (e.g., Trojan intrusions) and benign but serious ones (e.g., storage media degradation and hardware corruptions); detecting denial-of-service attempts and ensuring quality-of-service to legitimate users through load-balancing and redirection; automatic snapshotting and logging to allow for forensic analysis and recovery from failures and intrusions. In a cloud-based era where more data lives longer and is accessed over wide-area, insecure networks, this project help elevate the level of security of every user's data files. (NSF)

uID: A Strongly-Secure Usable Identity Ecosystem with Privacy

Radu Sion
sion@cs.stonybrook.edu

uID is a secure, usable, privacy-enabling digital identity ecosystem, able to integrate, and synergize with existing governmental, commercial and open-source identity and authentication solutions.

Designing tomorrow's digital identity solution is faced with unique challenges. Identity mechanisms overwhelmingly refer to and are used by people. They need to be usable and affordable, and address individual concerns of privacy and confidentiality. At the same time, to ensure trust they need to provide accountability and be strongly secure.

Further, it is important to realize

that no one platform can be a sole provider—a viable ecosystem will have standards with well specified APIs and conduits for interoperability that naturally foster a healthy market.

Finally, it is essential that these mechanisms interoperate and are efficient so as to not constitute a bottleneck when deployed.

While addressing all of the above challenges, uID will focus on two key goals: privacy protection and transaction unlink-ability. These properties are unfortunately conflicting and require a complex multi-layer research and development approach calling on multi-disciplinary expertise across all the layers of today's digital transactions. Simple "browser plug-ins" or "email-based" mechanisms alone are bound to fail by not considering the multiple cross-layer security challenges. (CEWIT)

Cloudtracker: Transparent, Secure Provenance Tracking and Security Policy Enforcement in Clouds

Radu Sion and Don Porter
sion@cs.stonybrook.edu
porter@cs.stonybrook.edu

As companies, governments, and individual users adopt increasingly diverse computing platforms, from outsourced cloud computations to personal laptops and mobile devices, enforcing uniform security policies across these platforms becomes unwieldy.

Similarly, regulatory compliance and business auditing requires tracking the history of this data in a comprehensive, secure, and platform-independent manner.

Unfortunately, technology has

not kept pace with these practical concerns, and several systems and security research challenges must be addressed to make this vision a reality.

There is a natural and under-explored connection between understanding the origins of data and using that data's history to enforce security policies.

To leverage this connection, this project is developing a comprehensive, general framework for automatically tracking the history of data and enforcing associated security policies in cloud computing environments. The research focuses on three key research challenges.

First, the project investigates novel applications of virtualization technologies to transparently infer data provenance by inspecting a guest operating system (OS) and applications. Second, this project is developing techniques to securely store, manage, and query provenance data at cloud scale. Finally, the project combines the first two technologies to transparently and collaboratively enforce security policies throughout the cloud and end-user systems. The prototype system is designed to allow individual users and organizations to rapidly adopt new technology platforms, from clouds to novel end-user systems, without having to worry about the interaction of these new systems with security policies and regulatory compliance concerns. (NSF)

Hardware Security for 3D ICs

Emre Salman
emre.salman@stonybrook.edu

3D ICs introduce unique challenges and opportunities for managing multiple aspects of hardware security, as investigated in this research.

Security

Our objective is to go beyond the conventional split manufacturing based approaches and treat 3D ICs as stand-alone entities. As such, we would like to identify 3D specific security attacks and develop appropriate countermeasures. For example, we developed a layout-level camouflaging methodology to thwart image analysis based reverse engineering attacks in monolithic 3D ICs. This method was evaluated with a fully placed and routed SIMON core, lightweight block cipher developed by NSA. Our results demonstrate that the monolithic 3D technology is highly effective for circuit camouflaging since both the area and power overheads are eliminated with a slight degradation in timing characteristics. (NSF-SaTC:STARSS)

Fast, Architecture-Neutral Static Binary Instrumentation

R.Sekar
sekar@cs.stonybrook.edu

Program instrumentation techniques form the basis of many recent soft-ware security defenses, including defenses against common exploits, security policy enforcement, application monitoring and debugging. As compared to source-code instrumentation, binary instrumentation is easier to use and more broadly applicable due to the ready availability of binary code. Moreover, source-code based instrumentation may be incomplete because some of it may be eliminated by compiler optimizations, and because some low level code added by linkers (or compilers) is not instrumented.

One of the major challenges in binary instrumentation is the complexity of modern instruction sets. Accurate instrumentation requires the semantics of all instructions to be captured, since all of the analyses and transformations performed by the instrument are based on this semantics. Any errors in modeling instructions will likely cause instrumented programs to fail. Clearly, this is a daunting task even for a single architecture: the Intel manual describing the x86 instruction set runs to over 1500 pages describing over 1100 instructions. When this task is multiplied across different architectures such as ARM, PowerPC, SPARC, MIPS, etc, the effort involved becomes impractically large. We are therefore developing a novel approach that avoids the need for modeling instructions by leveraging knowledge embedded in retargetable code generators in today's compilers such as GCC. This approach not only simplifies the development of instrumentation, but also makes it applicable to all architectures for which a code generator is available.

Another important advance made by our approach is that of enabling a rich set of optimizations to be performed on binary instrumentations, thereby significantly improving performance over today's techniques. Moreover, our approach enables the use of today's compiler back-ends for generating and optimizing instrumentations, thereby achieving architecture-independent instrumentation.

Today's binary instrumentation techniques have largely been based on dynamic (i.e., runtime) binary instrumentation (DBI). DBI techniques provide two key features needed for security instrumentation: (a) it should be applied to all application code, including code contained in various system and application libraries, and (b) it should be non-by-passable. Previous static binary instrumentation (SBI) techniques have lacked these features.

However, DBI techniques can incur very high overheads in several common usage scenarios, such as application startups, system-calls, and many real-world applications. We have therefore developed a new platform for secure static binary instrumentation (PSI) that overcomes these drawbacks of DBI techniques, while retaining the security, robustness and ease-of-use features. Our experimental results have demonstrated an order of magnitude improvement in performance over DBI techniques on many real-world applications. (NSF and ONR)

Development of Security Threat Control System with Multi-Sensor Integration and Image Analysis

Sangjin Hong
sangjin.hong@stonybrook.edu

This purpose of this project is to develop a heterogeneous sensor network platform for intelligent surveillance system applications. Multiple image sensors, such as hyper-spectral image sensors, are utilized in order to capture various undetectable images for discovering hidden information. Different sensors collaborate for creating consistent overall real-time information to be used to infer many abnormal surrounding situations. In the project, following researches are conducted. Multiple object association through estimation and prediction, low complexity embedded system design, large scale system modeling, multimedia database access strategy, and stochastic collaborative signal processing. (Korea MKE)

Security and Privacy in Geo-Social Networks

Radu Sion
sion@cs.stonybrook.edu

Location based social or geosocial networks (GSNs) have recently emerged as a natural combination of location based services with online social networks: users register their location and activities, share it with friends and achieve special status (e.g.,"mayorship" badges) based on aggregate location predicates.

Boasting millions of users and tens of millions of daily check-ins, such services pose significant privacy threats: user location information may be tracked and leaked to third parties. Conversely, a solution enabling location privacy may provide cheating capabilities to users wanting to claim special location status. In this paper we introduce new mechanisms that allow users to (inter)act privately in today's geosocial networks while simultaneously ensuring honest behavior. An Android implementation is provided. The Google Nexus One smartphone is shown to be able to perform tens of badge proofs per minute. Providers can support hundreds of millions of check-ins and status verifications per day. (NSF, ONR)

Security in Augmented Reality

Arie Kaufman
ari@cs.stonybrook.edu

Augmented Reality (AR) devices access and produce privacy sensitive data such as people's faces or surrounding voices and in general the entire user's visual scene.

Having third party applications on the AR device accessing the sensitive information is more concerning. In the naive approach, the user defines the application's level of access. For example, before installing an application on an Android phone, the OS informs the user that the application requires accessing certain information. We design and develop methods, algorithms, and frameworks that wisely trace the exchange of information between functions/applications and prevents leaking it. Illustrating the problem with an example: consider a user using an AR navigation and in the meantime seeing his credit card information. In this case, the OS or the framework should not allow the application to access the credit card information or prevent the application from leaking the credit card numbers. (NSF CVDI and ITSC)

Imaging and Visualization

Mesh-less Point Cloud Registration by Conformal Mapping

Arie Kaufman and Xianfeng David Gu
ari@cs.stonybrook.edu

With the mass-production of home-use 3D sensors, such as the Microsoft Kinect, 3D Scanning is becoming more popular. However, due to limitations of this scanners (e.g., low resolution, 2.5D) advanced registration is required for generating higher quality models, such as the user's avatar. These scanners create 3D point cloud models, which we need to register in order to build the high quality model or to search for a match in a database of point clouds. The classical Iterative Close Point (ICP) method assumes that there is a large percentage of overlap between the two models to be registered. On the other hand, traditional conformal mapping requires meshes for boundary estimation and does not extend well to point clouds. We introduce a mesh-less conformal mapping framework capable of registering noisy point clouds captured by Kinect without any mesh data using model component segmentation and skeleton tracking. (Internal Funding)

Inverse Reinforcement Learning for Human Attention Modeling

Minh Hoai Nguyen and Dimitris Samaras
minhhoai@cs.stonybrook.edu,
samaras@cs.stonybrook.edu

The process by which people shift their attention from one thing to another touches upon everything that we think and do, and as such has widespread importance in fields ranging from basic research and education to applications in industry and national defense. This research develops a computational model for predicting these human shifts in visual attention. Prediction is understanding, and with this model we will achieve a greater understanding of this core human cognitive process. More tangibly, prediction enables applications to anticipate where attention will shift in response to seeing specific imagery. This in turn would usher in 1) a new generation of human-computer interactive systems, ones capable of interacting with users at the level of their attention movements, and 2) novel ways to annotate and index visual content based on attentional importance or interest.

This project investigates a synergistic computational and behavioral approach for modeling the movements of human attention. This approach is based on an assumption that attentional engagement on an image (or video frame) depends on both the pixels that are being viewed and the viewer's previous state.

Based on this assumption, visual attention is posed as a Markov decision process, and inverse reinforcement learning is used to learn a reward function to associate specific spatio-temporal regions in an image, corresponding to the pixels at a viewer's momentary locus of attention, with a reward. Under this novel approach, the attention mechanism is treated as an agent whose action is to select a location in an image or image frame that will maximize its total reward. This model is being evaluated against a behavioral ground truth consisting of the eye movements that people make as they view images and video in the context of free viewing and visual search tasks.

Figure 5

Volumetric Mesh Mapping

Arie Kaufman and Xianfeng David Gu
ari@cs.stonybrook.edu

With the rapid development of volumetric acquisition and computational technologies, vast quantities of volumetric datasets exist in numerous applications, such as industrial inspection and medical imaging. The demands for processing such volumes are pressing to analyze the topology and geometry, such as volumetric mapping to canonical structures, volumetric registration, volumetric feature extraction, geometric database indexing, volumetric parameterization, and so on.

This project focuses on developing rigorous algorithms for computing the topology and geometry for general mesh volumes.

present

past ← → future

Observed attentional behavior   Predicted attentional behavior
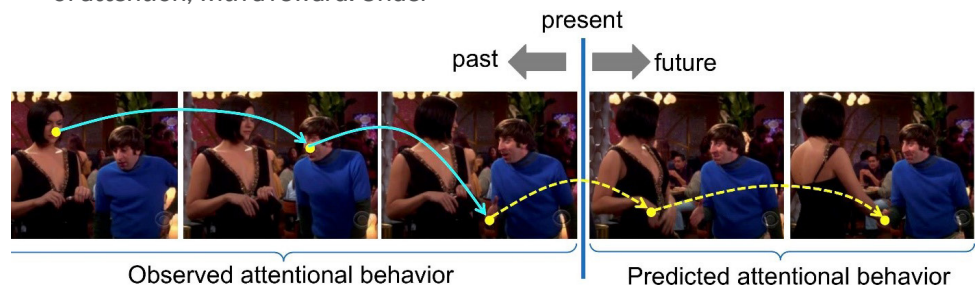
Figure 5

Specifically, we have been developing computational algorithms for Ricci flows. On the other hand, it is highly desirable to map one or more volumes to a canonical domain, to support database indexing and volume registration.

We have built a concrete set of software tools for computing and visualizing the topology and geometric structures for mesh volumes, including volumetric parameterization, volumetric registration, volumetric mapping to canonical structures, fundamental groups computation, and topological and geometric feature extraction.
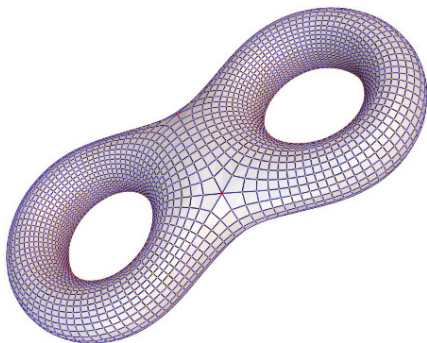
Figure 6

Engineering, science, medicine, computer graphics, vision, scientific computing, and mathematics will directly benefit from these tools, the research and education. These tools can be further used in: (1) industry: in CAD/CAM/CFD simulation and analysis, non-destructive testing of scanned parts, reverse engineering, and large geometric database indexing; (2) medical imaging for volumetric registration and fusion, comparison, shape analysis, abnormality and cancer detection; (3) computational fields, for weather prediction, air flow around vehicles, and toxin prediction, using volumetric computed datasets; and (4) other fields for confocal volume microscopy for cellular research, seismology for earthquake prediction and gas and oil exploration, radar and underwater sonography for terrain mapping and object detection, both civilian and military. (NSF)

....

Quadrilateral and Hexahedral Mesh Generation Based on Surface Foliation Theory

David Xianfeng Gu
gu@cs.stonybrook.edu

For the purpose of isogeometric analysis, one of the most common ways is to construct structured hexahedral meshes, which have regular tensor product structure, and fit them by volumetric T-Splines. We have developed a novel surface quadrilateral meshing method, which leads to the structured hexahedral mesh of the enclosed volume for high genus surfaces.

We prove the equivalence relations among colorable quad-meshes,

finite measured foliations and Strebel differentials on surfaces. This trinity theorem lays down the theoretic foundation for quadrilateral/hexahedral mesh generation, and leads to practical, automatic algorithms.

The algorithm pipeline is as follows: the user inputs a set of disjoint, simple loops on a high genus surface, and specifies a height parameter for each loop; a unique Strebel differential is computed with the combinatorial type and the heights prescribed by the user's input; the Strebel differential assigns a flat metric on the surface and decomposes the surface into cylinders; a colorable quad-mesh is generated by splitting each cylinder into two quadrilaterals, followed by subdivision; the surface cylindrical decomposition is extended inward to produce a solid cylindrical decomposition of the volume; the hexahedral meshing is generated for each volumetric cylinder and then glued together to form a globally consistent hex-mesh.

The method is rigorous, geometric, automatic and conformal to the geometry. It has been applied for real applications in automobile industry.

Figure 6

Imaging and Visualization

Early Event Detection

Minh Hoai Nguyen
minhhoai@cs.stonybrook.edu

The need for early detection of temporal events from sequential data arises in a wide spectrum of applications ranging from human-robot interaction to video security. While temporal event detection has been extensively studied, early detection is a relatively unexplored problem. This paper proposes a maximum-margin framework for training temporal event detectors to recognize partial events, enabling early detection. Our method is based on Structured Output SVM, but extends it to accommodate sequential data. Experiments on datasets of varying complexity, for detecting facial expressions, hand gestures, and human activities, demonstrate the benefits of our approach. To the best of our knowledge, this is the first paper in the literature of computer vision that proposes a learning formulation for early event detection.

Figure 7

Wireless Sensor Network Routing

David Xianfeng Gu
gu@cs.stonybrook.edu

This project focuses on design efficient, reliable and secure routing strategies for wireless sensor network routing. Given a sensor network, according to their distance, one can build a planar graph. By graph embedding method, the graph can be assigned geometric coordinates, such that all cells are convex. The convexity of all cells guarantees the delivery using greedy routing method. The embedding can be transformed to achieve load balancing. Furthermore, the curvature of the network can be designed to avoid congestion. (NSF)

Figure 8
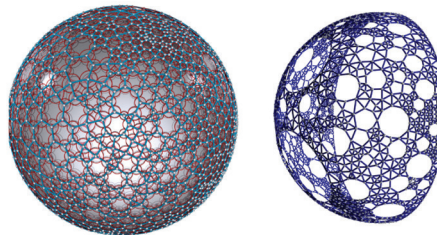
Figure 8

Generative-Adverserial Networks Based on Optimal Transportation Theory

David Xianfeng Gu
gu@cs.stonybrook.edu

Deep learning has achieved unprecedented success in many areas, generative models have become widely popular. However, as the theoretical understanding of their spectacular success remains in its nascent stages, it is crucial to lay down the theoretic foundation for deep learning. We investigate the theoretic interpretation of deep learning models, especially generative models. By converting components of deep learning from "black boxes" to "transparent boxes" as much as possible, our method has made contributions to the societal need for Transparent AI. We have developed geometric theories and computational algorithms for understanding deep learning, especially generative models, from the point of view of manifold embedding and optimal mass transportation. For manifold embedding, we study the capacity of neural networks to prevent over/under fitting; functional/mapping spaces of DNNs to avoid
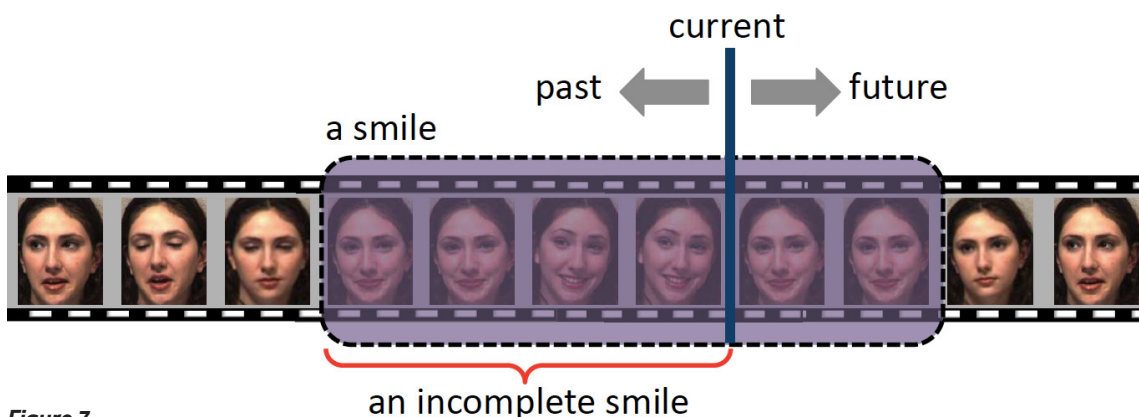
Figure 7

mode collapse; dynamics of linear regions of ReLU DNNs to ensure the convergence of the learning process.

For optimal transport, we study a geometric interpretation of deep learning models, and explain the "black-box" of generative models using optimal transport/convex geometry; the solution space of a generative model to approximate a probability distribution; the dimensionality and the Riemannian metric of the solution space, to measure the learning ability of generative models; the necessity of the competition between the generator and the discriminator, the simplified architecture and learning algorithm.

We have developed algorithms to generalize geometric learning algorithms to high dimensional spaces, improve the efficiency and accuracy in the practical level. For manifold embedding; we design algorithms to estimate the capacity of a DNN and the complexity of manifold, to predict over/under fitting; develop computational tools to allow users to monitor, analyze the dynamics of a learning process, visualize the cell decomposition and activation patterns and so on.

For optimal transport, we investigate different computational approaches to calculate the optimal mass transportation map and combine them in practice for various applications. We also investigate stochastic, hierarchical, Sliced OMT, and Relaxation methods. All these methods are accelerated using hardware, such as GPU, TPU and FPGA. The method is not only vital to the discovery and understanding of deep learning theory, but also lead to advances in two domains of interest, Facial Expression Analysis and Segmentation of Clinically Important Cellular Structures in digital histopathology images. Generative networks can be used to create copious amounts of realistic training data for problems that today cannot be addressed because of the infeasibility of large scale data collection. We use the theory to facilitate learning representation of data where the effects of modes of visual variation such as deformation and appearance can be disentangled.
Figure 9

Figure 9

ColorMapND: A Data-Driven Approach and Tool for Mapping Multivariate Data to Color

Klaus Mueller
mueller@cs.stonybrook.edu

This project addresses the challenge of color-mapping multivariate data. While a number of methods can map low-dimensional data to color, these methods do not scale to higher data dimensions. Our approach does not have these limitations. It is data driven in that it determines a proper and consistent color map from first embedding the data samples into an interactive circular multivariate information display and then fusing this display with a convex color space.

The data attributes are arranged in terms of their similarity and mapped to the circle boundary to control the embedding. Using this layout, the color of a multivariate data sample is then obtained from an interpolation of the map. The system we devised has facilities for contrast and feature enhancement, can deal with multi-field as well as multi-spectral data, and can produce heat maps, choropleth maps, and diagrams such as scatter-plots. An advanced version also supports 3D volumetric data displays.

Figure 10
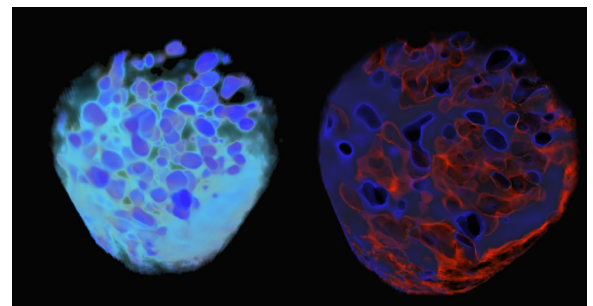
Figure 10

Imaging and Visualization

3D Facial Recognition

David Xianfeng Gu
gu@cs.stonybrook.edu

3D facial recognition has fundamental importance for homeland security. This project focuses on 3D facial surface recognition based on modern geometry and machine learning method. The human facial surfaces are captured using dynamic 3D camera based on phase shifting principle in real time with high resolution and high accuracy. The 3D facial surfaces are mapped onto the planar unit disk via Riemann mapping. The Riemannian metric on the original surface is encoded by the conformal factor on the disk. Then prominent geometric features are automatically selected by machine learning method.

A diffeomorphism of the planar disk is computed by optimizing a special functional, which describes the elastic deformation and bending of the shape. This optimal registration also induces a distance between two 3D facial surfaces. By using the distance among the faces, we can compare their similarities. Different geometric features are weighted in order to improve the recognition rate,

the weights are obtained by machine learning method automatically. Current system beats the state of the art. (NSF)

Figure 11

Conformal Wasserstein Shape Space

David Xianfeng Gu
gu@cs.stonybrook.edu

Surface based 3D shape analysis plays a fundamental role in computer vision and medical imaging. This project uses optimal mass transportation maps for modeling shape space. The computation of the optimal mass transport map is based on Monge-Brenier theory, in comparison to the conventional method based on Monge-Kantorivich theory, this method significantly improves the efficiency by reducing computational complexity.

This project develops the framework of Conformal Wasserstein Shape Space. Given a Riemannian manifold, the space of all probability measures on it is the Wasserstein space.

The cost of an optimal mass transportation between two measures

is the Wasserstein distance, which endows a Riemannian metric of the Wasserstein space. In this work, all metric surfaces with the disk topology are mapped to the unit planar disk by a conformal mapping, which pushes the area element on the surface to a probability measure on the disk.

Figure 12

Shape Analysis with Teichmüller Shape Space

David Xianfeng Gu
gu@cs.stonybrook.edu

Shape indexing, classification, and retrieval are fundamental problems in computer graphics. This work introduces a novel method for surface indexing and classification based on Teichmüller theory. Two surfaces are conformal equivalent, if there exists a bijective angle-preserving map between them.

The Teichmüller space for surfaces with the same topology is a finite dimensional manifold, where each point represents a conformal equivalence class, and the conformal map is homotopic to Identity. A curve in the Teichmüller space represents a

Figure 11
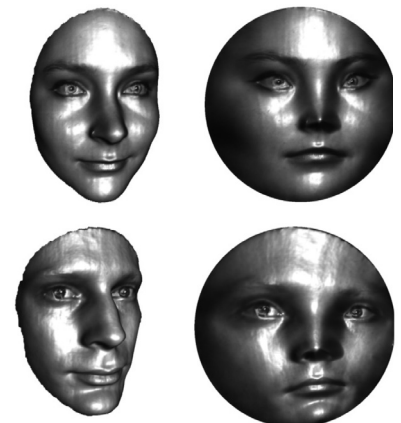
Figure 12

deformation process from one class to the other.

In this work, we apply Teichmüller space coordinates as shape descriptors, which are succinct, discriminating and intrinsic, invariant under the rigid motions and scalings, insensitive to resolutions. Furthermore, the method has solid theoretic foundation, and the computation of Teichmüller coordinates is practical, stable and efficient. This work develops the algorithms for the Teichmüller coordinates of surfaces with arbitrary topologies. The coordinates which we will compute are conformal modules represented as the lengths of a special set of geodesics under this special metric. The metric can be obtained by the curvature flow algorithm, the geodesics can be calculated using algebraic topological method. We tested our method extensively for indexing and comparison of large surface databases with various topologies, geometries and resolutions. The experimental results show the efficacy and efficiency of the length coordinate of the Teichmüller space.(NSF) *Figure 13*
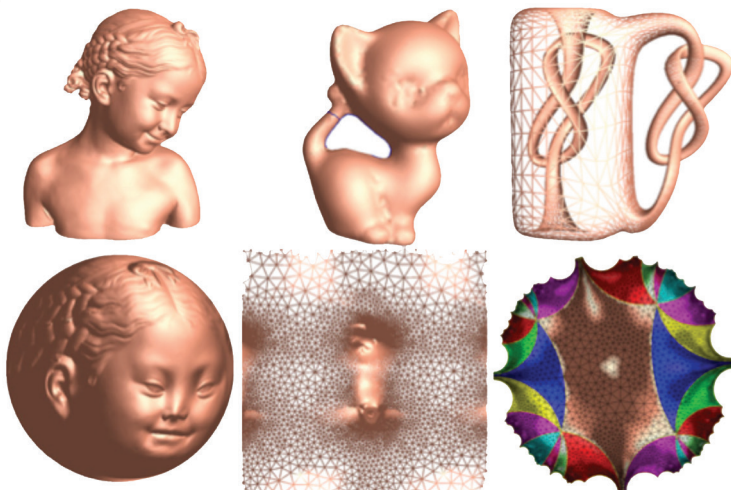
Figure 13

Novel Method for Vectorization of Arbitrary Natural Images and Its Applications

Hong Qin
qin@cs.stonybtook.edu

Vector graphics offer a compact and lossless image representation, with many advantages such as geometric edit-ability, resolution independence, significant saving of storage and network bandwidth, image display at drastically varying resolutions, and easy animation. This research initiative aims is to significantly advance the traditional boundary of image vectorization based on partial differential equations (PDEs) and their intrinsic connection with Green's functions and harmonic B-splines (serving as fundamental solutions for PDEs), which have not yet been explored for vector graphics, image modeling, image data fitting, and analysis. If successful, this research will deliver a novel vector image modeling methodology and its applications for image vectorization and authoring as well as solid texture and animation. At the core of this initiative's theoretic foundation are PDEs and their meshless closed-form solvers based on fundamental solutions. The novel representation is expected to outperform the conventional diffusion curve based and gradient mesh based representations. The new modeling scheme is capable of expressing arbitrary image in theory with arbitrary discontinuities. Consequently, this research will advance the state of the knowledge of current theory and practice of vector graphics.

Predicting Grade of Dysplasia of Pancreatic Lesions from CT scans

Arie Kaufman
ari@cs.stonybrook.edu

With the widespread use of high-resolution cross-sectional imaging, pancreatic lesions have become one of the most common incidental gastrointestinal disorders. While some of these lesions are benign, some, including intraductal papillary mucinous neoplasms (IPMN) and mucinous cystic neoplasms (MCN), have malignant potential and might progress to cancer. Given the potential risks and complications of pancreatic surgery, an accurate diagnosis is crucial for correct patient management. This project focuses on developing a machine learning 3D convent-based model for the classification of pancreatic lesions in CT images and predicting the grade of dysplasia in malignant types. Unlike previous works, our model efficiently leverages the 3D information in a CT scan to acquire a holistic picture of each case to generate more detailed predictions. We are designing this application through close collaboration and feedback from expert radiologists. (Marcus Foundation, NIH-REACH)

Imaging and Visualization

Learning Multi-Class Segmentations from Single-Class Datasets

Arie Kaufman
ari@cs.stonybrook.edu

Multi-class segmentation has recently achieved significant performance in natural images and videos. This achievement is due primarily to the public availability of large multi-class datasets. However, there are certain domains, such as biomedical images, where obtaining sufficient multi-class annotations is a laborious and often impossible task and only single-class datasets are available. While existing segmentation research in such domains use private multi-class datasets or focus on single-class segmentations, we propose a unified highly efficient framework for robust simultaneous learning of multi-class segmentations by combining single-class datasets and utilizing a novel way of conditioning a convolutional network for the purpose of segmentation. We have demonstrated various ways of incorporating the conditional information, perform an extensive evaluation, and show compelling multi-class segmentation performance on biomedical images, which outperforms current state-of-the-art solutions (up to 2.7). Furthermore, we have shown the applicability of our method also to natural images and evaluate it on the Cityscapes dataset. Our proposed framework can be applied to many other applications. (Marcus Foundation, NIH-REACH)

Virtual Pancreatography

Arie Kaufman
ari@cs.stonybrook.edu

Pancreatic cancer is the fourth leading cause of cancer related deaths in the United States with less than 10% survival rate over a 5-year period, and is often fatal due to late stage manifestation of symptoms. Consequently, many cases of early detection of pancreatic cancer are in the form of visible pancreatic lesions on incidental computed tomography (CT) scans. Visible features on the CT-scan such as internal and peripheral calcifications, lesion septation, and pancreatic duct dilation/cut-off and its relationship with the lesions provide visual clues for diagnosing malignant lesions. Currently, non-invasive screening of patients is performed through visual inspection of 2D axis-aligned CT images, though the relevant features are often not clearly visible. Virtual Pancreatography (VP) is a novel state-of-the-art visualization system for non-invasive diagnosis and classification of pancreatic lesions. VP is an end-to-end visual diagnosis system that includes: a machine learning based automatic segmentation of important anatomical structures such as the pancreatic gland and the lesions and possibly the primary pancreatic duct; a machine learning based automatic classification of lesions into several prominent types; and, specialized 3D and 2D exploratory visualizations of the pancreas and surrounding anatomy. Multiple visual configurations combine volume rendering with 2D views of raw CT intensities for effective and reliable diagnosis.

Perception-Based Camera Control in Virtual Reality

Arie Kaufman
ari@cs.stonybrook.edu

Virtual reality (VR) provides an immersive platform for exploring virtual worlds but can induce cyber-sickness due to the discrepancy between visual and vestibular cues. To avoid this problem, the movement of the virtual camera needs to match the motion of the user in the real world. Unfortunately, this usually is difficult due to the mismatch between the size of the virtual environments and the space available to the users in the physical space. The resulting constraints on the camera movement significantly hamper the adoption of virtual-reality headsets in many scenarios and make the design of the virtual environments very challenging. In this work, we study how the characteristics of the virtual camera movement and the composition of the virtual environment contribute to perceived discomfort. Based on results from user experiments, we devise a computational model for predicting the magnitude of the discomfort. We further demonstrate how the model can be used in a new dynamic path planning method that reduces perceptual sickness while maintaining the fidelity of the original navigation. We evaluate the effectiveness of our technique in improving perceptual comfort and task performance in user studies with a variety of applications. (NSF CVDI and ITSC)

FeatureLego: Volume Exploration Using Exhaustive Clustering of Super-Voxels

Arie Kaufman
ari@cs.stonybrook.edu

Volume exploration is often one of the first steps towards understanding spatial data and its contents. It is therefore an important problem in scientific and medical visualization.

Typical approaches harness a general feature extraction method followed by an interaction setup such as hierarchies or graphs for selection and composite visualization these features. General data-driven approaches to feature extraction are preferred to support a wide variety of data and modalities. Volume exploration frameworks may support user tasks such as search, selection, and combined visualization of semantic features; managing occlusion; and controlling visual parameters such as color, opacity, shading, and lighting. This project includes a volume exploration framework, FeatureLego, that uses a novel voxel clustering approach for efficient selection of semantic features. We partition the input volume into a set of compact super-voxels that represent the finest selection granularity. We then perform an exhaustive clustering of these super-voxels using a graph-based clustering method. Unlike the prevalent brute-force parameter sampling approaches, we propose an efficient algorithm to perform this exhaustive clustering. By computing an exhaustive set of clusters, we aim to capture as many boundaries as possible and ensure that the user has sufficient options for efficiently selecting semantically relevant features.

Furthermore, we merge all the computed clusters into a single tree of meta-clusters that can be used for hierarchical exploration. We have implemented an intuitive user-interface to interactively explore volumes using our clustering approach. Finally, we have shown the effectiveness of our framework on multiple real-world datasets of different modalities. (NSF CVDI)

Efficient Volume Exploration Using Active Contours

Arie Kaufman
ari@cs.stonybrook.edu

Active contours driven by Bhattacharyya distance is an established technique for 2D image segmentation. Given an initial segmentation mask, it computes the foreground and background distributions of the pixels over a feature space. The segmentation mask is then iteratively morphed to maximize the Bhattacharyya distance between the two distributions. This approach is suitable for volume exploration since it is general and does not depend on any prior knowledge about features. Previous implementations are restricted to 2D input images and 1D feature spaces (intensity values). We extend active contours approach to 3D volumetric data and n-dimensional feature vectors. The user can select from a list of local features such as intensity, median, gradient, curvature, etc. or define their own local features. This provides additional flexibility and capability to explore a large variety of input modalities. The feature extraction process is applied locally and on-the-fly which eliminates the need for extensive preprocessing and also supports larger size volumes. We enable this through an efficient GPU driven implementation for interactive volume exploration. (NSF CVDI)

Augmented Reality in Ubeity Space

Arie Kaufman
ari@cs.stonybrook.edu

The vastness and the immediate access to information via the internet and powerful search engines, is unparalleled today than ever before in the history of mankind. With the rapid advancement of mobile and wearable-device technologies, such as augmented reality (AR) devices, a new paradigm of information distribution and retrieval can be introduced based on ubeity (the state of being placed in a definite local relation). To this end, we propose a novel hierarchical data-structure that associates information based on the geographical location where the information is either located, generated, or related to. Such a "data-structure" allows a more intuitive way of sorting information, such as images, based on 'whereness'. With information now "sorted" by users or services at some granularity of location, users can access the information, subject to policies and privacies, and augment the information on users' mobile or wearable device as the user is physically present at the location or accessing a location remotely. Related technologies that are needed and is being developed include accurate indoor localization, placement of the information overlaid in the augmented space, and suitable interaction techniques. (NSF CVDI and ITSC)

Perception-based Immersive Visualization and Interaction

Arie Kaufman
ari@cs.stonybrook.edu

Immersive analytics for volume data with high resolution and large size is challenging as it requires high rendering performance without noticeable interaction latency. Recent advances in data acquisition further increase the demand for large-scale volume data management and visualization. In this work, we address two major challenges for immersive

Imaging and Visualization

analytics: (1) efficient in-core data updates; (2) 3D data interaction in an immersive environment. Specifically, we develop an efficient data structure based on human perception (e.g., gaze, foveated view and peripheral view, visual acuity) for large-scale volume data visualization. We also develop a new set of immersive interaction techniques for data labeling and data exploration based on gaze tracking and manual controlling. A series of perception experiments are conducted to validate our assumption in human perception during scientific analytics. We evaluate the performance of our technique in a variety of large-scale volume dataset and demonstrate an application of our method in neuron data registration. (NSF CVDI and ITSC)

....

Reality Deck: the World's Largest Visual Analytics Facility

Arie Kaufman
ari@cs.stonybrook.edu

Large, interactive, high-resolution displays have been demonstrated as a valuable tool for the exploration of massive amounts of data. Due to their size, they allow for physical navigation (walking around in space) rather than virtual navigation (manipulating a virtual camera with a controller). Such displays were limited to 300 mega-pixels in aggregate resolution. Additionally, they presented themselves as a single planar surface, reducing the potential physical navigation benefits that users can enjoy. We have built the Reality Deck, which is the next-generation immersive, large, interactive, super resolution display. It is a unique 416 tiled display visualization environment of size 30' x 40' x 11' high, that offers a total resolution of 1.5 billion pixels in a four wall horizontally immersive layout, while providing 20/20 visual acuity for the visualization space.

It is the first facility of its kind and improves the resolution by a factor of 5 compared to the next largest tiled display wall and by a factor of fifteen compared to other immersive environments, such as the CAVE. The high-resolution tiled LCD displays are driven by an efficient 20-node visualization cluster that utilizes six Nvidia RTX 6000 GPUs per node with four displays connected to each GPU. The cluster provides an aggregate of 848,872 CUDA cores, 13,248 ray-tracing cores, and 105,984 tensor cores, for an aggregate performance of about 3 PFLOPS peak single precision FP32 performance and 24 Tensor PFLOPS. The Reality Deck is a one-of-a-kind facility, which serves as a hybrid platform for core visualization and machine learning research, systems-level research for enabling the visualization and analytics of new types of data (such as gigapixel video) and finally as an exploration platform for real-world deep learning and visual analysis problems. We have implemented a number of interactive applications that leverage the super high resolution and super deep learning performance, that deal with a variety of large datasets and decision informatics, including gigapixel panoramic images, global GIS data,

molecular models, medical data and art collections. (NSF, CEWIT, NYS)

....

Volume Rendering in Augmented Reality

Arie Kaufman
ari@cs.stonybrook.edu

Volume rendering has always been an important topic in scientific visualization. Putting volume in an augmented reality setting will be both interesting and useful, for the fact that users can interact with the volume more naturally and explore the volume with associated data in real world. Recent advances in hardware greatly enhance the computing ability of mobile devices, which makes rendering small volume datasets on mobile devices a reality. However, for large-scale volume rendering, it is such a computation-consuming task that it can hardly be done without advanced graphics hardware. Due to the hardware limitation, directly rendering large-scale volume on

mobile devices is unrealistic. To interact with the large-scale volume, we are using both computation power of the desktop and flexibility of mobile devices. We can precisely render on the desktop, while at the same time users can use their mobile devices to explore part of the whole volume.

We plan to develop novel collaboration method and interaction techniques for the desktop-mobile device AR system. Collaboration of multiple users with mobile devices is the next step. (NSF CVDI)

....

Visualization for Climate Simulation Data

Arie Kaufman
ari@cs.stonybrook.edu

This is a cross-institutional and cross-disciplinary effort to enable the state of New York to be one step ahead of super-storm Sandy-like disasters. The goal is to bring together the expertise of various researchers at Stony Brook (lead by researchers at the School of

Marine and Atmospheric Sciences) and other universities in the fields of marine sciences, climatology, road network simulation, emergency response planning and visualization.

We are focusing on the visualization aspect of the project and are developing novel and scalable visualization techniques for climate simulation data that will be utilized in this project. Such visualizations would merge the simulation data with underlying road networks, elevation and other GIS sources, enabling emergency planners to be better prepared for future storms and emergencies.

These technologies will be deployable on traditional desktop computers but also scale up to gigapixel resolution facilities, such as the Reality Deck. (NYS)

Healthcare and Biomedical Applications

New Platforms for 3D Tissue Growth and Regeneration

David Rubenstein
david.rubenstein@stonybrook.edu

Our research is focused on the development of new techniques and technologies that can be used for the development of ex vivo tissues. Using combinations of printing, electro-spinning and gel casting we have a wide array of scaffolds that are compatible with and support the growth of various salient cell types of importance to regenerative medicine. We have focused on the development of methods for vascular tissue, bone, and liver production using mimetic technologies. For example, we have various techniques to improve the mechanical properties of biomaterials scaffolds fabricated using different techniques. In parallel, we have the ability to accurately control the topology and chemical composition of the formed scaffolds. With the successful completion of this work, we will move towards a solution to rapidly fabricate products that may be used in various regenerative medicine applications and we aim to develop a platform that can be applied to a tissue of interest with the input of various design constraints.

Improving Understanding and Treatment of Depression Using Brain Imaging

Christine DeLorenzo
christine.DeLorenzo@
stonybrook.edu

Depression and other mental illnesses are heterogeneous and complex. In order to improve outcomes for those suffering from these illnesses, we need to better understand their neurobiology. To accomplish this, we take comprehensive approach to probing brain biology and function. We focus on Positron Emission Tomography (PET) to visualize and quantify neuroreceptor systems in those with and without mental illness. Human PET imaging is complemented by human Magnetic Resonance Imaging (MRI), as well as rodent models and imaging. This is aided by collaborators in departments such as Neurobiology and Behavior and Stony Brook's unique imaging infrastructure including the simultaneous PET/MRI, one of few in New York State. Each of our studies involve clinical and technical objectives, both aimed at improving clinical care. For example, a currently ongoing study, "Use of PET/MRI in Major Depressive Disorder" involves the use of PET to quantify brain metabolism before and after antidepressant treatment. The simultaneous PET/MRI allows for simultaneous acquisition of MRI sequences (e.g., Diffusion Spectrum Imaging [white matter tracts], Arterial Spin Labeling [blood flow], Spectroscopy [concentration of brain metabolites]) during the PET scan. The clinical goals are to use imaging to (1) predict who will respond to the medication prior to treatment and (2) relate antidepressant-induced mood changes to imaged brain changes.

The technical goal is to apply engineering principles to enhance image acquisition and analysis, for example, by developing new technology and algorithms to improve the accuracy of PET-based measures, with minimal inputs. Due to the large amount of data generated by the in vivo imaging, processing, integrating and analyzing this data requires sophisticated techniques. However, the results can provide much needed insight into the human brain and mental illness.

Figure 14
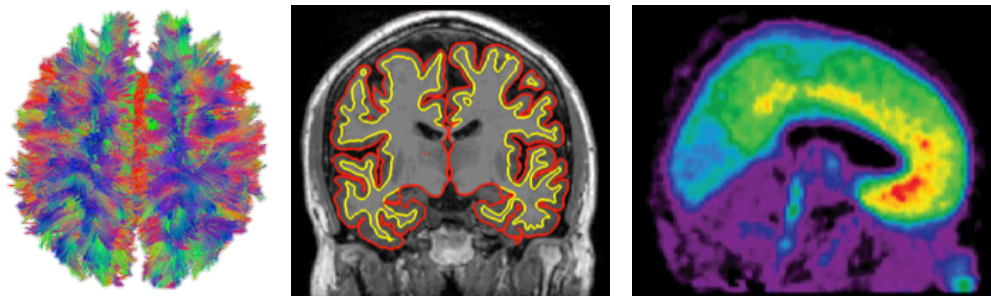
Figure 14 an example of some of the in vivo imaging modalities used to study mental illness including: Diffusion Spectrum Imaging (DSI, showing white matter tracts throughout the brain, left), structural Magnetic Resonance Imaging (MRI, showing the calculation of cortical thickness between the red and yellow surfaces, middle) and Positron Emission Tomography (PET, showing the density of serotonin 1A receptors, right).

Mobile Healthcare Device Platform

Wei Lin
wei.lin@stonybrook.edu

Our research focuses on the development of wearable healthcare devices. We've developed a hybrid platform that utilizes ARM Cortex-M series processor as the embedded system and the open source FreeRTOS as its real time operating system. The ARM processors are designed for mobile applications with low power consumption. The FreeRTOS provides a small foot print kernel for real time, multi-task applications. In order to integrate the FreeRTOS with the selected processor, a set of drivers were designed to bridge the FreeRTOS and the mobile processor.

They provide a uniformed software interface that allows the application code running on the FreeRTOS to easily control the hardware resources such as analog digital converter, I2C bus and universal asynchronous receiver/transmitter (UART). Application code can be divided into small modules for design simplicity. The software architecture maximizes the code re-usability and enables the quick switch of hardware design with little impact on the existing code. Its flexibility is very attractive for the development of mobile healthcare applications and can significantly reduce the development time of the prototype. This platform has been adopted in the mobile infant monitor device for the prevention of infant sudden death syndrome and the temporary pacemaker to treat ventilator induced diaphragm dysfunction of ICU patients.

Coronary Artery Biomechanics

Wei Yin
Wei.rubenstein@stonybrook.edu

Cardiovascular disease is the leading cause of death in the United Sates, and coronary artery disease is the most common type of cardiovascular disease. Shear stress induced by blood flow and tensile strain induced by blood vessel motion both play important roles in the initiation and progress of atherosclerosis, the major cause of coronary artery disease. Circulating platelets and vascular endothelial cells are very sensitive to their mechanical environment; any change can affect their functions and interactions significantly. The current projects in my research laboratory aim to investigate how altered shear stress/tensile strain conditions affect platelet and endothelial cell behavior and lead to cardiovascular disease initiation. Computational fluid structure interaction models, as well as in vitro, ex vivo and in vivo models, are used to investigate platelet and endothelial cell responses and interactions under physiologically relevant stress/strain conditions. Biomarkers associated with platelet and endothelial cell activation are of special interest to us. We also work on numerical models to simulate platelet coagulation kinetics and platelet adhesion to injured blood vessel wall under dynamic flow conditions.

Radiomics and Cancer Therapy

Allen Tannenbaum
Allen.tannenbaum@stonybrook.edu

In CEWIT, our group is performing several research projects in data analysis, with a strong emphasis on medical imaging. The mining of information and the analysis of medical imaging data taken from multiple modalities having different physical properties lies at the heart of the emerging field of radiomics. One wants to employ such multi-modal information to build models connecting the imagery to genetic and proteomic data as well as to various phenotypes. In addition to aiding in diagnosis, the challenge is use such information in prediction. It is here that the signal processing and control methodologies of statistical filtering and estimation theory may be essential. Given the noise in very high dimensional data and the lack of precise models, and the fact that the data may be time-varying, one need techniques for the quantification of uncertainty and model reduction.

Time-varying data includes longitudinal data, but many times the causal relationship from one time point to the next is not explicitly taken into account. This is where ideas from dynamical systems and control arise, and why we refer to this core as the "dynamic radiomics" core. In computer vision, control principles have been used now for some time: this is at the heart of controlled active vision, and visual tracking.

Particle filtering then becomes an essential player in this enterprise given the fact that no a priori noise model is assumed. Particle filtering is a sequential Monte Carlo method that includes Kalman filtering as a special case, when the underlying system dynamics are linear and the noise is Gaussian. This method allows one to take advantage of the longitudinal information in a principled manner. Further, particle filtering has proved to be an important tool in registration, segmentation, data assimilation, and

Healthcare and Biomedical Applications

data association. Because of the very large dimensionality of the systems involved (leading to problems in "Big Data"), one needs to use this type of Bayesian methodology in conjunction with techniques in nonlinear dimensionality reduction and data clustering.

Finally, we are developing an approach that will allow the introduction of more automatic and reproducible methods in radiation oncology that may aid in prognosis, radiation planning, and in predicting the effectiveness of the treatment. This will also take advantage of new approaches to image segmentation and registration that explicitly employ statistical filtering and control, and lead to interpretable and robust interactive methods.

Low-cost, Portable, Touchless Vital Signs Sensing System

Fan Ye
fan.ye@stonybrook.edu

This project studies how to develop a hardware/software solution that can provide long-term monitoring of vital signs (e.g., heartbeat and respiration rate) and tremor in a non-contact and privacy-preserving way without cooperative labor from individuals.

The patterns and changes in basic vital signs including heart beat and respiration rates are important and essential bio-indicators for assessments of onset, progression and resolution of diseases. Tremors-unintentional, rhythmic, fine-grained to-and-from movements of body parts-are prominent gauge in the diagnosis and care of diseases such as Parkinson's disease, Multiple Sclerosis. In early stages, people with these diseases experience often unnoticeable limb tremors. Longitudinal collection and evaluation of such data in one's natural habitat (i.e., home) can detect, predict improvement or decline in diseases, enabling preemptive and timely adjustments of intervention. Unfortunately, there exist no convenient, cost-effective solution for longitudinal collection of such data in-home.

We have developed VitalHub, which leverage Ultra-Wide Band (UWB) radio sensors aided by a privacy-preserving depth camera for multi-user vital signs and tremor monitoring. It tracks the body location and extracts fine-grained vital sign and tremor signals.

Experiments with 8 individuals show that VitalHub can track up to 6 users concurrently. It can measure the respiration and heart rates within 0.8 and 1.2 bpm errors, and detect tremors of small amplitudes (sub-mm) and frequencies (up to 10 Hz) at hands, legs and feet, which are hard to detect by other means.

Machine Learning-Guided Multi-scale Modeling of Platelet Dynamics Machine Learning

Yuefan Deng, Changnian Han, Prachi Gupta and Zijii Zhang
Yuefan.deng@stonybrook.edu

The coagulation cascade of blood may be initiated by flow-induced platelet activation and aggregation, which prompts clot formation in prosthetic cardiovascular devices and vascular disease processes. Upon activation, platelets undergo complex morphological changes. Activated platelets polymerize fibrinogen into a fibrin network that enmeshes red blood cells. In this research, we are aiming at a multi-scale modeling of the thrombus formation in shear flow, including platelet adhesion, aggregation and activation, by correlating with in-vitro results.

For in-vitro experiments, we introduce a novel data-driven deep-learning-based framework for characterizing platelet dynamics in shear flow to meet the need of assessing model validity by data collection and analysis. We propose a model can successfully predict platelet aggregation phenomena in micro scale details with nano scale accuracy. Meanwhile, we are designing a CNN-based network to automatically detect the moving cells out of the large stream of in-vitro results including platelets, red blood cells, etc. By the online-learning framework, real-time objects detection, segmentation, and classification can be achieved. For in-silico simulations, we develop a multi-scale model to simulate the dynamics of platelet aggregation by recruitment of unactivated platelets flowing in viscous shear flows by an activated platelet deposited onto a blood vessel wall. The binding of receptor and fibrinogen is modeled by a molecular-level hybrid force field consisting of Morse potential and Hooke law for the non-bonded and bonded interactions, respectively. The force field in two different interaction scales is calculated by correlating with the in-vitro experiments. We derived the relationship between recruitment force and distance between the mass centers of two platelets.

Besides modeling, simulations at such spatio-temporal scales are time-consuming. Traditional time stepping algorithms using the smallest time step size in order to capture the finest details lead to a significant waste of computing resources for simulating

coarse-grained portion of the problem. We propose an adaptive time stepping algorithm to intelligently adapt step sizes to the underlying biophysical phenomena. We establish a learning network to extract attributes and train the state trigger mechanism. Our algorithm presents a more efficient way for solving massive multi-scale problems. Our model and simulation framework can be further adapted to simulate initial thrombus formation involving multiple flowing platelets as well as deposition and adhesion onto blood vessels.

Markov Decision Analysis of Cancer Treatment Plans

Eugene A. Feinberg
Eugene.feinberg@stonybrook.edu

Based on the available data, this project constructs and studies Markov decision models for the analysis and comparisons of cancer treatment plans. It is currently focused on treatment plans for resectable pancreatic and gastric cancers. In particular, the project deals with comparisons of traditional adjutant therapies, when the surgery precedes the chemo (or chemoradiation) treatment, and neoadjuvant therapies conducting chemo or chemoradiation treatment first followed by the surgery.

High Performance Mobile Computing Using Field Programmable Gate Array

Wei Lin
Wei Lin@stonybrook.edu

Intensive data processing is a challenge in mobile devices because of the limitation of computation resources available in the embedded microprocessor. We have developed a hardware technology that utilizes the field programmable gate array (FPGA) to implement the sophisticated algorithms in digital circuits. The pipeline architecture enables the high throughput parallel data processing that exceeds the performance of high end desktop computers. We tested out solution on Xilinx SoC chip (zynq FPGA family) that has a duo core ARM 9 MCU and an FPGA fabric. The intensive computation tasks are implemented in the FPGA portion of the chip while the MCU is in charge of data communication and user interface. Our technology has been successfully applied in the optical spinal cord monitor that measure blood perfusion around the site of spinal cord surgery or trauma to provide the real time blood supply data as the indication of the spinal cord health and the prevention of the spinal cord injury due to the loss of blood supply. Diffuse correlation spectroscopy (DCS) is the optical technology to measure blood perfusion in tissue by deriving the blood flow index (BFI) from the autocorrelation of the light scattering back from the tissue. Two key computation bottlenecks of the technology are the autocorrelation of over one million data points and the nonlinear curve fitting of the autocorrelation curve to the theoretical model. Those efforts are for the derivation of one data point of BFI. We successfully implemented the correlator and the curve fitting analyzer of BFI in one FPGA chip. The performance is about 100 times faster than the software approach. The device on the chip solution significantly reduces the size of the DCS device and makes it a true mobile device for more clinical applications.

Front-end Electronics for a Self-Powered Implantable Sensor for Total Knee Replacement

Emre Salman
emre.salman@stonybrook.edu

Total knee replacement (TNR) is a very common surgery in the US where the annual volume is expected to exceed 1.3 million by 2020. It is highly important to be able to monitor the health of the joint after surgery so that follow-up revision surgeries are avoided. In this research, we develop the ultra-low power implanted electronics for a sensor designed to monitor the load on the joint. The project exhibits unique challenges where the output of the energy harvester behaves both as a data signal to be digitized and a power supply signal to be rectified. We are developing ultra-low power rectification, followed with efficient delta-sigma quantization methods. We are also investigating power management techniques that involve super-capacitors to reach the required power levels. (NIH-R21)

Rethinking Electronic Fetal Monitoring to Improve Perinatal Outcomes and Reduce Frequency of Operative Vaginal and Cesarean Deliveries

Petar Djuric
petar.djuric@stonybrook.edu

The essential role of electronic fetal monitoring (EFM) during labor is to prevent adverse outcomes due to fetal hypoxia and ischemia. This technology over the past 50 years has

Healthcare and Biomedical Applications

not been shown to decrease stillbirths or reduce the numbers of infants with cerebral palsy. The main objective of the proposed research is to use recent breakthroughs in machine learning to drive the development of predictive analytics to support and improve the interpretation of EFM data, especially under real world conditions and in real time where clinicians must make timely decisions about interventions to prevent adverse outcomes. It is anticipated that the proposed research will result in significantly decreased use of operative vaginal delivery and cesarean delivery while more precisely defining the fetus at risk for developing metabolic acidosis and long-term neurological injury. (NIH)

User-Guided Segmentation of Pancreas and Lesions

Arie Kaufmann
ari@cs.stonybrook.edu

Accurate segmentation of abdominal organs from medical images is an essential part of surgical planning and computer-aided disease diagnosis. Many existing algorithms are specialized for the segmentation of healthy organs. Cystic pancreas segmentation is especially challenging due to its low contrast boundaries, variability in shape, location and the stage of the pancreatic cancer. In this project, we have developed a semi-automatic segmentation algorithm for pancreata with cysts. In contrast to existing automatic segmentation approaches for healthy pancreas segmentation which are amenable to atlas/statistical shape approaches, a pancreas with cysts can have even higher variability with respect to the shape of the pancreas due to the size and shape of the cyst(s). Hence, fine results are better attained with semi-automatic steerable approaches. We use a novel combination of random walker and region growing approaches to delineate the boundaries of the pancreas and cysts with respective best Dice coefficients of 85.1% and 86.7%, and respective best volumetric overlap errors of 26.0% and 23.5%. Results show that the proposed algorithm for pancreas and pancreatic cyst segmentation is accurate and stable. (Marcus Foundation, NIH-REACH)

Comparison and Calibration of the RNA-Seq and Gene Micro-array Platforms with the Generalized Linear Errors-in-Variables Model

Wei Zhu and Ellen Li
Wei.zhu@stonybrook.edu,
Ellen.li@stonybrook.edu

Micro-array and RNA-Seq are two major platforms for large scale transcriptomic profiling with the ability to measure thousands of genes simultaneously. Both platforms have been widely used for a long time in Biomedical research: Micro-array is the traditional choice since mid 1990s and RNA-Seq emerges during the last decade as an attractive alternative. Despite Micro-array quantifies abundance with Fluorescence intensity while RNA-Seq with fragment count, these two platforms both quantify transcript abundance in a sample. Increasingly popular in the recent years, the RNA-Seq technology has been used in a number of pathological studies of a wide array of diseases previously studied via the Micro-array. Subsequently, whether measurements and analyses obtained from these two platforms are consistent or not has become an intriguing question and an important issue. Some comparative researches have generated parallel data set that implemented both platforms on the same samples. In these studies, Pearson's correlation and Spearman's correlation are computed for quantifying the comparison and a strong positive correlation is observed between Micro-array preprocessed intensity and RNA-Seq normalized count data in log scale.

Although a high correlation coefficient reflects a strong linear relationship between two platforms, it does not suggest observed gene expression are commensurate because its value remains the same regardless of slope and intercept of the line.

In this project, we develop the generalized linear errors-in-variables model as an exact calibration model between the Micro-array and the RNA-Seq measurement platforms. In addition, because the qRT-PCR is commonly regarded as the most reliable expression-profiling platform, we also apply the proposed method to determine whether the Micro-array or the RNA-Seq measurement is more consistent with the qRT-PCR measurement, of the same sample. Part of the research funding for this project came from NIH.

Context Factorization for Human Action Recognition

Minh Hoai Nguyen
minhhoai@cs.stonybrook.edu

This project investigates problems of human action recognition in video. A human action does not occur in isolation, and it is not the only thing recorded in a video sequence. A video clip of a human action also contains many other components, including

the background scene, the interacting objects, the camera motion, and the activity of other people. Some of these components are contextual elements that frequently co-occur with the category of action in consideration. The project develops technologies that separated human actions from co-occurring factors for large-scale recognition and fine-grain visual interpretation of human actions. The developed technologies can have many practical applications in a wide range of fields, ranging from human computer interaction and robotics to security and healthcare.

This research develops an approach to human action recognition by explicitly factorizing human actions from context. The key idea is to exploit the benefits of the information from conjugate samples of human actions. A conjugate sample is defined as a video clip that is contextually similar to an action sample, but does not contain the action. For instance, a conjugate sample of a handshake sample can be the video sequence showing two people approaching each other prior to the handshake. The handshake clip and the video sequence preceding it have many similar or even the same contextual elements, including the people, the background scene, the camera angle, and the lighting condition. The only thing that sets these two video clips apart is the actual human action

itself. A conjugate sample provides complementary information to the action sample; it can be used to suppress contextual irrelevance and magnify the action signal.

The specific research objectives of this project include: (1) collecting human action samples for many action classes; (2) developing algorithms to mine and extract conjugate human action samples; and (3) developing a framework that utilizes the benefits of conjugate samples for separating actions from context to learn classifiers for large scale recognition and fine grain understanding of human actions.

Figure 15

....

Medical (CT) Image Generation with Style

Klaus Mueller
mueller@cs.stonybrook.edu

Deep learning has shown great promise for a myriad of applications in CT imaging such as improving image quality in low dose acquisition, cross modality translations etc. However training of deep networks requires an abundance of clinical training data. This remains a challenge due to scarcity/privacy issues and the high interpatient anatomical variability. Also, often these datasets are not comprehensively annotated, owing to

the costliness and scarcity of expert annotation in the medical domain. Hence we have devised an approach that can increase the training data multiple folds with as few as ten training samples.

Our framework also ensures that the full sized generated CT images are anatomically correct and contain enough anatomical variation from training data.

....

Conformal Mapping for Medical Imaging

Arie Kaufman, Xianfeng David Gu, Dimitris Samaras, and Wei Zhu
ari@cs.stonybrook.edu

It is paramount in medical imaging to measure, compare, calibrate, register and analyze potentially deformed organ shapes with high accuracy and fidelity. However, this is extremely difficult due to the complicated shape of human organs. Different organs have different topologies and curvature distributions, and furthermore, the shape may deform due to disease progression, movement, imaging, surgery and treatment. We have used conformal geometry, a theoretically rigorous and practically efficient and robust method, to tack this challenge. The broad objective of this project is to develop conformal geometry as a

Figure 15

Action   Context

Healthcare and Biomedical Applications

primary tool in the vast biomedical applications of medical imaging.

We have been developing application of conformal surface flattening to a variety of organs, use of conformal mapping for volumetric parameterization, registration and fusion using conformal geometry, and statistical analysis and feature extraction using conformal geometry. The research design and methodology include developing and validating techniques to conformally flatten 3D organ surfaces to canonical parametric surfaces for colonic polyp detection, bladder cancer screening, and endovascular surgical planning for aortic aneurysm. We have further extended flattening to implement volumetric parameterization based on Ricci flow and then apply it to brain and colon structure segmentation, tumor evaluation, diffuse tensor field study. In addition, we have implemented shape registration and data fusion using a common canonical parameter domain. Brain data sets have been fused between and within subjects and modalities, as well as colon supine and prone have been registered for cancer screening. Finally, we have conducted statistical analysis and feature extraction using conformal geometry for drug addiction and Alzheimer's disease, where Fourier analysis on the canonical domains have transformed them to frequency domain. (NIH)

Human Cortical Surface Morphological Study

David Xianfeng Gu
gu@cs.stonybrook.edu

This project develops a general approach that uses conformal geometry to parameterize anatomical surfaces with complex (possibly branching) topology. Rather than evolve the surface geometry to a plane or sphere, we instead use the fact that all orientable surfaces are Riemann surfaces and admit conformal structures, which induce special curvilinear coordinate systems on the surfaces. Based on Riemann surface structure, we can then canonically partition the surface into patches. Each of these patches can be conformally mapped to a parallelogram. The resulting surface subdivision and the parameterizations of the components are intrinsic and stable. To illustrate the technique, we computed conformal structures for several types of anatomical surfaces in MRI scans of the brain, including the cortex, hippocampus, and lateral ventricles. We found that the resulting parameterizations were consistent across subjects, even for branching structures such as the ventricles, which are otherwise difficult to parameterize. Compare with other variational approaches based on surface inflation, our technique works on surfaces with arbitrary complexity while guaranteeing minimal distortion in the parameterization. It also offers a way to explicitly match landmark curves in anatomical surfaces such as the cortex, providing a surface-based framework to compare anatomy statistically and to generate grids on surfaces for PDE-based signal processing. (NIH) *Figure 16*

CT Registration: Arterial to Venous Phases

Arie Kaufman
ari@cs.stonybrook.edu

Contrast-Enhanced Computed Tomography (CT) is a non-invasive method that allows the evaluation of the enhancement of pancreatic lesions and provides useful findings for differentiating different types and assessing surrounding organs and vessels. For example, standard pancreas-specific imaging protocols include acquiring CT images 20 seconds (arterial phase) and 60 seconds (venous phase) after the contrast injection. Despite obvious benefits, the use of both phases in computer-aided image analysis systems is complicated by the scans being misaligned due to patient movement and anatomic motion between acquisitions, thus requiring costly and often imperfect registration or generating training annotations for images in both phases. This project focuses on estimating CT images of one phase given the corresponding CT image of another phase using machine learning approach of generative adversarial networks (GANs). (Marcus Foundation, NIH-REACH)
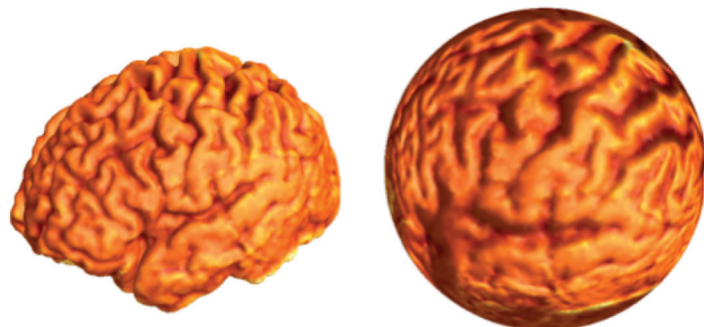
Figure 16

Classification of Pancreatic Cysts in CT Images Using Machine Learning

Arie Kaufman
ari@cs.stonybrook.edu

There are many different types of pancreatic cysts. These range from completely benign to malignant and identifying the exact cyst type can be challenging in clinical practice. In this project, we describe an automatic classification algorithm that classifies the four most common types of pancreatic cysts using computed tomography images. The approach utilizes both the general demographic information about a patient as well as the imaging appearance of the cyst. It is based on a Bayesian combination of the random forest (RF) classifier, which learns subclass-specific demographic, intensity, and shape features, and a new convolutional neural network (CNN) that relies on the fine CT texture information. Quantitative assessment of the proposed method was performed using a 10-fold cross validation on 134 patients and reported a classification accuracy of over 83%. (Marcus Foundation, NIH-REACH)

Visual Analysis of a Computer-Aided Diagnosis System for Pancreatic Lesions

Arie Kaufman
ari@cs.stonybrook.edu

Machine learning is a powerful and effective tool for medical image analysis to perform computer-aided diagnosis (CAD). Having great potential in improving the accuracy of a diagnosis, CAD systems are often analyzed in terms of the final accuracy, leading to a limited understanding of the internal decision process, impossibility to gain insights, and ultimately to skepticism from clinicians. In this project, we present a visual analysis approach to uncover the decision-making process of a CAD system for classifying pancreatic cystic lesions. This CAD algorithm consists of two distinct components: random forest (RF), which classifies a set of predefined features, including demographic features, and a convolutional neural network (CNN), which analyzes radiological features of the lesions. We study the class probabilities generated by the RF and the semantical meaning of the features learned by the CNN. We also use an eye tracker to better understand which radiological features are particularly useful for a radiologist to make a diagnosis and to quantitatively compare with the features that lead the CNN to its final classification decision. Additionally, we evaluate the effects and benefits of supplying the CAD system with a case-based "visual aid" in a second-reader setting. (Marcus Foundation, NIH-REACH)

AnaFe: Visual Analytics of Image-derived Temporal Features – Focusing on the Spleen

Arie Kaufman
ari@cs.stonybrook.edu

In this project, we develop a novel visualization framework, AnaFe, targeted at observing changes in the spleen over time through multiple image-derived features. Accurate monitoring of progressive changes is crucial for diseases that result in enlargement of the organ. Our system is comprised of multiple linked views combining visualization of temporal 3D organ data, related measurements, and features. Thus, it enables the observation of progression and allows for simultaneous comparison within and between the subjects. AnaFe offers insights into the overall distribution of robustly extracted and reproducible quantitative imaging features and their changes within the population, and also enables detailed analysis of individual cases. It performs similarity comparison of temporal series of one subject to all other series in both sick and healthy groups. We demonstrate our system through two use case scenarios on a population of 189 spleen datasets from 68 subjects with various conditions observed over time. (NSF CVDI)

Head-Cave: A Practical Immersive Setup for Virtual Colonoscopy

Arie Kaufman
ari@cs.stonybrook.edu

Virtual colonoscopy (VC) has become an important non-invasive screening process for colorectal cancer in recent medical practice. It is significantly less invasive, less time consuming process for many patients who would otherwise forgo a screening test simply to avoid discomfort. In the United States, thousands of VC tests are performed every year which is expected to increase in the future. The increased adoption of VC demands re-consideration for comfortable user interaction, time efficiency, and reading accuracy of radiologists using this software system.

We are developing an immersive desktop environment called the head-

Healthcare and Biomedical Applications

cave, which provides an immersive visual and interaction paradigm for performing virtual colonoscopy. While commodity hardware such as head mounted displays (HMDs) provide a robust VR platform, adoption in clinical practice is still lacking due to poor display resolutions and increased discomfort with usage over time. Our goal is to provide a practical setup of virtual reality (VR) elements for VC considering clinical work environment and working preferences of expert radiologists who perform these tests multiple times every day. Our setup includes a desktop based immersive environment created through an array of three to four 3D stereo enabled monitors and head tracking. User interaction will be supported through a combination of traditional and modern elements such as mouse/keyboard, touch screen, and hand gesture recognizing sensors. We have been working closely with expert radiologists to optimally map these interaction modes to different user activities in the head-cave VC system. The head-cave setup is anticipated to be useful for many other virtual examination and planning applications, including applications in biology and transportation. (NSF CVDI and ITSC)

Augmented Colonoscopy

Arie Kaufman
ari@cs.stonybrook.edu

In this project, we are seeking to bring together optical colonoscopy (OC) and virtual colonoscopy (VC), both are screening procedures for colorectal cancer. At the moment both procedures are disjoint even if one procedure occurs right after the other. The current task at hand is the transferring images in the OC domain to images in the VC domain and vice versa. The CycleGAN is the first thought that comes to mind for this image to image domain translation task, but it does not cater well to OC. This is due to the fact that the network does an L1 loss between real OC images and the OC images synthesized from VC images. This does not make sense, as the generator is expected to produce images with different textures and reflection patterns, and the direct comparison will hinder the network. We have created a machine learning Extended CycleGAN in order to handle this task. The Extended CycleGAN takes OC images and compares them in the VC domain, as a distance between the VC images reflects the difference in the structures of the images. The creation of VC images from OC can allow for further downstream applications, such as 3D reconstruction from OC. We further investigate applying some of the features of the tempCycleGAN to our Extended CycleGAN in order to also make it work in the temporal domain. (NSF CVDI)

CrowdDeep: Deep Learning from the Crowd for Nuclei Segmentation in Histopathological Images

Arie Kaufman
ari@cs.stonybrook.edu

The development of deep convolutional networks requires access to large quantities of high-quality annotated images for training and evaluation. As image annotation is a tedious task for biomedical experts, recruiting non-expert crowd workers can be an economical and efficient way to provide a rich dataset of annotated images. This work presents CrowdDeep, a novel technique to generate a large dataset of segmented nuclei in a short time-frame using crowd-sourcing. We train non-expert users to annotate nuclei in histopathological slides with the help of our system in an interactive manner. We present an approach to train a convolutional neural network (CNN) to minimize segmentation errors using a mixture of noisy annotated images from non-experts and clean expert-derived annotated images. The initial experimental results show improvements in segmentation accuracy by training the network with a large dataset containing both noisy and clean data, compared to a small dataset of only expert-derived annotations. (NSF CVDI)

Diagnosis and Classification of Cancer from Microscopic Biopsy Images

Arie Kaufman
ari@cs.stonybrook.edu

We develop a deep learning-based system for automatic diagnosis and classification of cancer using microscopic biopsy images. The detection and classification of cancer from microscopic biopsy images are challenging tasks because an image usually contains many clusters and overlapping objects.

Therefore, various steps should be involved in the algorithm including enhancement of microscopic images, normalization of the images, nuclei segmentation, features extraction, cancer detection and finally the classification. Before starting the classification task, we enhance and normalize the microscopic biopsy images, which results in improving the accuracy of nucleus segmentation as well as cancer classification.

After preprocessing the images, we segment the nuclei in the slide using an automatic algorithm which has been trained on a mixture of non-expert and expert annotations. In the second step, we detect the cancerous cells based on the shape and density of cells. Then, we primarily design a neural network to predict if a slide is cancerous or not and it will be extended to predict the type of cancer. Based on the size of the available training set, we apply transfer learning and/or image augmentation. (NSF CVDI)

Segmentation of 3D Neurons in Wide-Field Microscopy Brain Images

Arie Kaufman
ari@cs.stonybrook.edu

In this project, we aim at improving the accuracy of 3D neuron reconstruction and visualization by developing a convolutional neural network (CNN), which segments the neuronal structures in noisy wide-field microscopy brain images. Neurobiologists often prefer wide-field microscopes (WFM), since they generate microscopic images orders of hours faster while they are cheaper in comparison with confocal or electron microscopes. However, WFM images suffer from a degraded contrast between foreground and background voxels due to out-of-focus light swamping the in-focus information, low signal-to-noise ratio, and poor axial resolution. Therefore, separating the neuron voxels from background voxels in WFM images is a challenging task. We are developing a voxel classifier which predicts the probability of every individual voxel in an image being a part of a neuron or not. A major challenge of using CNN on neuronal image segmentation is

the large size of volumetric images, resulting in expensive computations. One general method to overcome the large size of images is performing patch-based prediction by labeling each patch based on its center pixel. However, in this technique, there is a huge amount of redundant computations. Another major challenge of using CNN for neuronal image segmentation is the variation in neuronic structures resulting in the necessity of using large filters for long thin neurons. To overcome these challenges, we design a novel fully convolutional neural network (FCN) using inception learning and residual learning which separates the neuron voxels from background voxels in noisy WFM images. (NSF CVDI)

NeuroConstruct: 3D Registration and Visualization of Neuronal Structures in Wide-Field Microscopy Brain Images

Arie Kaufman
ari@cs.stonybrook.edu

In the past two decades, advances to light microscopy have allowed for the acquisition of high-resolution 3D images of the brain across multiple spatial scales. In collaboration with neurobiologists, we are developing an end-to-end 3D system called NeuroConstruct that enables neurobiologists to register and visualize neuronal morphologies from serially sectioned mouse brain samples. Employing a coarse-to-fine approach, we have developed a novel method that maximizes the morphological continuity of neurite trajectories at the interface between successive sliced sections. To address the limitations of wide-

field microscopy brain images, we introduce a pipeline of simple and novel visualization-driven techniques that eliminates the occlusion and clutter due to out-of-focus blur. We propose a new kind of a distance transform algorithm, called the gradient-based distance transform function. Applying enhancement filters to the computed distance field, we generate an opacity map for the extraction of neurites (axons and dendrites) and cell bodies, from the registered raw data. Additionally, we utilize Stony Brook University Reality Deck, the world's largest immersive gigapixel facility, as a cluster for the processing and visualization of large, high-resolution, microscopy data. We provide neurobiologists with an interactive interface to naturally perform multi-scale exploration of the visualization modes generated using our pipeline. The combination of large scale and high resolution imaging with immersive environment based on the visualization tools provide domain scientists with detailed activity maps of neurological circuits. (NSF CVDI)

The New York State Graduate Medical Education Data Dashboard

Rong Zhao and David Armstrong
rong.zhao@stonybrook.edu

Since 1998, the Center for Health Workforce Studies (CHWS) of the University at Albany has been conducting an annual survey of all residents and fellows completing training in New York State. The goal of this survey is to provide the graduate medical education (GME) community with useful information about the outcomes of training and the demand for new physicians. CEWIT is working with CHWS to develop a database and

Healthcare and Biomedical Applications

web portal to make these valuable data more accessible to stakeholders at GME institutions across the state. This data dashboard application allows authorized users to visualize and download information about their program(s). Users may also view the latest reports based on the New York Resident Exit Survey and update information about their institution. Additionally, administrative users at CHWS can use this software to update institution information, manage users' access to the data, and communicate with institution users. (HRI)

HPSA Technical Assistance and Designation Application Tracking System

Rong Zhao, Robert Martiniano and Nafin Harun
rong.zhao@stonybrook.edu

Under contract with the New York State Department of Health (NYSDOH), the Center for Health Workforce Studies (CHWS) of the University at Albany is responsible for developing Health Professional Shortage Area (HPSA) applications for New York State, providing technical assistance to organizations and individuals around shortage areas, and providing technical assistance to health care professionals interested in state or federal service obligated programs. CEWIT, working with CHWS, has developed a database and web portal to track both HPSA applications and the technical assistance efforts of CHWS staff and to provide status reports to both NYSDOH and the federal Health Resources and Services Administration (HRSA). Based on user feedback collected over the past two years, CEWIT is currently updating the tracking system, enhancing its user experience, and developing

new features such as comprehensive activity logging and analytics and case management capabilities. (HRI)

New York State Health Workforce Planning Data Guide

Rong Zhao and Robert Martiniano
rong.zhao@stonybrook.edu

Working with the Center for Health Workforce Studies (CHWS) of the University at Albany, CEWIT has developed an interactive, web-based Health Workforce Planning Data Guide, designed to support collaborative health workforce planning between health care providers and health care educators among others. This system allows its users to access a wide range of data at various granularities, including population demographics, health status, health outcomes and health professions. Users can visualize the data using maps, graphs, and tables statewide, by region, or by county, compare related data elements, and create customized reports.

Developing this software application provides enhanced visibility to the Data Guide, which is a valuable resource for a wide range of stakeholders, including SUNY schools, New York State Department of Health (NYSDOH), healthcare providers and provider networks across New York State, and local public health departments, among others.

We are currently upgrading the Data Guide by deploying a more robust cloud-based infrastructure updating existing data elements, adding new data elements, enhancing the user experience and providing more versatile visualization and longitudinal analysis capabilities. (SUNY)

New York State Nurse Practitioners Data Dashboard

Rong Zhao and Nafin Harun
rong.zhao@stonybrook.edu

Since 2015, all Nurse Practitioners (NPs) in New York State have been required by law to provide additional information on demographic, education, and practice characteristics to the state at the time of their triennial recertification. As part of the efforts to disseminate information on NPs, the New York State Nurse Practitioners website is being developed in a collaboration between the Center for Health Workforce Studies (CHWS) of the University at Albany and CEWIT. The objective of the New York State Nurse Practitioners Data Dashboard is to provide public access to aggregated information on NPs actively practicing in New York State, making valuable health workforce data available to educators, researchers, providers, planners, policy makers and other stakeholders. The NP website includes counts of NPs by status and practice specially, visualizations and interactive tools that allow users to view the information by different geographies, as well as downloading tools designed to provide access to aggregated data on NPs.

It also provides tools for CHWS staff to track, manage and analyze the data collected from the recertification process. Based on user feedback collected since last year, CHWS and CEWIT are continuously improving the functionality and robustness of this data dashboard, as well as implementing more advanced data management and visualization tools to further enhance the user experience. (HRI)

Aging and Longevity

Controlling Cholinergic Circuits To Curtail Senescence-Related Cognitive Decline

Lorna Role
lorna.role@stonybrook.edu

Loss of acetylcholine (ACh) and reduction in the detectability of multiple cholinergic markers are strongly associated with cognitive decline and age-related dementia. Despite the established role of acetylcholine signaling in memory, none of the currently available cholinergic therapeutics yields better than modest effects in enhancing recall.

In light of recent advances in molecular genetic techniques and the development of novel approaches for selective neural stimulation, it is time to explore new, targeted treatments that go beyond traditional pharmacotherapeutics. Our goals are: 1: To pinpoint the critical mechanisms that underlie senescence-induced cognitive decline, and 2: To develop novel functional interventions that will slow and/or reverse the loss of cognitive acuity characteristic of age-related neurodegeneration.

First, we recognize that the development of innovative approaches to reversing memory loss requires an understanding of cholinergic signaling over a range of functional states - from those with varying degrees of age-related cognitive decline, to cognitively normal healthy adults, to "super agers" (cognitively healthy octogenarians) - using higher resolution functional mapping of the cholinergic system.

By combining state-of-the-art basic science approaches to genetic mapping of memory-engaged cholinergic circuits with the latest in human imaging methods and improved visualization techniques, we propose an unprecedented deep, and truly translational, dive into defining the role of the cholinergic system in this range of cognitive capabilities.

Second, although detailed knowledge into the cholinergic system is essential to our success, the real goal of this proposal is to develop novel molecular tools for the selective activation of individual cholinergic neurons that are critically engaged in the formation and recall of specific memories. This concept is an extension of our recent (unpublished) work that reveals an essential role for cholinergic signaling in the process of memory recall. The next step is to put this knowledge to work by developing molecular switches for selective functional stimulation of memory-specific cholinergic axonal projections.

Our analysis will focus first on the entorhinal cortex (EC), a region known to be strongly affected early-on in Alzheimer's disease (AD). Specifically, we will use the latest structural and functional mapping techniques to delineate the time course and mechanisms that underlie the demise of the circuits between the basal fore-brain cholinergic nuclei (BFCN) and the EC.

In parallel and complimentary studies in humans and in an animal model of accelerated aging, we will assess the relationship between structural and functional integrity of the cholinergic BFCN-EC circuits and cognitive performance. In particular, we will first focus on spatial recognition memory, as it is affected early-on in AD, is controlled by the EC and c an be assessed similarly in rodents and humans.

Interactions Between Aging-Related Protein Processes

Gábor Balázsi
gabor.balazsi@stonybrook.edu

Protein mis-processing and aggregation are common in neurodegenerative diseases related to aging, including Alzheimer's disease (AD). AD patients suffer from two types of protein mis-processing: deposition of the amyloid (Aβ) protein plaques and the aggregation of the micro-tubule associated protein tau filaments. Both processes depend on protein levels. However, the interactions between these two processes are not well understood. There is a need to determine: Can Aβ plaque formation promote tau aggregation and vice versa? And how will these interactions affect aging-related cellular processes?

For this investigation, we propose to introduce the Aβ-encoding gene AB42 and tau-encoding gene TAU under independently tunable, chemically controllable promoters from this laboratory first in yeast and then in mammalian (neuronal) cell lines.

We have developed negative feedback-based "linearizer" synthetic biological tools (synthetic gene circuits) for precise protein level control that have linear inducer dependence of average gene expression; and uniform protein levels (minimal cell-cell variability) at all induction levels.

By independently tuning the levels of the two proteins Aβ and tau, we will investigate their synergistic phenotypic effects (plaque formation, aggregation and cell viability). There have been several different studies that suggest that there is a direct connection between AB42 and GSK3.

Aging and Longevity

Abundant sources suggest that AB42 interacts with RCAN1, which interacts with PPP3CA/Calcineurin, which interacts with Tau. Therefore, the main interaction occurs between AB42/GSK3 and AB42/RCAN1. Since these genes or their equivalents exist in budding yeast, we will start by studies in yeast and then continue in mammalian cell lines.

Figure 17

. . . .

Proteostasis Collapse: a Mechanism of Aging in Cells

Kenn Dill
dill@laufercenter.org

What is the mechanism of declining proteostasis in aging cells? Proteostasis, the homeostasis of the proteomic, is the process by which the cell maintains the folding and disaggregation health of its proteins. The decline and collapse of proteostasis with age and in neurodegenerative diseases is poorly understood. Cellular aging is too general and universal a biological process to be explained by any one particular gene, protein, pathway, regulatory circuit, or cell type, because they are too localized, detailed, specific and deterministic. Proteostasis collapse is a fundamental cell-wide, generic and stochastic process that could be a basic force in aging.

To the matter of how proteomes collapse in aging cells, we bring three significant new academic developments: (i) Raleigh's new high-throughput IMS-MS screening method for finding small-molecule inhibitors of amyloid, (ii) Balazsi's synthetic linearizer gene circuits that can control — through external signals — the expression levels of target genes, such as of chaperons (iii) Dill's proteostasis systems biology (PSB) modeling that joins together protein folding, aggregation and oxidation physics with a large scale and systems-biology model of the complex multi-chaperone circuits of proteostasis. The well-known Gompertz curve shows that the aging rate itself increases with age. So, aging is not simply about cumulative damage; there is positive feedback.

We hypothesize that proteins accumulate oxidative damage, becoming mis-folded and aggregated, adding load to a chaperoning and synthesis system that is itself experiencing degrading effectiveness with age, leading to a vicious spiral that ends catastrophically. Important predictions to be tested in combinations of C. elegans and S cerevisiae are: (A) that up-regulating chaperones can delay the collapse phase of cell aging (but not just of a single chaperone system; their coupling is complex), (B) that the known strong effects of temperature and oxidants on worm aging can be explained by the shifting balance of mis-folded and folded proteins, and (C) that a nonlinear increase in mis-folded and aggregated proteins with age might be the mechanistic origin of some aging phenotypes, and (D) the predicted differential expression of chaperones with age.

. . . .

Identifying, Managing, and Presenting Personalized Connections for Enhanced Cognitive Stimulation

Shmuel Einav
shmuel.einav@stonybrook.edu

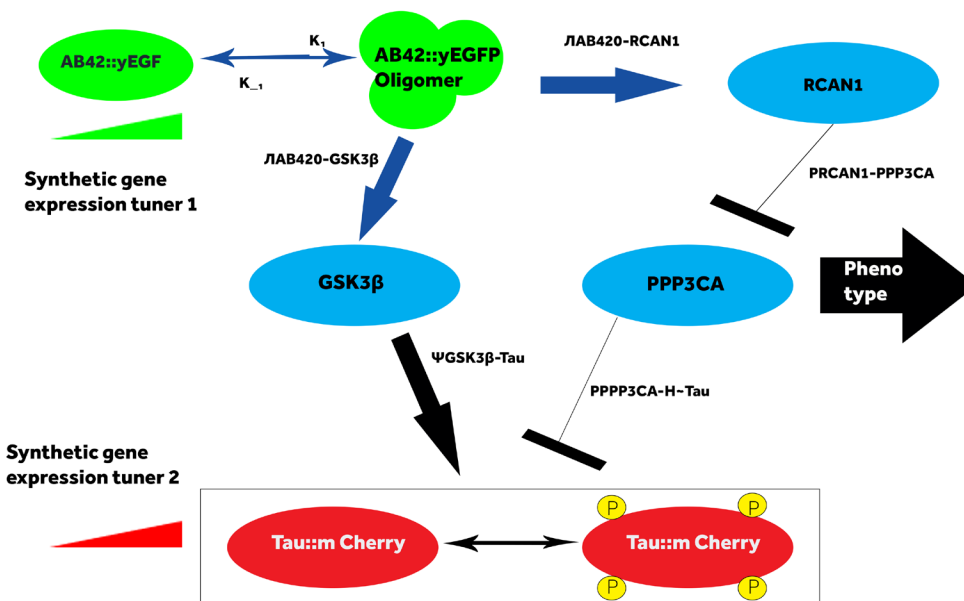Alzheimer's disease is commonly preceded by mild cognitive impairment (MCI) that can progress for years

Figure 17 Using synthetic gene circuits (linearizers) to precisely control the levels of the proteins Aβ and tau, we will investigate plaque and aggregate formation and cell viability.

before becoming diagnostic. In this study, we propose to develop methods for identifying and presenting personalized connections between the user and contents. An interactive multimedia application and an immersive virtual reality environment will then be constructed to offer MCI patients an enhance cognitive stimulation experience.

Our proposed system presents contents in multiple forms, including image, video, audio, text, as well as virtual reality, augmented reality, or mixed reality, based on analysis of the 5W's, namely, when, where, who, what and why. In particular, our study introduces methods to organize the key events of a person's life as time capsules in a gamified cognitive stimulation environment with proper and personalized context, important and familiar actors (i.e., persons of interest), meaningful descriptions of the events, location of the events, and other relevant information that may assist the user in recalling past events of personal interest. A semi-automated tagging process is implemented to associate meta-data with each of these information items. The process begins with the user's selecting of an information item, such as an image or video, followed by identification of the 5 W's and other meta-data, and tagging the information item with such meta-data. Using the meta-data, the

system places the information item into appropriate containers, such as childhood, family, career, education, or hobby. Once an image or video is tagged, the system will employ a learning algorithm to analyze and tag other information items that are considered relevant to the given item. The relevancy may be determined by meta-data, content, or both. This learning process allows the system to iteratively and incrementally identify personal connections between the user and information items collected during the lifetime of the user The user's feedbacks, such as tag selections, browsing sequences, and other interaction patterns, play an important role in this learning process. If the user cannot recall the event or the personal connection to the event, hints (e.g. another photo of the same event, location and time of the event, or other people associated with the event) will be provided. Understanding the user's response patterns and emotional changes allows the system to adjust the delivery of questions and hints and provide a personalized cognitive training experience.

In summary, our proposed study includes the following components: 1) identifying, managing and presenting personalized connections between user and contents in many forms; 2) developing information containers that are meaningful to the user for identifying and managing the

connections; 3) automating the analysis of content structures and semantics using machine learning techniques; 4) constructing user-friendly, interactive applications and virtual reality environments to offer a meaningful, enjoyable, and socially engaging cognitive stimulation experience; and 5) performing usability testing by observing and evaluating user's interactions with the system during cognitive stimulation sessions.

....

Developing Three Dimensional Digital Pathology and Biomarker Data Analytics to Understand Brain Tumor Micro-Environmental Dynamics

Fusheng Wang
fusheng.wang@stonybrook.edu

Glioblastoma (GBM) is the most common and lethal primary brain tumor, with an incidence of about 10,000 new cases per year in the United States and an age-adjusted incidence rate of 3.2 per 100,000 population. It is the most malignant form of glioma causing 3-4% of all cancer-related death. Due to its high invasive properties and its subtle location within the brain substance, complete surgical resection of GBM is impossible, leading to inevitable disease recurrence.

The fatality of this disease is directly attributable to its accelerated rate of radial expansion once the GBM histology is achieved, at which point its growth rate reaches five to ten times that of low grade gliomas.

Regardless of molecular subtype, all diffuse gliomas proceed through this rapid growth phase with a median
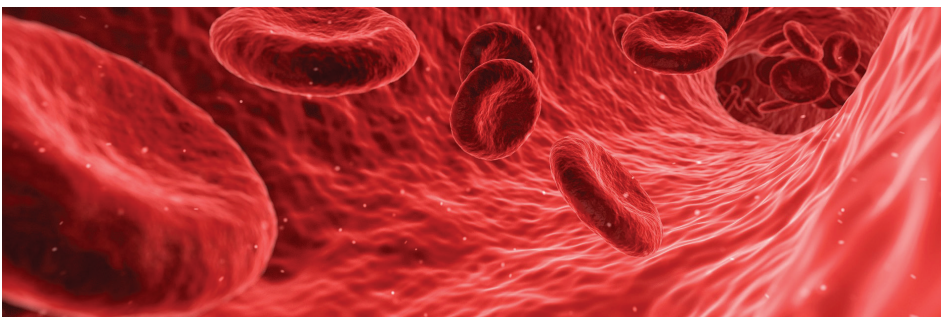
Aging and Longevity

survival of only 12.6 months. The understanding of this explosive growth is limited, hampering the development of novel treatment options directed at underlying mechanisms. Our project advances the study of GBM infiltration mechanisms by developing innovative 3D imaging and spatial analytics methods for histopathology and biomarker images. We will characterize GBM TME spatial patterns in a single three-dimensional (3D) tissue space where both pathological phenotypes and spatially mapped pathophysiological biomarkers will be integrated for a more precise and comprehensive TME analysis. Currently, almost all tissue-based computational cancer investigations are limited to two-dimensional (2D) phenotypic structure analysis and have only modest capacity for processing 3D histology objects. Due to the lack of training data sets of pathology object landmarks and high dependence on a large number of tumor pathophysiological factors, it is challenging to learn mappings from 2D to 3D landmark points, making stereographic methods inapplicable to such analyses. Additionally, we cannot assume vessels are cylindrical objects due to drastic neoplastic vessel degradation.

Similarly, we cannot model necrotic volumes with basic 3D geometry structures, as their shapes are non-linearly determined by insulted vessel shapes, hypoxia density distribution, cell apoptosis rate, and glial cell motility. Thus, we propose to develop scalable and efficient image processing and spatial analytical methods to recover original 3D phenotypic features and spatial topology signatures from the information-lossless 3D tissue space with high resolution serial microscopy slides. We also propose to integrate spatially

registered biomarker data with 3D pathology phenotypic information for more accurate and comprehensive GBM TME study. Our proposed new 3D enabling cancer technologies will accelerate the understanding of tumor TMEs responsible for fatal progression at an extremely high resolution, and incubate new tools to integrate pathology phenotypic features with pathophysiology biomarkers for precision cancer research.

....

Accurate Computational Prediction of Immune Response of MHC/TCR System

Dima Kozakov
dmytro.kozakov@stonybrook.edu

The basic idea of personalized cancer vaccines is to inject the patient with a cocktail of peptides (called neoantigens) containing driver cancer mutations of a patient's own cancer. If those peptides will bind the patients own Major Histocompatibility Complex (MHC) protein, and if that in turn will bind the T-Cell Receptors (TCR) from the patients own repertoire, this will strengthen the patient's organism reaction to already known neoantigens and teach it to target the new ones. The approach is extremely hot right now; see for example recent Nature review.

The specific goal of the proposed project is to develop an improved computational model for prediction of peptide binding to MHC proteins, and the resulting binding of the MHC. peptide complex to the TCRs using machine learning with structural data. This approach can be used for the optimal choice of peptide cocktail for the formulation of the personalized vaccine. The uniqueness of our

approach is that we have developed fast and accurate approaches for protein-peptide and protein-protein docking (best in class), which we will use for the generation of structures of complexes, which in turn can be treated as 3D "images" representing the interaction, to be used as the input for machine learning. Specifically we propose to use Convolutional Neural Nets (CNN) deep learning, a highly effective approach for image recognition. Public databases contain large amounts of data on MHC-peptide and MCH-Peptide-TCR interactions (~0.5M data points) - but we hope as part of the project to generate additional binding data points using a high throughput experimental approach, to further increase the accuracy of the learning algorithm.

If Google/Calico wants to make it collaborative they could e.g. assist by providing computer time and machine learning experts as this is a large scale project, however we feel we already have adequate expertise in our team. Such a method would be of significant interest to industry, and we have the right unique tools to realize it.

In fact I have been recently approached by two Boston startups related to this area to explore such a project using the docking tools developed in my lab - specifically Neon Therapeutics (mentioned in the attached review) and Boston Gene.

....

Computer-Aided Design of Inhibitors Targeting Triple-Mutant ErbB-Family Kinases

Todd Miller and Robert C. Rizzo
todd.miller@stonybrook.edu,
rizzorc@gmail.com

Molecularly targeted therapeutics are a powerful cutting edge paradigm for

the personalized treatment of cancer. The primary goal of this proposal is computer-aided discovery of small molecule inhibitors that target specific mutant proteins involved in breast cancer. In 20-30% of breast cancer there is over-expression or somatic mutation of the human protein epidermal receptor (HER2), also called ErbB2, which makes it a major drug target. ErbB2 belongs to the epidermal growth factor receptor (EGFR) family, in which all four members are receptor tyrosine kinases. Within this family, HER1/EGFR (a main target for non-small-cell lung cancer) has high sequence identity (~78%) and structural similarity to ErbB2. L858R is a somatic mutation in the kinase domain of EGFR that commonly occurs in tumor samples and result in elevated kinase activity and ligand-independent cell growth in cultured cells. A secondary mutation, the T790M "gatekeeper", has been biochemically shown to lead to acquired resistance to reversible inhibitors such as gefitinib and erlotinib that bind the catalytic site (ATP-binding pocket).

Fortunately T790M resistance can be overcome by irreversible inhibitors (e.g. HKI-272,9 afatinib) which covalently bond to residue C797. However, another mutation, C797S, occurs in response to treatment which again causes resistance. Together these three EGFR mutations (L858R/T790M/C797S) depict the painful battle against deregulated activity and drug resistance for patients suffering from cancer. The analogous mutations in the ErbB2 protein are L866R, T798M and C805S. Although these have been less studied than EGFR, the ErbB2 triple mutant is expected to be clinically relevant. We hypothesize the ErbB2 triple mutant will show deregulated

activity compared to wild type and will show resistance to existing reversible inhibitors (e.g. erlotinib) and irreversible inhibitors (e.g. HKI-272). The primary objective is to identify small-molecule inhibitors that specifically target these ErbB-Family mutants and confirm deregulated activity and resistance to known inhibitors. Our specific aims are: 1) Identify small-molecule inhibitors that specifically target ErbB triple mutants using atomic-level computational approaches, and 2) Experimentally quantify kinase activity and inhibition (IC50) of selected compounds against the kinase mutants and develop the most promising leads. The results will be prioritized using different rank-ordering metrics including our new footprint-based scoring criteria with the 200-300 top-scoring compounds being purchased for experimental testing and characterization. As compounds are experimentally verified active, related analogs will be computationally scrutinized, and, as the project progress, refined to improve potency with good drug-like characteristics. Compelling preliminary results obtained by Rizzo and Miller targeting ErbB2 provide important proof-of-concept that the proposed work targeting the important triple mutant has a high probability of success.

. . . .

Machine Learning Methods to Generate Patient Specific Correlations of Multiple Biomarkers

Joel H. Saltz and Ken Shroyer
joel.saltz@stonybrookmedicine.edu, kenneth.shroyer@stonybrookmedicine.edu

Tissue-based biomarker studies are playing increasingly central roles in making cancer diagnoses more precise, steering treatment, and characterizing crucial roles of the nuanced tumor micro-environment. For the most part, pathologists have interpreted tissue-based biomarker studies by utilizing scoring schemes to estimate quantity, distribution, and patterns of staining. The ability to accurately characterize multiplex IHC studies in whole slide images is still very much an open problem that is of great interest from both a clinical and basic research standpoint.

Our group has an active research effort in the development of whole slide imaging tools and analytics. The tools support iterative deep learning algorithm training applied to whole slide cancer image tissue classification, where the method is described in the context of a deep learning based lymphocyte-detection method. In this work, we analyze digitized whole slide images across thirteen TCGA tumor types to relate spatial and molecular tumor immune characterizations for ~5,000 patients. We relate the image analysis derived quantization and pattern of lymphocyte infiltration to results obtained from detailed and highly nuanced molecular studies carried out by the TCGA Pan Cancer Atlas Immune group. This work is a component of a broader set of studies funded by National Cancer Institute (NCI) and by our cancer center to develop core tools and algorithms for tissue histopathologic analytics; integration of clinical, radiologic, and pathologic data; and cancer surveillance with quantitative digital pathology image analysis tools.

The next step in the evolution of our work will leverage our expertise in

Aging and Longevity

digital pathology and deep learning to create and validate machine learning methods to accurately quantify and characterize patterns of multiplex immunohistochemistry (IHC) staining. We aim to create the next generation of computational IHC testing by using quantitative methods to analyze multiplex IHC immunoassays that can detect multiple biomarkers with consistent staining patterns on a single slide simultaneously while preserving tissue, maximizing data, and conserving time.

The overall goal is to create a robust, reproducible, and diagnostically reliable computational pipeline to generate patient-specific correlations of multiple biomarkers with image analysis-based extracted features of malignant cells for each cancer case per patient in order to guide existing treatment options and identify new potential targets for therapeutic intervention during treatment failure and relapse in the coming era of personalized medicine. By combining the power of multiplex IHC with the computational power of machine learning and digital analyses to quantify and correlate the spatial relationships and distribution of these markers, cellular pathways and protein signaling networks can be directly observed and lead to unprecedented insights into cancer.

Early Diagnosis of Pancreatic Cancer

Arie Kaufman
ari@cs.stonybrook.edu

Pancreatic ductal adenocarcinoma (PDAC) is a major cause of cancer deaths and is projected to become the second leading cause by 2020. It has an extremely poor prognosis with an overall five-year survival rate of only 8%.

The elderly are particularly at risk, as most cases of PDAC are encountered in patients of age 60 and older and are asymptomatic until a late stage when treatment is not viable. There is a need for early screening and diagnosis of pancreatic cysts, premalignant lesions of the pancreatic ducts that have the potential to become PDAC, to find the disease at a potentially treatable stage and thus increase the health span of these patients. Our goal is to develop a reliable screening procedure by providing enhanced computational tools using advanced machine learning and computer visualization techniques for the diagnosis of PDAC through inspection of CT scans, the primary imaging modality for pancreas screening. Building upon our preliminary virtual pancreatography (VP) prototype, our proposed work includes three pre-clinical aims and one optional clinical study.

Aim 1: Strategy for pre-screening for pancreatic cancer. Due to the asymptomatic nature of PDAC, biomarkers hold the possibility of providing the first insight for possible diagnosis. Since a single biomarker is insufficient and a combination will likely be required, we propose to build a classifier based not only on K17 but also on other key components of the most aggressive molecular subtype of PDAC. Additionally, abdominal CT data are often acquired for other reasons, and these can be analyzed by our system to provide PDAC screening for such incidental scans.

Aim 2: Automatic segmentation and classification of pancreatic lesions. For a patient selected for additional screening, a specific CT scan of the pancreas will be acquired and analyzed by the proposed system, which consists of three components: segmentation, classification and visualization (Aim 3). We propose to develop a deep learning algorithm for fully automatic segmentation of the pancreas and its lesions, which builds upon our semi-automatic segmentation prototype. We will also build upon our analysis of our current classification system for further refinement.

Aim 3: Visualization tools for diagnosis and following disease progression. Radiologists typically inspect CT scans by tediously scrolling through multiple slices. We will develop a visualization platform to volume render the segmented components in 3D, allowing the clinician to intuitively view the pancreas, lesion(s), and duct from any orientation with or without the surrounding anatomy. We will also planarize the primary duct so it can be viewed within a single reconstructed slice, providing easier and accurate understanding of the duct and lesion relationship. We will also develop techniques for visualizing features of the lesions, such as calcification and thin walls, which will help in classification.

Aim 4: Integration and evaluation of clinical VP system. We will evaluate how our developed clinical VP system fits into the current decision tree for how lesions are worked up and treated. Our system will interface with radiologists' PACS and RIS systems and be integrated into their image interpretation work-flow. We currently have 206 de-identified CT scans with ground truth pathology, and we will acquire additional datasets for a total of over 450 patients.

Internet of Things

Access Control and Management of Enterprise Scale Internet-of-Things

Fan Ye
fan.ye@stonybrook.edu

Internet-of-Things technology turns mundane physical entities (e.g., door locks, thermostats, lights, surveillance cameras) into smart objects that can interact with human users and other devices. Existing solutions largely focus on small size homes. This project explores enterprise scale IoT systems, where heterogeneity (users and devices may have various attributes), scale (more than tens of thousands of users and devices), huge operation amount per day, and frequent user/device churns make home-based solutions inapplicable.

We have designed, prototyped and evaluated solutions for multi-level service discovery, fine-grained access control, and automatic precondition command execution in enterprise IoT systems. Argus is a physical proximity based service discovery system offering 3-level IoT device visibility, public for devices discoverable to everyone; differentiated where the subset/variant of services depends on user attributes; and covert visibility where the subset/variant depends on user attributes that are sensitive and never disclosed in any form to preserve user privacy. Experiments in a testbed show that Argus has one order of magnitude less updating overhead, and it is very responsive, taking less than a second to discover 20 nearby services of different types, agile for satisfactory user experience. Heracles achieves robust, fine-grained access control at enterprise scale.

It adopts a capability-based approach using secure, unforgeable tokens that describe the authorizations of users, to either individual or collections of objects in single or bulk operations. It provides centralized policy and distributed execution desired in enterprise environments. Extensive security analysis and performance evaluation on a testbed prove that Heracles achieves robust, responsive, fine-grained access control in large scale enterprise environments.

Third, we notice that in IoT there are situations where before users can execute commands on IoT devices, certain conditions must be met for sake of safety, correctness or efficiency. E.g., before a fire sprinkler sprays water, the power outlets within the spraying range should be cut off to avoid electric shocks. Thus, a series of other commands need to precede the user commands in a correct order to make those conditions true. We have developed APEX, a system automatically deducing, satisfying all the preconditions of the user commands. It has two strategies: conservative and aggressive. Evaluation on a testbed proves that the conservative strategy sustains high execution success rates despite resource contention, while in realistic enterprise environments, the aggressive strategy may execute significantly faster, saving up to seven seconds and reducing about 46% of conservative strategy's time cost.

Geometric and Topological Analysis for Trajectory Privacy

Jie Gao
Jgao@cs.stonybrook.edu

The maturing of mobile devices and systems provide an unprecedented opportunity to collect a large amount of data about real world human motion at all scales. The rich knowledge contained in these data sets can have a huge impact in many fields ranging from transportation to health care, from civil engineering to energy management, from e-commerce to social networking. While the applications are paradigm-transforming, recent studies show that the trajectory data can raise serious privacy concerns in revealing personally sensitive information such as frequently visited locations or social ties.

These concerns become the major hurdle in utilizing these data sets. This project systematically studies the issue of anonymizing trajectory data, from the bottom layer of trajectory sensing and data collection, to the middle layer of trajectory representation and anonymity, to the application layer of how the anonymized trajectory data can be used. By the nature of trajectories as being time stamped sequence of points, in this project novel geometric and topological algorithms that directly work on distributed sensors collecting the trajectories are developed for achieving the objective. Queries to such decentralized sensors are made to ensure no sensitive information is released. The intellectual contribution lies in the following aspects.

1) The topological representation of trajectories, i.e., how trajectories pass around obstacles and landmarks in the domain is adopted. The topological representation is compact and descriptive, introducing novel discrete and combinatorial problems to study. 2) A novel framework is developed for distributed sensors to directly learn, classify and compare the topological types of the target trajectories, using harmonic one-

forms and Hodge decomposition from algebraic topology. The new framework can substantially reduce the communication cost within the network, while maintaining the requirement of user privacy from the very beginning of sensing and data collection. 3) A family of anonymization algorithms using different ideas are developed, by altering the way to connect the time-stamped points into trajectories, by adjusting the topological resolution to reach a balance between data anonymity and utility, and by sensing and recording randomized hash data to answer popular trajectory queries. 4) The trajectory data sets are often huge, so algorithms for handling large scale trajectory data sets are developed in both centralized and decentralized settings. (NSF)

AC Computing for Wirelessly Powered IoT Devices

Emre Salman
emre.salman@stonybrook.edu

We have developed a charge-recycling based computing paradigm for RF powered applications such as RFIDs, wireless sensor nodes, and future IoT devices. Contrary to existing methods that rely on rectification and regulation, we can directly use the harvested AC signal to power the computational blocks. Our method eliminates the significant power loss due to rectification, particularly at low input power levels. Furthermore, charge-recycling operation is significantly more efficient than conventional static CMOS. Our results demonstrate more than an order of magnitude increase in energy efficiency. Thus, we can provide significant on-site computational capability, which is highly desirable for RF-powered devices. Furthermore and more interestingly, the unprecedented increase in efficiency brings us closer to utilizing ambient RF power, as we are currently investigating. Our method has direct applications in multiple domains such as IoT security, computational RFIDs, and near-field brain implantable devices. (NSF-CPS, Simons Foundation)

Smart Energy and Urban Systems

Institute of Gas Innovation and Technology (I-GIT)

Devinder Mahajan, Christopher Cavanagh (National Grid), Shawn Jones (CSE) devinder.mahajan@stonybrook.edu

A new independent initiative to design and advocate proposed policies that enable or support the development of technologies that expand the value of societal investments in natural gas related infrastructure including, transmission, distribution, data, local renewable production and end-use applications for space conditioning, process, electric generation, vehicular transportation and greenhouse gas mitigation. Founded in 2018, I-GIT is a partnership between Advanced Energy Research and Technology Center (AERTC) and National Grid. The CSE joined in 2019. The first charge of I-GIT is to assess pathways that would lead to decarbonize 80% electricity in New York State by 2050, following the State's road map and National Grid's Northeast 80x50 Pathway. I-GIT is also investigating hydrogen, produced from renewable off-shore wind, injection in natural gas pipelines. Multiple projects are underway under I-GIT umbrella in collaboration with CEWIT.

Power to Gas (P2G) Demonstration on Long Island

Devinder Mahajan, Shawn Jones (CSE), Satya Sharma devinder.mahajan@stonybrook.edu

This is a joint project between the Center for Sustainable Energy (CSE), I-GIT, and CEWIT. As the share of renewables increase, the intermittency of these energy sources must be stabilized in electricity networks to realize their full potential. The power-to-gas (P2G) option offers a pathway to store renewable power from solar, wind or hydro to be converted into renewable hydrogen to directly power fuel cells for electricity or to be further reacted with captured CO_2 to form renewable natural gas (RNG) that could be injected into the pipeline. The potential of P2G to store energy that surpasses battery storage by at least two orders of magnitude, could be realized. The big data management will be provided by CEWIT. The P2G concept is relatively new and regional data is crucial to expand its application in New York State. In Europe, there are 70 operating P2G plants but in the U.S., there are a few test units. This project will demonstrate the integration of a suite of technologies (water electrolyzer, CO_2 extraction system, hydrogen storage system, and a fuel cell) to serve as a test bed for advanced P2G applications. The project also includes identifying optimal combinations to establish baselines for each source of renewable energy (solar, wind, hydro) and educating the public, including international audiences, about the tools and techniques available with the P2G test bed located on Long Island. (I-GIT, CSE)

Off-Grid Wood Waste to Energy (WTE) Demonstration: Long Island Resilient CHP

Devinder Mahajan, John Livermore (CSE), Julia Hasty (All Power Labs) devinder.mahajan@stonybrook.edu

Long Island's geographic location puts the region at risk of losing power during intense storms due to flooding, downed power lines and debris strewn all across the island. Local utility PSEG Long Island's records reflect that the last major windstorm (Super-storm Sandy) to hit Long Island produced approximately 40,000 cubic yards of wood debris and left sections of Long Island without power for up to two weeks. This amount of wood has 420,000 MW-equivalent if converted into power so wood waste-to-energy is the basis of a proposed demonstration project on Long Island. This is a joint project with All-Power Labs (APL), Center for Sustainable Energy (CSE) and National Grid.

At the heart of the project is an innovative commercial mobile power pallet, dubbed "PP30", a 30kW gasifier provided by APL and fueled by waste wood on Long Island. The APL system (PP30™) is a modular unit that could be deployed routinely to avoid landfilling of debris and to provide emergency power. Under the proposed project, one APL 30 kW power pallet will be initially staged at a designated site. Though the PP30 unit has documented success gasifying wood chips in California, the consortium seeks to gather performance, emissions, reliability and afford-ability data using a feedstock mix produced from operations on Long Island. The business model envisions revenues from the sale of power, and to produce a potentially usable solid by-product, biochar. A mobile data collection system to monitor a series of units will be managed by CEWIT.

Institute of Gas Innovation and Technology
AT STONY BROOK UNIVERSITY

Smart Energy and Urban Systems

Development of Smart Grid Energy Management Technologies

Eugene Feinberg
eugene.feinberg@stonybrook.edu

This project consists of three tasks. The first task deals with data analysis, design, and implementation of the advanced metering infrastructure (AMI) data management system to be integrated into a larger energy management system (EMS). The AMI data management system should be capable to support a range of customer service applications including dynamic pricing, demand response, billing, and outage detection. The second task deals with applications of time-series analysis and artificial neural networks to medium and long-term electric load forecasting models. These models have adaptability capabilities for capturing changes in consumption patterns over time. The third task develops fast service restoration algorithms for electric distribution networks. (Energy-IT)

Smart Composites for Energy Harvesting

T. A. Venkatesh
t.venkatesh@stonybrook.edu

Smart piezoelectric materials, by virtue of their coupled electromechanical characteristics, have been recognized for their potential utility in many applications as sensors and actuators, from medical ultrasound devices for prenatal care, micro/nano-positioners for atomic force microscopes and sonar hydrophones to non-destructive testers and ink-jet print heads.

Considerable research efforts in the past years have resulted in the development of several monolithic piezoelectric materials such as lead zirconate titanate (PZT) and barium titanate, with enhanced coupled properties. However, despite the enhancement in their piezoelectric properties, monolithic piezoelectric materials generally exhibit certain limitations. For example, they are mechanically brittle as most of the piezoelectric materials are ceramic-type materials and their functionality is generally unidirectional as the poling characteristics of the piezoelectric material allow them to sense or actuate in one direction (i.e., in the dominant poled direction) only. Because of these limitations, the range of applicability of monolithic piezoelectric materials is limited. A composites' approach to piezoelectric materials can potentially overcome the limitations of monolithic piezoelectric materials. The overall objectives of our research efforts are: (i) To obtain a comprehensive understanding of the fundamental properties of smart piezoelectric composites; and (ii) To design novel smart materials based devices and structures for sensing and actuating functions as well as for energy harvesting applications. (CEWIT)

Energy Choices: Using an Agent-Based Modeling Simulation and Game to Teach Socio Scientific Topics

Lori Scarlatos
Lori.scarlatos@stonybrook.edu

In our modern world, where science, technology and society are tightly interwoven, it is essential that all students be able to evaluate scientific evidence and make informed decisions. Energy Choices, an agent-based simulation with a multi-player game interface, was developed as a learning tool that models the interdependencies between the energy choices that are made, growth in local economies, and climate change on a global scale. We have pilot tested Energy Choices in two different settings, using two different modes of delivery. In our research, we are continuing development of the simulation-to increase the number of parameter choice -and the game-to make it more engaging for student players. We are also investigating the creation of a general interface framework that can be applied to other games built upon socio-scientific agent-based simulations.

Smart Grid Android Manager

Radu Sion
sion@cs.stonybrook.edu

The Mobile Smart Energy Manager is an Android/iPhone based Energy Manager that runs on smart phones and tablets. It will connect to local autonomous micro-grid information gateways and components (such as solar panel relays), the macro-grid operators and utility companies' portals, smart home components, and the internet, to provide real-time energy-related information, debit and consumption data, billing and the ability to manage smart-home devices in real-time on-demand. The Mobile Smart Energy Manager will also allow a unified user control platform for the integration of various external smart grid data processing and visualization plug-ins. Examples of such plug-ins include: (1) Big data analytics

visualization of micro-grid and macro-grid energy data, (2) Connectivity conduit to external data sources (3) Visual devices such as the reality deck display and the SCADA smart grid control center (4) Networking plug-ins to interface with any additional custom wireless protocols designed as part of the SGRID3 project. (DOE)

Laser Scribing Technology for High Efficiency Building Integrated Thin-Film Solar Modules

David J. Hwang
david.hwang@stonybrook.edu

Practical challenge in utilizing renewable solar energy is reducing the cost per watt and simultaneously improving conversion efficiency to compete with current fossil fuel technology. Building-integrated photovoltaic (BIPV) has attracted significant attention due to the extended surfaces available for the PV installation including walls, roofs, facades and windows in buildings. Research goal of this project is to provide viable solutions for inter-cell connection and see-through patterning based on laser scribing technology towards efficient BIPV. For high efficiency, monolithic thin film CIGS-Perovskite tandem modules are pursued, and the BIPV architectures are realized either by fabricating the flexible PV modules to be attached to the arbitrary building surfaces, or by fabricating see-through PV modules to meet the requirements of semi transparency. Key tasks include optimization of laser scribing parameters with minimal dead zone and electrical shunt, and superior module aperture areal efficiency and PV efficiency. Flexible arrangements

compatible with roll to-roll process are developed for high productivity. (Korea Institute of Energy Technology Evaluation and Planning, Korea Institute of Science and Technology).

Hybrid Additive Manufacturing of High-Efficiency Thermoelectric Devices for Waste Heat Recovery

David J. Hwang and Jon Longtin
david.hwang@stonybrook.edu,
jon.longtin@stonybrook.edu

Thermoelectric generators (TEGs) are an attractive means to produce electricity, particular from waste heat applications. Unfortunately, commercial TEGs are only available in flat, rigid formats, and of limited size often requiring manual assembly steps, increasing cost and defect rates. Furthermore, many engineering components are not flat (e.g. exhaust pipes). The emphasis of this study is to develop concepts to fabricate TEG's directly onto exhaust and waste-heat components through various hybrid additive manufacturing routes including (i) combined thermal spray and laser micro-machining, (ii) laser-sintered ink dispensing to directly deposit thermoelectric materials

onto arbitrarily shaped surfaces, and (iii) radiation curing strategy to solve a major manufacturing bottleneck. Our developments can also be extended to other applications, such as electricity power plants (fossil and nuclear), diesel locomotive engines, ship engines and even nuclear safety. This work is also collaboration with UV/EB Curing Center at SUNY-ESF. (NYSERDA, DOE/NSF, DOE NEUP)

In Operando Characterization of Hybrid Layered Semi-Conducting Materials for Photovoltaic and Photodetector Applications

T.A. Venkatesh and Mircea Cotlet
t.venkatesh@stonybrook.edu

This research develops new hybrid nanomaterials based on 0D, 2D semiconductor materials with potential utilization in photodetector, photovoltaic and sensor applications. It focuses on assembly, ex-operando optical characterization and in-operando optoelectronic characterization of field effect transistor devices incorporating such hybrids, with particular reference to inter-facial phenomena.
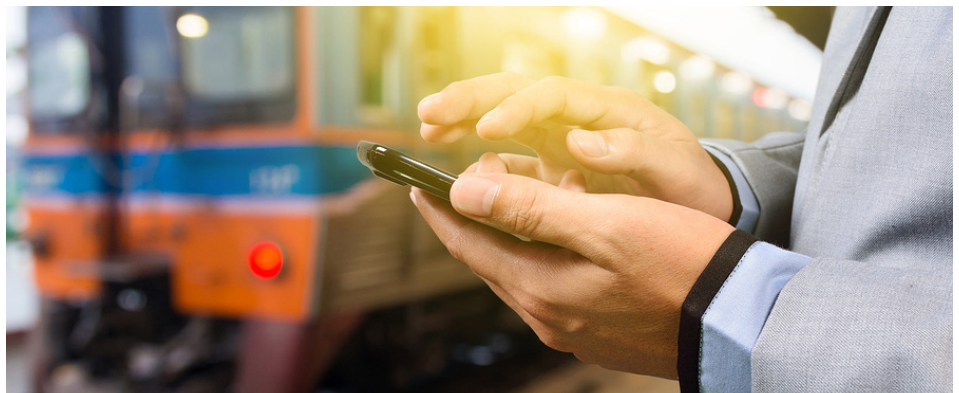
Algorithms and Big Data

Accelerator Control Systems

Thomas Robertazzi and Kevin Brown
thomas.robertazzi@stonybrook.edu
brownk@bnl.gov

Brookhaven National Laboratory, in close proximity to Stony Brook University, supports a number of major accelerator systems. This includes the Relativistic Heavy Ion Collider (RHIC). A large and complex control system is used to operate these facilities. Joint work between Stony Brook's Electrical and Computer Engineering faculty, Computer Science faculty, and BNL's Collider-Accelerator Department seeks to improve the robustness of these control systems. This has taken on two forms. One form is the introduction of a simulation framework which aims to improve the reliability of software in the control system. The second form has been the analysis of fundamental performance bottlenecks in distributed client/server control systems. For example, heavy traffic can cause servers to overload and even crash. In the context of this, several reinforcement-based learning algorithms are proposed to optimize a number of goals.

Computational Geometric Approaches to Geospatial Optimization

Joe Mitchell
Joseph.Mitchell@stonybrook.edu

We consider a variety of optimization problems that arise in facility location, vehicle routing, embedded sensor networks, and other geospatial problems. Of particular interest are problems involving multiple and various vehicles, including mobile robots, drones/UAVs, autonomous

cars, trucks, boats, etc. A specific class of problems is that of optimizing the movement of a set of cooperating heterogeneous vehicles, which may be ground, sea, or air-based, to achieve mission plans, including deliveries, reconnaissance, search and rescue, and military mission plans. (DARPA)

Approximation Algorithms for Geometric Optimization

Joe Mitchell and Esther Arkin
Joseph.Mitchell@stonybrook.edu

We apply the methodologies of computational geometry to design, analyze, implement, and test algorithms for problems that arise in several application areas, including geometric network optimization, air traffic management, sensor networks, robotics, geometric modeling, and manufacturing. The main goal is the development of fundamental advances in approximation algorithms for geometric problems. Additionally, the project will strive to foster and deepen collaborations with researchers and domain experts in application areas and industry, in order to formulate their algorithmic needs precisely and to make available algorithmic tools, insights from theoretical results, and software from experimental investigations.

The specific repertoire of problems includes geometric network optimization (optimal routing and network design in geometric contexts, including TSP variants, vehicle routing, constrained spanning trees, minimum-weight subdivisions, optimal route planning with various constraints, and survivable network design);air traffic management (optimal use of airspace in the face of dynamic and uncertain constraints induced by weather and

traffic congestion, sectorization (load balancing), and optimization of flow management structures for the National Airspace System); and sensor networks and coverage (sensor deployment, localization, data field monitoring, and coverage for stationary or mobile (robotic) sensors). (NSF)

Geometric Networks

Joe Mitchell and Esther Arkin
Joseph.Mitchell@stonybrook.edu

We study geometric networks, which represent interconnections between entities that arise in physical domains or geometric spaces. Networks are all around us and are an important part of the technology in our daily lives. Examples of geometric 43 44 Algorithms networks include wired/wireless communication networks, transportation systems, power grids, sensor networks, and geometric graphs that arise in information visualization. Geometric networks often have special structure that allows their analysis and optimization to be done more efficiently than is possible in general (non-geometric) networks. We study the design and analysis of energy-efficient wireless communication networks. In particular, we investigate wireless communication networks involving directional antennas and/or network improvement methods.

We also study several related optimization problems in geometric networks that arise in other applications, including sensor networks, transportation science, air traffic management, vehicle routing in robotics, covering tours, and exploration/mapping. The primary objective of the project is to develop new algorithmic

solutions to a cohesive collection of geometric optimization problems that fall in the common category of network problems. The goal is to develop theoretically sound, provable methods, but with a strong awareness towards practicality and implementation. (US-Israel Bi-national Science Foundation).

Algorithms in Support of Flight Trajectory Analysis

Joe Mitchell
Joseph.Mitchell@stonybrook.edu

This project applies geometric algorithms to the analysis of large databases of trajectory data for moving agents. Given the exploding amount of time trajectory data collected, there is a need to be able to process, analyze, mine, and understand trajectories that arise from vehicles (cars, buses, trains, aircraft), pedestrians, animals, autonomous vehicles/sensors, etc. Using methods of computational geometry and geometric data structures, we design methods and tools to process trajectory data in the space-time domain, allowing a high level of understanding of structures in patterns and rapid query access to large databases. (Sandia National Labs)

The Visual Causality Analyst: An Interactive Interface for Causal Reasoning

Klaus Mueller
mueller@cs.stonybrook.edu

Uncovering the causal relations that exist among variables in multivariate datasets is one of the ultimate goals in data analytics. Causation is related to correlation but correlation does not imply causation. While a number of casual discovery algorithms have been devised that eliminate spurious correlations from a network, there are no guarantees that all of the inferred causations are indeed true. Hence, bringing a domain expert into the casual reasoning loop can be of great benefit in identifying erroneous casual relationships suggested by the discovery algorithm. To address this need we have devised the Visual Causal Analyst – a novel visual causal reasoning framework that allows users to apply their expertise, verify and edit causal links, and collaborate with the causal discovery algorithm to identify a valid causal network. It empowers users to gain a good understanding of the landscape of causal structures particularly when the number of variables is large. Our framework is also novel in that it can handle both numerical and categorical variables within one unified model and return plausible results.

D-BIAS: A Human in the Loop Methodology for Algorithmic Bias Assessment and Mitigation

Klaus Mueller
mueller@cs.stonybrook.edu

Algorithmic decision making (ADM) is becoming omnipresent as a tool to guide professionals in making decisions in a wide spectrum of applications, such as hiring, admissions, social care, law enforcement, and others. ADM is based on observational data and a set of algorithms that operate on them. Initially conceived as a mechanism to eliminate human bias from a decision process, there is an increasing recognition that ADM is also not without bias, mostly due to the data. As a result, people can be treated unfairly due to their presence in a certain group, or even as an individual. Bias in the data relates to societal constructs, and algorithmic techniques cannot be expected to understand these complicated relationships. We propose a visual analytics approach that leverages human understanding to manipulate data and mitigate the effects of bias. We use causal analysis and correlation to identify sources of bias and debias it. Our visual tool identifies semantic relations between the attributes of the data, and it uses them to aid the decision maker (DM) in understanding the factors in the dataset that are contributing to the bias. The DM can then use his or her domain knowledge and institutional goals to make alterations to the bias reduction scheme such that it fits with the ground reality.

An Optimally-Weighted Aggregating SGD Method for Scalable Parallelization

Yuefan Deng, Zeyang Ye and Pengzhan Guo
Yuefan.deng@stonybrook.edu

We investigate the stochastic optimization problem for deep learning to enable a scalable parallel computing algorithm. The key of the study involves reformatting the objective function for the stochastic optimization by proposing a new update rule by weighted aggregating stochastic gradient decent (SGD), instead of using the center variable. The new rule introduces a weighted aggregation scheme based on the performance of local workers to determine their relative contributions.

Algorithms and Big Data

Besides speeding up convergence, this new rule enables parallelization. We explore the parameters for optimal weighting.

For benchmark, we conduct the experiments based on our new scheme against those of mainstream algorithms including the elastic averaging SGD in training deep neural networks for classification tasks. The current results confirm the superiority of our scheme in accelerating the training of deep architecture and scalable parallelization. Our scheme can be generalized to all kinds of objective functions in the training process of more complex deep neural networks. This research project is expected to significantly improve the SGD efficiency in parallel computing by introducing a smarter and more scalable scheme design.

Theory and Algorithms for Discrete Curvatures on Network Data

Jie Gao
Jgao@cs.stonybrook.edu

New developments in technologies of embedded systems, sensors, and wireless communications provide great potential to improve the safety and security of the physical and social environment we live in. These technologies can help identify and mitigate unfortunate accidents, emergency events, and malicious attacks. This project seeks to develop mathematical tools and algorithms based on discrete curvatures for the purpose of understanding and detecting community structures and anomalies in networks that can be of crucial value in many applications. The project considers high level mobility patterns, community structures, and anomalies as well as finer details such as who is where. The mathematical tools to be developed will be useful in other networks (for example, protein-protein interactions in biological networks).

This project will investigate mathematical problems arising the analysis of real-time spatial and temporal human mobility data. The focus will be on the community detection problem on graphs by using discrete Ricci curvatures and discrete curvature flows on graphs. The problem is to extract stable groups in human mobility patterns, which will serve as the traffic norm for detecting abnormal patterns that can be tied to criminal or terroristic events. To detect these stable groups, or communities, the main observation is that community structures in a network resemble well known geometric phenomena such as thick-thin decompositions in Riemannian geometry. Inspired by Riemannian geometry and the success of Hamilton-Perelman's Ricci flow program, this work investigates how to use discrete curvatures and discrete curvature flows to detect community structure in a network. Preliminary investigations show that the proposed method has great potential and can detect communities with high accuracy. This potential prompts PIs to examine computationally feasible definitions of discrete Ricci curvatures on weighted networks. The important work of Ollivier on discrete Ricci curvature is the starting point of this investigation. The drawback of Ollivier's curvature is that it is computationally expensive - almost impossible to compute the proposed discrete curvature flow on large networks containing more than a million nodes. As such, the main task in this work is to find computationally feasible Ricci curvatures where the discrete curvature flow can be computed in real time for large networks. The affirmative resolution of this work will be useful in pure mathematical research and computer science. The work will also develop software for practical use. (NSF)

Algorithm Diversity for Resilient Systems

Scott D. Stoller and Y. Annie Liu
Stoller@cs.stonybrook.edu,
liu@cs.stonybrook.edu

Diversity can significantly increase the resilience of systems, by reducing the prevalence of shared vulnerabilities and making vulnerabilities harder to exploit. Work on software diversity for security typically creates variants of a program using low-level code transformations. This project is the first to study algorithm diversity for resilience. Our approach to creating algorithm diversity is to start from a high-level executable specification and generate different algorithms that satisfy it. This builds on our extensive prior work on a systematic approach to generating efficient implementation from specifications, based on high-level invariants and the fundamental principle of incremental computation.

To measure the benefits of algorithm diversity, we are also formulating diversity metrics and implementing them for Python and DistAlgo. These include static metrics that quantify differences between the code of two programs, and dynamic metrics that quantify differences between specific aspects of the runtime behavior of two programs. Experiments with centralized algorithms and distributed algorithms confirm that algorithm diversity can achieve more diversity

than low-level code diversity, and the two together can achieve even more. (ONR)

Autoregressive Distributed Lag Model with Time Varying Parameters for Time Series Analysis

Wei Zhu
Wei.zhu@stonybrook.edu

The Autoregressive Distributed Lag (ADL) model has been used for decades to model the relationship between economic and financial variables. The popularity of the ADL model stems from its threefold advantage over other time series models. In this project, we extend the ADL model by allowing changes in both inter-variable relationships and innovation volatilities. The proposed time-varying ADL model improves the forecasting performance by revealing the variations of model parameters and exploiting the most up-to-date state of the underlying dynamic system. The time-varying ADL model is more adaptable to financial time series that commonly features structural instability especially when undergoing unexpected shocks such as a recession or policy change. In one application, we have adapted the proposed ADL model to CCAR (Comprehensive Capital Analysis and Review) stress testing - to generate the conditional forecasts of Pre-Provision Net Revenue (PPNR) of financial holding companies with large assets. In another application, we have examined the utility of time-varying ADL model for investment stock portfolio selection. These real world applications confirmed the advantages of the proposed new model. Currently, we are extending this model to other

research directions, in particular, to examine the utility of this model in predicting the trend and patterns of our K-12 education system such as teacher retention and student performance. This project is partly funded by NSF.

Novel Data-adaptive Analytics for Manifold

Hong Qin
qin@cs.stonybrook.edu

This project's overarching objective is to trail-blaze a data-adaptive theory for manifold information modeling and analysis (inspired by Hilbert-Huang transform (HHT) on one-dimensional signals), and to apply this new theory to manifold geometry/texture/appearance analysis, synthesis, and visualization, with a unique emphasis on data-driven analytics and informatics. HHT has exhibited its initial success in handling non-linear and non-stationary time series in one-dimensional space. It comprises empirical mode decomposition (EMD) and Hilbert spectra analysis. Despite its growing popularity in many scientific and engineering fields, such as geophysics, marine science, and climate studies, technical challenges and unsolved research issues still prevail when trying to bridge the large gap between HHT and data-adaptive space-frequency analytics for manifold data. Additionally, Hilbert spectra analysis (based on instantaneous frequencies, local amplitude, and local phase at each point) remains an open research problem for data processing and analysis on manifold. If successful, this research on the EMD and Hilbert spectra computation on manifold will make significant contributions towards completely data-adaptive space-frequency analytics. Beyond

the state-of-the-knowledge in visual informatics, the research outcomes of this project are expected to contribute to information processing and analytics of other data types/domains such as scalar/vector/tensor field in 3D volume, 4D time-varying medical data, higher-dimensional data with curved and regular structure, or graphs of irregular structure. It is evident that, the big data challenges call for a paradigm shift from model-centric methodologies (e.g., Fourier analysis, wavelet analysis) to data-driven, fully-adaptive approaches for analyzing geometric, non-geometric, volumetric, and higher-dimensional datasets relevant to all scientific fields wherever numerics are prevalent. Broader participation and penetration would include applied mathematics, physical sciences, earth and space sciences, engineering, digital medicine, and communications.

ROC Random Forest for Classification of Imbalanced Data

Wei Zhu
Wei.zhu@stonybrook.edu

The imbalanced class problem in classification is highly relevant in many realistic scenarios such as the detection of a disease or the prediction of bankruptcy. One solution is to design specific algorithms incorporating the unbalanced classes in the training process of a classifier. In our FinTech Laboratory, we have developed the multi-class classification tree based on the area under the ROC curve (AUC) to resolve the imbalanced classification problem. More recently, we have extended the ROC tree to the ROC random forest by utilizing the volume under surface (VUS). Currently, we are working

Algorithms and Big Data

on better incorporating categorical variables into our prediction models, and to examine the performance of this novel class of classifiers for customer screening per loan applications. This research has been funded by a SPIR grant.

Control and Sequential Optimization of Stochastic Systems

Eugene A. Feinberg
Eugene.feinberg@stonybrook.edu

This project studies the structure of optimal and nearly optimal policies for control and sequential optimization of stochastic systems and develops algorithms for finding such policies. It also provides complexity estimations for developed and previously known algorithms to select the best ones. The project investigates the systems with complete and incomplete state observations. For problems with incomplete state observations, the important question is whether their belief transition probabilities continuously depend on belief distributions and actions. In terms of applications, this project deals with inventory control and supply management problems. (NSF)

Efficient Correction for EM Connectomics with Skeletal Representation

Arie Kaufman
ari@cs.stonybrook.edu

Machine vision techniques for automatic neuron reconstruction from electron microscopy (EM) volumes have made tremendous advances in recent years. Nonetheless, large-scale reconstruction from teravoxels of EM volumes retains both under- and over-segmentation errors. In this paper, we present an efficient correction algorithm for EM neuron reconstruction. Each region in a 3D segmentation is represented by its skeleton. We employ machine learning technique of deep convolutional networks to detect and correct false merge and split errors at the joints and endpoints of the skeletal representation. Our algorithm can achieve the same or close accuracy of the state-of-the-art error correction algorithm by querying only at a tiny fraction of the volume. A reduction of the search space by several orders of magnitude enables our approach to be scalable for terabyte or petabyte scale neuron reconstruction. (NSF CVDI)

Design Eco-system for Robot Motion Designs

Anurag Purwar
Anurag.purwar@stonybrook.edu

In this project, we are developing algorithms and a mobile app for robot motion design using planar linkage-based mechanisms. This is part of an NSF STTR phase I project to develop a low-cost, open, and design-driven robotics education framework called Robotics Launchpad for K-12 schools, colleges, and makers and hobbyists.

The U.S. Department of Labor estimates that by 2018, 2.5 million STEM jobs will go unfilled, which at an average pay of $85,000 per year for such jobs translates to $200 billion in lost wages. Robotics education, which brings several STEM topics together, has a great potential to create a workforce that is prepared to tackle the technical challenges of the 21st century.

However, the robotics education in schools is coding-and manual-driven and does not provide significant design experience. The mobile app will provide unique capabilities to help students design robot motions using linkage mechanisms; see Figure below as an example of how the app can be used to design a linkage mechanism

(a)   (b)   (c)   (d)
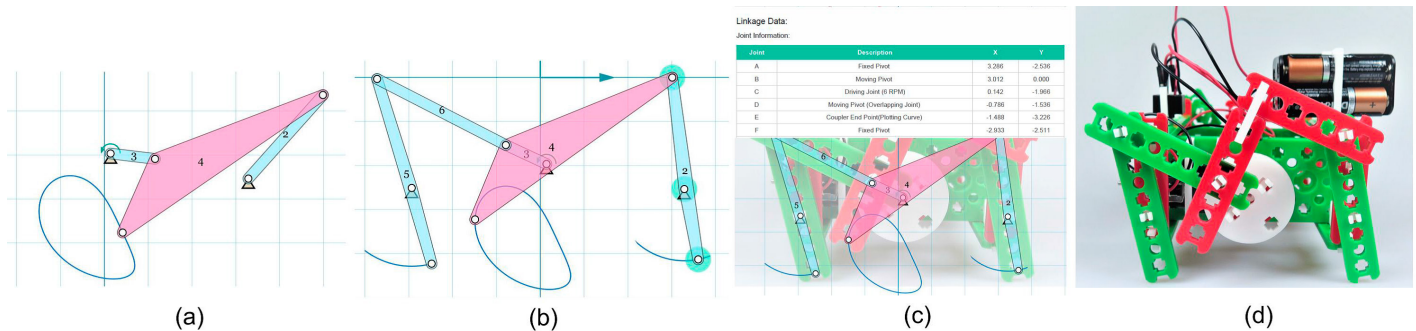
Figure 18 An example of design innovation cycle: (a) design a four-bar linkage where the coupler point traces the path of foot point during walking, (b) extend it to a six-bar for additional support for the front and rear legs, (c) export dimensions from the app to assemble the robot using the hardware kit, and (d) attach motors and wire them using a battery pack to complete the robot. Once the robot is complete, students can take a picture of the linkage; import it in the app and draw a linkage over it to validate their design. See the robot in action at http://youtu.be/jU7hFrNwaxo. More on Robotics Launchpad: http://www.snappyxo.com

for a walking robot and then physically prototyped using the hardware kit. This combination of an easy to use hardware and intuitive motion design app sets this framework apart from any kits in the market.

Figure 18

....

Fast and Accurate Clustering Algorithm for Counting Bar-code Sequences

Song Wu
Song.wu@stonybrook.edu

Bar-code sequencing (bar-seq) is a high-throughput, and cost effective method to assay large numbers of lineages or genotypes in complex cell pools. Because of its advantages, bar-seq use is quickly growing – from using neutral random bar codes to study the evolution of microbes or cancer, to using pseudo-bar codes, such as shRNAs, sgRNAs, or transpose insertion libraries, to simultaneously screen large numbers of cell perturbations. However, the computational pipelines for bar-seq have not been well developed. Available methods, that use prior information and/or simple brute-force comparisons, are slow and often result in over-clustering artifacts that group distinct bar codes together. In this project, we developed a tool called Bartender, which is an ultrafast and accurate clustering algorithm to detect bar codes and their abundances from raw next-generation sequencing data. To improve speed and reduce unnecessary pairwise comparisons, Bartender employs a divide-and-conquer strategy that intelligently sorts bar-code reads into distinct bins before performing comparisons.

To improve accuracy and reduce over-clustering artifacts, Bartender employs a modified two-sample proportion test that uses information on both the cluster sequence distances and cluster sizes to make merging decisions. Additionally, Bartender includes a "multiple time point" mode, which matches bar-code clusters between different clustering runs for seamless handling of time course data. For both simulated and real data, Bartender clusters millions of unique bar codes in a few minutes at high accuracy (>99.9%), and is ~100-fold faster than previous methods. Bartender is a set of simple-to-use command line tools that can be performed on a laptop.

....

CAREER: A Unified Framework for Designing Efficient Resource-Oblivious Parallel Algorithms

Rezaul Chowdhury
rezaul@cs.stonybrook.edu

This project will develop fundamental theory and efficient tools to facilitate the design of parallel algorithms that make no mention of any hardware parameters in the code, but still run efficiently across a wide spectrum of parallel computing platforms ranging from small laptop computers to gigantic supercomputers. These algorithms will be called resource-oblivious algorithms

These algorithms will enable efficient parallel code regardless of the ever-changing underlying hardware platforms allowing more focus on the correctness of implementations without the need of optimizing for a particular hardware. Porting code from one machine to another will be reduced or eliminated. In short, algorithms will be implemented once and continue to run efficiently. One

way of designing efficient resource-oblivious parallel algorithms for the multilevel cache-hierarchy of a multi-core machine is to reverse the direction of information flow between the program and the hardware — instead of making choices based on hardware parameters, such as cache sizes and number of cores, the program now simply assists scheduling by telling the scheduler about the properties of the algorithm being ran. This project will extend the notion of resource-obliviousness to networks of hybrid compute nodes containing both multi-core processors and many core co-processors.

A unified framework will be built bottom-up starting with programs that run solely on processors (stage 1) or mainly on co-processors (stage 2) of a single node, followed by hybrid programs utilizing both processors and co-processors of the same node (stage 3), and ending with programs for networks of hybrid nodes (stage 4).

Each stage will have four outcomes.

- Algorithmic models of resources separately for evaluation and for execution of algorithms.

- Efficient resource-oblivious algorithms for a suite of representative problems with theoretical bounds on the evaluation model.

- Schedulers that guarantee practical program performance predicted by the evaluation model.

- Implementation and experimental evaluation of the schedulers and the algorithms on real machines.

This research will make a wide variety of computational science applications easier to develop and maintain. The research results will be disseminated through new and existing courses on analysis of algorithms, parallel

Algorithms and Big Data

programming and supercomputing using resources made available through the NSF-funded XSEDE program, as well as workshops targeting both computer science and computational science audiences.

··

Center for Visual and Decision Informatics (CVDI): A National Science Foundation Industry/ University Cooperative Research Center

Arie Kaufman and Rong Zhao
ari@cs.stonybrook.edu
rong.zhao@stonybrook.edu

The Center for Visual and Decision Informatics (CVDI), an NSF supported Industry/ University Cooperative Research Center (I/UCRC) established in 2012, works in partnership with government, industry, and academia to develop the next-generation data visualization, advanced analytics, and machine learning tools and techniques that enable decision-makers to significantly improve the way their organization's information is organized and interpreted. CVDI is a multi-institution center including Stony Brook University, University of Virginia, Drexel University, University of Louisiana at Lafayette, University of North Carolina at Charlotte, and Tampere University in Finland. Its mission is to unleash the transformative potential of data science in a wide range of application domains, such as information technology, healthcare, pharmaceutical, biotechnology, commerce, retail, finance, insurance, media, entertainment, transportation, logistics, manufacturing, defense, security, education, and public administration. Our research is strategically focused on design and evaluation methods, algorithms, architectures, software, visualization techniques, mathematical and statistical foundations, and benchmarking of complex systems that facilitate large-scale data analytics, visualization and decision support. The ultimate goal of the CVDI is to develop new technologies that can be applied by its industry partners to create value across a wide range of sectors. Since its inception, CVDI has attracted over 9.9 million funding for its RandD activities and completed more than 85 projects. The center has also produced more than 160 publications, seven patent applications and six licensing agreements based on its RandD projects. For more information please visit cvdi.stonybrook.edu. (NSF)
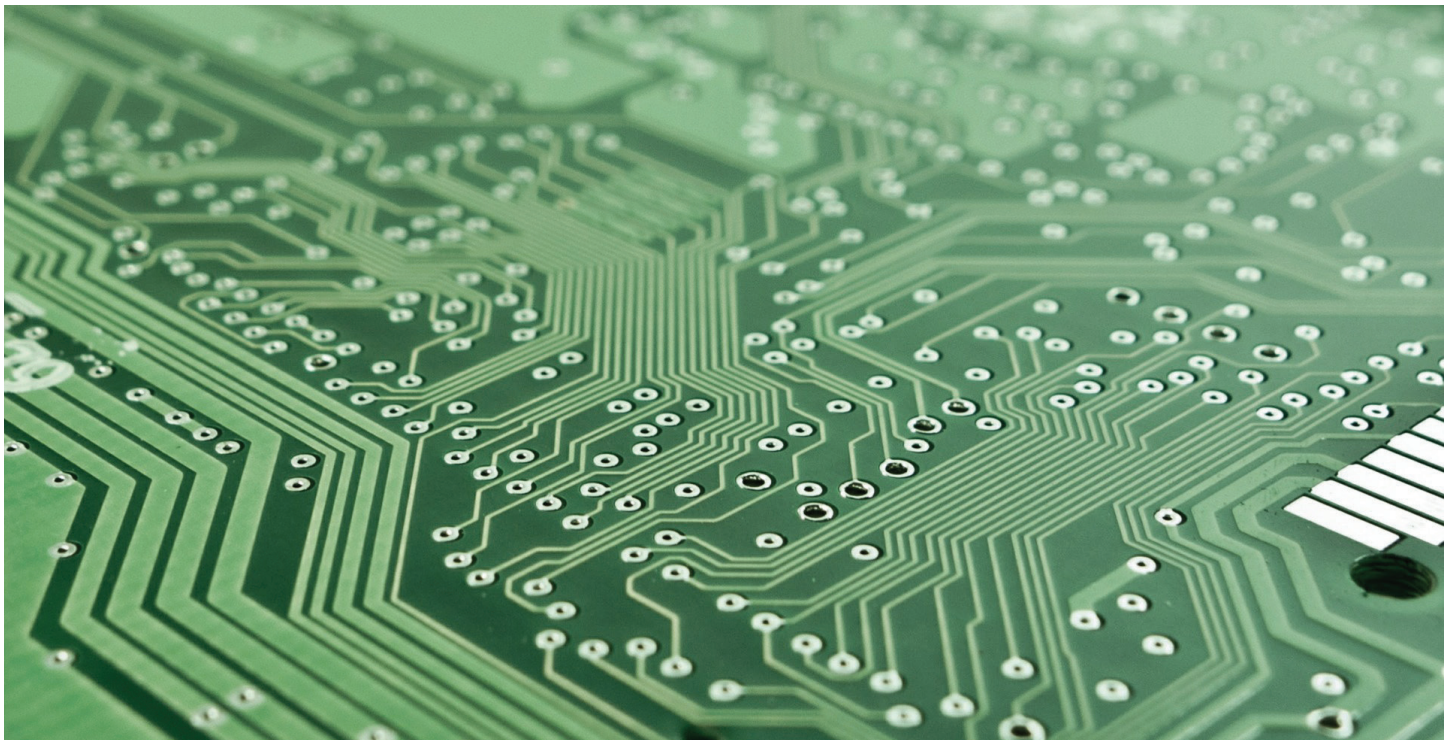
Machine Learning

Supervised Machine Learning for Catalyst Characterization and Design

Anatoly I. Frenkel
anatoly.frenkel@stonybrook.edu

Rapid growth of methods emerging in the last decade for synthesis of "designer" catalysts – ranging from the size and shape-selected nanoparticles to mass-selected clusters, to precisely engineered bimetallic surfaces, to single site and pair site catalysts - opened new opportunities for tailoring the catalyst structure for the desired activity and selectivity. It also sharpened the need for developing new approaches to the operando characterization, ones that identify the catalytic active sites and follow their evolutions in reaction conditions. Commonly used methods for determination of the activity descriptors in this new generation of catalysts, based on the correlation between the changes in catalyst performance and its structural and electronic properties, are hampered by the paucity of experimental techniques that can detect such properties with high accuracy and in reaction conditions. Out of many such techniques, X-ray absorption spectroscopy (XAS) stands out as an element-specific method that is very sensitive to the local geometric and electronic properties of the metal atoms and their surroundings and, therefore, is able to track catalyst structure modifications in operando conditions.

Despite the vast amount of structure-specific information (such as, e.g., radial distribution function of neighbors to selected atomic species) stored in the XAS data, extracting it from the spectra is not straightforward, especially in the conditions of low metal weight loading and harsh reaction environment. Recent developments in XAS data analysis achieved by employing supervised machine learning (SML) methods for structural characterization of catalysts enable the "inversion" of experimental spectra and extracting structural descriptors.

The structure of the catalyst can be tracked in real time and in reaction conditions, as shown in the cartoon. New opportunities for catalysis research that the SML methods enable, such as high throughput data analysis, and their applications to other experimental probes of catalyst structure, are now becoming possible.

Figure 19

Divisible Loads: Scheduling and Performance

Thomas Robertazzi
thomas.robertazzi@stonybrook.edu

Work is continuing on developing theory and applications for processing and communicating very large partition-able loads on parallel processors in a minimal amount of time and/or with a minimal amount of monetary cost. Divisible/partition-able computing and communication loads are loads that can be arbitrarily divided for distribution and processing among a networked collection of processors.

One seeks to optimally divide the load in a time and/or cost optimal manner among the processors and links in a scheduled manner. Not every computing/communication load is divisible but divisible loads represent an important class of loads. Divisible loads arise in such fields as image processing, big data, data intensive computing and embarrassingly parallel problems.

Work that started at Stony Brook in 1988 has developed, through the efforts of researchers throughout the world, a tractable and elegant analytical theory of divisible load scheduling optimization that has relations to electric circuit theory and queuing theory.

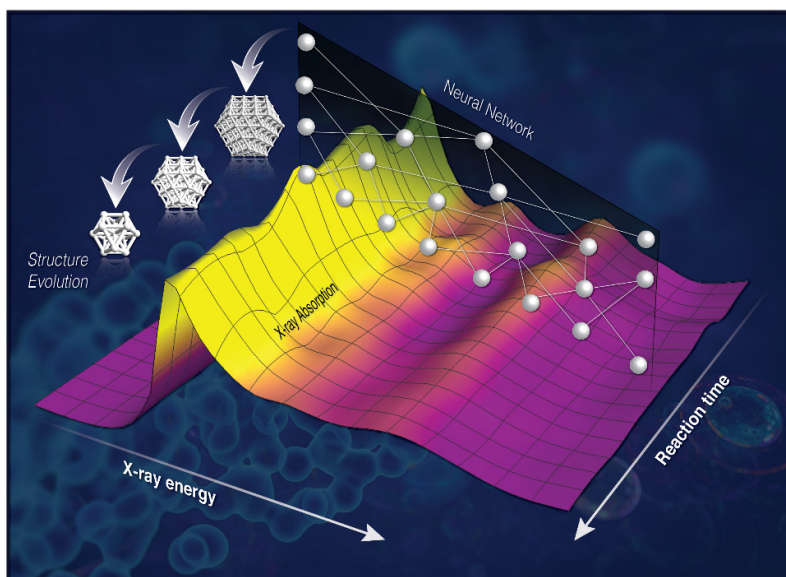Recent work at Stony Brook, includes signature searching and computer

Figure 19

Machine Learning

performance models. In signature searching one has a networked collection of processors search through very large amounts of data for data patterns of interest (i.e. signatures) that indicate the presence of an event or object of interest. Signature searching has applications to sensors, geophysical exploration, network security, bio-medicine and image processing. The relationship between important computer performance models, such as Amdahl's law, and divisible load scheduling is also being explored.

In an age of big data, divisible loads are likely to be of interest for some time.

The Data Context Map: Fusing Data and Attributes into a Unified Display

Klaus Mueller
mueller@cs.stonybrook.edu

Common visualization methods for multivariate data make it difficult if not impossible to observe the data points in the context of the attributes. Our data context map allows these types of comprehensive layouts. We achieve it by combining two similarity matrices typically used in isolation – the matrix encoding the similarity of the attributes and the matrix encoding the similarity of the data points. The resulting layout places the data objects in direct context of the attributes and hence we call it the data context map. It allows users to simultaneously appreciate (1) the similarity of data objects, (2) the similarity of attributes in the specific scope of the collection of data objects, and (3) the relationships of data objects with attributes and vice versa. The contextual layout also allows data regions to be segmented and labeled based on the locations of the attributes.

This enables, for example, the map's application in selection tasks where users seek to identify one or more data objects that best fit a certain configuration of factors, using the map to visually balance the trade offs.

Figure 20

The Intelligent Dashboard

Klaus Mueller
mueller@cs.stonybrook.edu

Creating effective dashboards for the visual analysis of multivariate data can be a tedious task. Our Intelligent Dashboard is an automated dashboard building system that can be used to design effective dashboards with ease. It first analyzes widely available text corpora to identify the most relevant analysis tasks in the underlying data's subject space. Following, the system looks for interesting statistical patterns between attributes in the multivariate data and assigns a score to them. Next, a machine learning model maps the data and their attributes to the most appropriate data analysis tasks. Finally, the dashboard is assembled by putting all of these together into a consistent ensemble of visualizations for good storytelling. Our system also learns generalizable visualization domain knowledge from these analyses to extend its capabilities to other data subjects without the need for the primary analysis steps.

Indoor Floor Mapping and Navigation Services

Fan Fe
Fan.ye@stonybrook.edu

Lacking of floor plans is a fundamental obstacle to ubiquitous indoor location-based services (LBS) that show people where they are and provide step by step guidance to reach desired destinations indoors. People spend more than 80% of daily life in the indoor environment. Thus a scalable, low cost yet accurate solution for indoor floor mapping is critically needed.

We have developed a number of solutions that leverage smart-phones, progressively cut down the amount of data and user efforts, improve the accuracy for floor maps. BatMapper is the state-of-the-art, which uses acoustic signals and analyzes the echoes bouncing off objects such as walls, doors to construct maps with just a few minutes of data.
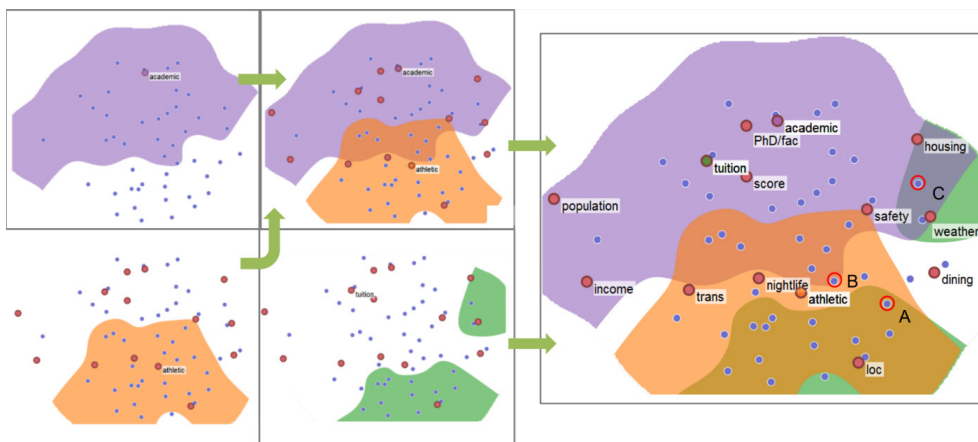

Figure 20

Experiments in real buildings show BatMapper achieves 1–2cm distance accuracy in ranges up around 4m; a two to three minute walk generates fine grained corridor shapes, detects doors at 92% precision and 1 2m location error at 90-percentile; and tens of seconds of measurement gestures produce room geometry with errors < 0.3m at 80-percentile, at one to two orders of magnitude less data amounts and user efforts than prior work.

The system is undergoing commercial pilot tests and expected to be more widely applicable to many business scenarios.

ICE: An Interactive Configuration Explorer for High Dimensional Parameter Spaces

Klaus Mueller and Erez Zadok
mueller@cs.stonybrook.edu

There are many applications where users seek to explore the impact of the settings of several categorical variables with respect to one dependent numerical variable. For example, a computer systems analyst might want to study how the type of file system or storage device affects system performance.

We devised a novel approach, the Interactive Configuration Explorer (ICE), which directly addresses the need of analysts to learn how the dependent numerical variable is affected by the parameter settings given multiple optimization objectives. No information is lost as ICE shows the complete distribution and statistics of the dependent variable in context with each categorical variable.

Analysts can interactively filter the variables to optimize for certain goals such as achieving a system with maximum performance, low variance, etc.

Our system was developed in tight collaboration with a group of systems performance researchers.

Figure 21

A Data-Centric Wireless Communication Paradigm

Fan Fe
Fan.ye@stonybrook.edu

Edge environments like connected/autonomous vehicles are increasingly instrumented with more quantities and varieties of high data rate sensors (e.g., cameras, lidars, radars). Emerging futuristic applications (e.g., see-through augmented reality, video streaming among vehicles) require many-to-many sharing of sensing data among peer vehicles at extreme throughput and scale. Existing short range wireless communication technologies (e.g., 802.11 and 802.11p/DSRC) focus on conventional address-based, point-to-point communication, incurring high complexity and inefficiency for many-to-many data sharing. This project seeks to design, prototype, and evaluate a translational wireless communication infrastructure that enables vehicular data sharing at extreme throughput and scale. It
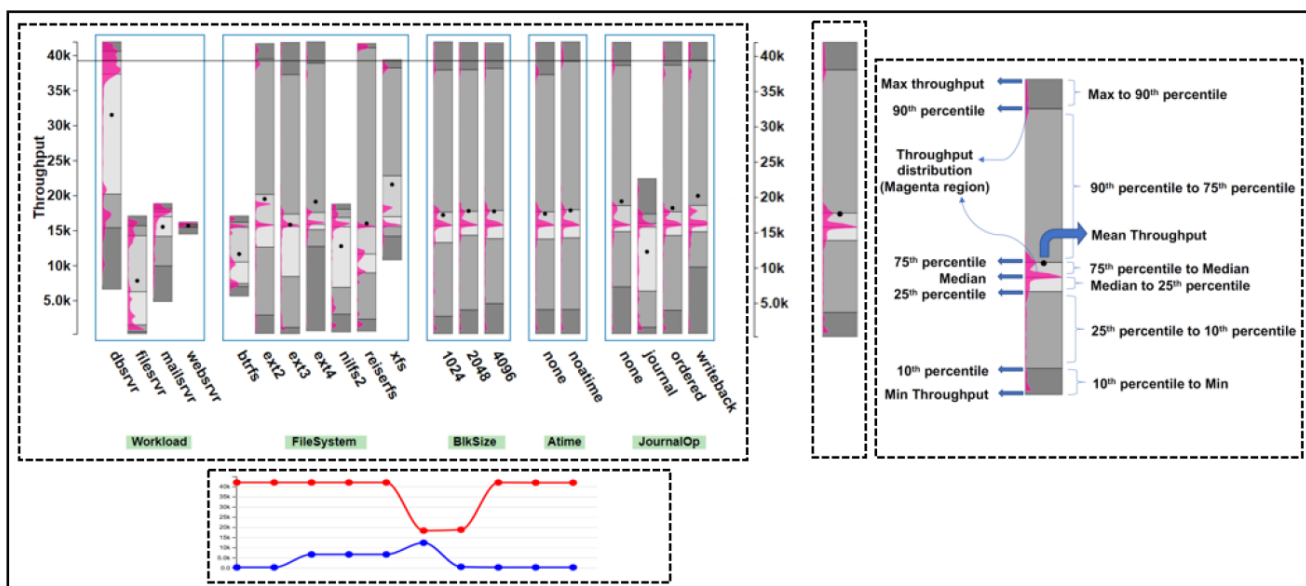
Figure 21

Machine Learning

takes a bold, fundamentally different data-centric approach to wireless communication where semantically meaningful data descriptors (instead of addresses) are used to decide what data (instead of whose data) the radio should receive. By matching data against descriptors, multiple vehicles can exchange data with multiple others at unprecedented throughput and scale. We have designed and developed V-MAC, a radio with data-centric medium access that filters frames by data content instead of MAC addresses.

This eliminates latencies in neighbor discovery and group formation, and allows vehicles to retain any useful overheard content. V-MAC supports robust, efficient and scalable multi-cast with consistent low losses despite multiple receivers of different reception qualities. Experiments on a Raspberry Pi and commodity WiFi dongle based prototype shows that V-MAC cuts down cross stack latency from PHY to user space from 10ms to 73μs (128x faster), reduces loss rate from 10-90% in WiFi ad hoc to <10% among ten receivers indoors, from 50-90% to 0-20% for miniature and real vehicles, while adding merely10μs latency for filtering against up to two million data items, and inter-operable with regular WiFi nodes.

Advanced Computing Systems

Web Scale Graph Embeddings

Steven Skeina
skiena@cs.stonybrook.edu

Deep learning models thrive on massive amounts of training data, and there is no larger data source generally available than the open web. Yet this resource is underutilized for academic research, due to scale and the unwieldiness of working with it.

Embeddings are concise feature representations derived in an unsupervised manner from massive datasets. We have pioneered popular techniques for building embeddings from graphs (Deepwalk) and multilingual text corpora (Polyglot), using traditional data resources such as Wikipedia. Training embeddings on the full web opens up exciting new possibilities, but it is an open problem how to efficiently construct high-quality embeddings on such a scale.

We will develop improved algorithms for several aspects of graph embeddings, including (1) fast, memory efficient algorithms based on projection methods, (2) hierarchical approaches to graph embedding appropriate for distributed algorithms which can scale up to the full web graph captured by the Common Crawl, and (3) dynamic graph embeddings to capture the evolution of networks. These embeddings will power studies of media analysis, web development, and network security.

From Clarity to Efficiency for Distributed Algorithms

Y. Annie Liu and Scott D. Stoller
liu@cs.stonybrook.edu

Distributed algorithms are at the core of distributed systems, which are increasingly indispensable in our daily lives. Yet, developing practical implementations of distributed algorithms with correctness and efficiency assurance remains a challenging, recurring task.

This project develops formal semantics and powerful optimizations for a very high level language for distributed algorithms. The language, DistAlgo, allows the algorithms to be expressed easily and clearly, making it easier to understand them precisely and to generate executable implementations. The semantics are for exact understanding of the language. The optimizations are for generating efficient implementations, without manually coding the algorithms and applying ad hoc optimizations.

We evaluate the language and the translation and optimization methods by applying them to important and difficult distributed algorithms, by using DistAlgo in courses and in other research projects. DistAlgo has been used by hundreds of students in graduate and undergraduate courses in over 100 different course projects, implementing the core of network protocols, distributed coordination services, distributed file systems, distributed databases, parallel processing platforms, security protocols, and more. DistAlgo is open-source and free, available from http://distalgo.cs.stonybrook.edu/. (NSF)

Advanced Computing Systems

High-level Cryptographic Abstractions

Y. Annie Liu
liu@cs.stonybrook.edu

Secure applications critically depend on correct use of cryptographic operations. However, interfaces of cryptographic libraries have remained clumsy and complicated to use, and require an understanding of low-level cryptographic algorithms to use correctly. This project develops high-level abstractions consisting of simple cryptographic primitives and declarative configuration. These abstractions can be implemented on top of any cryptographic library in any language. We implement these abstractions in Python, and use them to program a variety of well-known, sophisticated security protocols, including Signal, Kerberos, and TLS.

Programs using our abstractions are much smaller and easier to write than using low-level libraries, and are safe against the vast majority of cryptographic misuse discovered in the literature. In experiments so far, our implementation incurs a small overhead, less than five microseconds for shared key operations and less than 341 microseconds (<1%) for public key operations. Our abstractions and implementation have been used by even high-school research students to implement a variety of security protocols. (ONR and NSF)

Knowledge Representation and Reasoning: What about the Paradoxes?

Y. Annie Liu and Scott D. Stoller
liu@cs.stonybrook.edu

Artificial intelligence (AI) is a hot topic, especially with machine learning permeating so many areas, from image processing to playing Go to medical diagnosis, triggering also fear of the future as machines beat best humans for increasingly more challenging tasks.

What kind of knowledge can machines or humans learn, and how do they reason? Can humans acquire all, do better, and be in control? This goes back to knowledge representation and reasoning (KRR) that have been at the core of AI. Despite over 100 years of extensive studies, Russell's paradox remains a show stopper: Suppose a human or machine only knows that a barber shaves those who do not shave themselves. Does the barber shave himself?

Prior KRR methods can have subtle implications and can disagree significantly, due to negation in circular reasoning. This project develops a tiny framework that explains a full range of differences and ensures desired solutions in all cases. It also allows easier and more efficient reasoning.

Specifically, we develop founded semantics, a simple new semantics for logic rules, and its straightforward extension to another simple new semantics, constraint semantics, that unify the core of different prior semantics.

The new semantics support unrestricted negation, as well as unrestricted existential and universal quantifications. They are uniquely expressive and intuitive by allowing assumptions about the predicates and rules to be specified explicitly. They are completely declarative and relate cleanly to prior semantics. In addition, founded semantics can be computed in linear time in the size of the ground program. (NSF)

Exploring Ethics with Interactive Sci-Fi Comics

Lori Scarlatos, Tony Scarlatos and Paul Fodor
Lori.scarlatos@stonybrook.edu

Ethics has become a critical consideration for everyone in the engineering and applied science disciplines. Yet even after taking the required ethics courses, many engineering and computer science students still have misconceptions about ethics, such as thinking that it is simply a matter of right versus wrong. Our project focuses on getting students to understand ethical frameworks and explore ethical dilemmas through branching comic-book stories that they both read and create. We have experimented with this approach in the CS Ethics course, using Comic-BEE to create stories for the students to explore, and then having the students create their own stories. Students who go through this process, of considering choices and their consequences, tend to develop a greater understanding of what it means to be ethical. We are currently working on ways of achieving this by having students write the code for their own stories using Ren'Py, an open-source engine for creating visual novels.

Games as Survey Instruments

Lori Scarlatos
Lori.scarlatos@stonybrook.edu

Surveys are a popular method for collecting data about people's attitudes, feelings, and opinions. However, surveys sent to the general public tend to have poor response

rates. Even for surveys given in face to face settings, some respondents do not consider their answers carefully, or give answers that they think the testers want to hear.

Our project focuses on the development of games that are a replacement for surveys. In our test of one such game, we found that the data collected from our game was very close to what was collected from a traditional survey. We also found that people were more likely to want to play the game: people who filled out the survey before playing the game were more likely to finish both as compared to people who played the game and then were asked to complete the survey. We also discovered that the repeated actions required by the game helped to identify respondents who made their choices arbitrarily. We are now looking at developing and validating additional games to collect survey data, ranging from psychological profiles to market data.

Interactive Immersive Web Browser

Arie Kaufman
ari@cs.stonybrook.edu

Traditional web browsers are based on one computer with single or multiple computer screens. The limited size and resolution of the single computer-based web browser set up is not suited for visualization for massive data sets. When dealing with the massive data sets with a web browser, the detail information will be obtained by zooming, but zooming will lead to the loss of the whole picture context. Furthermore, even cluster based large high resolution display systems, such as Reality Deck, provide better visualization for massive data sets, but

the current web browsers could not apply themselves on these systems. We have been developing new web browser techniques that are fit for usage within large high resolution environments. Based on the cluster based visualization systems, the new web browser would enable the users to explore massive data sets online, with detail information and the whole picture context simultaneously afforded by the cluster based visualization facilities. (NSF CVDI and CVDI member companies)

Secure Provenance in High-End Computing Systems

Radu Sion and Erez Zadok
sion@cs.stonybrook.edu
zadok@cs.stonybrook.edu

Data provenance documents the inputs, entities, systems, and processes that influence data of interest-in effect providing a historical record of the data and its origins. The generated evidence supports essential forensic activities such as data-dependency analysis, error/compromise detection and recovery, and auditing and compliance analysis.

This collaborative project is focused on theory and systems supporting practical end-to-end provenance in high-end computing systems. Here, systems are investigated where provenance authorities accept host-level provenance data from validated provenance monitors, to assemble a trustworthy provenance record. Provenance monitors externally observe systems or applications and securely record the evolution of data they manipulate. The provenance record is shared across the distributed environment.

In support of this vision, tools and systems are explored that identify policy (what provenance data to record), trusted authorities (which entities may assert provenance information), and infrastructure (where to record provenance data). Moreover, the provenance has the potential to hurt system performance: collecting too much provenance information or doing so in an inefficient or invasive way can introduce unacceptable overheads. In response, the project is further focused on ways to understand and reduce the costs of provenance collection. (NSF)

Eco Hadoop: Cost and Energy-Aware Cloud and HPC Computing

Radu Sion
sion@cs.stonybrook.edu

The project examines novel services built on top of public cloud infrastructure to enable cost-effective high-performance computing. Eco Hadoop explores the on-demand, elastic, and configurable features of cloud computing to complement the traditional supercomputer/cluster platforms. More specifically, this research aims to assess the efficacy of building dynamic cloud-based clusters leveraging the configurability and tiered pricing model of cloud instances. The scientific value of this proposal lies in the novel use of untapped features of cloud computing for HPC and the strategic adoption of small, cloud-based clusters for the purpose of developing/tuning applications for large supercomputers.

Through this research, we expect to answer key research questions regarding: (1) automatic workload-

Advanced Computing Systems

specific cloud cluster configuration, (2) cost-aware and contention-aware data and task co-scheduling, and (3) adaptive, integrated cloud instance provisioning and job scheduling, plus workload aggregation for cloud instance rental cost reduction. If successful, this research will result in tools that adaptively aggregate, configure, and reconfigure cloud resources for different HPC needs, with the purpose of offering low-cost RandD environments for scalable parallel applications. (NSF)

20GHz+ Energy-Efficient Superconductor Flux Quantum Processor Design

Mikhail Dorojevets
mikhail.dorojevets@stonybrook.edu

New flux quantum logics with extremely low power consumption at frequencies of 20-50 GHz make superconductor technology one of the candidates for use in future energy efficient systems for critical national applications. In the first phase of the project, the PI's SBU team has developed viable design techniques, including asynchronous wave pipelining, and set of design tools and cell libraries for VLSI-scale superconductor design. Several RSFQ chips with the complexity of up to 10K Josephson Junctions (JJs) have been successfully designed using those techniques and demonstrated operation at 20 GHz. During the next stage of the project supported by ARO and IARPA (2012-2018), the cell-level design and energy efficiency evaluation of a complete set of 32-bit processing units and Bloom filters for cybersecurity have been done for superconductor circuits implemented with a Reciprocal Quantum Logic (RQL) using SBU RQL VHDL-based

design and evaluation tools. The goal of the on-going work is to develop a new architecture and design of a 20GHz+ event-driven processor to be fabricated at MIT Lincoln Lab using the latest superconductor technology.

National File System Trace Repository

Erez Zadok
erez.zadok@stonybrook.edu

This project sustains and enhance a national repository of traces that capture usage of file-systems, storage and input/outputs with the goal of helping researchers to address performance bottlenecks. As part of ongoing activities, new tools essential to the successful use of traces will be added. Also, new mirror sites are planned to expand efficiency and utility of the trace repository. The national repository of file systems traces is of immense importance to the software engineering community. By updating the traces for the new and emerging applications, this project provides a vital service to this community.

Disk based file system is a performance bottleneck for many computer applications. Thus, it is useful to have a set of activity traces for storage systems to understand the access characteristics. Such traces may potentially help optimize flash-based solid-state storage and other emerging non-volatile memory based storage technologies.

As the storage technologies evolve with network storage, cloud-based storage, geo-replicated storage, the trace repository provides a set of targets for storage systems optimization.

Optimizing and Understanding Large Parameter Spaces in Storage Systems

Erez Zadok and Klaus Mueller
erez.zadok@stonybrook.edu

Computer systems contain large software that gets more complex every day. They include many configurable parameters: "knobs" you can adjust and switches you can turn on/off to adjust performance, energy consumption, etc. For example, a smart-phone's settings include many on/off features and "sliders" one can adjust; but trying all the possible combinations to, say, improve performance and reduce battery consumption is very time consuming. This project aims to optimize computer systems by (1) automatically exploring many parameter combinations and (2) helping humans see visual indications of how these parameters work and better understand complex systems. This project will: (1) develop techniques to optimize storage systems, because they are the slowest part of any computer; (2) combine features from existing optimization and machine learning techniques; (3) improve the search for optimal settings by deciding when to stop and restart searching as well as considering the cost of changing system settings; (4) develop human driven visual techniques to explore extremely large sets of option combinations to better understand them and further direct the optimization process; and (5) evaluate all these techniques on real world storage systems.

Computer storage systems have gotten so complex that no human can fully optimize them, especially when

circumstances change. This project will (1) help automate the optimization of storage systems, improving their performance and energy use; (2) advance the state of the art in hybrid optimization and machine learning techniques; (3) develop and release interactive visualization systems that let humans understand, view, and direct a search process to promising directions; (4) train and educate graduate and undergraduate students; and (5) produce results that are applicable to other computer system optimization problems.

....

NLP- Assisted Formal Verification of the NFS Distributed File System Protocol

Erez Zadok, Scott Smolka and Niranjan Balasubramanian
erez.zadok@stonybrook.edu

The Internet's success and growth is owed to standards that ensure computers can talk to each other. These standards are human-written, technical design documents that take years to develop and implement. Alas, such design documents are often imprecise, and their software implementations do not always conform to their designs.

This project aims to speed up the process of designing and implementing Internet standards using: (1) Artificial Intelligence (AI) techniques to automatically process design documents so that flaws in them can be detected and reported and (2) runtime analysis of software implementations to detect deviations from their respective designs.

This project will: (1) conduct a major case study involving the complex, distributed Network File System version 4 (NFSv4) protocol; (2) develop a theoretical model of NFSv4's expected behavior using Natural Language Processing (NLP) AI techniques; (3) analyze the model to detect inconsistencies; (4) check the model against another reference implementation that is known to be correct; and (5) monitor an actual running NFS implementation for compliance with our verified theoretical model.

NFS is a popular and growing protocol that enables users to access their files and data across any network. The NFSv4 design documents are fairly complex and over 500 pages long. This project will (1) help accelerate NFSv4's ongoing design, development, and adoption; (2) advance the state of the art in NLP/AI techniques to understand human-written design documents; (3) advance the state of the art in formally modeling and verifying such designs; (4) train and educate graduate and undergraduate students; and (5) produce results that are applicable to many other Internet standards.

Researcher's Biographies

Esther Arkin

Esther Arkin received her BS in Mathematics, from Tel-Aviv University, Israel, 1981, and her MS and PhD in Operations Research, Stanford University, August, in 1983 and 1986. Before joining Stony Brook's Applied Mathematics and Statistics department in 1991, she was a Visiting Assistant Professor, in the School of Operations Research and Industrial Engineering, Cornell University, and a Visiting Associate Professor, Department of Statistics and Operations Research, Tel-Aviv University, Israel (2/95-6/95).

Niranjan Balasubramanian

Niranjan Balasubramanian is affiliated with the Department of Biomedical Informatics and Center of Excellence in Wireless and Information Technology (CEWIT). He received his PhD from University of Massachusetts, Amherst, where he was a part of the Center for Intelligent Information Retrieval (CIIR). Before he started his PhD studies, he was a software engineer at the Center For Natural Language Processing (CNLP) at Syracuse University. He completed his Masters degree in Computer Science at the University of Buffalo in 2003. Prior to joining the computer science department at Stony Brook in spring 2015, Dr. Balasubramanian was a post-doctoral researcher in the Turing Center at the University of Washington in Seattle.

Gábor Balázsi

Gábor Balázsi received his undergraduate Physics degree at the Babeş-Bolyai University in Kolozsvár (Cluj), Romania. From 1997 to 2001 he completed a Physics PhD at the University of Missouri at Saint Louis, USA, as a PhD student with Professor Frank Moss. His PhD research was on perturbation propagation and synchronization in normal and epileptic neurons and glial cells. In 2002, he continued as a postdoctoral fellow in Systems Biology at Northwestern University in Chicago, studying gene- regulatory network response to environmental changes with Professors Zoltán N. Oltvai and Albert-László Barabási. In 2005 he became a postdoctoral fellow in Synthetic Biology at the Center for Biodynamics at Boston University with Professor James J. Collins. There he designed synthetic gene circuits to study how cellular diversity promotes drug resistance. He continued and expanded these efforts in his own laboratory over the last thirteen years (eight of which were at the University of Texas MD Anderson Cancer Center in Houston, Texas), building a growing library of synthetic gene circuits first in yeast, and then in cancer cells. As the Henry Laufer Associate Professor of Physical and Quantitative Biology at Stony Brook University he leads an interdisciplinary research group, which utilizes synthetic gene circuits to control gene expression in yeast and human cells. The goal of his laboratory is to combine synthetic biology and physical modeling to understand fundamental biological processes underlying microbial drug resistance and cancer progression. Dr. Balázsi was one of the recipients of the 2009 NIH Director's New Innovator Award, which was created to "stimulate highly innovative research and support promising new investigators." His research group is part-experimental and part-computational, fostering interdisciplinary training while advancing the frontiers of quantitative biology.

Rezaul Chowdhury

Rezaul Chowdhury received his PhD from the Department of Computer Sciences, UT Austin, working with Professor Vijaya Ramachandran, and defending "Cache-efficient Algorithms and Data Structures: Theory and Experimental Evaluation." Prior to joining SBU in 2011, he was in Boston working with Professor Sandor Vajda's Structural Bioinformatics Group at Boston University, and Professor Charles Leiserson's SuperTech Research Group at MIT. Before moving to Boston, he was a postdoctoral fellow at the Center for Computational Visualization (CVC), Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin. He worked with Professor Chandrajit Bajaj. Chowdhury now leads the Theoretical and Experimental Algorithmics (TEA) Group where they concentrate on both algorithm design and algorithm engineering. He holds a joint appointment with the Institute for Advanced Computational Sciences (IACS).

Christine DeLorenzo

Christine DeLorenzo is a biomedical engineer with expertise in clinical/translational studies and medical imaging. She is currently PI for studies involving imaging of individuals with major depressive disorder (MDD), funded through the NIH and private foundations. As the Director of the Center for Understanding Biology using Imaging Technology (CUBIT) at Stony Brook University (SBU), her research focus involves accurate quantification of the state-of-the-art imaging modalities available at SBU, including positron emission tomography (PET) acquired on SBU's simultaneous PET/MRI. With

a web-based image pipeline for image transfer, quality assurance, manual editing (when needed), analysis and user approvals, CUBIT provides validated and accurate image-derived measures for neuroimaging applications. As Director, she manages an image processing and IT team, ensures that image analysis is accurate and supervises implementation of novel algorithms into CUBIT's analysis pipeline. Her ultimate goal is to use neuroimaging to improve understanding, diagnosis and treatment of mood disorders, reducing the burden of these diseases for patients and their families.

Yuefan Deng

Yuefan Deng entered Columbia University in 1983, after graduating from Nankai University with honors, through a special scholarship program CUSPEA organized by the Chinese-American Nobel laureate Professor T. D. Lee. His PhD thesis on simulating gauge theory with supercomputers, completed in 1989, was supervised by Professor Norman Christ After a brief stay at NYU's Courant Institute of Mathematical Sciences as Professor James Glimm's postdoctoral fellow, he joined the faculty of Stony Brook University where he was promoted to full professor of applied mathematics in 1999. As visiting or adjunct professor, Dr. Deng worked at the Hong Kong University of Science and Technology (1.5 years), the IBM T. J. Watson Research Lab (1.0 year) where he joined the group of Deep Blue, Columbia University (1.0 year), Nanyang Technological University (0.5 years), University of Singapore National University of Singapore (1.0 year), and the Singapore's Agency of Science, Technology and Research (1.0 year). Dr. Deng has published more than 100 papers in physics, applied mathematics, biomedical engineering, and biosciences plus three books: one on ODEs, one on PDEs, and another on parallel computing. He holds thirteen patents in China or US or both. He has supervised and is supervising 30 doctoral students and has taught 10,000 undergraduate students. He is the recipient of the State University of New York Chancellor's Award of Excellence in Teaching (2016) and the Stony Brook University's CEAS Dean's Award of Excellence in Teaching (2016). His research is supported by the US DOE, NIH, NSF, New York State, and ARO, as well as China's Ministry of Science and Technology.

Petar Djuric

Petar M. Djurić received his BS and MS degrees in Electrical Engineering from the University of Belgrade, in 1981 and 1986, respectively, and his PhD degree in Electrical Engineering from the University of Rhode Island (1990). Prof. Djurić has served on numerous technical committees for the IEEE and has been invited to lecture at universities in the United States and overseas. His SPS activities include Member-at-Large of the SPS Board of Governors (2011-2013), Vice President-Finance (2006-09), Area Editor of Special Issues – *IEEE Signal Processing Magazine* (2002-05), Associate Editor – *IEEE Transactions on Signal Processing* (1994-96 and 2003-05), Chair – SPSSignal Processing Theory and Methods Technical Committee (2005-06), and Treasurer – SPS Conference Board (2001-03). He has been on the Editorial Board of several journals and has served on numerous committees of professional conferences and workshops. He is the first Editor-in-Chief of the *IEEE Transactions on Signal and Information Processing over Networks* (2015-2018).
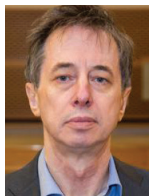
Mikhail Dorojevets

Mikhail Dorojevets received his MS degree in Physics and Electronic Engineering from the Institute of Physics and Technology, Moscow in 1982. He received his PhD degree in Computer Engineering from the Russian Academy of Sciences in 1988. From 1982 to 1995, he was a principal designer in several projects on high-performance systems design for Russian science, defense, and industry. For his work on the MARS-M supercomputer with decoupled multi threaded architecture built by 1988, he received the Outstanding Achievement Diploma jointly awarded by the USSR Academy of Sciences, and the USSR State Committee on Science and Technology. Since 1996, he has been an Assistant and (since 2000) an Associate Professor of the Department of Electrical and Computer Engineering, at Stony Brook University (SBU). His research interests include parallel computer architectures, and all aspects of microprocessor and computer design. Since 2010, Mikhail Dorojevets and his design team at Stony Brook have been working on the architecture and complete logical and physical layout design of a 30 GHz 16-bit RSFQ processor implemented with ISTEC 1.0m 10 km/cm2 superconductor technology in a joint project with colleagues at the National Universities of Yokohama and Nagoya (Japan). Mikhail Dorojevets's current research work is focused on the development of energy-efficient superconductor processors with ultra-low-power consumption and cyber security. Since 2004, Mikhail Dorojevets has been a member of Department of Defense Superconducting Technology Assessment Panel.

Shmuel Einav

Shmuel Einav, Professor of Biomedical Engineering, is a world-distinguished expert in the cardiovascular circulatory system and the field of biomedical engineering. He is best known for his studies on blood flow through heart valves, coronary circulation, blood-tissue interaction, and flow and turbulent characteristics in occluded arteries. The focus of his research is the role of Hemodynamics in the initiation of atherosclerosis, the dynamics of cardiovascular flows, and the influence of flow and the associated shear stress on vascular endothelial biology. In recognition of his significant achievements and important contributions to science, biomedicine and technology, he has been elected as a Fellow of the American Institute

Researcher's Biographies

for Medical and Biological Engineering (AIMBE), Fellow of the Biomedical Engineering Society (BMES), Fellow of the International Federation for Medical and Biological Engineering (IFMBE) and a Fellow of the American Society of Mechanical Engineers (ASME).

Eugene A. Feinberg

Eugene A. Feinberg is currently the Distinguished Professor in the Department of Applied Mathematics and Statistics at Stony Brook University. He is an expert on applied probability, stochastic models of operations research, Markov decision processes, and on industrial applications of operations research and statistics. He has published more than 150 papers and edited the Handbook of Markov Decision Processes. His research has been supported by NSF, DOE, DOD, NYSTAR (New York State Office of Science, Technology, and Academic Research), NYSERDA (New York State Energy Research and Development Authority) and by industry. He is a Fellow of INFORMS (The Institute for Operations Research and Management Sciences) and has received several awards including 2012 IEEE Charles Hirsh Award for developing and implementing smart grid technologies, 2012 IBM Faculty Award, and 2000 Industrial Associates Award from Northrop Grumman. Dr. Feinberg is an Associate Editor for Mathematics of Operations Research, Stochastic Systems, and Applied Mathematics Letters. He is an Area Editor for Operations Research Letters.
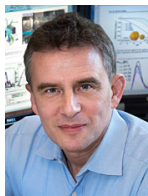
Paul Fodor

Paul Fodor obtained his PhD and MS in Computer Science from Stony Brook University in 2011 and 2006, and BS in Computer Science from Technical University of Cluj-Napoca in 2002. He also worked at IBM T.J. Watson Research in the team that built the IBM Watson system that played on the Jeopardy! TV quiz show.

Anatoly I. Frenkel

Anatoly Frenkel is a Professor in the Department of Materials Science and Chemical Engineering at the Stony Brook University and a Senior Chemist (Joint Appointment) at the Division of Chemistry, Brookhaven National Laboratory, having joined in the Fall of 2016. Prior to his appointment at SBU, he has held a number of different positions, including Associate and then appointed Full Professor and Chair, Physics Department at Yeshiva University, a Research Scientist and Principal Investigator in Materials Research Laboratory of the University of Illinois at Urbana-Champaign. He received MS degree from St. Petersburg University and PhD degree from Tel Aviv University (with Prof. A. V. Voronel), all in Physics, followed by a postdoctoral appointment at the University of Washington (with Prof. E. A. Stern). His research interests focus on development and applications of in situ and operando synchrotron methods to solve a wide range of materials problems, with most recent emphasis on catalysis, electromechanical materials, filtration materials, quantum dots, physico-chemical properties of nanoparticles, as well as machine learning methods for structural analysis and design of nanomaterials. He is a founding Principal Investigator and the Spokesperson for the Synchrotron Catalysis Consortium at Brookhaven National Laboratory. He is a Fellow of the American Physical Society and a Fellow of the Empire Innovation Program at the New York State. He is the author of over 330 peer-reviewed publications, which have been cited nearly 17,000 times, and has given over 300 invited lectures.

Jie Gao

Jie Gao is an associate professor at the Department of Computer Science, Stony Brook University. She received her PhD degree from the Department of Computer Science, Stanford University in 2004 and B.S. degree from the Special Class for the Gifted Young at University of Science and Technology of China in 1999. She spent the academic year 2004-2005 at Center for the Mathematics of Information, California Institute of Technology.
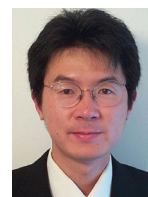
Xianfeng David Gu

David Xianfeng Gu got his PhD in Computer Science from Harvard University in 2002 supervised by the Fields medalist, Prof. Shing-Tung Yau. Dr. Gu is one of the major funders of an emerging inter-disciplinary field: Computational Conformal Geometry, which combines modern geometry, topology theory with computer science. Dr. Gu and his collaborators laid down the theoretic foundations and systematically developed the computational algorithms, and applied conformal geometric method in many fields in engineering and medicine, such as Computer Graphics, Computer Vision, Visualization, Geometric Modeling, Networking, Artificial Intelligence, Medical Imaging and Computational Mechanics and so on.
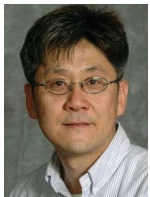
Minh Hoai Nguyen

Minh Hoai Nguyen received a PhD in Robotics from Carnegie Mellon University and a Bachelor of Engineering from the University of New South Wales. Before coming to Stony Brook, Minh Hoai was a post-doctoral research fellow with Andrew

Zisserman at Oxford University. He was also a Kurti Junior Research Fellow at Brasenose College.

Sangjin Hong

Sangjin Hong received his BS and MS in Electrical Engineering and Computer Science from the University of California at Berkeley and his PhD in Electrical Engineering and Computer Science from the University of Michigan at Ann Arbor. Since then he has been at the Stony Brook University – State University of New York, where he is a Professor of Electrical and Computer Engineering. He has previously worked as a systems engineer at Ford Aerospace and Communications Corp., as a technical consultant at Samsung Electronics Inc., and as a research fellow at the University of Michigan at Ann Arbor. He is a Senior Member of IEEE and a Member of Eta Kappa Nu and Tau Beta Pi Honor Societies. His research interests include: Low-power VLSI design of multimedia wireless communications and digital signal processing systems, including SOC design methodology and optimization.

David J. Hwang

David Hwang received his BSE and MSE from the Seoul National University and his PhD from the University of California Berkeley. His research interests are on micro and nanoscale heat transfer, and laser-assisted processing and diagnostics of materials of wide range application area. Examples include micro/nanoscale laser material processing and diagnostics, time resolved diagnostics by ultrafast lasers, laser scribing of thin film solar cell, via-hole fabrication for back contact solar cell, selective crystallization and doping for solar cell, micro/nanoscale chemical species analysis by laser induced breakdown spectroscopy, optical near-field and scanning probe based processing and diagnostics, scalable nanomanufacturing, laser-assisted direct synthesis of nanomaterials

system, solar hydrogen production, optical near-field coupled with electron microscopes for in-situ monitoring of light-matters interaction, and three-dimensional hybrid opto-fluid device fabrication for biological applications.

Arie Kaufman

Arie Kaufman is a Distinguished Professor and CEWIT Chief Scientist, the Director of the Center of Visual Computing (CVC) , and a Distinguished Professor of Radiology at Stony Brook University. He joined the faculty at Stony Brook in 1985 and was appointed Chair from 1999 to 2018. He also held posts at the Hebrew University, Tel-Aviv University, Florida International University, Ben-Gurion University, and Columbia University.
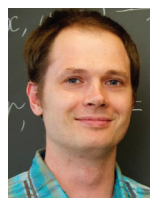
Dima Kozakov

Dima Kozakov received his MS in Applied Mathematics and Physics from Moscow Institute of Physics and Technology, and PhD from the Biomedical Engineering Department at Boston University. Dima was a research faculty at Boston University prior to joining Stony Brook University. Dima's research interests lie at the intersection of applied mathematics, physics and computational biology. He focuses on two main goals. The first is the development of mathematically elegant, computationally efficient and physically accurate algorithms for modeling macromolecular structure and function on the genome scale. The second is the application of novel methods to improving the understanding of biological problems and to the design of therapeutic molecules with desired biological and biomedical properties.

Ellen Li

Ellen Li is the Chief of the Division of Gasteroenterology and Hepatology, Director of Inflammatory Bowel Disease Center

Molecular Genetics and Microbiology, at the School of Medicine at Stony Brook University. Ellen Li's focus is on prevention, diagnosis, and treatment of digestive diseases. Her current research interests are on defining the role of the gut microbiome in digestive diseases, particularly in inflammatory bowel diseases, colon cancer, and functional GI disorders, such as irritable bowel syndrome. Her specialty as a scientist is in translational research, which focuses on taking scientific discoveries and translating them into ways to help improve people's health.
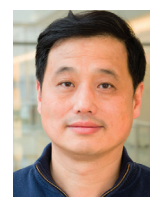
Wei Lin

Wei Lin received his BS degree in Biomedical Engineering in 1986 and MS degree in Mechanical Engineering in 1989 from Shanghai Jiaotong University, Shanghai, China. He received PhD degree in Mechanical Engineering from Stony Brook University in 2001, Stony Brook, NY. Dr. Lin joined the Department of Applied Mechanics in Fudan University, Shanghai, China as the assistant professor in 1989 and lecturer in 1993. He joined the Department of Biomedical Engineering as the assistant professor in 2004 and associate professor in 2013. His current research interests include medical instrumentation, wearable medical devices and high performance computing in embedded systems. Dr. Lin is a member in IEEE and Biomedical Engineering Society.

Y. Annie Liu

Y. Annie Liu is a Professor of Computer Science at Stony Brook University. She received her BS from Peking University, MB from Tsinghua University, and PhD from Cornell University. Her primary research is in languages and algorithms, and focuses specially on systematic methods for design and optimizations. The methods are centered around incrementalization—the discrete counterpart of differentiation in calculus. She has strong other interests

Researcher's Biographies

in interactive environments, real-time and embedded systems, database, knowledge representation and reasoning, distributed systems, and security. She has published in many prestigious venues, taught in a wide range of Computer Science areas, and presented over 100 conference and invited talks worldwide. Her awards include a State University of New York Chancellor's Award for Excellence in Scholarship and Creative Activities.

Ji Liu

Ji Liu received the PhD degree in electrical engineering from Yale University in 2013. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Stony Brook University. Prior to joining Stony Brook University, he was a Postdoctoral Research Associate at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. His current research interests include distributed control and computation, decentralized machine learning, distributed optimization, multi-agent systems, social networks, epidemic networks, and cyber-physical systems.

Jon Longtin

Jon Longtin joined the Mechanical Engineering Faculty at Stony Brook in Fall 1996. His research interests include using the unique qualities of laser light for novel measurement of the thermophysical behavior of liquid and solids. Examples include measuring the surface temperature and concentration of a liquid and determining a liquid's thermophysical properties, e.g., specific heat and thermal conductivity. Another of his interest areas is fundamental and applied aspects of ultra high intensity, short-pulse laser interactions with liquids and

solids for laser material processing, where behavior far different than that observed at lower, more moderate laser intensities can occur. Longtin is the author of over thirty publications, including invited chapters in *Micro-scale Energy Transport* and *Annual Review of Heat Transfer*. He is currently a member of the ASME K-15 committee, Heat Transfer in Manufacturing and Materials Processing. Longtin is the recipient of the 1996 JSPS Postdoctoral Fellowship, 1997 NSF CAREER and PECASE awards, and the 1998 Excellence in Teaching Award. His research is funded by the National Science Foundation, NASA, and other sources.

Devinder Mahajan

Devinder Mahajan is Professor and Graduate Program Director of Chemical and Molecular Engineering and serves as Director of the newly established Institute of Gas Innovation and Technology (I-GIT) in the R&D Park, at Stony Brook University. His research includes monitoring energy policies and development of low-carbon technologies to address sustainability and climate change issues. He has published over 300 papers, delivered over 110 lectures, edited eight special journal volumes, and holds fifteen patents. Dr. Mahajan received his B.Sc. (Honors School) and MSc (Honors School) from Panjab University, Chandigarh; PhD from the University of British Columbia, Canada in Chemistry and completed his post-doctoral training at Brookhaven National Laboratory (BNL), New York. He joined the staff of BNL in 1983 and moved to Stony Brook University in 2002 and concurrently held a joint Scientific staff appointment with BNL until 2015.

Todd Miller

Todd Miller has a PhD from the Rockerfeller University, 1988. He is currently a professor and chairman of the Department of Physiology and Biophysics at Stony Brook University. He founded and runs the Miller Laboratory at Stony Brook University. The major research goals of the laboratory are: (1) to understand how tyrosine kinases recognize their target proteins in cells; (2) to determine the regulatory mechanisms that control tyrosine kinase activity; and (3) to develop strategies to block the action of oncogenic tyrosine kinases.

Joseph Mitchell

Joe Mitchell is an internationally recognized leader in the area of algorithms, particularly in computational geometry – the study of the design, analysis, and implementation of efficient algorithms to solve geometric problems. His particular interest is in optimization problems, approximation algorithms, geometric networks, and applications of algorithms to problems in robotics, transportation, autonomous vehicle routing, sensor networks, computer graphics, manufacturing, geographic information systems, and computer vision. He has worked extensively in the area of applications of geometric algorithms in Air Traffic Management.

Klaus Mueller

Klaus Mueller has a BS in Electrical Engineering from the Polytechnic University of Ulm, Germany, and an MS in Biomedical Engineering and a PhD in Computer Science, both from The Ohio State University. Apart from his appointment at the Computer

Science Department at Stony Brook University, he also holds adjunct faculty positions at the Biomedical Engineering Department and the Radiology Department, and he is an adjunct scientist at the Computational Science Center at Brookhaven National Laboratory. His research is sponsored by NSF (including the Career award in 2001), NIH, DOE, DHS, and private industry and research labs.

Nick Nikiforakis

Nick Nikiforakis is affiliated with the National Security Institute. He received his PhD in Computer Science from KU Leuven in Belgium. He received his MSc, in Parallel and Distributed Systems and BSc in Computer Science from the University of Crete, Greece.

Michalis Polychronakis

Michalis Polychronakis is affiliated with the National Security Institute. He received the BSc ('03), MSc ('05), and PhD ('09) degrees in Computer Science from the University of Crete, Greece, while working as a research assistant in the Distributed Computing Systems Lab at FORTH-ICS. Before joining Stony Brook, he was an Associate Research Scientist at Columbia University.

Don Porter

Donald Porter is affiliated with the Computer Science Department of the University of North Carolina at Chapel Hill. He earned his PhD and MS in Computer Science from The University of Texas at Austin in 2010 and 2007, and BA in Computer Science and Mathematics from Hendrix College in 2003.

Anurag Purwar

Anurag Purwar is currently an Associate Editor of the *ASME Journal of Computing and Information Science in Engineering* and of *International Journal of Mechanics Based Design of Structures and Machines.* Dr. Anurag Purwar's research interests are in machine design area with a focus on kinematic design of robots and mechanisms, CAD/CAM, and application of Computational Geometry, Virtual Reality (VR), Computer Graphics and Visualization in Design Engineering. He has published more than 50 peer-reviewed conference and journal papers and his research has been funded by National Science Foundation (NSF), NY-state SPIR, NY-state Center for Biotechnology, Sensor-CAT, SUNY Research Foundation, industry, Stony Brook University, and SUNY Office of Provost.

Hong Qin

Hong Qin is a Full Professor (with tenure) of Computer Science in Department of Computer Science at State University of New York at Stony Brook (Stony Brook University), where he is also a member of SUNYSB's Center for Visual Computing. He received his BS (1986) degree and his MS degree (1989) in Computer Science from Peking University in Beijing, China. He received his PhD (1995) degree in Computer Science from the University of Toronto.

During 1989-1990, he was a research scientist at North-China Institute of Computing Technologies. During 1990-1991, he was a PhD candidate in Computer Science at the University of North Carolina at Chapel Hill. After he received his PhD degree in Computer Science at the University of Toronto in September 1995, he was a research scientist in Department of Computer Science at the University of Toronto during September-December 1995. From December 1995 to 1997, he was an Assistant Professor of Computer and Information Science and Engineering at the University of Florida. From September 1997 to July 2001, he was an Assistant Professor of Computer Science at State University of New York at Stony Brook. From August 2001 to July 2006, he was an Associate Professor (with tenure) at State University of New York at Stony Brook.

Thomas Robertazzi

Thomas Robertazzi has done research in recent years on exascale data processing, quantum computing, scheduling and electric power systems. He has authored, co-authored or edited more than a half dozen texts and monographs. He currently teaches courses on networking, reliability and quantum systems. He has also been the instructor in courses on scientific and engineering issues affecting society. Dr. Robertazzi received the PhD from Princeton University and a BEE degree from The Cooper Union. He is a Fellow of the IEEE for contributions to parallel processing scheduling.

Lorna Role

Lorna Role earned a BA degree in Applied Mathematics in 1975 and a PhD in Physiology in 1981, both from Harvard University. After three years of postdoctoral study with Dr Gerald Fischbach she was appointed an Assistant Professor in the Department of Anatomy and Cell Biology in the Center for Neurobiology and Behavior at Columbia University. She was promoted to Associate Professor with tenure in 1992, and in 1996 was promoted full Professor. From 1996 to 2008, she also held the position of Research Scientist at the New York State Psychiatric Institute. In 2001 she was elected a member of The American College of Neuropsychopharmacology (ACNP). In 2006 she was awarded the Sidney R. Baer, Jr. Prize for Innovative Research related to Schizophrenia (NARSAD) and in 2007– 2008 received a second NARSAD Distinguished Investigator Award. In 2008, she joined the State

Researcher's Biographies

University of New York at Stony Brook as Professor and Chair of the Department of Neurobiology and Behavior. In 2009 she was advanced to Fellow of the ACNP and in 2010 was the recipient of an NIH Directors Pioneer award. In 2011 she has become a member of the American Association for the Advancement of Science and in 2016 she has been named a SUNY Distinguished Professor.

David Rubenstein

David Rubenstein is an Associate Professor and the Director of the Graduate Program for Biomedical Engineering. His research interests lie in developing new biometric scaffolds for micro vascular tissue engineering applications and parthenogenesis of cardiovascular diseases. Tailored scaffolds are fabricated with electro-spinning and 3D printing and then tested for optimal mechanical, chemical and topographical cues. Cellular and molecular biology approaches towards elucidating convergent inflammatory and thrombotic pathways are used to develop new therapeutics and identify new therapeutic targets for vascular disease intervention.

Emre Salman

Emre Salman is an Associate Professor of Electrical and Computer Engineering at Stony Brook University (SUNY), New York. His broad research interests include analysis, modeling, and design methodologies for high performance, energy efficient, and secure integrated circuits with a broad range of applications including high speed processors, low power mobile computing, Internet-of-things, and bio-implantable devices. His research is supported by multiple programs within National Science Foundation (NSF), Semiconductor Research Corporation (SRC), National Institute of Health, and Simons Foundation. Emre received NSF CAREER Award in 2013, Outstanding Young Engineer Award from IEEE Long Island in 2014, Best Paper Award from SRC TECHCON in 2016, and Technological Innovation Award from IEEE Region 1 in 2018.

Joel Saltz

Joel Saltz is a leader in research on advanced information technologies for large scale data science and biomedical/scientific research. He has developed innovative pathology informatics methods, including: the first published whole slide virtual microscope system; pioneering pathology computer-aided diagnosis techniques; and methods for decomposing pathology images into features and linking those features to cancer "omics," response to treatment and outcome. He has broken new ground in big data through development of the filter-stream based DataCutter system, the map-reduce style Active Data Repository and the inspector-executor runtime compiler framework. He has also been an active contributor in clinical informatics, having developed predictive models for hospital readmissions, point of care laboratory testing quality assurance systems, decision support systems for electrophoresis interpretation and graphical user interfaces to support clinical data warehouse queries. Dr. Saltz has been a pioneer in establishing the field of biomedical informatics; he founded and built two highly successful departments of biomedical informatics, one at Ohio State University and one at Emory University. In 2013, he came to Stony Brook as Vice President for Clinical Informatics and Founding Department Chair of Biomedical Informatics – to create a living laboratory for biomedical informatics and to create a third unique biomedical informatics department dually housed in the School of Medicine and the College of Engineering. Dr. Saltz is trained both as a computer scientist and as a physician through the MSTP program at Duke University. He has deep experience in computer science, having served on the computer science faculties at Yale University and the University of Maryland. He completed his residency in clinical pathology at Johns Hopkins University and he is a practicing, board-certified clinical pathologist.

Dimitris Samaras

Dimitris Samaras earned his PhD in Computer Science from University of Pennsylvania in 2001; MS in Computer Science from Northeastern University in 1994, and Diploma in Computer Engineering and Informatics from University of Patras, Greece in 1992.

Tony Scarlatos

Tony Scarlatos is the Director of the Multimedia Lab. He received his MA from Adelphi University in 1995 and BFA from Pratt Institute in 1982. Tony Scarlatos' research interests include computer-based training, distance learning, multimedia, computer-human interaction, and multi-modal interfaces. With his combined background in the visual arts, education, and computer science, these are synergistic pursuits. The unifying goal of his research is to facilitate faster comprehension and longer retention for learners using technology. This work has been supported through a number of grants: Long Island Lighting Company RandD Initiative (1993 and 1994), New York State Department of Education (1997), and a Department of Commerce TIIAP grant (1998). His research is currently being supported by a National Science

Foundation grant, "CRCD: Innovative Approaches to Computer-Human Interfaces" (2002-2006).

Lori Scarlatos

Lori Scarlatos received her PhD in Computer Science from the State University of New York at Stony Brook in 1993, MS with Distinction in Computer Science from Pratt Institute in 1984, and BFA with Honors, Art and Design from Pratt Institute in 1982. In all Lori Scarlatos' various research projects, her focus has been on how computers can help people to see, understand, and learn. This includes research on educational games and simulations; human computer interaction, with an emphasis on tangible user interfaces and physical computing; computer graphics, including level-of-detail surface modeling and animation systems; information visualization; multimedia; and computer science education.

R. Sekar

R. Sekar obtained his PhD in Computer Science from Stony Brook University in 1991, and B. Tech in Electrical Engineering from Indian Institute of Technology, Madras in 1986. R. Sekar's main research focus is on software and systems security. It is driven by practical problems, and emphasizes building real systems. It draws on principles and techniques from programming languages and compilers, operating systems, algorithms, networks, and artificial intelligence to address problems such as: Software vulnerability mitigation (buffer overflows, SQL injection, XSS)-Malware and untrusted code defense-High-performance intrusion detection (network and host-based)-Attack isolation and recovery-Self-healing and self-regenerative systems- and monitoring and management of distributed systems.

Satya Sharma

Satya Sharma came to Stony Brook University as the Executive Director of CEWIT and a faculty member of Department of Mechanical Engineering in 2003. Before joining Stony Brook, Dr. Sharma was Senior Vice President at Symbol Technologies from 1995 to 2003, overseeing various divisions of the company including World-Wide Operations, Mobile Computing and Wireless Engineering, and Quality and Process Improvements. Prior to his tenure at Symbol, he was Director of AT&T Power Systems and an adjunct professor at Southern Methodist University. Dr. Sharma has managed technology, led product development, managed marketing and financial functions, led operations and taken a leadership role in organizational transformation. He led Lucent Technologies to win the Deming Prize in 1994, making it the first and still the only American manufacturing company to have this honor. While leading Symbol's Mobile Computing and Wireless Engineering, the company won the National Medal of Technology in 2000. Dr. Sharma also led Lucent and Symbol to win the Shingo Prize in 1992 and 2003 respectively. Dr. Sharma has more than 70 technical publications and conducts research in a wide variety of disciplines including wireless and mobile computing, quality management, and materials science. He holds a PhD degree in mechanical engineering from University of Pennsylvania and an MBA degree from Ohio State University.

Radu Sion

Radu Sion is affiliated with the National Security Institute. He is an Associate Professor of Computer Science at Stony Brook University (on leave). He has also founded and is currently running Private Machines Inc. Radu Sion obtained his PhD in Computer Science from Purdue University in 2004, MS and BS in Computer Science from "Politehnica" University of Bucharest in 1999 and 1998, respectively.

Steven Skiena

Steven Skiena is a Distinguished Teaching Professor of Computer Science and Director of the Institute for AI-Driven Discovery and Innovation at Stony Brook University. His research interests include the design of graph, string, and geometric algorithms, and their applications (particularly to biology). He is the author of six books, including "*The Algorithm Design Manual*," "*The Data Science Design Manual*", and "*Who's Bigger: Where Historical Figures Really Rank*". Skiena received his PhD in Computer Science from the University of Illinois in 1988, and the author of over 150 technical papers. He is a former Fulbright scholar, and recipient of the ONR Young Investigator Award and the IEEE Computer Science and Engineer Teaching Award. More info at cs.stonybrook.edu/~skiena/

Scott Smolka

Scott A. Smolka earned his PhD in Computer Science from Brown University in 1984; MA and BA in Mathematics from Boston University in 1977 and 1975, respectively. Scott A. Smolka is a recipient of the Fellow of the European Association of Theoretical Computer Science, since 2016; Research Excellence Award, Department of Computer Science, Stony Brook University, 2012; Best Paper Award, Second International Conference on Runtime Verification (RV'11), 2011; 2008-2009 President/Chancellor's Award for Excellence in Scholarship and Creative Activities, 2009; Computer Science Department Certificate of Appreciation for departmental, university and community service, especially his leadership role in CS@35, the day-long event/fundraiser celebrating the Computer Science Department's 35th anniversary, 2006.

Researcher's Biographies

Milutin Stanaćević

Milutin Stanaćević received the BS degree in Electrical Engineering from the University of Belgrade, Belgrade, Serbia, in 1999, and the MS and PhD degrees in Electrical and Computer Engineering from Johns Hopkins University, Baltimore, MD, in 2001 and 2005, respectively. In 2005, he joined the faculty of the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY, where he is currently an Associate Professor. Dr. Stanaćević is a recipient of the National Foundation Career Award and IEEE Region 1 Technological Innovation Award. He was an Associate Editor of the IEEE Transactions on Biomedical Circuits and Systems and now serves on several technical committees of the IEEE Circuits and Systems Society. His research interests include the mixed-signal VLSI circuits, systems, and algorithms for parallel multi-channel sensory information processing with emphasis on real-time acoustic source localization and separation, micro-power biomedical instrumentation and RF backscatter-based communication and sensing.

Scott D. Stoller

Scott Stoller is affiliated with the National Security Institute. He received his PhD in Computer Science from Cornell University in 1997 and BA in Physics, summa cum laude from Princeton University in 1990. Scott Stoller's primary research interests are design, analysis, optimization, testing, and verification of software, with focuses on computer security, concurrency, and incremental computation.

Allen Tannenbaum

Allen Tannenbaum is affiliated with the Department of Applied Mathematics and Statistics. He obtained his PhD from Harvard University. Allen Tannenbaum research focuses on Medical image analysis; computer vision; image processing; systems and control; controlled active vision; mathematical systems theory; bioinformatics; and computer graphics.

T.A. Venkatesh

T. A. Venkatesh is currently an Associate Professor in the Materials Science and Chemical Engineering department at Stony Brook University. He received his doctoral degree in Materials Science and Engineering and a minor in business administration from the Sloan School of Management, at the Massachusetts Institute of Technology, Cambridge, MA. Through modeling and experimental approaches, his research group's efforts are directed towards understanding the fundamental structure-property relationships in advanced materials over multiple (structural or functional) length-scales, the appropriate contexts for which are formulated by materials issues in the Nanotechnology, Biotechnology and Energy Technology areas. His research activities have been supported by federal agencies (National Science Foundation and Department of Energy), state agencies (NYSERDA) and industry. He has published over 50 papers and his publications have received over 2500 citations.
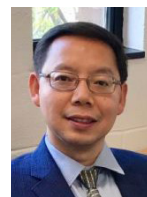
Fusheng Wang

Fusheng Wang received his PhD in Computer Science from University of California, Los Angeles, and MS and BS in Engineering Physics from Tsinghua University, China. Prior to joining Stony Brook University, he was an assistant professor at Emory University. He was a research scientist at Siemens Corporate Research (Princeton, NJ) before joining Emory University.

Xin Wang

Xin Wang is currently an associate professor of the department of Electrical and Computer Engineering. Prior to joining the department, she was an assistant professor in the department of computer science and engineering of SUNY at Buffalo between 2003 and 2005, and was a Member of Technical Staff in the area of mobile and wireless networking at Bell Labs Research, Lucent Technologies, New Jersey between 2001 and 2003. Her research interests include mobile and ubiquitous computing, wireless communications and network systems, networked sensing and fusion, networked autonomous systems, detection and estimation. She received her BS and MS degrees in telecommunications engineering and wireless communications engineering from Beijing University of Posts and Telecommunications, Beijing, China, respectively, and her PhD degree in electrical engineering from Columbia University, New York, NY. She received the prestigious NSF CAREER award in 2005.

Song Wu

Assistant Professor Song Wu runs a statistics lab within the Department of Applied Mathematics Statistics and in the College of Engineering and Applied Sciences at Stony Brook University. Physically, the lab is located in the Center of Excellence in Wireless and Information Technology (CEWIT). Wu received his BS in Biological

Sciences from the University of Science and Technology of China and his PhD in Statistics from the University of Florida.
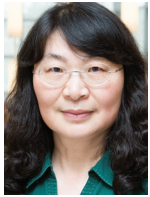
Yuanyuan Yang

Yuanyuan Yang received the BEng and MS degrees in computer science and engineering from Tsinghua University, Beijing, China, and the MSE and PhD degrees in computer science from Johns Hopkins University, Baltimore, Maryland. Dr. Yang is a SUNY Distinguished Professor in the Department of Electrical and Computer Engineering and Department of Computer Science, and Associate Dean for Diversity and Academic Affairs, College of Engineering and Applied Sciences, at Stony Brook University. Prior to joining Stony Brook in 1999, she had held a tenured faculty position at University of Vermont. Dr. Yang is internationally recognized for her contributions in networking and parallel and distributed computing systems areas. She was elected as an IEEE Fellow in 2009 "for contributions to parallel and distributed computing systems." Her current research interests include cloud computing, data center networks, edge computing, and wireless/mobile networks. Dr. Yang currently is the Associate Editor-in-Chief for the *IEEE Transactions on Cloud Computing*. She has served as the Associate Editor-in-Chief for the *IEEE Transactions on Computers* and an associate editor for the *IEEE Transactions on Computers* and *IEEE Transactions on Parallel and Distributed Systems*. She is also an associate editor for the *Journal of Parallel and Distributed Computing*. Dr. Yang has published more than 350 scientific papers in leading refereed journals, conferences and book chapters. She is an inventor/co-inventor of seven U.S. patents in the area of interconnection networks.
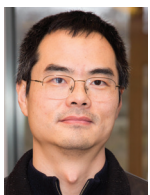
Fan Ye

Fan Ye received his PhD from the Computer Science Department of UCLA in 2004 and then joined IBM T. J. Watson Research Center as a Research Staff Member. Currently he is an Assistant Professor in the ECE department of Stony Brook University. His research interests include mobile sensing platforms, systems and applications, Internet-of-Things, indoor location sensing, wireless and sensor networks. He has published over 60 peer reviewed papers that have received over 7000 citations according to Google Scholar. He has 21 granted/pending US and international patents/applications. He was the co-chair for the Mobile Computing Professional Interests Community at IBM Watson for two years. He received IBM Research Division Award, five Invention Achievement Plateau awards, Best Paper Award for International Conference on Parallel Computing 2008. He has served as panelist multiple times for NSF and Hong Kong Government Research Councils, on program and organizing committees for conferences including ACM Mobicom, ACM Sensys and IEEE Infocom.
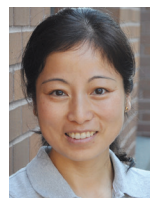
Wei Yin

Wei Yin received her MS in Biomedical Engineering from the University of Akron in 2001 and her PhD in Biomedical Engineering from Stony Brook University in 2004. She is an associate professor in the Biomedical Engineering department at Stony Brook University. Her research focus is on cardiovascular disease is the leading cause of death in the United Sates, and coronary artery disease is the most common type of cardiovascular disease.

Erez Zadok

Erez Zadok is affiliated with the Smart Energy Technologies Cluster. He completed his PhD in Computer Science from Columbia University in 2001. He directs the File Systems and Storage Lab (FSL) at the CS department.
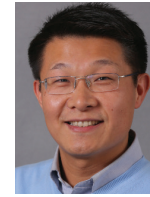
Rong Zhao

Rong Zhao is Director of Software Systems Division of CEWIT, a faculty member of Department of Computer Science, and an adjunct faculty member of the College of Business and Department of Technology and Society at Stony Brook University. At CEWIT, he is responsible for leading the research and development of technology-based solutions. He is Managing Director of the Center for Visual and Decision Informatics (CVDI), which is an Industry/University Cooperative Research Center (I/UCRC) designated by the National Science Foundation (NSF). At CVDI, he is in charge of managing the center operation and developing collaborative projects in big data management, analytics and visualization. He is also a coordinator for the New York State Strategic Partnership for Industrial Resurgence (SPIR) program.

Wei Zhu

Wei Zhu is a statistician whose diverse research projects include design and analysis of experiments (for dose response studies and clinical trials), longitudinal and contemporaneous pathway discovery and analysis (for biological pathways and financial networks), errors in variable modeling (for measurement platform/instrument comparisons and calibrations), and robust regression analysis. Her former students have positions in academia and industry as quality control managers, biostatisticians, risk managers and financial analysts. Professor Zhu is affiliated with the SBU Center for Finance (https://www.stonybrook.edu/cf/).

**CENTER OF EXCELLENCE IN WIRELESS
AND INFORMATION TECHNOLOGY (CEWIT)**

Stony Brook University's Research and Development Park
1500 Stony Brook Road
Stony Brook, NY 11794-6040
Phone: (631) 216-7000
Email: info@cewit.org

cewit.org