COMPUTER SCIENCE AT COLUMBIA UNIVERSITY

The Department of Computer Science at Columbia University

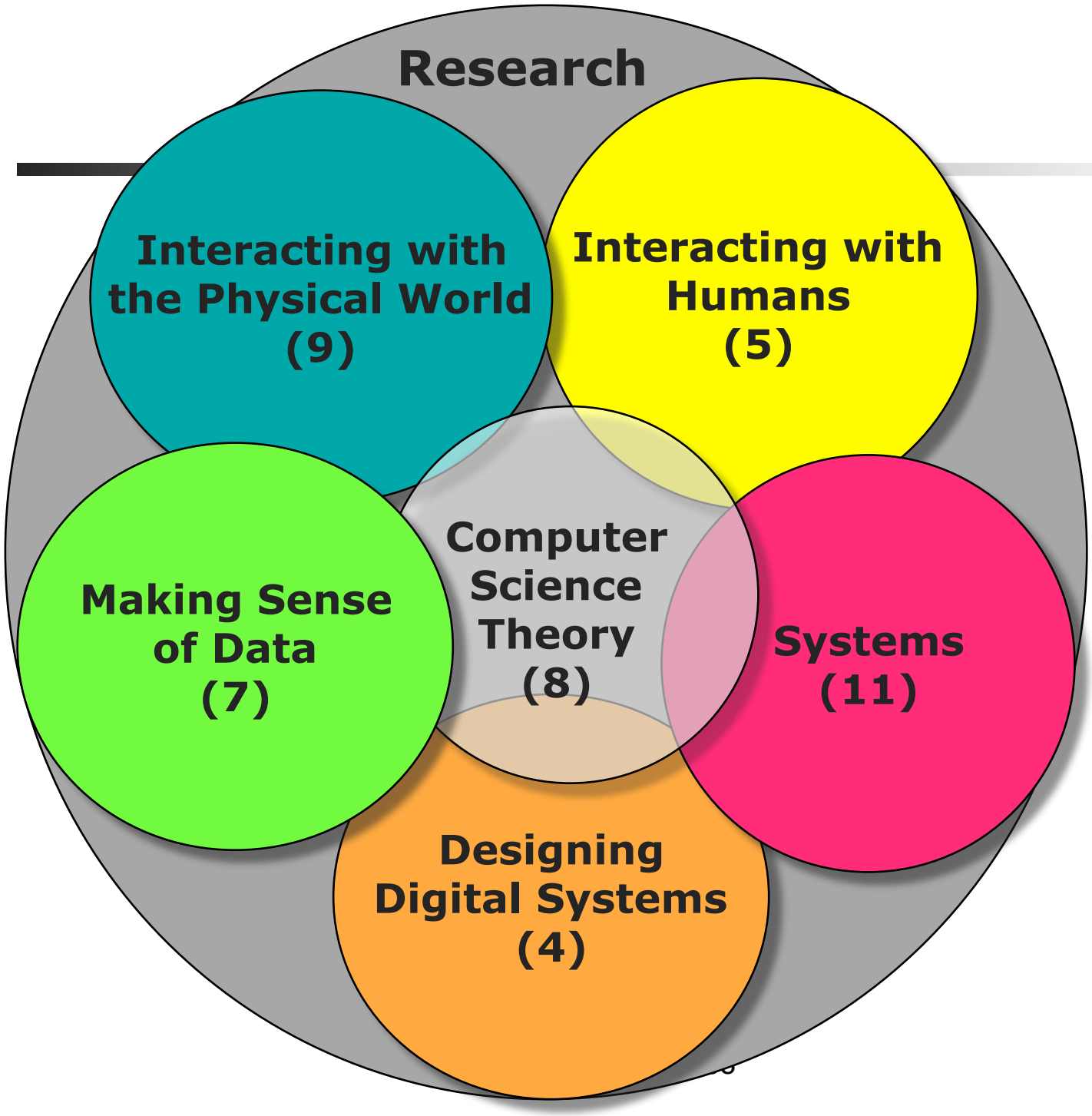
Henning Schulzrinne, Chair
Dept. of Computer Science
Columbia University
2008

Columbia Computer Science in Numbers

- ~34 full-time faculty and lecturers
  - + visitors, postdocs, adjunct faculty, joint appointments (EE, IEOR), …
- 103 PhD students (13 new arrivals)
- ~200 MS students (137 new arrivals)
- 60 CS + CE undergraduate juniors & seniors

CS@CU

Faculty: 35 (33 tenure track, 1 lecturer, 1 joint)


Research
Interacting with the Physical World (9)
Interacting with Humans (5)
Making Sense of Data (7)
Computer Science Theory (8)
Systems (11)
Designing Digital Systems (4)

CS@CU

Research areas

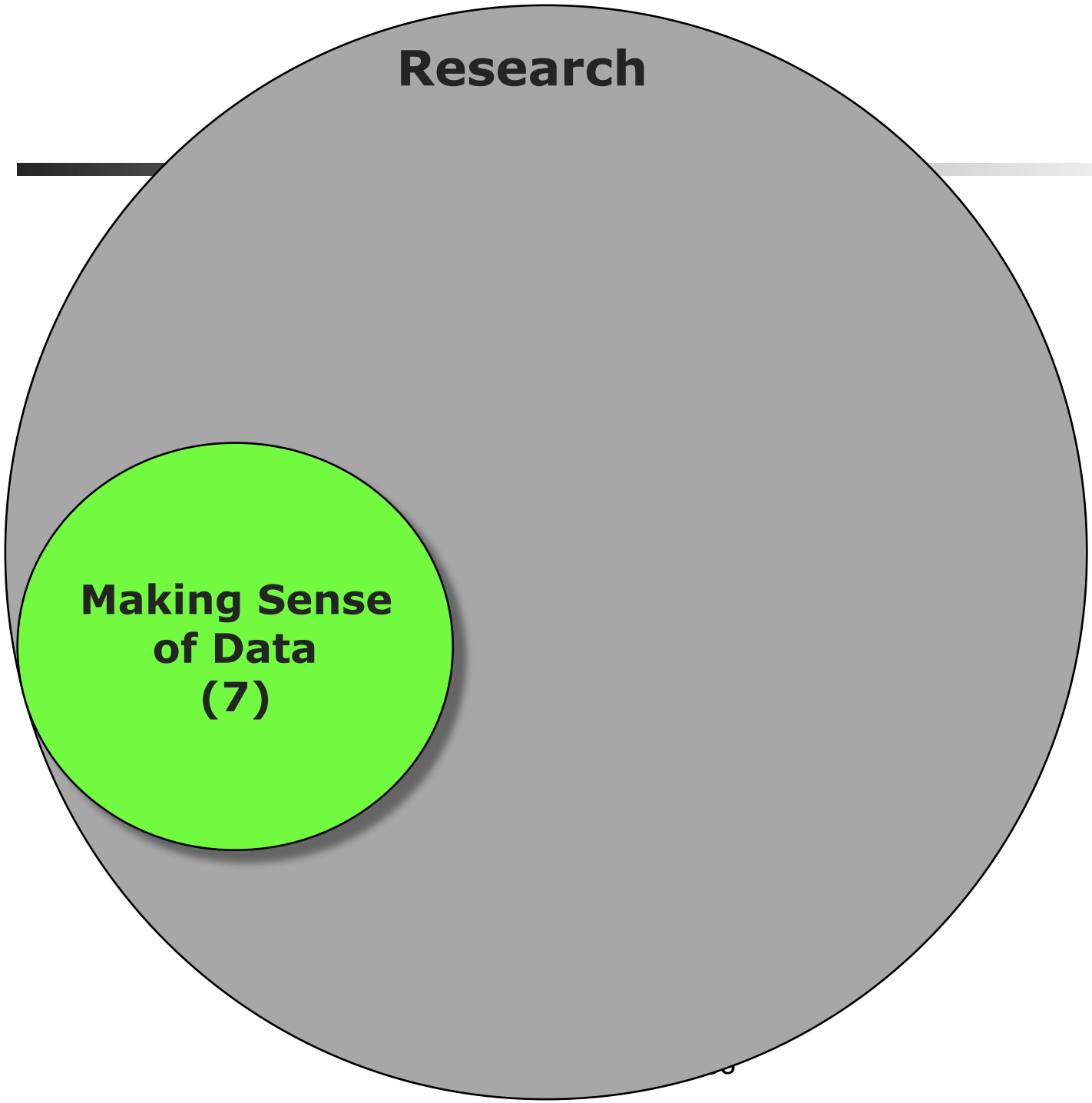
Interacting with the Physical World	*graphics, robotics, vision*	*Allen, Belhumeur, Feiner, Grinspun, Grunschlag, Jebara, Kender, Nayar, Ramamoorthi*
Interacting with Humans	*user interfaces, natural language and speech processing, collaborative work, personalized agents*	*Feiner, Hirschberg, Kaiser, Kender, McKeown*
Systems	*networks, distributed systems, security, compilers, software engineering, programming languages, OS*	*Aho, Bellovin, Edwards, Kaiser, Keromytis, Malkin, Misra, Nieh, Schulzrinne, Stolfo, Yang, Yemini*
Designing Digital Systems	*digital and VLSI design, CAD, asynchronous circuits, embedded systems*	*Carloni, Edwards, Nowick, Sethumadhavan*
Making Sense of Data	*databases, data mining, Web search, machine learning applications, computational biology*	*Cannon, Gravano, Jebara, Kaiser, Pe'er, Ross, Servedio, Stolfo*
Computer Science Theory	*cryptography, quantum computing, complexity, machine learning theory, graph theory, algorithms*	*Aho, Galil, Gross, Malkin, Servedio, Traub, Wozniakowski, Yannakakis*

CS @CU

CCLS: A Research Center in CS

The Center for Computational Learning Systems (CCLS) aims to be a world leader in learning and data mining research and the application of this research to natural language understanding, the World Wide Web, bioinformatics, systems security and other emerging areas. CCLS will emphasize interdisciplinary efforts with other departments at Columbia, and will leverage Columbia's CS Department's strengths in learning, data mining and natural language processing, extending the effective size and scope of the Department's research effort.

CS@CU


Research
Making Sense of Data (7)

Columbia's Database Group

http://www.cs.columbia.edu/database

Databases, data mining, information retrieval, web search

Faculty

Luis Gravano

Ken Ross

Mihalis Yannakakis

Ph.D. Students

John Cieslewicz

Wisam Dakka

Alpa Jain

Julia Stoyanovich

CS @CU

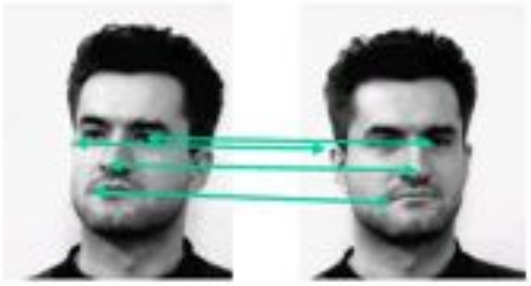
Some Projects in Gravano's "Subgroup" **http://www.cs.columbia.edu/~gravano**

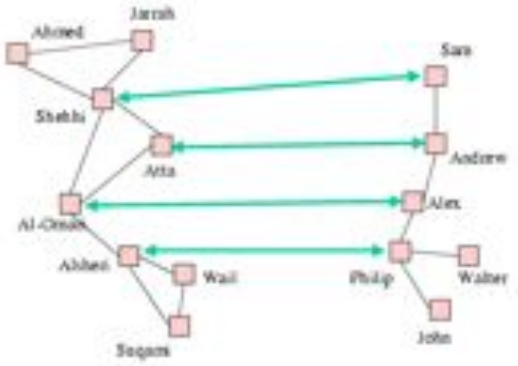
- Snowball, an information-extraction system
  http://snowball.cs.columbia.edu
- QProber, a system for classifying and searching "hidden-web" databases
  http://qprober.cs.columbia.edu
- SDARTS, a protocol and toolkit for metasearching/distributed information retrieval
  http://sdarts.cs.columbia.edu
- RANK: "top-*k*" query processing
  http://rank.cs.columbia.edu

CS @CU

Machine Learning Lab

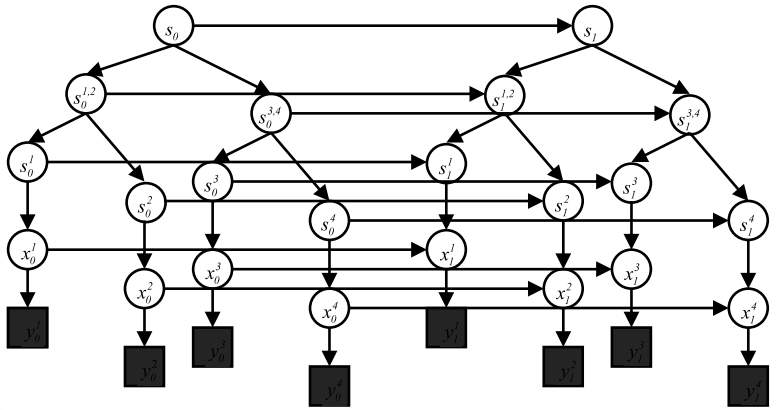
- Prof. Tony Jebara www.cs.columbia.edu/learning
- Computational statistics and algorithms for finding patterns in data and making predictions
- Theme: how to extend statistics to novel, multidimensional and structured data
- Data: images, text, time series, social nets

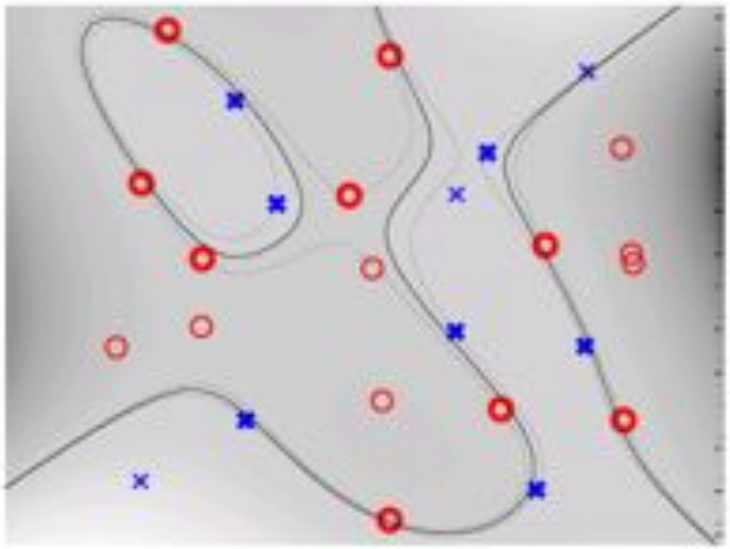


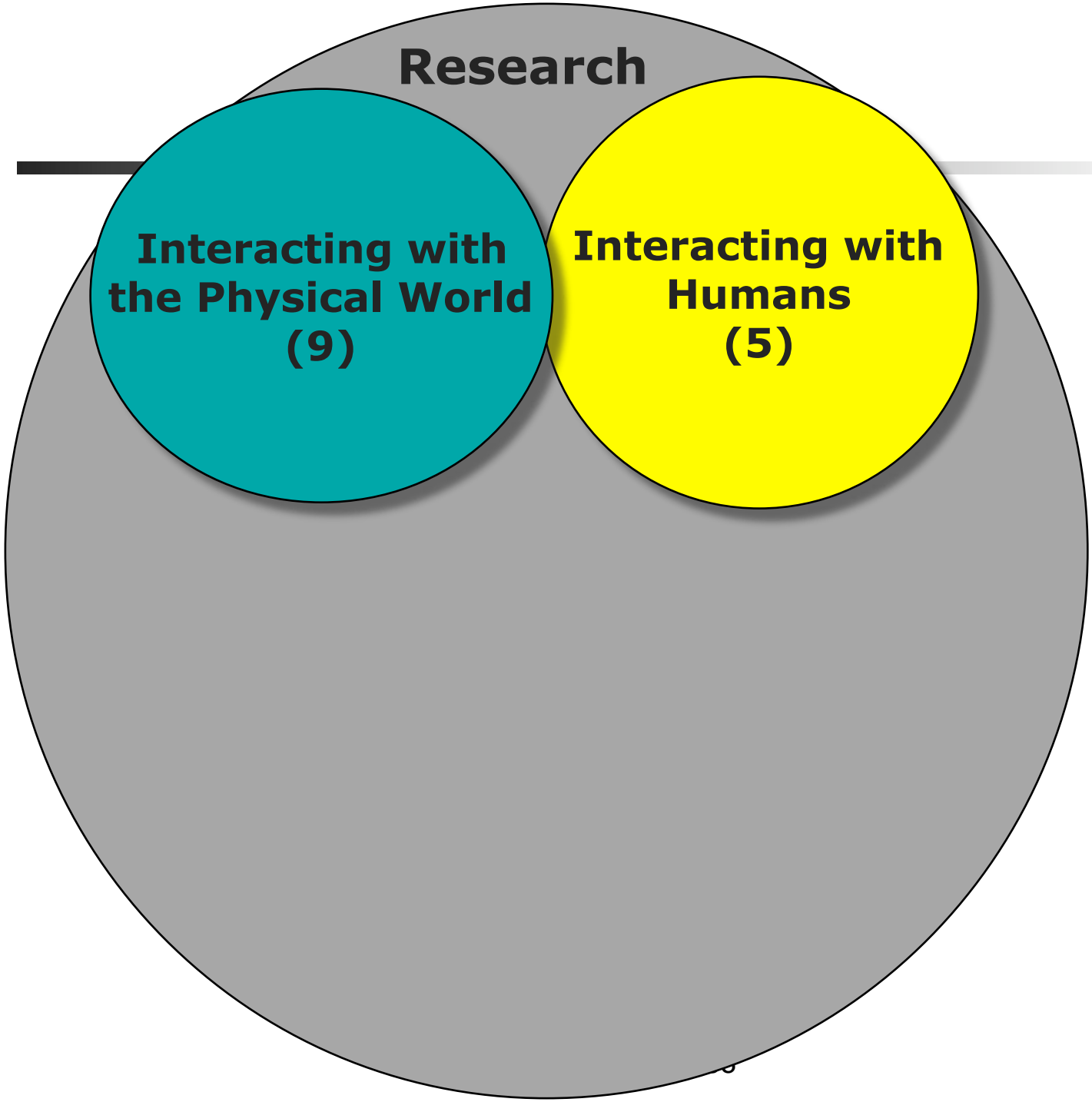


Machine Learning Lab

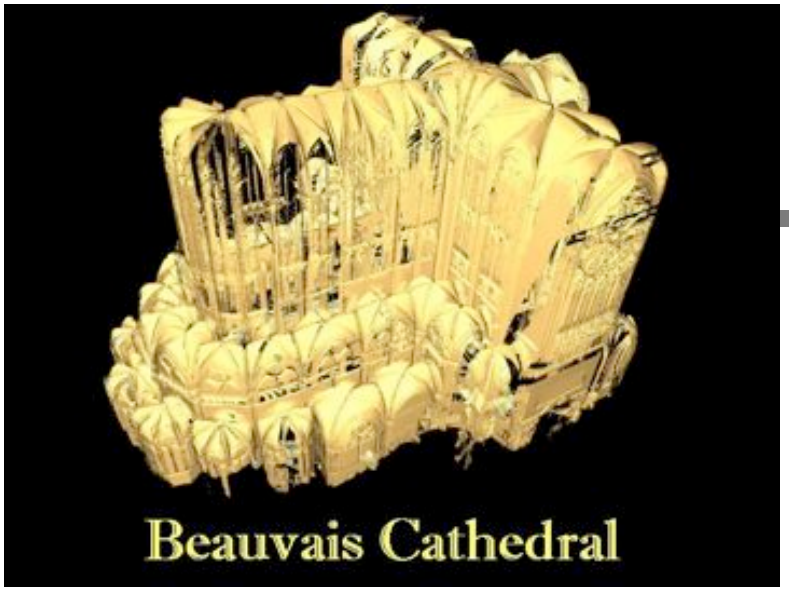
- Tools: Bayes Nets, Support Vector Machines, Representation, Invariance





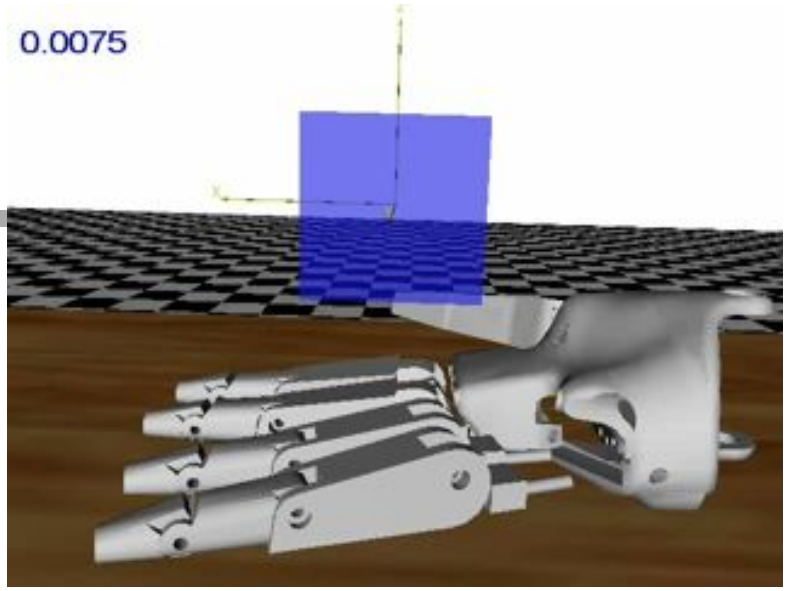

Research
Interacting with the Physical World (9)
Interacting with Humans (5)

CS@CU

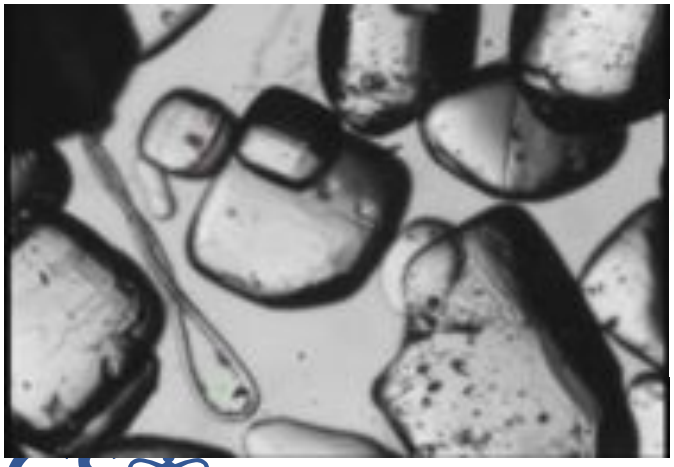


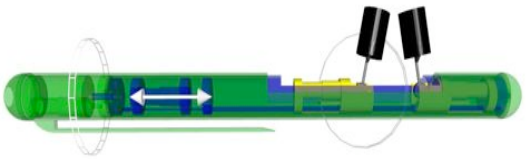
3-D Site Modeling

Columbia University Robotics Group



Graspit! Simulator

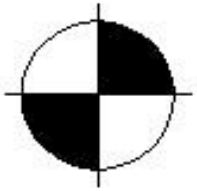




Robotic Crystal Mounting

Mobile Robotics

Columbia University
Robotics Group



Prof. Peter Allen

- Current Projects:

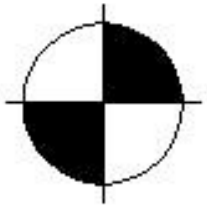
1. 3-D Modeling: Combining laser scanning and computer vision to create photorealistic models. Current NSF ITR project includes scanning Beauvais Cathedral in France and ancient ruins in Sicily
2. Robotic and human hand simulation using our Graspit! simulator which includes full dynamics, grasp quality measures, and grasp learning
3. Microscale protein crystal mounting using visual control. Microscope camera used to track/pick up very small crystals for x-ray diffraction
4. AVENUE mobile scanning robot: automating the site modeling process using GPS, wireless network, computer vision and range scanning
5. New insertable stereo cameras with pan, tilt and translation for minimally-invasive surgery

- People:

Postdocs: Atanas Georgiev and Andrew Miller
GRA's: Paul Blaer, Alejandro Troccoli, Ben Smith
M.S.: Rafi Pelosoff, Alex Haubald

CS @CU

Columbia University
Robotics Group



Goal: Creating intelligent machines and systems

Collaborative Research:

- Molecular Biology (crystal mounting)
- Art History (3D Modeling)
- Biomechanics (human hand simulation)
- Surgery (next-generation surgical imaging)

One of the labs affiliated with CVGC (Columbia Vision and Graphics Center)

Research opportunities include a wide range of software, hardware and systems projects.

Expertise in robotics, graphics, or vision is helpful

CS@CU

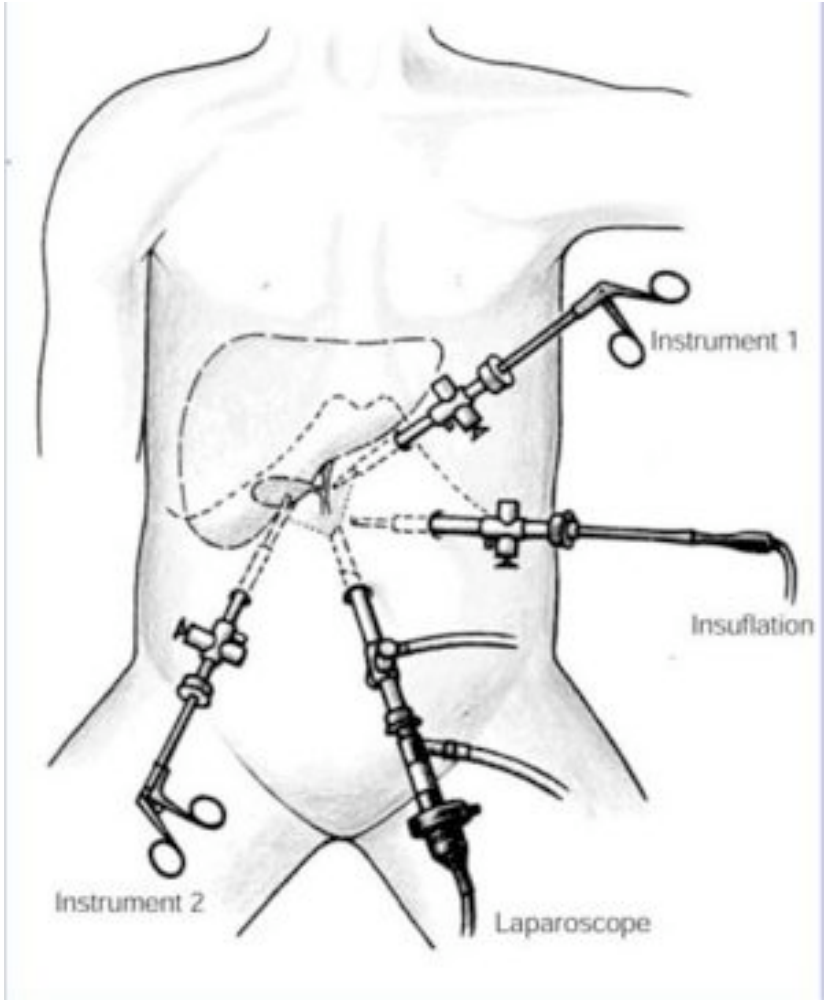
COMPUTER SCIENCE AT COLUMBIA UNIVERSITY

Insertable Imaging and Effector Platforms for Robotic Surgery

Peter Allen
Dennis Fowler (Dept. of Surgery)
Andrew Miller

http://www.cs.columbia.edu/robotics

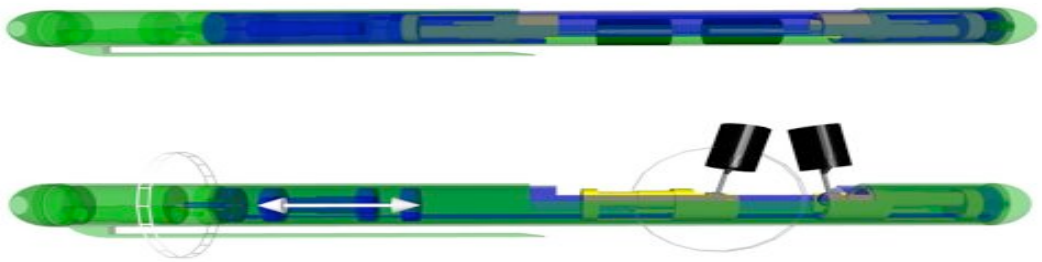
Current Laparoscopic Paradigm



- Multiple holes/insertion points
- Ports needed for each camera, instrument involved
- Limited range of motion at incision
- Pushing long sticks into small openings is still the idea!!!
- Assistant(s) needed to control camera
- Monocular viewing
- Works well - but can we do better?

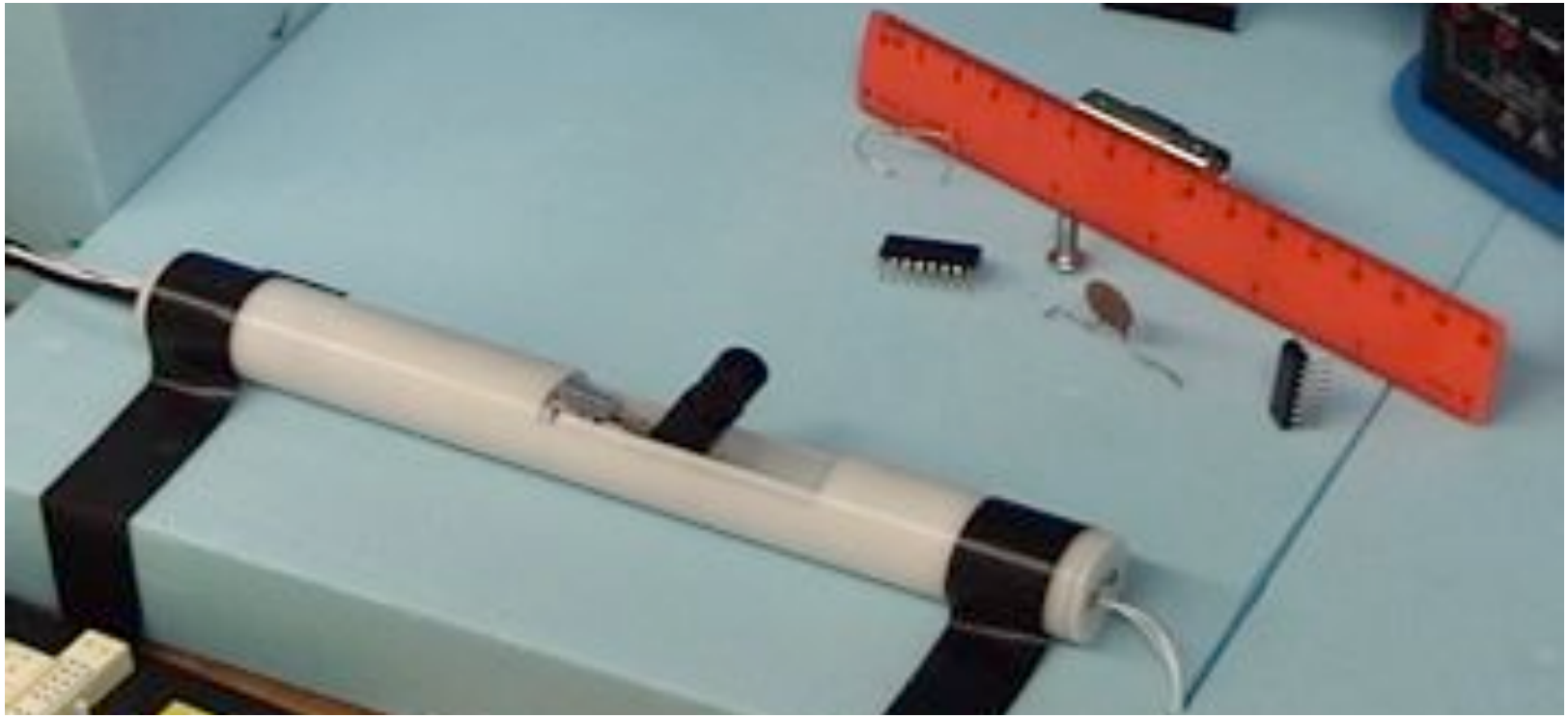
Next Generation Imaging Device

- Insertable unit
- 5 Degrees-of-freedom: 2 pan, 1 tilt, 2 translate
- Stereo Cameras
- More mobility for imaging
- Frees up incision port for other tooling



Single Camera Prototype

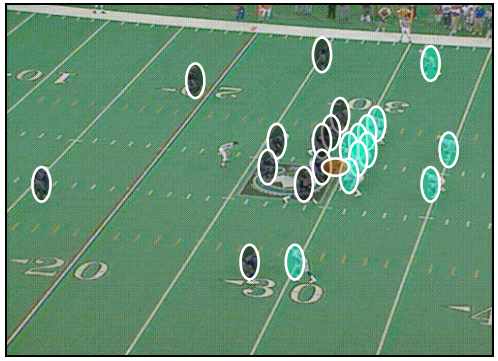
Diameter: 18mm; Length: 19cm
Camera opening: 5.8cm
Pan: 120°; Tilt: 130°; Translation: 5cm

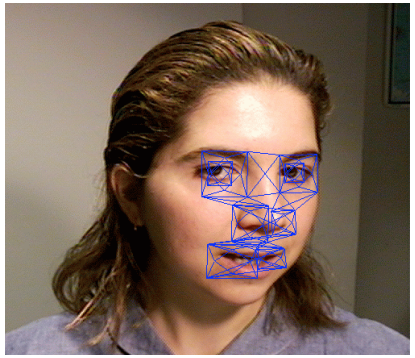


CS@CU

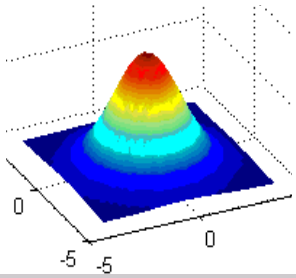
Video

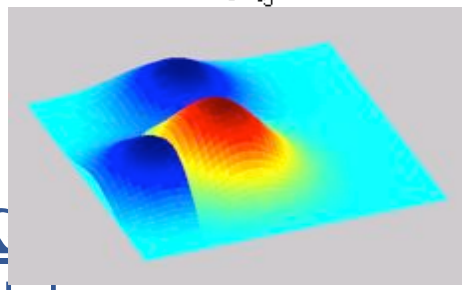
Computer Vision, Tracking People and Understanding Video

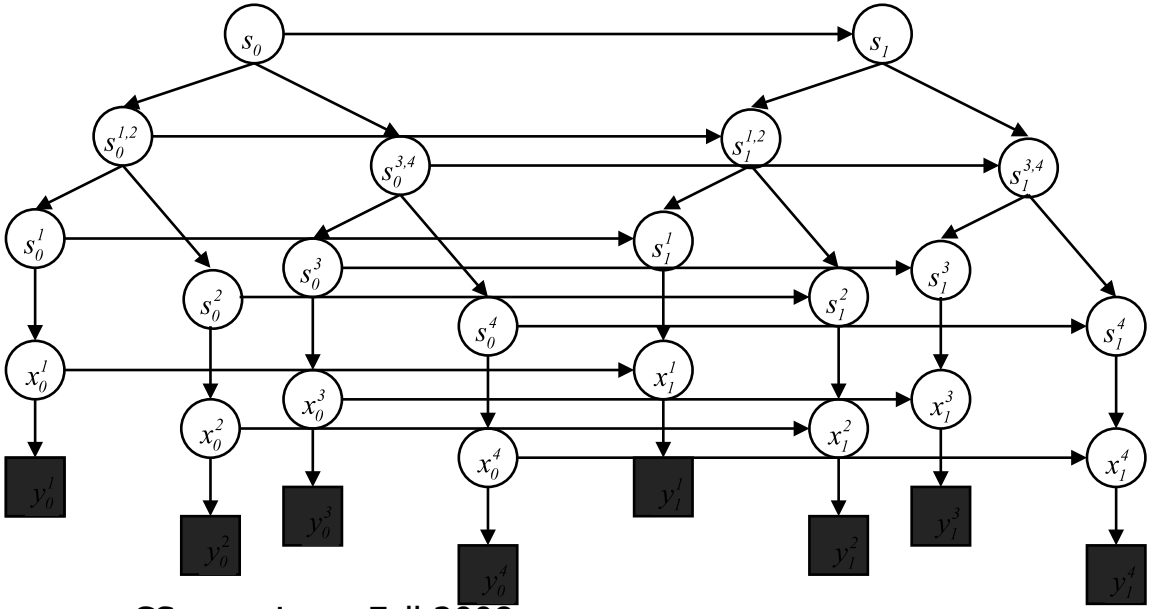



Discriminative Graphical Models

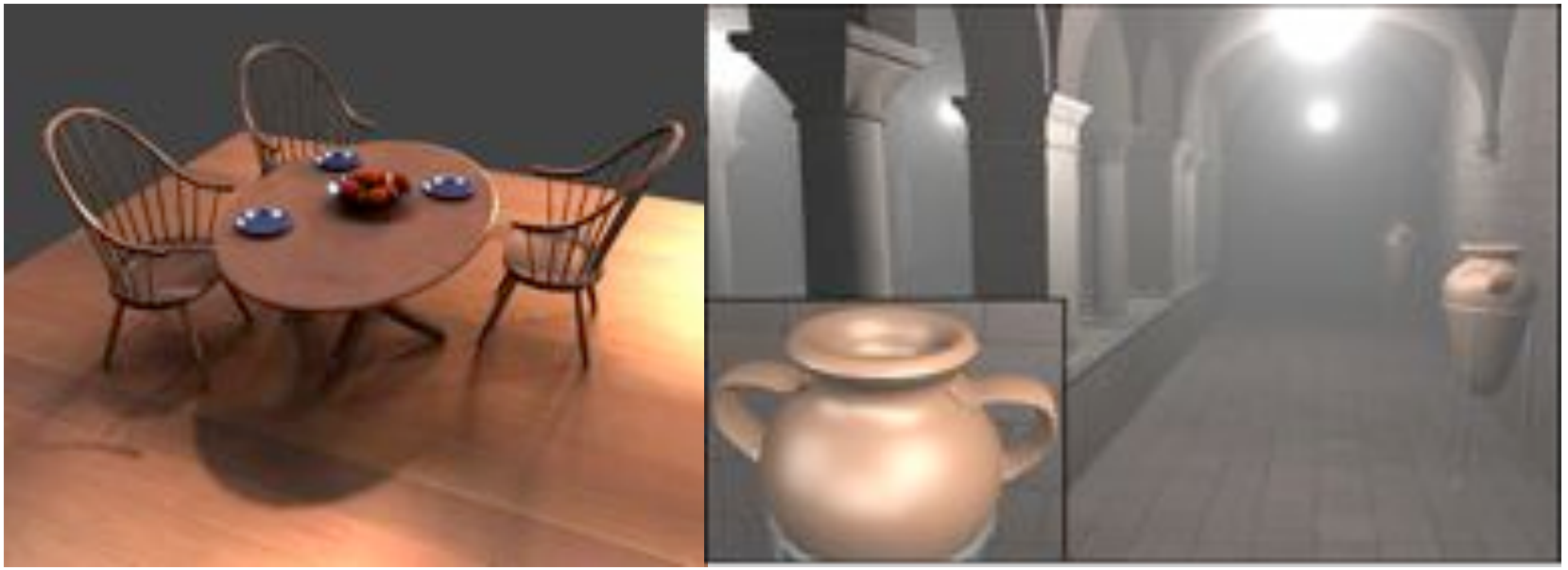






Computer Graphics Group

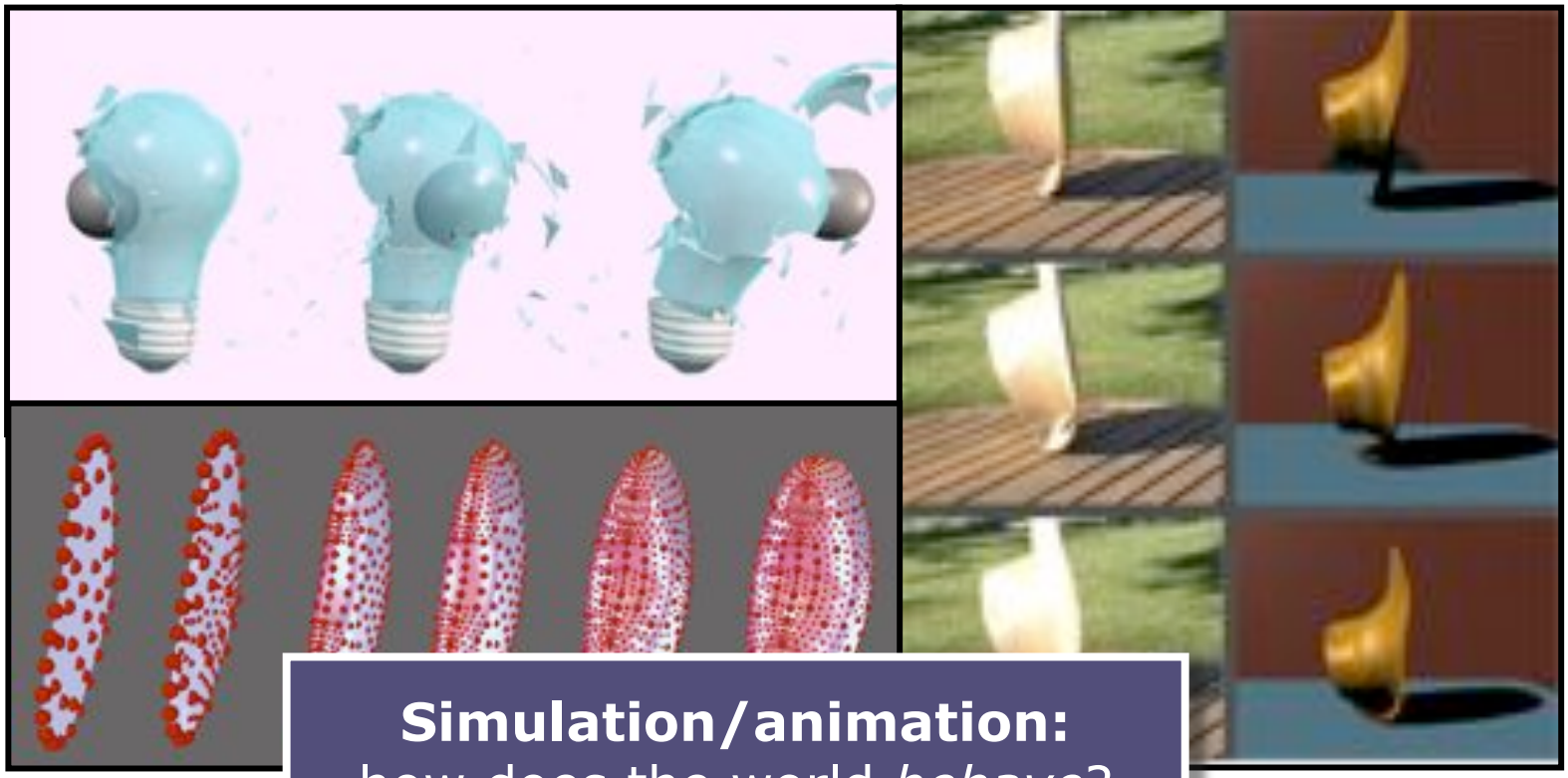
- Profs. Grinspun & Ramamoorthi
- Fundamental methods and math



Rendering:
how does the world *appear to us*?

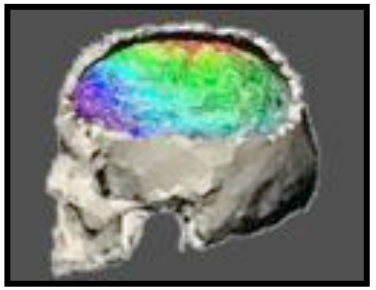
CS @CU

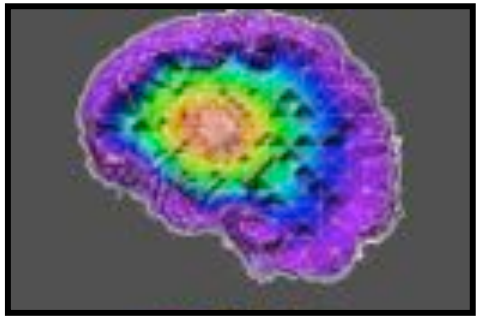
Computer Graphics Group



Simulation/animation:
how does the world *behave*?

Computer Graphics Group





geometric modeling:
representing and computing on geometric objects

Computer Graphics and User Interfaces Lab

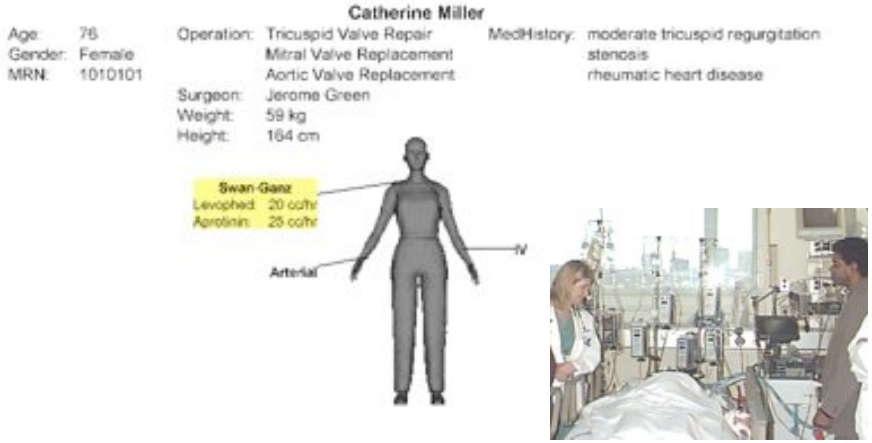
S. Feiner, H. Benko, G. Blaskó, S. Güven, D. Hallaway,
E. Ishak, S. White

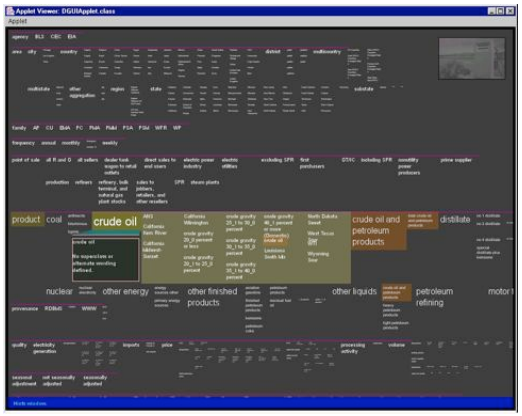
- Wearable UIs
- Augmented reality
- Virtual reality

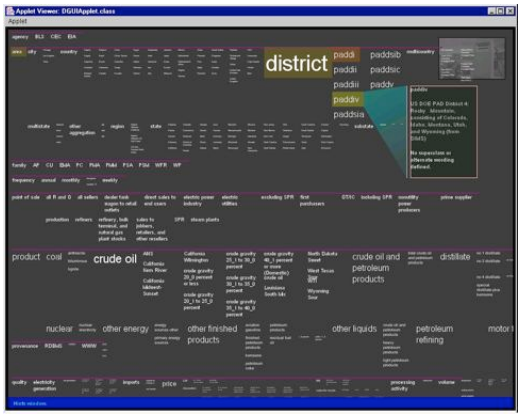
Computer Graphics and User Interfaces Lab

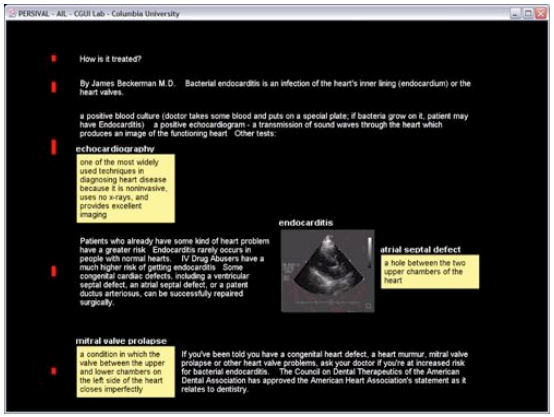
S. Feiner, H. Benko, G. Blaskó, S. Güven, D. Hallaway,
E. Ishak, S. White

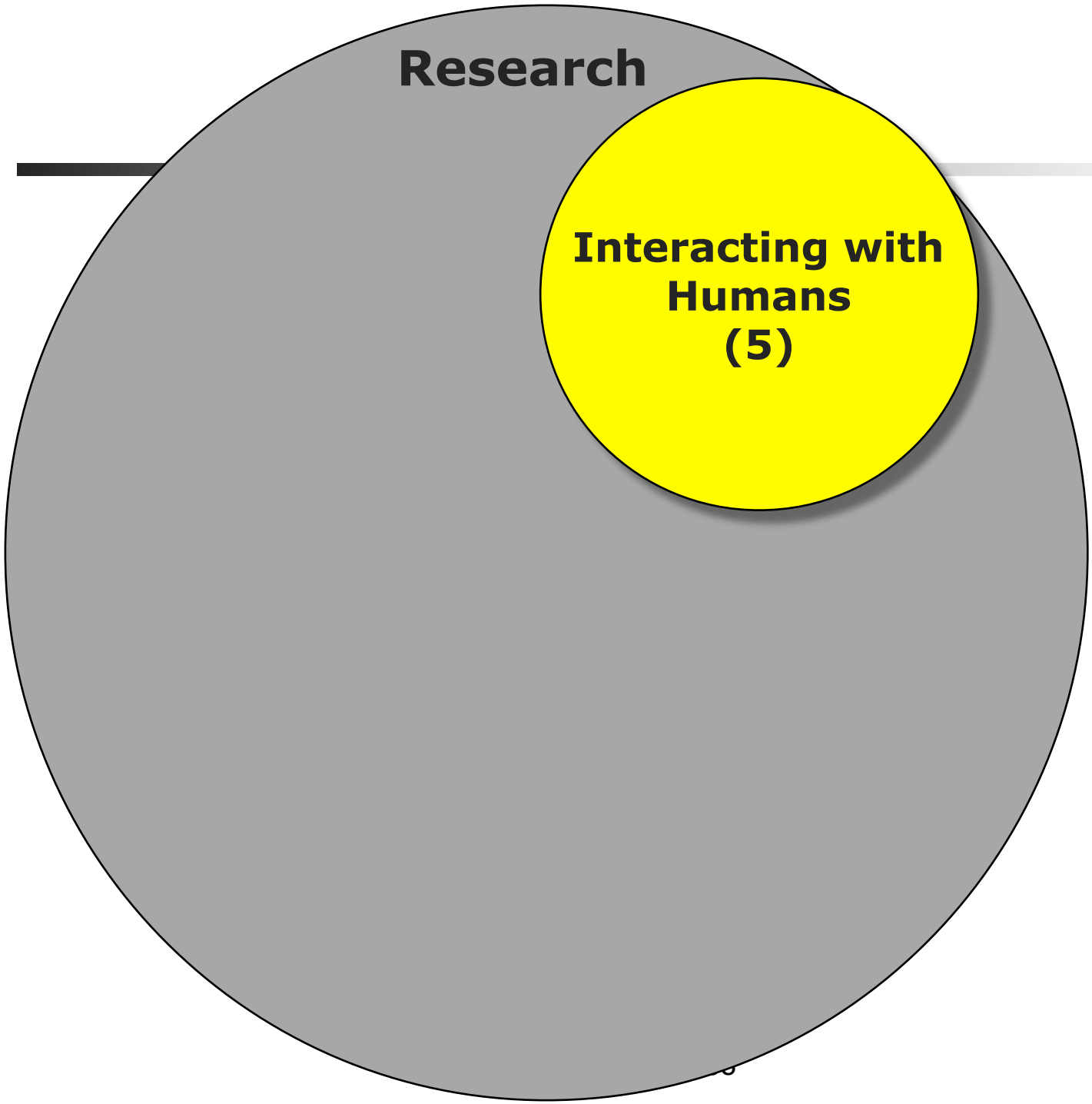
- Automated generation of graphics
- Display layout
- Coordination with text generation











CS @CU

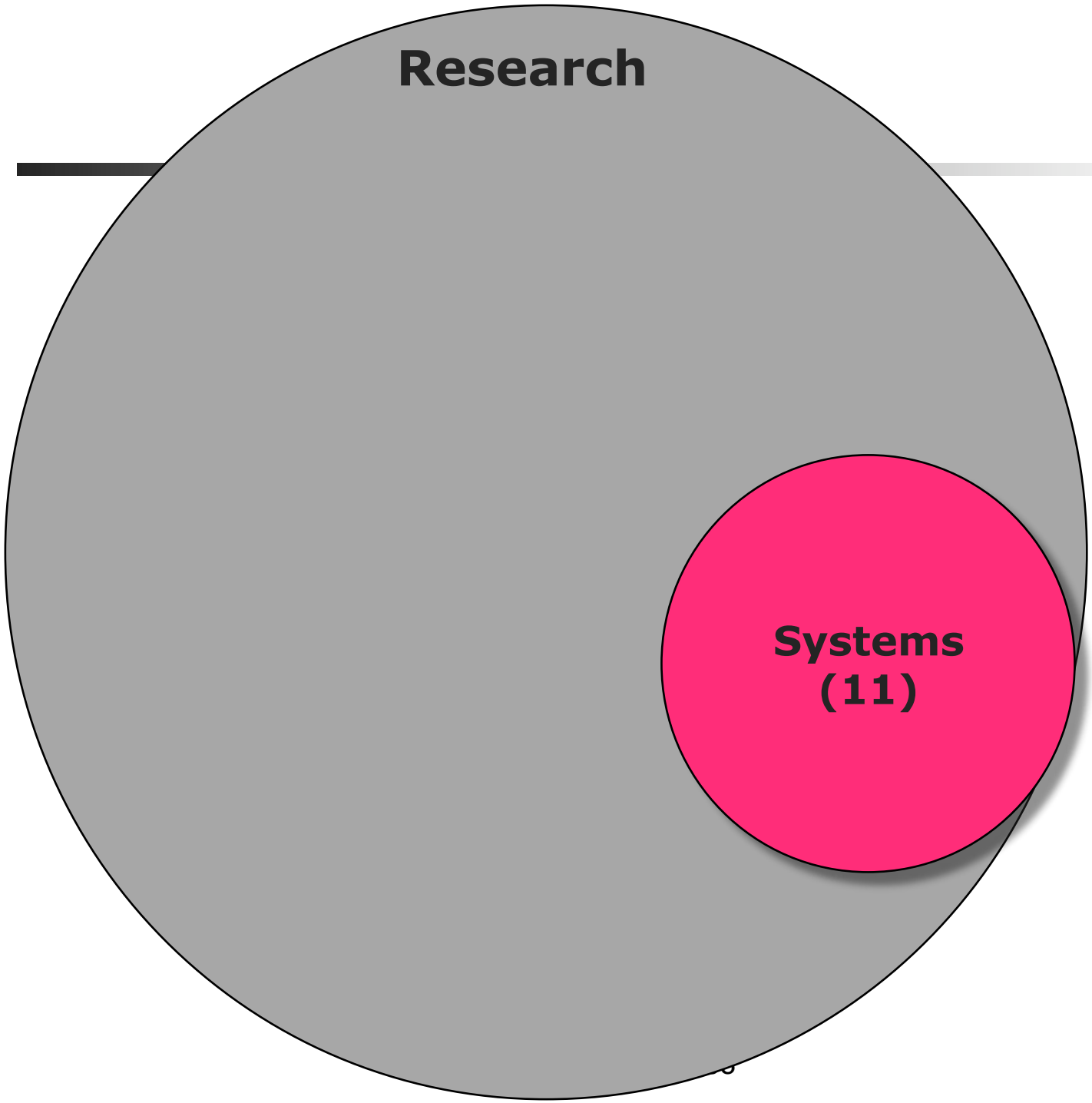
Spoken Language Processing Lab

Who we are:

Julia Hirschberg, Stefan Benus, Fadi Biadsy, Frank Enos, Agus Gravano, Jackson Liscombe, Sameer Maskey, Andrew Rosenberg

What we do:

- *Recognize and generate different speaker states – emotions (anger, uncertainty), charisma, deception*
- *Summarize spoken 'documents'*
- *Study spoken dialogue systems*
- *'Translate' prosody between English and Mandarin*


Research
Systems
(11)

Gail Kaiser:
Programming Systems Lab

enable (vt) : *to make possible, practical, or easy*

PSL

PROGRAMMING SYSTEMS LAB
COLUMBIA UNIVERSITY

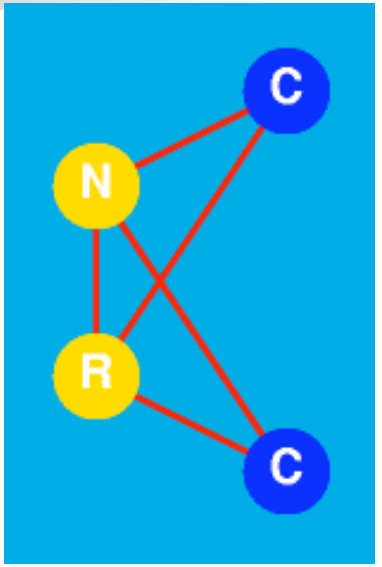
http://www.psl.cs.columbia.edu/

- Develop and empirically evaluate methodologies and technologies to enable *"better, faster, cheaper"* deployment and maintenance of large-scale software systems
- Seeking PhD, MS or advanced undergraduate students with substantial "real world" experience in any of compilers, operating systems, databases, computer security, networking, system administration
- Also seeking students interested in applied machine learning, power engineering, compbio (no experience required, just sincere interest)

- self-managing systems ("autonomic computing")
- software testing for emerging applications (e.g., for machine learning algorithms, bioinformatics databases, electrical distribution systems)
- novel architectures for special-purpose pub/sub event systems
- computer security
- software development environments and tools
- Multi-disciplinary projects

CS@CU

Networking research at Columbia University



- Columbia Networking Research Center
- spans EE + CS
- 15 faculty – one of the largest networking research groups in the US
- about 40 PhDs
- spanning optical networks to operating systems and applications
- theory (performance analysis) to systems (software, protocols)

CS @CU

Network Computing Laboratory

http://www.ncl.cs.columbia.edu

- Operating Systems
- Distributed Systems
- Scheduling and Resource Management
- Thin-Client and Network Computing
- Web and Multimedia Systems
- Performance Evaluation

CS@CU

Network Computing Laboratory
Recent Research Projects

- Zap: Transparent process migration
- VNAT: Mobile networking
- GR3: O(1) proportional share scheduling
- Thinc: WAN remote display protocol
- Certes: Inferring web client response times

CS @CU

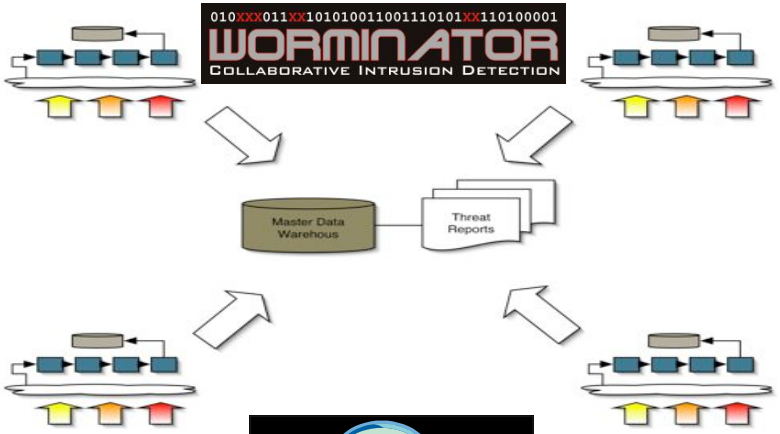
Columbia Intrusion Detection Lab (Sal Stolfo)

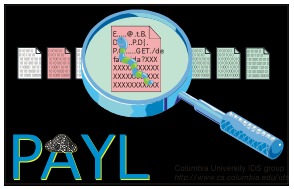
- Attackers continue to improve techniques undeterred –
  - Present COTS security defenses are porous and suffer from the false negative problem
- There is no one monolithic security solution; security is a design criteria at all layers of the stack and across multiple sites
- Behavior-based computer security will substantially raise the bar
- Columbia conducts a broad spectrum of research related to securing critical infrastructure in close collaboration with industry and government with attention to practical and deployable results
- Visit: http://www.cs.columbia.edu/faculty
  - http://www.cs.columbia.edu/ids
  - http://worminator.cs.columbia.edu

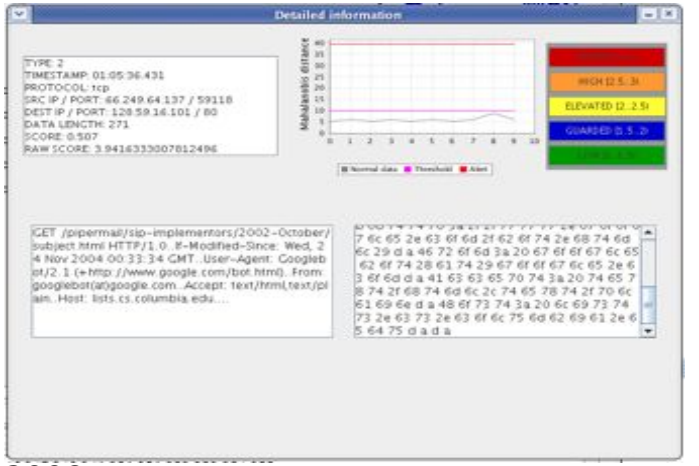
CS @CU

Columbia Intrusion Detection Lab: Anomaly Detection for Zero-Day Attack

- Worminator
  - Cross Domain Security Alert Sharing infrastructure
  - Modeling of attacker intent, and precursors to attack
- PAYL – Payload Anomaly Detection
  - Behavior-based detection of "abnormal" traffic
  - Zero-day exploits detected in network packet data flows
- EMT – Email Mining Toolkit
  - Forensic analysis of email logs for profile and model generation
  - Comparison of profiles/models
  - Detect malicious users/groups and aliases







COMPUTER SCIENCE AT COLUMBIA UNIVERSITY

EMT: Email Mining Forensic Analysis

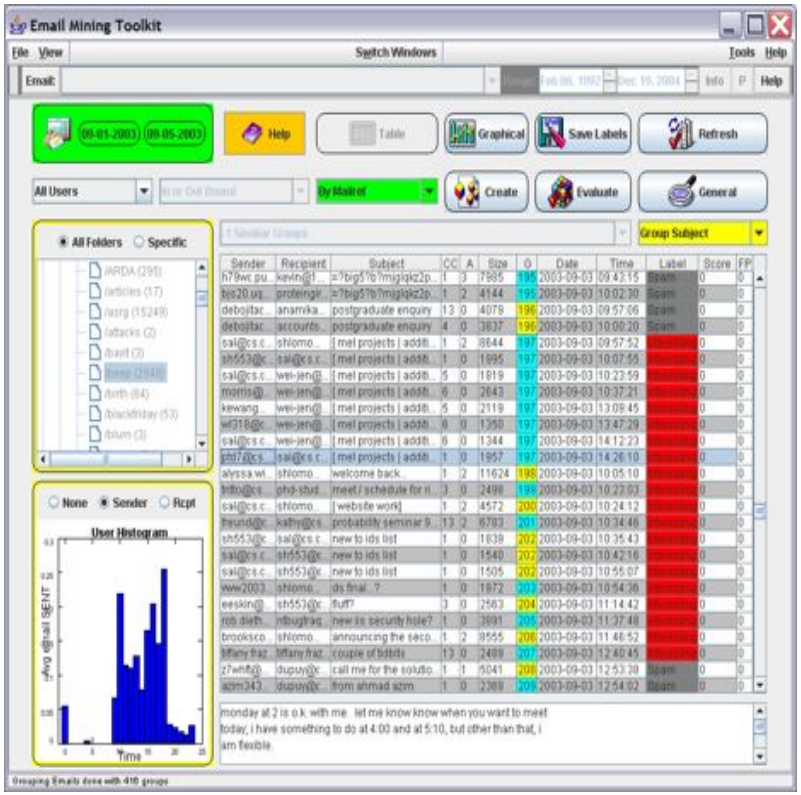
Prof Sal Stolfo
Columbia University
Computer Science Department

212.939.7080/sal@cs.columbia.edu

EMT Forensics

- Automatic system to acquire email data for study in a forensic environment
- Scalable to 100,000's of emails and attachments
- Automatically supports forensic tasks to be completed in seconds with analyst control over all variables and features
- Java-based application for email collection, analysis, and reporting in one integrated solution
- Pluggable architecture with API for easy customized extensions

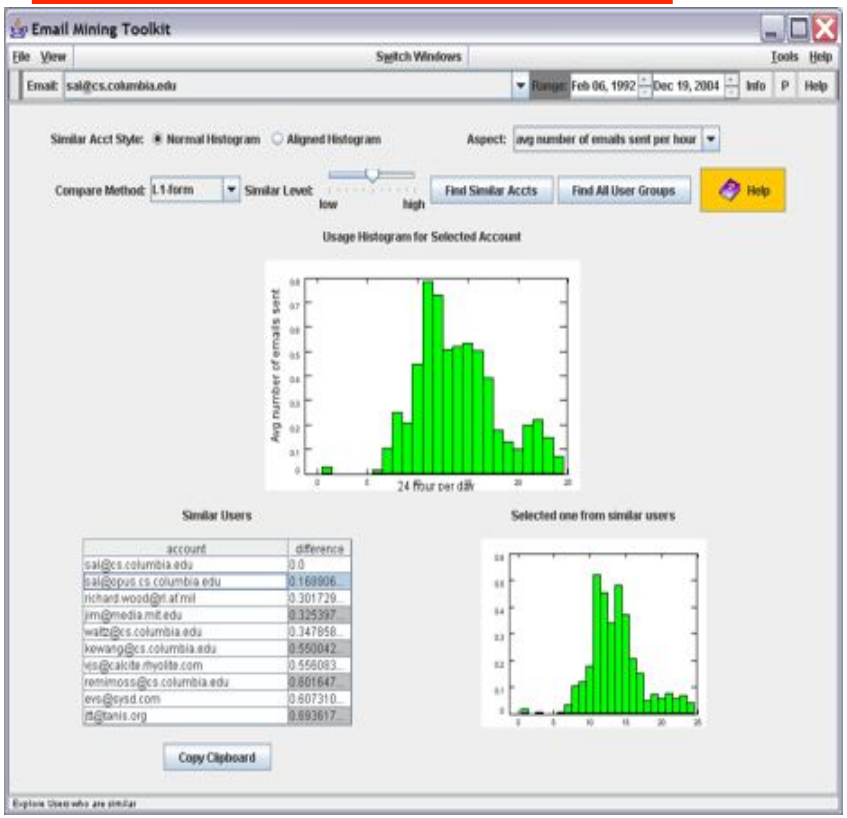
Main View of Email Archive



What might EMT do…

- Forensic analysis tasks for regulatory compliance
  - Which accounts are most important
  - Which accounts are behaving anomalously
  - Interesting behaviors between members of a social clique (clique violation or usage violation)
  - Who belongs to very many cliques

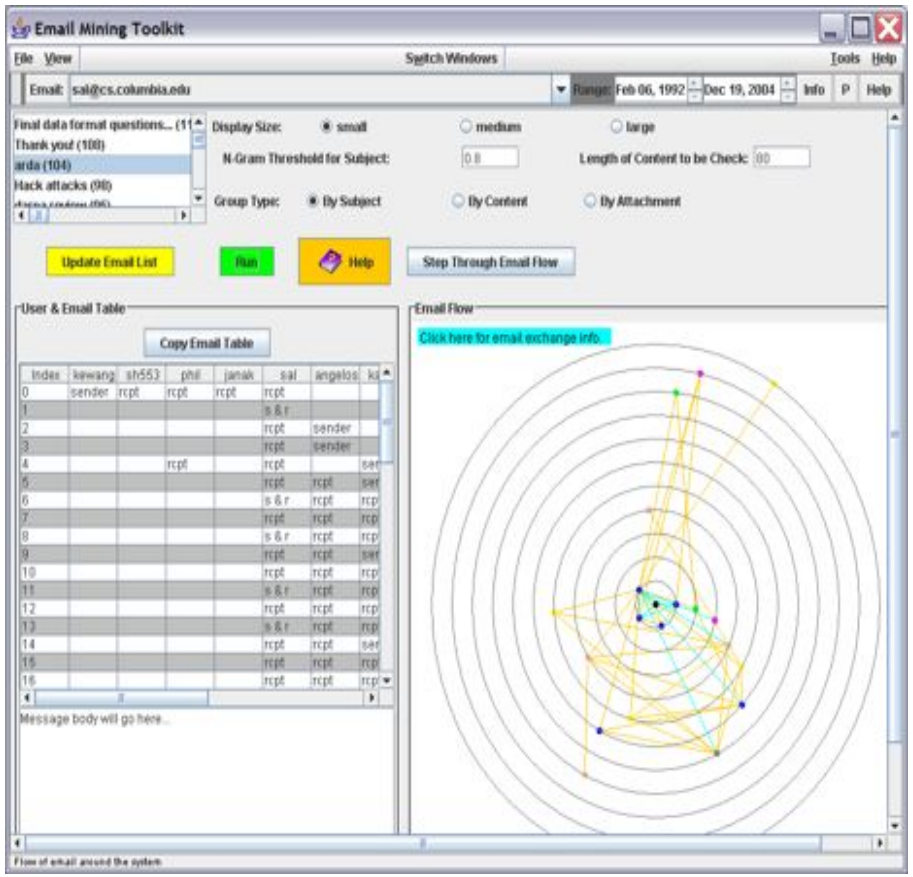
Who are the most important people in an organization and how do they behave?



What might EMT do…

- Managing organization information flow
  - Who communicates regularly with whom
  - Who has read my email
  - How does email flow through my organization

How does email flow over time?



Network Security Lab
Prof. Angelos D. Keromytis

- Applied research in security, networking, operating systems
  - Emphasis on systems and on building stuff
- Main research projects
  - Self-healing software and software security
- Application on countering network viruses/worms
  - Network denial of service
- Currently 6 Ph.D. students (Cook, Locasto, Burnside, Stavrou, Sidiroglou, Androulaki)
- Closely affiliated faculty: Stolfo, Bellovin, Ioannidis (CCLS), Yung

http://nsl.cs.columbia.edu/

CS @CU

NSL Projects

- Self-healing software
  - Enable legacy software to learn from its failures and improve itself over time, without human intervention!
- Network Worm Vaccine
  - Limit worm infection rate via anomaly detection engine and automatic patching of vulnerable software, based on self-healing concepts
- Resilience Against Denial of Service Attacks
  - Use network overlays as a mechanism for separating good and "bad" traffic
- High-speed I/O: The Operating System As a Signaling Mechanism
  - New OS architecture - remove memory and CPU from data path
- Efficient Cryptography
  - Design and implementation of ciphers for specific environments - use of graphics cards, variable size block ciphers, IXP processor
- Collaborative Distributed Intrusion Detection
  - Identifying global attack activity as well as "low and slow" scans via shared intrusion alerts across administrative domains

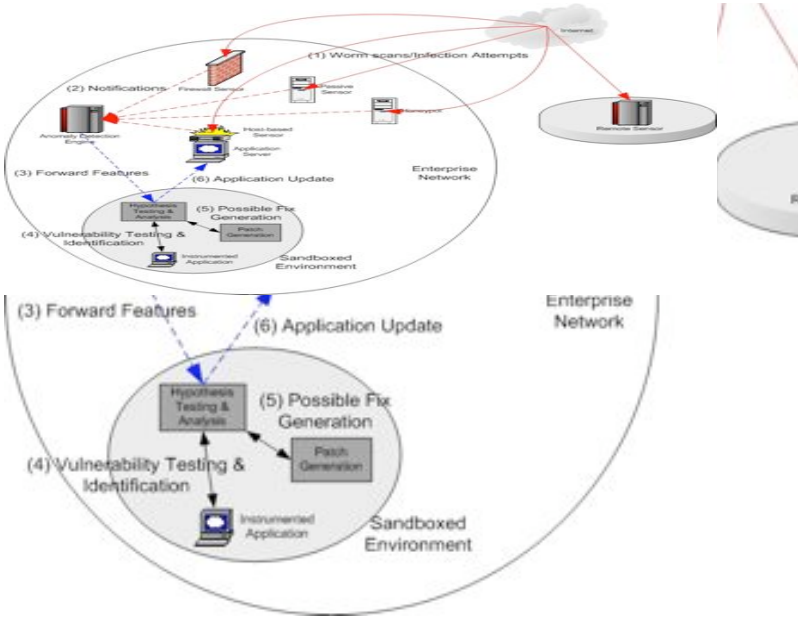
CS @CU

Self-healing Software Systems

- Novel techniques for software that repairs its failures based on Observe-Orient-Decide-Act (OODA) loop
- Demonstrated concept with two experimental prototypes
  - One aimed at the problem of worms
  - One aimed at software survivability in general
- Application Communities: enable large numbers of identical applications to collaboratively monitor their health and share alerts
  - Software monocultures are useful!

CS @CU

Self-patching Architecture

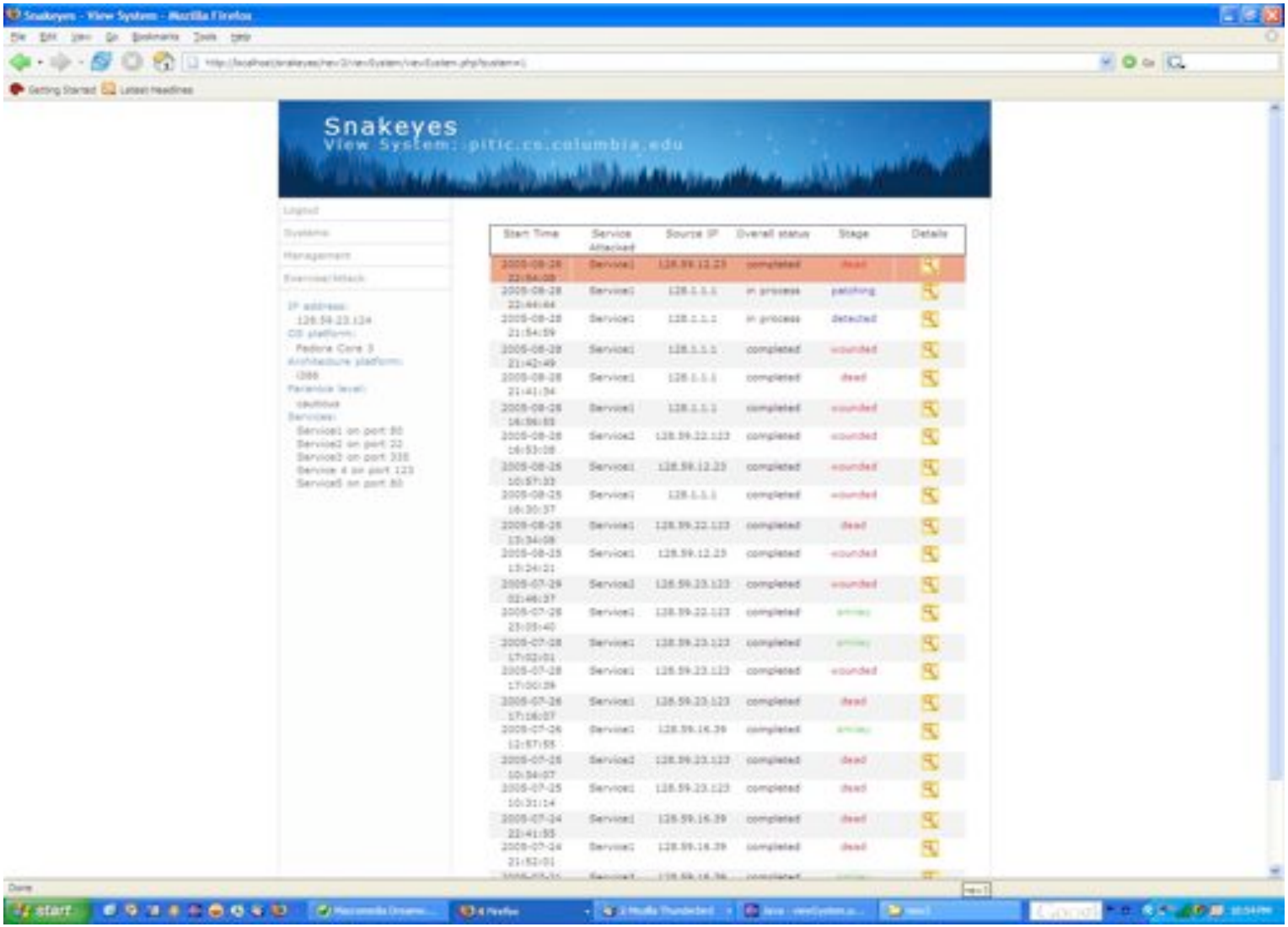


- Systems approach to creating software that:
  - Detects new attacks/failures
  - Automatically generates and applies appropriate fixes
- Developed *error virtualization* as a generic "band-aid" technique
- Prototypes for open-source and binary-only environments
- Efficient security and high availability mechanism with little performance penalty
- Spin-off: Revive Systems Inc.

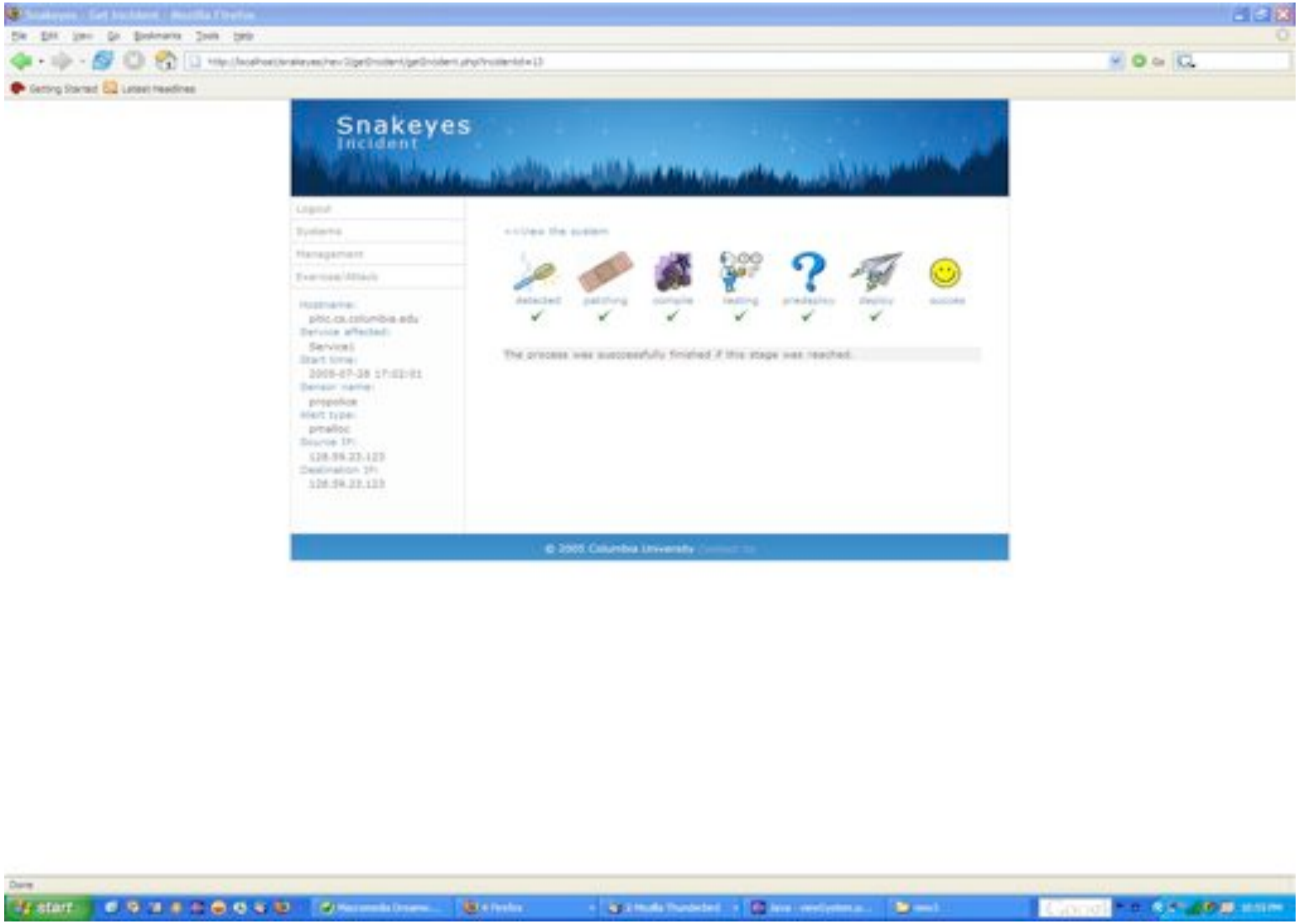
CS @CU

Network Worm Vaccine

Network Worm Vaccine



Network Worm Vaccine



IRT real-time laboratory (IRT)

http://www.cs.columbia.edu/IRT

- Internet multimedia protocols and systems
  - Internet telephony signaling and services
    - application sharing, 911 systems
  - Ubiquitous communication
  - Peer-to-peer IP telephony
- Wireless and ad-hoc networks
  - VoIP hand-off acceleration
- Quality of service
  - multicast, scalable signaling, …
- Service discovery and location-based services
- DOS prevention and traceback

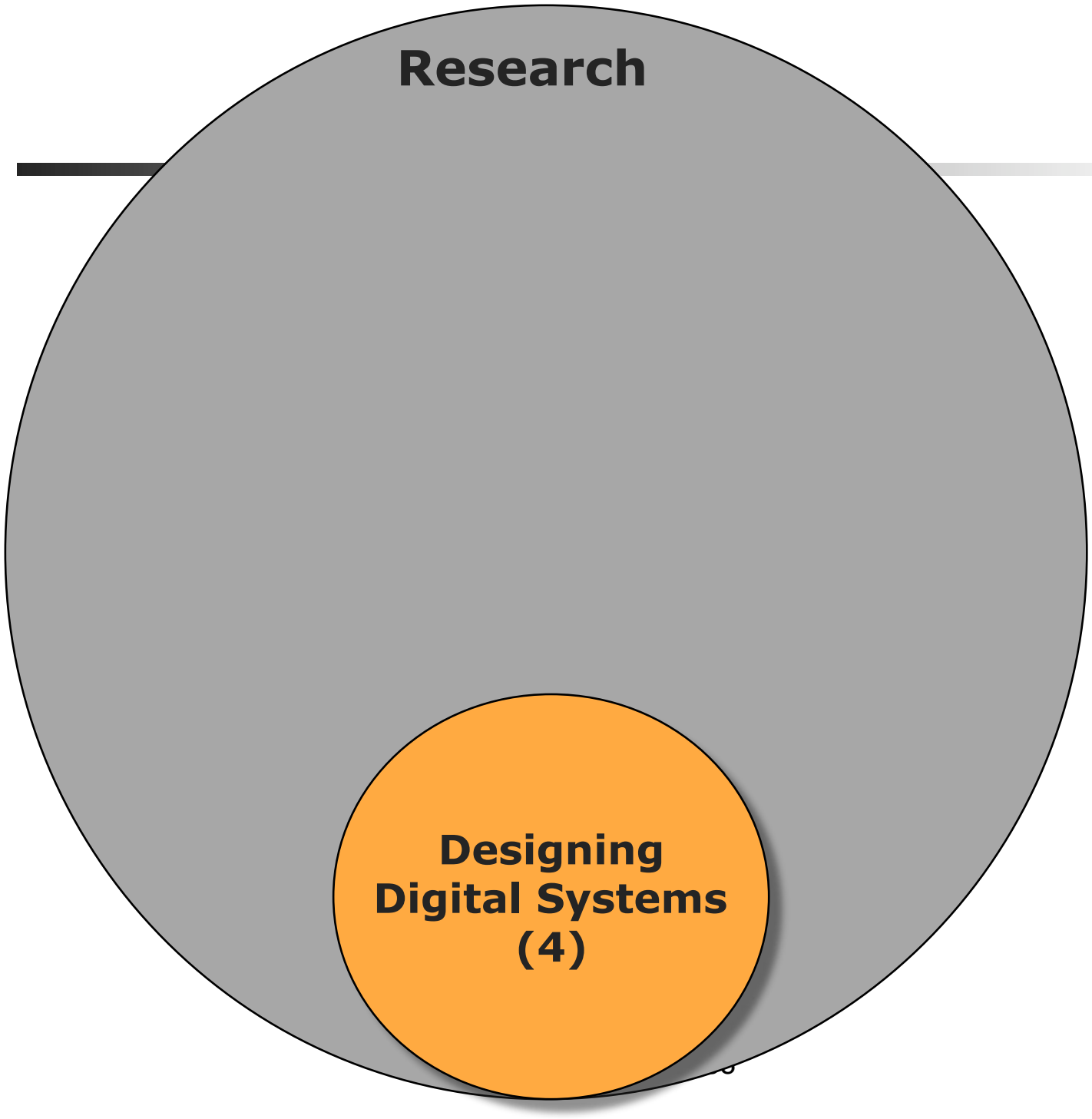
CS @CU

Distributed Network Analysis (DNA)

Prof. Vishal Misra, Dan Rubenstein

- Expertise in mathematical modeling of communication/network systems
- Also do prototyping/experimentation to validate theory
- Topics:
  - Resilient and Secure Networking
  - Wireless (802.11, Mesh)
  - Sensor Networks
  - Overlay and P2P Networking
  - Server Farms
- Analytical Techniques
  - Stochastics
  - Algorithms
  - Control Theory, Queueing Theory, Information Theory
  - Whatever else might be needed…

CS @CU



CS@CU

Asynchronous Circuits & Systems Group
http://www.cs.columbia.edu/~nowick

- Prof. Steven Nowick (nowick@cs.columbia.edu)
- Research in *clockless* digital systems
  - Most digital systems are *synchronous* = have a global clock
  - Potential benefits of asynchronous systems:
    - Modular "plug-and-play" design: assemble components, no global timing concerns
    - Low power: no burning of clock power, components only activated on demand
    - High speed: not restricted by fixed clock speed
  - Challenges: new techniques needed
    - New "CAD" (computer-aided design) software tools to aid designers
    - New circuit design styles

CS @CU

Asynchronous Circuits & Systems Group

- CAD Tools:
  - Software tools + optimization algorithms
  - Allow automated 'push-button' circuit synthesis + optimization
  - For individual controllers (state machines), for entire systems (processors)
- Circuit Designs:
  - New techniques to design asynchronous circuits (adders, multipliers)
  - Interface circuits: for mixing synchronous + asynchronous subsystems
  - Very high-speed pipelines: several GHz

CS @CU

CS@CU COMPUTER SCIENCE AT COLUMBIA UNIVERSITY

Designing Scalable and Robust Heterogeneous Computer Systems

Prof. Luca Carloni
Prof. Steven M. Nowick

{luca, nowick}@cs.columbia.edu

Department of Computer Science
Columbia University
New York, NY, USA

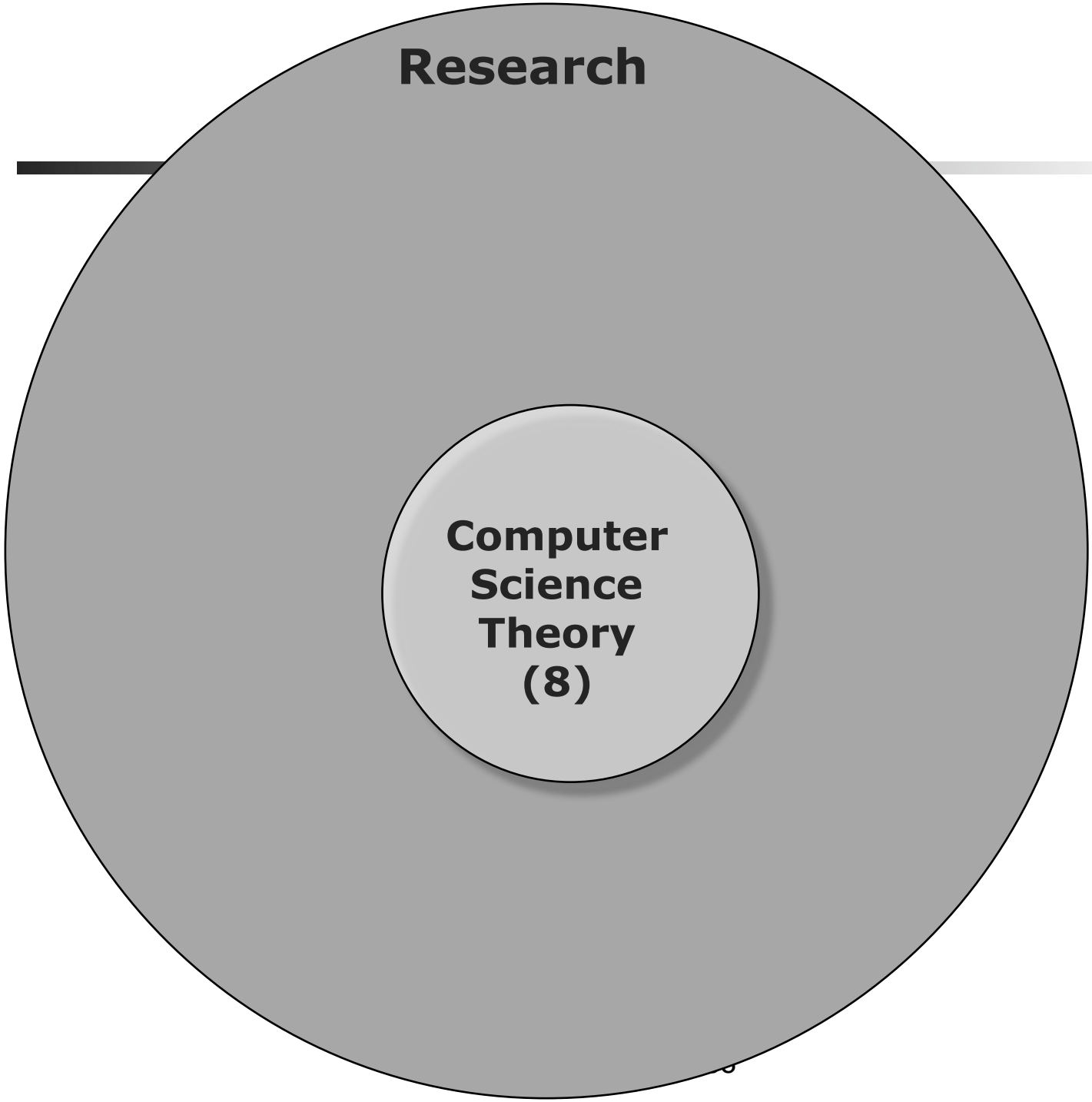
CS overview - Fall 2008

Scalable Heterogeneous Computer Systems (Prof. Nowick & Carloni)

Challenges in Future-Generation Computer Systems:

- System complexity (1 billion transistors/chip, multiple processors/chip), design time, lack of reusability
- Variability: large unpredictable communication delays, process variation, global clock distribution
- Lack of CAD tool support: system-level synthesis and optimization, performance analysis, verification
- Heterogeneous timing: robust interfacing of multiple-clock domains, mixed asynchronous/synchronous

1. CAD Tools/Design Methodologies for Asynchronous + Mixed-Timing Systems *(Prof. Nowick)*
   - Provide complete asynchronous design tool suite
   - Targeted for use in military & consumer electronics
   - Some support for GALS (globally async/locally sync) and mixed-timing systems
   - Tools for heterogeneous system-level performance analysis, automated partitioning and optimization
2. CAD Tools/Design Methodologies for "Latency-Insensitive" Synchronous Systems *(Prof. Carloni)*
   - Develop methodology for "elastic" synchronous systems – robustly handle large communication delays
   - Modular robust-by-construction assembly: synchronous computing nodes (with wrappers) + adaptable channels
   - Communication structure: support dynamic variability, flow control
   - Tool development: for synthesis and optimization, physical design


Research
Computer Science Theory (8)

Tal Malkin: Cryptography

- Crypto group $\subset$ Theory group $\cap$ Secure Systems Lab
- Crypto = construct computation and communication efficient schemes maintaining desired functionality even in adversarial environment
  - (e.g., public key encryption, secure computation, authentication, contract signing, voting, e-commerce, …)
- Motivation and Goals ← security, privacy, social, financial, political needs
- Solutions ← rigorous, theoretical approach
- Research themes:
  - Definitions (identify, conceptualize, formalize goals)
  - Protocol design (efficiency and provable security)
  - Foundations (complexity, assumptions, limits)
  - → Search for both positive and negative results

CS @CU

Tal Malkin: Examples of Research Topics

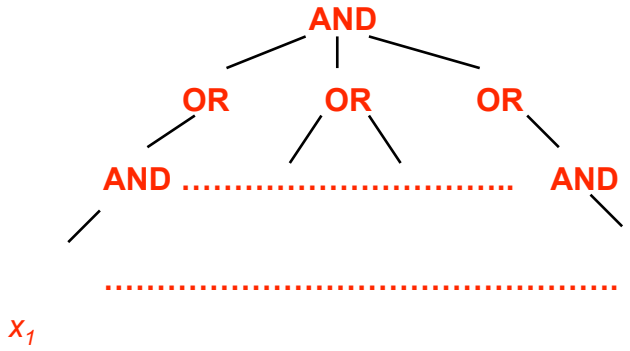
- Protecting against temporal or partial key exposure: key-evolving (e.g., forward-secure) schemes to mitigate damage of key leakage.
- Protecting against key manipulation or tampering attacks: algorithmic defense against physical attacks on keying material.
- Private information retrieval: keep user's interests private even from database holder.
- Relations among cryptographic primitives: reductions and oracle separations; minimal assumptions for cryptographic tasks.
- Secure computation of approximations, completeness for multi-party computation, multicast encryption, anonymous routing, intrusion detection, steganography, …
- For more information: take crypto class this fall, contact Prof. Malkin, check out http://www.cs.columbia.edu/~tal

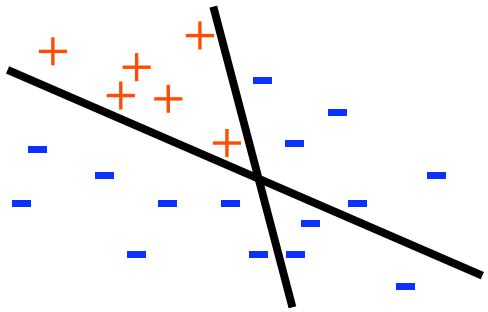
CS @CU

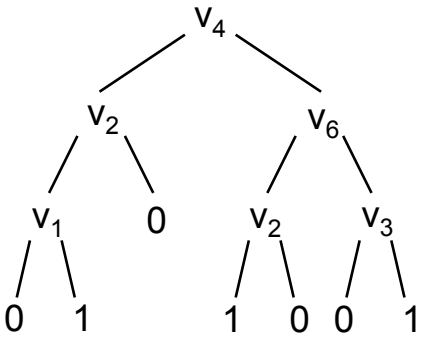
Rocco Servedio: Theory of Computing

http://www.cs.columbia.edu/~rocco

Main research goal: design and analyze provably correct and efficient learning algorithms for interesting and important classes of functions







Boolean formulas

geometric concepts

decision trees

Rocco Servedio: Theory of Computing

- Main approach: explore & exploit connections between computational learning theory and other areas of CS theory
- Complexity theory: representation schemes studied in complexity theory (Fourier representations, polynomial threshold functions) are useful for learning
- Cryptography: basis for robust hardness results for learning problems
- Quantum computation: quantum algorithms can efficiently solve learning problems which classical algorithms provably cannot

CS @CU