CUCS **Resources**

You can subscribe to the CUCS news mailing list at
http://lists.cs.columbia.edu/mailman/listinfo/cucs-news/

CUCS colloquium information is available at
http://www.cs.columbia.edu/lectures

Visit the CUCS alumni portal at **http://alum.cs.columbia.edu**

Columbia University in the City of New York
Department of Computer Science
1214 Amsterdam Avenue
Mailcode: 0401
New York, NY 10027-7003

ADDRESS SERVICE REQUESTED

NON-PROFIT ORG.
U.S. POSTAGE
PAID
NEW YORK, NY
PERMIT 3593

CS@CU

NEWSLETTER OF THE DEPARTMENT OF COMPUTER SCIENCE AT COLUMBIA UNIVERSITY VOL.4 NO.1 FALL 2007

Graphics & Computer Vision Innovator **Wins ACM SIGGRAPH Research Award**

Ravi Ramamoorthi

Columbia Professor Cited for Ground Breaking Work on Visual Appearance of Objects

Professor Ramamoorthi and colleagues have been working on methods for interactive image synthesis or rendering of realistic photograph-quality scenes. Images like this can be created in 3-5 seconds per frame, with varying lighting and view. Note a variety of "photorealistic" effects, including complex natural lighting, intricate soft shadows of the chairs, and glossy reflections from the bowls on the table. The technique is based on a new mathematical framework of wavelet triple product integrals, that has broad relevance in a variety of computer graphics, numerical analysis and multimedia applications.

Press release from ACM that announced the award.

***San Diego**, August 6, 2007*—ACM SIGGRAPH will award its 2007 Significant New Researcher Award to **Ravi Ramamoorthi** for his seminal contributions to the visual appearance of objects. Dr. Ramamoorthi, an associate professor of computer science at Columbia University, has developed mathematical and computational models that have led to a deeper understanding of visual representation by digitally recreating or rendering complex natural appearance. Some of his models have already been adopted by industry.

The SIGGRAPH Significant New Researcher Award is given annually to a researcher who has made a recent significant contribution to the field of computer graphics and is new to the field. Dr. Ramamoorthi will receive his award at SIGGRAPH 2007, August 5-9, at the San Diego Convention Center, San Diego, CA.

Dr. Ramamoorthi's work combines foundational mathematical analysis with novel practical algorithms to address long-standing problems in graphics and computer vision. His

(continued on next page)

research explores practical applications ranging from real-time photorealistic rendering, to addressing complex illumination, materials, and shadows in computer vision. His renowned SIGGRAPH 2001 paper and Ph.D. dissertation used ideas from signal processing to establish a firm mathematical framework to describe reflection, which has led to a deeper theoretical understanding of light transport as it relates to the visual appearance of objects.

Dr. Ramamoorthi received a B.S.degree in engineering as well as M.S. degrees in computer science and physics from the California Institute of Technology in 1998. He was awarded a Ph.D. in computer science from Stanford University in 2002, and joined the faculty of Columbia University in August of that year. A prolific contributor to ACM SIGGRAPH publications and conferences, Ramamoorthi has published his research in several other leading international journals on graphics.

About ACM: ACM, the Association for Computing Machinery (*www.acm.org*), is an educational and scientific society uniting the world's computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

About ACM SIGGRAPH: The ACM Special Interest Group on Computer Graphics and Interactive Techniques *(www.siggraph.org)* is an interdisciplinary community interested in research, technology, and applications in computer graphics and interactive techniques. Members include researchers, developers and users from the technical, academic, business, and art communities. SIGGRAPH provides information to the computer graphics community through its annual conference, publications and the SIGGRAPH Video Review.

Dana Pe'er: Decoding Genetic Variations and Regulatory Networks

Dana Pe'er

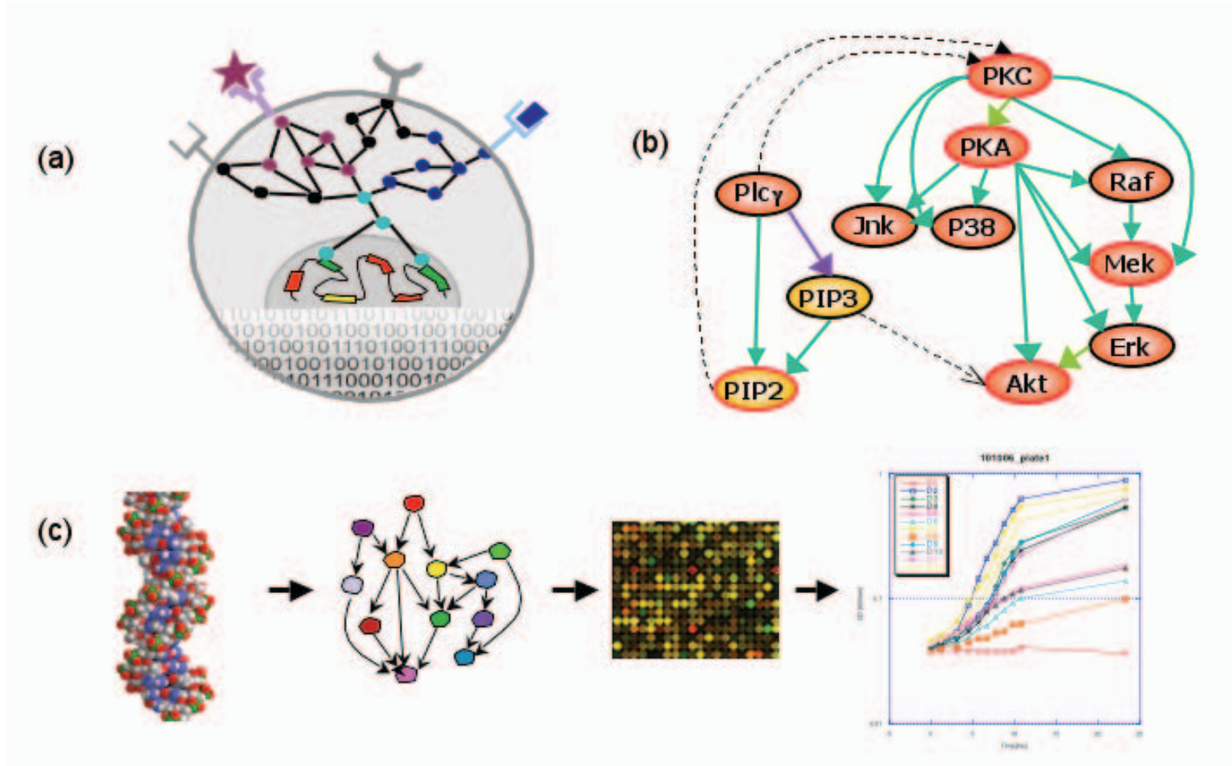
NIH New Innovator Dana Pe'er (pro-nounced pay-er) is looking forward to building her lab team and working on the next phase of her research, which seeks to illuminate how a cell's regulatory network processes signals, and how this signal processing goes wrong in cancer. As one of the world's leading computational biologists, Pe'er develops highly sophisticated computational "machine learning" methods that analyze genomic data and detect patterns that underlie interactions and influences between molecules in a cell.

With the NIH award funding, Pe'er and her team will seek to understand the general underlying principles governing how cells process signals, how molecular networks compute, and how genetic variations alter cellular functioning.

Specifically, she wants to understand how changes in DNA codes modify a cells response to its internal and external cues, which then leads to changes throughout the entire body. These changes, or malfunctions, can cause anything from autoimmune disease to cancer.

"Cancer is a very individual disease—unique in how it develops in every person," said Pe'er, who came to Columbia less than a year ago with her husband, Itsik Pe'er, also a computational biologist at the University. "Our research is aimed towards personalized cancer diagnosis and treatment. For each individual patient, we wish to detect the key mutations that cause cancer, understand how these combine to cause the malignant behavior and pinpoint where and how to target drug intervention, leading to better therapies and drug development. I'm hoping this will be my impact."

Written by Clare Oh
Article reprinted with permission from Columbia News



(a) We view the cell's molecular network as a "computer", receptors sense the environment, signaling pathways integrate all signals and process and appropriate response, these activate transcription factors which execute this response in an orchestrated manner. Our lab aims to understand how molecular networks function at a global systems scale.

(b) By applying Bayesian network structure learning algorithms to high-throughput biological data, we successfully recon-structed known interactions and influences in human signaling pathways, as well as suggesting new influences which we experimentally confirmed.

(c) The primary focus of our lab currently is to understand how changes in the DNA sequence perturbs the regulatory network, alters signal processing and is manifested in complex cellular phenotypes such as growth rates in cancer tumors.

Young Investigator Awards

Our roving reporter **Sean White** interviewed **Professor Dana Pe'er** and **Professor Ravi Ramamoorthi**, to learn more about their research and motivations.

Dana Pe'er

How did you get interested in computer science?

I think from a very young age, I always knew that I wanted to mix biology, math and computer science, maybe when I was in 7th grade. Biology is really beautiful, very descriptive and very hardcore physics and it boils down to math. I felt as though a lot of patterns in biology need to be pushed forward with math and computer science.

When I finished my Army training, there was this whole genome revolution and DNA was being sequenced and suddenly we began looking at molecules. There were some new technologies coming out, microarray technologies, which really allowed you to take snapshots of all the RNA in the cell at once. I really wanted to try and understand how cells compute.

We've got the parts list and suddenly people don't know what to do with it. Microarrays were generating these huge data sheets for the activity of ten thousand genes across a hundred conditions and biologists were used to looking at five or six numbers and analyzing them. But what are you going to do with a matrix of ten thousand by a hundred? Are you going to stare at that and find a pattern? It really felt like something computer science could help with.

What motivates your research?

I'm motivated by biological questions. I'm motivated by questions that look at systems as a whole, and how they function in a more perceptual, principled way. Beyond just drawing a wiring diagram of what the network is, I want to understand some fundamental principles of how it works, how different changes in the DNA really propagate, how it's designed.

Do you have a method or approach to your research?

The place that I have a real advantage is to go places, that without new state of the art computation, you couldn't go. I take Bayesian networks and make some adaptations to them, or take regularized regression and make some adaptation to it. I didn't invent it, but I do make non-trivial changes to sophisticated machinery. The ability to take state of the art algorithms, know how to use them properly and know how to make non-trivial changes to them to make them fit the biology, which is very different than what they've been specifically designed for, is something that most bioinformaticians don't have. So when I choose a problem, I try and play to my relative strength.

Tell me about your current research challenges.

My current favorite project is where we study how molecules interact in yeast. I work in yeast because you can really ask simple questions in yeast and actually get at them. Once I develop the concepts in yeast, I've gone from yeast to human three times. One of the things biology has failed to study properly is combinatorics. They always like looking at a single set. So I designed the experiment where you have multiple combinatorial changes to the DNA. You measure how that changes the regulation and how that changes the fitness.

There's also an element of active learning. I developed my own technology to do active learning in these biological networks. That's a project that's close to finishing right now. Making random changes in the DNA has been easy but making fine, exact sophisticated changes to the DNA has been pretty much impossible. We've developed a new technology to be able to do that.

I wanted (my research) to be relevant to the world. I try and always have an arm in the lab that takes these things and tries to apply them to something real, something meaningful, which in my case is cancer because it is a disease where regulation goes wrong.

How do you see your field evolving?

You have to work at multiple levels. I've been criticized because in biology, it's so complex you have to focus. If you look only in the small scale, then you're never going to see the big picture, you're never going to see how it all comes together. You're not going to understand everything by looking at small scale. On the other hand if you take too many steps back, and look at it too abstractly, you get the entire mess without understanding what little bits and pieces of it are doing. Therefore getting the two perspectives and constantly jumping back and forth between the bird's eye view and the microscopic view, really allows me to think of new ideas.

Ravi Ramamoorthi

How did you get interested in computer science?

I think the first time I got hooked into computers was when I saw a CASIO PB-110 pocket computer in 3rd grade. Later, when I was in the 6th grade, my father bought me a ZX Spectrum home computer that I got totally into. I tried to write a number of games (in 2D in BASIC), inspired by the (certainly simple by today's standards) various commercial games available on it... in many ways, that's when I really got interested in graphics.

What motivated you to focus on your current research?

When I was at Stanford, working on my PhD in computer graphics, I thought rendering or image synthesis (creating realistic images on a computer) was really fundamental to the field—it's based on a sound theory of simulating physical light transport, and is basic to actually making images (or creating graphics). Over the last five years, I realized that computer graphics image synthesis is increasingly using high-dimensional datasets to tabulate visual appearance (a broad term to describe the way objects or people "look", fundamental to graphics image synthesis). In the process, functions like lighting, reflectance properties of materials, spatially and time-varying appearance are often measured or tabulated. So, a major focus is on the appropriate mathematical representations and computational models for visual appearance, with relevance to a broad range of practical applications. It was this body of work that the ACM SIGGRAPH Awards Committee recognized.

Do you have a method or approach that you like to use in your research?

My goal has been to look for fundamental results, be they mathematical or computational, often focusing on the correct mathematical representations and signal-processing tools. In this way, I believe one can lay firm scientific foundations for the field, while impacting a wide range of practical applications. Besides the immediate impact in graphics and vision, my work has for instance also settled long-standing questions in radiative transport in physics (I had a paper published in the Journal of the Optical Society of America in 2001).

I think a good paper and a research question would make an important insight and fundamental long-term contribution to our scientific understanding, coupled with a useful practical result or implementation. Of course, this is a difficult goal to strive for.

What do you consider the most interesting challenges of your current research?

I think the mathematical and computational challenge is how best to represent, manipulate, edit, render with, and compress complex visual appearance datasets. At its core it is a data representation and signal-processing problem. Some ideas draw from techniques in other fields; some ideas (like our wavelet triple product algorithm from SIGGRAPH 2004) are essentially new and potentially applicable to other domains. Because of the fundamental nature of the questions posed by this research, it has really been an interesting journey over the last several years.

From a practical perspective, the challenge in visual appearance is how do you acquire completely realistic models of lighting, reflectance, spatial variation etc from the real world and be able to display them interactively so you can really feel you are in a real environment. This is something very hard to achieve; even in offline computer graphics imagery and movies, faces for instance often look fake in some subtle way. More broadly, how can you obtain completely realistic simulations of the visual world (including motion and object geometry)? Will there come a time when a high-school student can create a completely realistic virtual world in one afternoon?

Do you have a favorite project that you've worked on?

I'm proud of a number of contributions, mostly those that have introduced some fundamentally new mathematical idea or computational insight. My SIGGRAPH 2001 papers and PhD thesis introduced the notion of reflection from a surface as a spherical convolution, that brought a host of new mathematical tools into the field that have now become widespread. I really like the SIGGRAPH 2004 paper on wavelet triple products, as it was a new and deep mathematical question, that hadn't really been addressed previously even in the numerical literature (and may have further impact in that domain). This year, we've made a number of quite significant theoretical advances building on previous results, which really show how rapidly the scientific foundations have advanced over the last five years.

How do you see your subfield evolving?

Since our goal at some level is to simulate the real world fully, graphics is constantly drawing from and giving back to a number of different fields, including physical simulation, studies of human and animal motion, geometric representations in mechanics, and lighting simulation, as well as associated hardware and numerical methods. It's an exciting time to be in the field, since graphics now influences or has the potential to impact almost all the other areas of computer science and more broadly many science and engineering disciplines, besides drawing from them (both ideas and researchers).

Message from the Chair

Henning Schulzrinne, Professor & Chair

Thank you for reading our fall 2007 edition of the Columbia Computer Science newsletter!

The Department is proud to introduce two new additions to our faculty, strengthening core areas of research and instruction in our department. We will introduce both in the next edition of our newsletter in more detail, but here's a short preview. Simha Sethumadhavan (Ph.D., The University of Texas at Austin) joins the department in January 2008 as an assistant professor. His research interests are in computer architectures that scale to Exaflops and beyond, and software support for features found in these architectures. He was part of the successful Tera-OP, TRIPS processor team and his most significant contribution was the design and implementation of the first-ever fully partitioned primary memory system. His dissertation proposes methods to scale the hardware used for disambiguating memory references which has been a key impediment to exploiting instruction-level parallelism for several decades now. His other research interests include application-specific architectures for emerging workloads, implementation of distributed systems and novel computing technologies.

Junfeng Yang will receive his Ph.D. in Computer Science from Stanford University in January 2008 and will join us next summer. He currently works at Microsoft Research. During his graduate studies, he built tools that detected serious data-loss errors in real-world storage systems and security vulnerabilities in the Linux kernel. His research interests span the areas of operating systems, security, and software engineering, focusing on practical error detection in large software systems.

Our MS program, both in its traditional on-campus form and through our distance learning program (CVN), has long been a core part of the department, offering advanced computer science education with the encouragement to pursue research projects. This year, we have welcomed our largest incoming MS class in years, with 114 new students.

For all of our students, I believe that being able to convincingly express their technical ideas in writing and public speaking is a core competency for aspiring computing professionals. This goes significantly beyond producing grammatically correct English, a skill that the American Language Program at Columbia University teaches to our international students. In the past, our Women in Computer Science (WICS) group organized seminars on writing topics for the department; commonly, PhD advisors spend a fair amount of their time editing the papers of their students. For my students, I have attempted to collect guidelines* and common mistakes, but all of these attempts seem insufficiently systematic and are hard to sustain. This year, with the help of colleagues at UCSB, we organized a multi-week workshop, taught by Janet Kayfetz. Students wrote diaries and technical articles, and presented practice talks. In small groups, students critiqued each other's writing samples. The feedback has been sufficiently positive that we intend to repeat this experiment next year. We are also looking at other options that turn aspiring researchers into competent writers. I would be particularly interested in hearing what other departments are doing to address this issue.

As always, I look forward to hearing from you as our readers.

**http://www.cs.columbia.edu/~hgs/etc/writing-style.html*

Featured **Articles**

Meddling with Lawyers

By **Steven M. Bellovin**

In Lord of the Rings, Gildor, an Elf, advised Frodo, "Do not meddle in the affairs of wizards, for they are subtle and quick to anger." Many people feel the same about lawyers. That said, I work with lawyers a lot, to our mutual benefit.

Much of the work I do involves public policy, especially things like privacy, cryptography, and wiretapping. These are complex issues even from a policy perspective. However, what is possible—and what the risks are—is very dependent on technology. Lawyers, even smart ones—and most of the ones I deal with are very smart —are often unaware of important technical issues.

As technologists, we have the same right, qualification—and responsibility—to speak out on such issues. However, we have no more rights, qualifications, or responsibilities than anyone else ***if*** the issue is purely policy. Where we differ from the general population is on the technical side. There, we have special knowledge that most people don't have; that makes us more ***qualified*** to speak on certain points—but only when they're technical points, not policy.

I confess to ambivalence about wiretapping. It's inherently very intrusive (hence the need for "minimization", the elimination from the record of all calls by others who share the same phone line); it also provides some of the very best evidence against people I really don't like. The current US wiretap law was actually a pretty good compromise, in that it limited wiretapping to serious crimes where other investigative techniques have failed or are too risky, but recent changes to the law, such as CALEA (the Communications Assistance to Law Enforcement Act), are more dangerous.

CALEA mandates that all phone switches contain built-in wiretap interfaces. It was designed with an eye towards a certain set of benefits, without regard to the quite-considerable risks. That's where computer scientists have value to add in the debate, above and beyond that of J. Random Citizen: we can warn policymakers what could go wrong. And go wrong things did. Few computer security professionals were surprised to learn that similar features on a Greek cellphone switch were subverted by parties unknown; as a result, more than 100 Greek citizens, including the prime minister, had their phones tapped.

We've had better luck with voting machines. Though many states rushed to adopt direct-recording electronic voting machines in the wake of the 2000 presidential election vote-counting fiasco, they are now realizing that risks inherent in such machines. As result, they're now switching to machines that produce an auditable paper trail, just as most computer scientists had suggested.

All that said, some lawyers really do get it. A Court of Appeals opinion striking down the Communications Decency Act showed that the judges really understood the end-to-end principle, a fairly subtle argument in systems design. I mentioned this to a friend who had been at the hearing; he said that the judges were asking questions based on it in real-time.

So—when lawyers come calling, don't run away. You can have an effect—and it's fun.

Adventures in Mobile Networking: What I Learned From Building and Demoing a Testbed of Networked Roombas.

By **Joshua Reich**

Several weeks ago I traveled to Montreal to compete in the joint ACM Mobicom/Mobihoc student demo competition. Of course I didn't go alone, my robots came too, and I think I can say that we were all thrilled to go (it was their first time on an airplane after all). It was a long and winding journey that took us there and from this journey, I have learned many valuable lessons (hopefully not all obvious), which I will now attempt to convey to you.

Like all good, and not so good, stories, this one has a beginning. Late fall of last year I was meeting with my advisors, Vishal Misra and Dan Rubenstein. We were discussing a particular Delay Tolerant Networking (DTN) technique on which I had been working. We knew we'd like to apply this technique to some problem, test it, and hopefully demonstrate it's efficacy—otherwise no one would be persuaded to care about what I discovered. (As it later turned out, what I had discovered was, at least for the time being, a dead end). The conversation went something along these lines:

Vishal: We should test this.

Me: I agree, I could throw together a simulation.

Vishal and Dan: Simulation would be a nice start, but it's a bit passé.

Me: Well I have some experience with robots…

Dan: How about Roombas? People like them!

Vishal: The robotic vacuum cleaners.

Me: They are quite neat, um but how do they fit the problem?

Dan: We could use them in…

Me: I'm not convinced, that doesn't really seem to utilize the techniques we've been working on.

Dan: Okay how about…

(Loop a couple of times)

Me: Yeah I could see that, I'm still not sure it's the best fit with this technique, but if you think it's a good idea, I'll start looking into it.

As it turned out, my reservations weren't entirely unfounded; the Roombas never really did fit that problem nicely. But more importantly, it turns out my advisors had given me the opportunity to build a system that fit many other problems nicely!

Lesson 1: *listen carefully to your advisors and/or colleagues and don't get too attached to your idea of the month. You didn't marry it after all.*

Lesson 2: *if you are going to invest lots of time building a complex system, build one that can be used to explore many different interesting scenarios, both because the problem that motivated you to build it might have been a dud and because it could lead you to lots of other interesting problems!*

Fast forward several months: I'm working in my office, trying to get a particularly recalcitrant router (the router runs a Linux distribution for embedded devices called OpenWrt and controls the Roomba—or so the theory goes) to do something stunningly simple, wondering how I got myself in this mess. By this point it's clear the previous work wasn't going to unfold in the expansive way I had hoped and that it didn't fit with the Roombas either. So I had, a week earlier, written it up to file away for some future time. I did so with the hope it might prove useful, or perhaps spontaneously metamorphosize itself into a tech report. Meanwhile I was stuck with misbehaving robots and only a partially conceived plan of how I would get them to work. What precisely I would use them for when I accomplished this was something I was actively avoiding thinking about. Then I get an email from Vishal: "does anyone in the group have results appropriate for the MAMA workshop—work in progress or results that may not be quite conference worthy?" I pause for a moment, I really need to keep working on these robots and don't really have time to spare as I'm off to a summer internship soon. Then I shake my head and run for the file cabinet as an old Rabbinic saying runs through my head "if not now, when?"

Lesson 3: *write up all your research even if it's not clear when or where you'll publish it. Jump at the opportunity if an appropriate venue arrives. Who knows where it will take you?*

One month later, San Diego: I walk into the dining room of a crowded conference. I've just finished my workshop talk, or maybe I've still to do it. Doesn't really matter the point is I'm here, I don't know anyone, and I need to sit somewhere for lunch—sort of like the first day of high school actually. Ah, there are Dan and Vishal. Maybe, I'll go sit with them. The table's crowded but there should be enough room for one more. I sit down, introduce myself, listen to the conversation. I eat yogurt and lettuce. Oftentimes that's all you get when you're both strictly vegetarian and kosher—I don't like attracting attention that way, but it seems inevitable if I don't want to starve. I try to make small talk, am not terribly successful, then:

Suman: …we're looking for more student demos at Mobicom.

Vishal: oh Josh has a robot system… Roombas… networks. He could demo.

Me: *pause, thinking (I'm starting my internship this Monday and finishing two weeks before the demo, I've only got one partially built robot and no demo application. This could be a particularly embarrassing way of committing professional suicide if I can't keep things on track.), then I hear those darned Rabbis again.* Yeah, definitely I'd love to! I'm pretty sure I can throw something together.

Suman: excellent send me an abstract.

Lesson 4: *eat lunch with your advisors (but don't be a pest), listen to your Rabbis (at least the imaginary ones in your head—please feel free to substitute in Priest, Imam, Guru, Inner Spirit or whatever fits your belief system).*

Monday, August 27th, Columbia: I've just returned to campus. A bit burnt out, finished my internship (60-70 hours/week) on Friday, I'm very tempted to go to the beach, but for now the only waves I'll get to examine are ones produced by the sin and cosine functions. I had advised two students over the summer and came in to the office on one or two weekends. Not nearly as much had gotten done as I'd hoped—seems to be a given in life, eh? So I sat down with a determined look on my face, ready to somehow forge the disparate pieces of work, code, and robot I had left behind at the summer's beginning. And I did, but not with quite the vigor for which I had hoped. It wasn't that I failed to work and produce and plod forward, it was that I was plodding not springing forward. I was frankly a bit worried about this, I needed to move faster, what was I going to do? Nothing. I was tired. But I could keep going at a plod, and I did.

Lesson 5: *keep working, no matter how slow progress is.*

Towards the middle of the week before the demo, I began to pick up speed, but it still wasn't quite enough. Friday afternoon I was done, toasted, cooked, exhausted. I had a demo. I knew it wasn't great. I could demo the equipment and the software, but the application I'd come up with wasn't implemented, nor were its constituent parts even close to being integrated. I showed the demo to Vishal. He said it would be fine, but maybe I could find a way to get the application working. Instead of a complete console GUI, why not just have the robots spin and beep when they passed messages? It was a great idea. "Maybe, I just don't think it's feasible in the time left, I'll see if I can do something but I doubt it…", I said. I was depressed. I couldn't work on the Sabbath, and I was flying out for the demo on Monday morning, early. Plus I had bought tickets to Farm Aid with my wife for Sunday and didn't want miss it. I needed to sleep.

Morning: I'm sleeping, but I'm thinking, about the demo. Could it be done? Haltingly, "yes, maybe, no yes it could". I started trying to look at the code as I recollected it in my head. Yup, I'd do this, and that. This piece could be hooked into that code base. Damn, that would be great! What about the beeping? And the dancing? I could probably do that in another 2 hours, need to modify some lower-level stuff, implement some new routines. How about the visual recognition? Maybe that way? Yeah, perhaps. Perhaps I could emulate downscaling the transmission… all told—30 hours? 25? 57? Doable though. Maybe. No, I could do it. By now I'm awake and energized. I'm ready to work as long and hard as I need to, only one problem—10 more hours of the Sabbath left. I've rarely wanted to break it this much, but I reluctantly decide to wait.

10.5 hours later: I'm working and I keep working with occasional

breaks for food and sleep until 5PM the next day. Then I go to Farm Aid with the wife. I'm feeling sick (sore throat) and tired, but Willie Nelson rarely comes to New York and he's getting old. The Rabbis remind me of their favorite saying. We return at 11PM having had a wonderful evening—Willie, The Allman Bros., Dave Matthews, Counting Crows (although I care for them less). Linda's happy, so am I. Work for another couple, pack, watch some Battlestar Galactica (great show), sleep for 3. 6AM off to the airport robots in tow, application (mostly) working (and it's really cool).

Lesson 6: *sometimes an unreasonable deadline can be helpful. Push yourself.*

Lesson 7: *solicit as much feedback as you can get and listen to it, but maybe after sleeping.*

Lesson 8: *You can trade sleep for religious observance, Dave Matthews, and a happy spouse.*

Montreal: was great. I spent most of the time from arrival through Tuesday afternoon in my room fixing last minute problems, testing, and retesting, remotely compiling changes on Columbia's servers and slapping them onto my robots, which would then break unexpectedly. Those of you in research know how this one goes. The afternoon before the demo, I took off, walked around and just savored being in Montreal. It was a wonderful, sun-filled afternoon. I won't forget it anytime soon. I found a veggie restaurant and had dinner. Then I grabbed some OJ and yogurt (no lettuce this time) for the next day. Back to work.

Lesson 8: *You went somewhere, look around. Enjoy it, otherwise what's the point?*

Morning of, I got downstairs early, but not as early as I should have. Some of the demo spaces were already filled, and none were as commodious as I'd been expecting. Also the lighting was terrible. I wandered about and tried to figure out where people would be entering the space, where they were most likely to look, and which spaces were obscured. I also needed at least two walls to keep my robots from wandering off (I only had two IR barriers). I almost took two tables, across from one another against a wall. But I decided that would be impolite and didn't do it (someone else actually did do this later on). Instead I took a somewhat less well-lit, off-to-the-side space. But it was in the first row towards the entry way and in front of a second set of doors that might be opened (in the end they weren't) and it had the walls I needed. I took a little time to see what other folks were doing, many of them had business cards, handouts, etc. I didn't. Stupid.

Lesson 9: *Get there early, really early. Scout the space, find an advantageous location (maybe don't worry about being a hog, maybe do—I haven't decided).*

Lesson 10: *Have handouts with your name and contact info at the ready.*

While it was still relatively quiet, I tested one of my robots in the space itself (the whole bunch were over 100 lbs and I'd rather debugging occur in the privacy of my room—which was up the stairs and down the hall). The visual detection component was toasted by the poor lighting. I was worried. Without this piece, the demo would be really lame. What would the robots always recognize that I could whip up on the spot? The direct beam of a flashlight was the answer that sprung to mind. I got one from the janitors, reprogrammed the robots, and while it wasn't ideal, it worked.

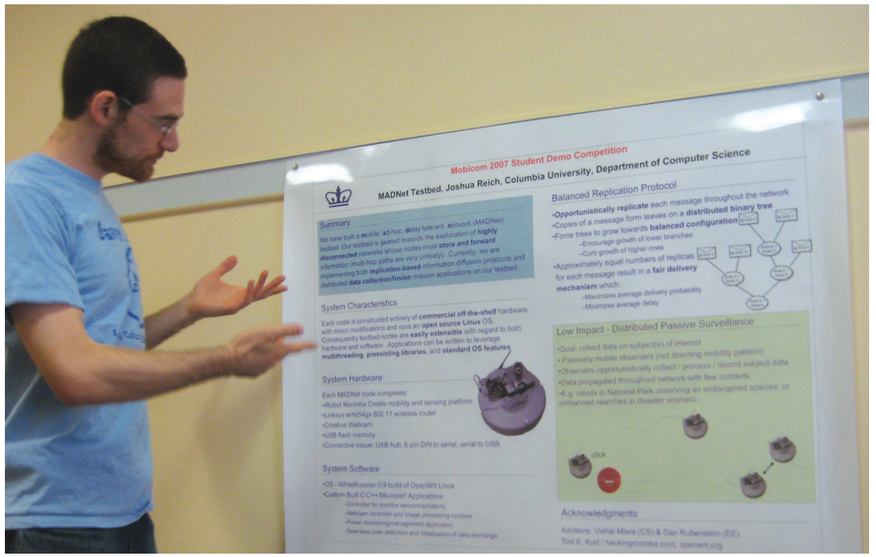
Lesson 11: *Test your demo in the space beforehand, be prepared to improvise.*

3PM: Show time. I still wasn't ready though. Another last minute bug had popped up and here I was frantically typing on the terminal. When people come by, I apologize, explain there is a glitch and that they should please come back in a couple of minutes, or enjoy my beautiful poster. 3:12PM: it's close enough, I'll just wing it. The next 2.5 hours are full of telling spectators about my demo, shining a flashlight at robots, sweating when they stop moving. I'm replacing batteries and rebooting hardware (when it busts, which it does often) all the while answering questions and taking each opportunity to show folks how the system works. After all, battery replacement is a nice opportunity to discuss lifetime, power-draw, performance, and component integration; while software bugs open into a nice discussion about the software development process in OpenWrt, issues with debugging and the multi-threaded, socket-based, distributed software that makes our 'bots tick. When all else failed, I'd let the robots spin around in crazy circles and give animated presentations in front of the poster board—it seems to me that attendees actually found these very valuable. I also had two units sitting on the table that people could pick up and fiddle with. And sometimes things worked smoothly. Those were great 5-minute spans. I didn't get to use the bathroom or drink anything, nor did I really get to see much of the other demos and I completely forgot to take out my camera. But while no one else from my group was at the conference, my friend Yufang was, and he told me about the other demos, gave me feedback, and moral support. If I had only told him about the camera!

Lesson 13: *Be ready on time, but if you aren't don't panic. People will look at your poster, and look at your demo, they also look at you—be a car salesman, but with out the greasiness. Try to turn every problem into an opportunity.*

Lesson 14: *Go to the bathroom beforehand, have some water handy, bring a friend, and don't forget to take out your camera!*

So you must be wondering what the demo actually did. My research explores the challenges faced by mobile, networked systems. This includes a spectrum of scenarios: from those in which nodes can manage to retain fairly robust connectivity,



to situations where nodes are highly disconnected (in this latter situation, nodes can only propagate information by storing and forwarding one another's messages). The purpose of my demo was two-fold: firstly, it was to showcase the **m**obile **a**d-hoc **d**elay-tolerant **ne**tworking **t**estbed we've built (we call it a Roomba MADNeT), secondly, to demonstrate a proof-of-concept application utilizing this testbed. The proof-of-concept application we demoed was an opportunistic surveillance application. Imagine a national park sometime in the future—a very large area, some service robots moving around, some vehicles with mounted network-enabled data-gathering equipment, perhaps even some units worn by hikers and rangers. All of these units could take pictures of grizzlies and other interesting animals, but wouldn't be in control of their own mobility patterns, nor would they generally be within transmission range of each other. How would these nodes get their data to interested researchers? Our demo application solved this problem by leveraging a data propagation technique we had previously developed, and you've already heard the story of how it was implemented (see our upcoming MC2R paper for more details). The aforementioned lighting problem prevented our nodes from recognizing their target accurately, hence our substitution of the flashlight for our original target, a red Roomba. When one node saw the flashlight it would demonstrate this by stopping and beeping. Other nodes did likewise when receiving messages. As in the national park example this happened only gradually and dependent on nodal movement patterns, since our nodes only rarely were within range of one another (we dialed down the broadcast strength in software). In this way, onlookers could see the propagation of information across the network and interact with it at the same time. It was actually quite interesting to watch, and, along with luck, was probably the reason a steady stream of visitors came to our demo area.

By the time things wrapped up I felt wrung-out, exhilarated, and ready for dinner. Thankfully I had Yufang (by this point the robots were sleepy, grumpy and poor company). We packed up the equipment and went out to celebrate. What were we celebrating? Having done a good job? Having survived? I'm still not sure. The next day I returned to New York, thinking the next couple of weeks would provide for a moderate degree of downtime. Then I got a call from Vishal (it was moments from sundown and I'd be off email for 3 days whilst observing the Jewish new year): "Had I checked my email?" "Yeah, I saw that I was one of the four finalists, I was truly thrilled." "No, you won." "I won? The competition?" "Yeah, you won the entire thing, best demo". I was a bit overwhelmed. It was a great start to the new year. In addition to being tremendously gratifying, this experience seems to be propelling me towards other exciting adventures, research opportunities, and problems I'm eager to begin solving. Who knows where they will lead me? I did miss the award presentation dinner though, so perhaps I'll close by thanking all the people who are supporting me through this continuing journey —my advisors Dan and Vishal, my colleagues in the Distributed Network Analysis lab and the department, the robots, Yufang, and of course, my family and wife.

A Day in the Life of a **Model Professor**

By **Stephen Edwards**

In an attempt to do something that I had never done before, I spent a day modeling for an ad agency in Taiwan that was doing a promotion for McDonald's.

In an attempt to do something that I had never done before, I spent a day modeling for an ad agency in Taiwan that was doing a promotion for McDonald's. This was my wife Nina's fault, of course: she had visited Leo Burnett (the agency) to try to promote her work and one thing led to another. They were looking for a western model for a promotion and discovered she had an American husband. She sent off a few pictures of me and since they didn't know any better, they gave me the one-day job.

In fact, this whole episode can be attributed to Nina's networking skill. The fully story goes like this: Nina went to Italy to attend a children's book conference. There, she met the head of the society of children's book illustrators in Taiwan, who invited her to give a lecture. Nina accepted, and at her lecture in July, she met an illustrator who had worked for Leo Burnett, the ad agency. This illustrator referred Nina to the right people at Leo Burnett, who happened to notice that Nina had an unusual (i.e., western) last name and asked if her husband was a westerner since they were looking for a western model.

They first brought me in for a "fitting" at 11:30. It was then that I learned what they wanted me for: they wanted an "eskimo" (actually, an Alaskan) for a poster they were doing for McDonald's stores in Taiwan in September, 2006. Great. It's 100F degrees outside and they want me to dress up in a really warm jacket, hat, and scarf. The fitting went fairly quickly (they just wanted to make sure the colors they

had chosen were reasonable and that I looked something like what they were looking for), so Nina and I went to visit some friends until 6:00 PM, when they needed me back at the photo studio for the actual shoot.

There's actually quite a lot of irony here: having grown up in Minnesota, which is routinely colder than Alaska in the winter, I'm actually a reasonable stand-in for an Alaskan. I certainly have the experience of wearing really warm jackets like the one I'm pictured in.

They sat me down, applied makeup (!) and then started shooting. I guess I did OK, but they kept wanting me to smile bigger, which is hard when you can't hear anything from the tight hat and you're boiling under the hot lights. They turned the air conditioner up full blast and pointed a fan at my feet and it was still very hot.

After four hours, they finally got what they wanted. There were about eight other people in the studio all working on this including the photographer, the hard-working stylist, who was constantly retouching my makeup, fussing with the crazy hood she had sewn onto the jacket, an art director or two, somebody from McDonald's, and a bunch of assistants.

The hood was a good story: they wanted a furry hood, but the jackets they could find didn't have a hood that was furry enough. So they got a mink stole and wrapped it around the hood on my head. I think the stylist quickly sewed it in place, but we had to keep moving it around.

It was an interesting experience, but I wouldn't want to make a living being a model. My face ached for days from trying to smile wider.

The "fish" campaign for which this photoshoot was done turned out to be fairly substantial. Not only was I on a poster, but also on large banners at McDonalds across Taiwan, and the McDonald's Taiwan website. *(http://www.mcdonalds.com.tw/)*

They put up the posters, banners, etc. on September 1st, 2006. Nina and I went for a walk through downtown Chunli to visit the McDonald's there. We found two that had my photo up: a large one at a major intersection that had two banners outside and my poster inside, and a smaller one that just had the banner. As we were visiting, some Taiwanese high school girls looked at me and asked Nina, "Is that him?"

If anybody from McDonald's in Taiwan (or elsewhere) reads this, know that I enjoyed working with such a professional group, which was a big part of why I wanted to do this. Plus, it makes for a good story.

"My face ached for days from trying to smile wider."

Professor **Peter Allen**, Computer Science, **Dennis Fowler**, Surgery, and Professor **Nabil Simaan**, Mechanical Engineering, have been awarded a 3 year NIH Exploratory Research Grant to develop an "Image Guided in vivo Tooling Platform for Minimal Access Surgery". This project focuses on the development of a newly conceived insertable robotic effector platform and the integration of that platform with a recently developed insertable, remotely controlled camera system to be used for minimal access surgery. The project will involve the actual design and construction of the platform for tools and the integration of the imaging platform (insertable camera system) with the tools into a fully functional image guided system for minimal access surgery. This may also include the addition of various sensors on the tools, so that the resultant data stream from both the imaging platform and the tools can be processed to control the intervention. The overall aim is to develop a disruptive technology that includes an insertable image source, a wide range of surgical tools, and a computer to integrate the function of all components.

Ph.D. student **Hila Becker** was selected as a finalist in the 2007 Anita Borg Memorial Scholarship competition. The Anita Borg Institute for Women and Technology and Google sponsor the 2007 Google Anita Borg Memorial Scholarship. The Google Anita Borg Memorial Scholarship was established to honor the legacy of Anita Borg and her efforts to encourage women to pursue careers in computer science and technology. Finalists receive a cash award. For the 2006-2007 academic year, the institute received over 250 applications from students at 115 different universities across the country. Eligible students must be going into their final year of study at a US university or college. Selection criteria includes academic performance, letters of recommendation, answers to short essay questions and interviews with members of the review committee. After three rounds of review, the committee selected 50 finalists.

Ph.D. student **Matei Ciocarlie** won the best student paper award at the 2007 World Haptics Conference in Tsukuba, Japan for his paper "Soft Finger Model with Adaptive Contact Geometry for Grasping and Manipulation Tasks" (co-authored with **Claire Lackner** and Professor **Peter Allen**). The prize carries an award of $1,000.

Professor **Stephen Edwards** was featured in an EE Times article about a 'Wild and Crazy Idea' session at the 2007 Design Automation Conference (DAC) held in San Francisco in June.

Professor **Steven Edwards** was awarded an NSF grant (joint with Professor Edward Lee of UC Berkeley) on "Precision Timed Architectures". This project proposes to reintroduce timing predictability as a first-class property of embedded processor architectures.

Professor **Eitan Grinspun** won an NSF CAREER award, titled "Multiresolution Foundations for Physics-Based Computer Animation and Interactive Engineering Design". This project will develop new computer simulation techniques that help engineers, scientists and artists to understand the behavior of complex mechanical systems. Software that helps to develop intuition will help engineers to produce better designs, spur scientists to poise more likely hypotheses, and give artists better control over the process of computer animation.

Professor **Jonathan Gross**'s newest book, Combinatorial Methods with Computer Applications, was published in late October 2007. The topics include recurrences, sums, binomial coefficients, permutations, number-theoretic techniques, graph enumeration, and combinatorial designs. The cover features the painting "Peaceable Kingdom", by the American artist Edward Hicks.

Professor **Julia Hirshberg** was elected to the University Senate.

Professors **Gail Kaiser** and **Jason Nieh** were awarded a grant from the NSF to study "Autonomic Mechanisms for Reducing System Downtime due to Maintenance and Upgrades". The project investigates and develops autonomic mechanisms for reducing system downtime due to software maintenance and upgrades. The project addresses operating system upgrades and also application upgrades, focusing on stand-alone binary-executable applications. The main goal is to lessen the possibility that patches and updates will "break" expected functionality of the environment that worked fine together with the old version—overall maximizing availability and reliability both during and after maintenance while imposing little management overhead.

Professor **John Kender** received a highly-competitive NSF-Information and Intelligent Systems Division grant titled "Analysis and Display of Semantics of Structured and Unstructured Videos". This supplements an earlier award from the Disruptive Technology Office. This project explores three new related approaches to making the indexing and retrieval of videos more efficient, meaningful, and humanly navigable, even when the videos have little editor-imposed structure. The first is the exploration and refinement of a novel, highly efficient machine learning technique for data-rich domains, which selects small and fast subsets of multimedia features that are most indicative of a given high-level concept. Speed-ups of three decimal orders of magnitude are possible. The second is the development of new methods and tools for refining user concepts and domain ontologies for video retrieval, based on statistical analyses of their collocation and temporal behavior. The goals are the determination of video synonyms and hypernyms, the verification of temporal shot patterns such as repetition and alternation, and the exploitation of a newly recognized power-law decay of the recurrence of content. The third is the demonstration of a customizable user interface, the first of its kind, to navigate a library of videos of unedited and relatively unstructured student presentations, using visual, speech, facial, auditory, textual, and other features. These features are shown to be more accurately and quickly derived using the results of the first investigation, and more compactly and saliently presented using the results of the second.

Professor **Angelos Keromytis** was awarded a 3-year grant from the NSF CyberTrust program to investigate the use of a new communication paradigm against network denial-of-service attacks. This project will investigate a new communication paradigm, named PacketSpread, which makes feasible the use of capability-like mechanisms on the current Internet, without requiring architectural modifications to networks or hosts. Professor Keromytis was also awarded a different 2-year grant by the NSF CyberTrust program, where he will investigate the use of new security policy models and mechanisms that integrate various protection and reaction elements in an enterprise network in a coherent and effective defense system.

Professor **Angelos Keromytis** won a 5-year grant with the University of Pennsylvania and Georgia Tech, titled "Foundational and Systems Support for Quantitative Trust Management". The goal of this project, funded by the Office of Naval Research (ONR) under the Multi-University Research Initiative (MURI) program is to develop Quantitative Trust Management as the basis for a scalable decentralized approach to dynamic, mission-based access control (MBAC). Dynamic trust management techniques will address the inabilities of current capabilities to maintain security policies at the operational tempo required for network-centric warface, to scale to emerging nation-state threats, and to manage heterogeneous computing and network elements supporting Service-Oriented Architectures (SOA).

Professors **Angelos Keromytis**, **Jason Nieh** and **Sal Stolfo** won a 5-year MURI grant, funded by the Air Force, titled "Autonomic Recovery of Enterprise-wide Systems After Attack or Failure with Forward Correction", joint with GMU and Penn State University. This project will develop autonomic recovery and regeneration mechanisms that will enable commodity systems to detect attacks, corruptions, and failures, then self-regenerate to a known good state, for both program and data, while increasing the reliability and security of the software to be more resistant and less vulnerable to attack.

Professor **Tal Malkin** received a grant from Mitsubishi Electric Research Laboratories (MERL) to study privacy-preserving learning.

Professor **Vishal Misra** was elected to the board of directors of ACM SIGMETRICS. SIGMETRICS is the ACM Special Interest Group (SIG) for the computer/communication system performance community.

Professor **Shree Nayar**'s work on computational photography was prominently featured in an April 2007 edition of Science News.

Associate professor **Jason Nieh** was awarded tenure. Congratulations, Jason!

In Spring 2006, Professor **Steven Nowick** was invited by Duane Armstrong, a manager at NASA Goddard Space Flight Center (Greenbelt, MD), to collaborate in developing a new advanced measurement chip for space applications. The project is exploring the potential benefit of clockless (i.e. asynchronous) digital circuitry for time-of-flight laser measurement. Over the past year, using one of Prof. Nowick's asynchronous digital circuit styles for controllers (called "burst-mode"), and his automated computer-aided design (CAD) tool "Minimalist", they have successfully designed a working chip. The chip is being incorporated by NASA into several prototype applications, where it will be evaluated for use in future space missions.

Ph.D. student **Janak Parekh** spoke on behalf of doctoral candidates at Convocation on May 14 as part of Columbia's 2007 graduation exercises. This is the Convocation for doctoral candidates in the schools of Architecture, Business, Engineering, Journalism, Law, Nursing, Physicians and Surgeons, Public Health, Social Work, and Teachers College. Vice President **Lee Goldman** and a faculty member also spoke at this celebration.

Professor **Dana Pe'er** is among 29 researchers selected from more than 2,100 applicants to receive the 2007 National Institutes of Health (NIH) Director's New Innovators Awards. The awards recognize the promising work of scientists who are in the early years of their career and who have not previously received an NIH regular research, or R01, grant. *(See the article in this edition of the newsletter.)*

Professor **Ravi Ramamoorthi** was named the winner of the 2007 Significant New Researcher Award by ACM SIGGRAPH, the ACM special interest group on computer graphics. The citation reads: "The Significant New Researcher Award is awarded annually to a researcher who has made a recent significant contribution to the field of computer graphics and is new to the field. The intent is to recognize people very early in their career who have already made a notable contribution and are likely to make more. ACM SIGGRAPH is proud to recognize Ravi Ramamoorthi for his groundbreaking work on mathematical representations and computational models for the visual appearance of objects". *(See the article in this edition of the newsletter.)*

Professor **Ravi Ramamoorthi** was named a winner of the 2007 Office of Naval Research Young Investigator Award, one of only two awardees in Computer Science.

Professor **Ravi Ramamoorthi** received an NSF grant titled "Theory and Algorithms for High Quality Real-Time Rendering and Lighting/Material Design in Computer Graphics". The proposed research is on developing the theoretical foundations and next generation practical algorithms for high quality real-time rendering and lighting/material design.

Research scientist **Owen Rambow** of the Center for Computational Learning Systems (CCLS) received an NSF grant titled "Email, Social Networks, and Organizations: Investigating How We Use Language to Create and Navigate Social and Organizational Relations". The project studies three aspects of human linguistic communication: the language used in the communication (for example, whether formal or informal), the topology of the social network of the communicators (for example, whether the the speaker is embedded in a single tight-knit group), and the roles the communicators occupy in an organization (for example, whether the speaker is an upper-level manager, or an administrative assistant).

Ph.D. student **Joshua Reich** won the student demo contest at 2007 ACM Mobicom/hoc for his Roombanet system. (Vacuum cleaners roaming around the courtyard have been a familiar sight in the CS department for some time.) His project was titled "MadNETt: DTNs on Roombas". Mobicom is the flagship wireless conference and the 2007 conference was joint with Mobihoc. *(See the article in this edition of the newsletter.)*

Professor **Henning Schulzrinne** was elected Vice Chair of ACM SIGCOMM. SIGCOMM is the ACM Special Interest Group (SIG) on Data Communications; it is ACM's professional forum for the discussion of topics in the field of communications and computer networks, including technical design and engineering, regulation and operations, and the social implications of computer networking.

Professors **Rocco Servedio** and **Tal Malkin** received an NSF CyberTrust grant on cross-leveraging cryptography and learning theory. The project proposes a detailed study of the connections between cryptography and learning. Very roughly speaking, cryptography is about manipulating and encoding information so that it is difficult to reconstruct the initial information, while learning theory is about efficiently extracting information from some unknown object. This duality means that ideas and results from each area can potentially be leveraged to make progress in the other area. The first main goal of the project is to obtain new cryptographic results based on the presumed hardness of various problems in computational learning theory. Work along these lines will include constructing and applying cryptographic primitives such as public-key cryptosystems and pseudorandom generators from learning problems that are widely believed to be hard, and exploring the average-case learnability of well-studied concept classes such as decision trees and DNF formulas. The second main goal of the project is to obtain new learning results via cryptography. The PIs will work to develop privacy-preserving learning algorithms; to establish com-

Department News & Awards (continued)

putational hardness of learning various Boolean function classes using tools from cryptography; to obtain computational separations between pairs of well-studied learning models; and to explore the foundational assumptions of what are the minimal hardness assumptions required to prove hardness of learning.

Ph.D. student **Yingbo Song** was named as a SEAS "Great TA" for the Fall 2006 semester. Congratulations, Yingbo!

Professors **Sal Stolfo**, **Angelos Keromytis** and **Shlomo Hershkop** won a 2-year grant from the Department of Homeland Security (DHS), as part of a larger project with the Institute for Information Infrastructure Protection (I3P). The project, titled "Human Behavior, Insider Threat, and Awareness", will focus on investigating, developing, and experimentally evaluating methods and models for detecting malicious insider activity and behavior on a host computer system.

Ph.D. student **Sean White**, M.S. student **Dominic Marino** and Professor **Steve Feiner** won the Best Note award at CHI 2007 for their short paper, "Designing a Mobile User Interface for Automated Species Identification". CHI 2007 is the top conference in human-computer interaction, and was held in San Diego, April 28-May 3, 2007.

Alumni News

Former PhD Student **Montek Singh** (Ph.D., 2001) has been promoted with tenure to associate professor of computer science at the University of North Carolina, Chapel Hill. His research is in high-performance asynchronous and mixed-timing digital circuits and architectures.

Larry Leftin and colleagues at Honeywell were recently awarded a patent in control theory: "Methods and systems for implementing an iterated extended Kalman filter within a navigation system" (United States Patent 20070213933). This patent applies to all navigation systems, especially those whose hardware exhibits non-linearities.

Eric Siegel (Ph.D., 1997) has spent the last four years founding a data mining and predictive analytics boutique firm in San Francisco. The company, Prediction Impact, Inc. *(www.predictionimpact.com)*, provides consulting services to its clients by forming specialized teams from its network of senior consultants. The firm also offers commercial training programs; more information about these is available from the website. On the personal side, Eric is married to Maria de Fatima Callou, a fashion designer he met five years ago while visit Jacques Robin (Ph.D. 1995) in Brazil. "Thanks for networking me to true love!" Eric proclaims to the CS department, and, especially, Dr. Robin.

Recent & Upcoming **PhD Defenses**

Ricardo Baratto

THINC: A Virtual and Remote Display Architecture for Desktop Computing

Continuing advances in the performance and ubiquity of networks have enabled the proliferation of technologies that extend our computing environment beyond the boundaries of a single computer. As one of these technologies, remote display has enjoyed increased popularity in recent years due to the number of benefits associated with it. However, its mass acceptance has been hampered by the inability of current technology to provide a high fidelity visual and interactive experience for end users across the vast spectrum of graphical and multimedia applications commonly found on traditional desktop computers.

This dissertation introduces THINC, a virtual display architecture for high performance remote display on local and wide area network environments. THINC virtualizes the display by introducing a simple device driver that looks and behaves like traditional, hardware-specific drivers. On top of the driver, THINC implements novel optimizations for efficiently translating high level application display requests into a set of simple, low-level commands, and techniques to prioritize and guarantee that only relevant commands are delivered to clients. We have compared THINC's performance to a set of current, state of the art commercial remote display systems. Our results demonstrate that THINC can deliver good interactive performance even when using clients located around the world, and provides much faster response time than other systems in WAN environments.

Building on THINC's core architecture, we present two systems that address the need for multimedia and mobile user support in remote display systems. First, we extend our virtualization approach to audio devices to provide audio capture and playback, and leverage existing video acceleration interfaces to efficiently deliver synchronized audio/video desktop content. Second, we present pTHINC, a PDA remote display solution that leverages server computing power to run full-function desktop environments and only send simple screen updates to the PDA for display. pTHINC uses server-side scaling of display updates to provide high-fidelity and seamless mobility across a broad range of different clients, screen sizes, and orientations. pTHINC also leverages existing PDA control buttons to improve system usability and maximize available screen resolution for application display.

Finally, this dissertation demonstrates how THINC's virtual display approach can be used for desktop computing systems beyond remote display. First, we show how display virtualization can be used to encapsulate the state of a desktop session and decouple it from the underlying hardware and operating system instance, allowing it to be transferred across computers. Second, we present a system that leverages THINC's inherent ability to redirect display output to transparently record all desktop content. Coupled with a text capture mechanism, this system enables users to playback, browse, and search the recorded data, allowing them full visual access to any information that has ever been displayed on their desktop.

Arvid Bessen

On the Complexity of Classical and Quantum Algorithms for Numerical Problems in Quantum Mechanics

Our understanding of complex quantum mechanical processes is limited by our inability to solve the equations that govern it except for simple cases. Numerical simulation of quantum systems is therefore essential to understand, design and improve quantum systems. We would like to understand the computational complexity of quantum mechanical problems and investigate the effect the introduction of quantum algorithms has on these problems.

In this thesis we explore different computational problems from quantum mechanics, study their computational complexity and try to find ways to realize the best possible algorithm. We investigate the reasons behind the improved performance of Shor's quantum factorization algorithm and we study the quantum phase estimation algorithm used by Shor's algorithm. We apply our techniques to the Sturm-Liouville eigenvalue problem and prove lower bounds for it.

We study the importance of destructive interference in quantum computing and how it makes quantum algorithms more powerful than classical stochastic computation. In order to do that we consider the complexity class StoqMA of algorithms that only use classical gates, but are given a quantum "witness" and show that there exists a complete problem for this class, which we call the stoquastic local Hamiltonian problem.

Finally we study a computational problem from lattice quantum chromodynamics (QCD). In most popular algorithms that treat problems in QCD the (gauged) Dirac matrix has to be inverted numerous times. We study a newly proposed multigrid method, adaptive smoothed aggregation that has promise to overcome these difficulties. We show that while classical CG's convergence becomes worse as the matrix becomes almost singular, adaptive smoothed aggregation will still perform well.

Hanhua Feng

Scheduling: From Optimality to Configurability

We begin by studying optimal scheduling policies in a single-server system and optimal dispatching policies in systems with multiple parallel queues and servers. With different assumptions of system characteristics (e.g. workload patterns) and performance requirements (e.g. mean response time and mean slowdown), the optimal scheduling policies differ vastly. As an example, in a system with a single queue, we show that a scheduling policy can be the best scheduling policy for one kind of service-time distributions and at the same time the worst for another, and that a scheduling policy can be the best policy in terms of mean slowdown but the worst in terms of mean response time.

Since there is no such thing as ubiquitous optimality, we put our efforts into two directions. First, inspired by the observation that the variability of service time is a dominant factor on the optimality of scheduling policies, we set constraints on service-time variability and investigate the best and worst scenarios in terms of mean response time for several scheduling policies on a single-server queue. Second, we propose a configurable policy that can be tuned with a parameter. This configurable policy is not only analyzed in both deterministic and stochastic configurations, but also implemented in Linux as a kernel scheduler. We justify by experiments our claim that the ubiquitous optimality can now be achieved by configurability.

Elena Filatova

Unsupervised Relation Learning or Event-Focused Question-Answering and Domain Modelling

In this thesis, we investigate the problem of identifying, within a text, relations that capture information important for event-focused document collections. The presented solutions work with events of various granularity and show how to use these relations to improve the performance of a number of natural language processing applications.

For a set of related event-focused documents, we intro-

duce a notion of a shallow semantic network based on the relations between the important elements discovered in these documents. This shallow semantic network captures the most important relations among the objects, people, and other elements that are involved in the events described in the input document collection. We present experimental evidence that such a relation-based representation of event-focused documents is superior to techniques that rely on term frequencies for the task of information selection.

For a set of document collections describing similar events within the same domain, we design and implement a completely automatic, data-driven procedure for inducing domain templates. These domain templates capture facts that are important for most domain instances. We then devise a procedure for identifying commonalities across different subdomains. We experiment with a special case of a domain, a biography domain, and identify commonalities across activities used for descriptions of people belonging to different occupations. We also propose a methodology for creating domain hierarchies.

We apply our methods for identifying relations to the question-answering task. We design and implement a two-pronged approach for answering open-ended event-related questions. The first approach relies on automatically created domain templates and is used when the event mentioned can be identified as one of a particular class of events (e.g., earthquakes, presidential elections, etc.). The second, complementary approach is based on a shallow semantic network, which we extract from the documents relevant to the question.

We also suggest a formal model for efficient information packaging that is based on mapping the information selection task on the set cover complexity problem. Using this mapping, we outline and implement information selection algorithms that are provably optimal polynomial-time approximations for information selection tasks that have a limit on the output size.

Kshitiz Garg

Rain in Vision and Graphics

Rain produces sharp intensity fluctuations in images and videos which severely degrade the performance of outdoor vision systems. Removing the visual effects of rain is therefore important in order to make outdoor vision robust to rain. In contrast, in graphics, rain effects are desirable as they are often used to convey scene emotions in movies and in other graphics applications, such as games, to enhance realism. Developing efficient algorithms for realistic rendering of rain effects is however challenging. This thesis, studies rain from the perspective of vision and graphics. We have two main goals (1) to understand the visual appearance of rain and (2) to develop efficient algorithms both for handling its effects in vision and for its realistic rendering in graphics.

We begin by modeling the visual appearance of rain. We describe the appearance by modeling the photometric intensities produced by individual raindrops and the space-time correlation produced by the motion of a large number of raindrops. We then use these appearance models to develop two complementary approaches to handle the effects of rain on vision systems. The first approach uses a post-processing algorithm for detection and removal of rain from videos that have already been captured. In the second approach we use our appearance model to show that the unique physical properties of rain its small size, high velocity and spatial distribution makes its visibility depend strongly on camera parameters. This dependence is used to reduce the visibility of rain during image acquisition by judiciously selecting camera parameters. Conversely, camera parameters can also be chosen to enhance the visibility of rain. This ability can be used to develop an inexpensive and portable camera-based rain gauge that provides instantaneous rain-rate measurements.

In graphics, we have developed two models for realistic rendering of rain. In the first model we consider the complex appearance patterns produced by close by raindrops. We show that these patterns are produced due to the rapid shape distortions (i.e. oscillations) that a raindrop undergoes as it falls and develop a physics-based model to faithfully render these complex appearance patterns. This model is then used to develop an efficient algorithm for rendering rain in images and videos. In the second work, we focus on realistic rendering of material-dependent splashing of raindrops. Here we measure raindrop splashes for 22 different real world materials. These measurements are then used to build a compact stochastic model that accurately captures the material and inclination dependent splashing behavior of raindrops. Additionally, it allows user to render physically plausible splashes for novel materials that have not been measured.

Cheoljoo Jeong

Optimization Techniques for Robust Asynchronous Threshold Networks

Digital systems design is changing rapidly as integrated circuit technology evolves. The arrival of deep submicron technology brings about increasing design challenges in terms of obtaining circuit reliability, increasing need for low-power design, and the handling of system complexity. It is critical to cope with these challenges in both circuit design and in automated computer-aided design (CAD) flows.

One attractive alternative is to use robust asynchronous circuits which gracefully address some of these problems. Asynchronous circuits communicate through handshaking in a distributed way rather than by adopting a centralized control such as the global clock. Asynchronous design has been the focus of renewed interest and research activity because of its potential benefits of low power consumption, low electromagnetic interference, robustness to delay variations, and modularity of designs. However, they typically suffer from high area and latency overhead, as well as lack of CAD tools and optimization techniques. Also, careful design is required since hazard-freedom is essential for their correct operation.

This thesis presents optimization techniques for a class of asynchronous circuits, called asynchronous threshold networks, which is one of the most robust asynchronous circuit styles. Two techniques are proposed: (i) technology mapping and cell merger, and (ii) relaxation. The technology mapping and cell merger algorithm optimizes the circuits using a specific technology library while the relaxation technique relaxes the overly-restrictive style of the circuits. Both of these techniques optimize area or delay of the circuits while still fully-preserving their robustness properties. Also, the proposed algorithms were implemented in automated CAD tools. Experiments were performed on industrial circuits and largest MCNC benchmarks with significant improvement in both area and delay.

Jackson Liscombe

Prosody and Speaker State: Paralinguistics, Pragmatics, and Proficiency

Prosody—suprasegmental characteristics of speech such as pitch, rhythm, and loudness—is a rich source of information in spoken language and can tell a listener much about the internal state of a speaker. This thesis explores the role of prosody in conveying three very different types of speaker state: paralinguistic state, in particular emotion; pragmatic state, in particular questioning; and the state of spoken language proficiency of non-native English speakers.

Paralinguistics. Intonational features describing pitch contour shape were found to discriminate emotion in terms of positive and negative affect. A procedure is described for clustering groups of listeners according to perceptual emotion ratings that foster further understanding of the relationship between acoustic-prosodic cues and emotion perception.

Pragmatics. Student questions in a corpus of one-on-one tutorial dialogs were found to be signaled primarily by phrase-final rising intonation, an important cue used in conjunction with lexico-pragmatic cues to differentiate the high rate of observed declarative questions from proper declaratives. The automatic classification of question form and function is explored.

Proficiency. Intonational features including syllable prominence, pitch accent, and boundary tones were found to correlate with language proficiency assessment scores at a strength equal to that of traditional fluency metrics. The combination of all prosodic features further increased correlation strength, indicating that suprasegmental information encodes different aspects of communicative competence.

Michael E. Locasto

Integrity Postures for Software Self-Defense

Software currently lacks the capability to respond intelligently and automatically to hacking attempts in a way that preserves both its availability and integrity. Unfortunately, most current software protection techniques typically abort a process after an intrusion attempt or a fault violates the integrity of an application's data or execution artifacts, thereby transforming an arbitrary exploit or failure into a self-induced denial of service attack—resulting in a compromise of both the application's integrity and its availability. Although many consider this approach safe, it is unappealing because systems remain susceptible to the original fault upon restart and risk losing accumulated state.

What is needed, then, are mechanisms that preserve both the internal and external integrity postures of a software application so that it can continue to provide service in the face of a new attack or similar future instances of that attack. Preserving external integrity can prevent the recurrence of a particular fault or vulnerability, and it can be accomplished by filtering input that triggers or exercises the underlying error. Preserving internal integrity can be accomplished by making informed, targeted changes to a program's state in case a new attack is able to bypass the external integrity monitor.

In order to support the maintenance of both internal and external integrity postures, I introduce the notion of "policy-constrained speculative execution", where the policy in question is an "integrity repair policy" based on my extensions to the Clark-Wilson Integrity Model. Speculatively executing slices of an application, or "microspeculation", is similar to hardware speculative execution, except that acceptance of a particular slice of execution is predicated on the integrity repair policy rather than the result of a branch conditional. These policies contain a collection of constraints that my system leverages to restore an internal integrity posture while an attack or fault occurs. Repair policy effectively customizes a software system's response to attack.

To demonstrate the feasibility and effectiveness of this approach to self-healing, I created a runtime environment, STEM, that supervises an application's execution and enforces integrity repair policies in both binary-only and source-available environments.

In order to preserve the external integrity posture of a software system, I combined STEM with an existing network content anomaly detector to create FLIPS, a system capable of automatic exploit signature generation. No system supplies perfect security, but we can provide intelligent, well-formed recovery mechanisms and automatically invoke them. An integrity repair model assists in bridging the gap between current systems and systems that can automatically self-heal.

Hassan Malik

Efficient Algorithms for Clustering High Dimensional Data using Interesting Patterns

Recent advances in data mining allow for exploiting patterns as primary means for clustering large collections of data. In this thesis, we present three advances in pattern-based clustering technology, and a related advance in pattern frequency counting. In our first contribution, we analyze numerous deficiencies with traditional pattern significance measures such as support and confidence, and propose a web image clustering algorithm that uses an objective interestingness measure to identify significant patterns, yielding measurably better clustering quality.

In our second contribution, we introduce the notion of closed interesting itemsets, and show that these itemsets provide significant dimensionality reduction over frequent and closed frequent itemsets. We propose GPHC, a sub-linearly scalable global pattern-based hierarchical clustering algorithm that uses closed interesting itemsets, and show that this algorithm achieves up to 11% better FScores and up to 5 times better Entropies as compared to state of the art of agglomerative, partitioning-based, and pattern-based hierarchical clustering algorithms on 9 common datasets.

Our third contribution addresses problems associated with using globally significant patterns for clustering. We propose a pattern-based hierarchical clustering algorithm that builds the cluster hierarchy without mining for globally significant patterns. This algorithm allows each instance to "vote" for its representative size-2 patterns in a way that ensures an effective balance between local and global pattern significance, produces more descriptive cluster labels, and allows a more flexible soft clustering scheme. With results of experiments performed on 40 standard datasets, we show that this algorithm almost always outperforms state of the art hierarchical clustering algorithms and achieves up to 15 times better Entropies, without requiring any tuning of parameter values, even on highly correlated datasets.

Our final contribution deals with the problem of finding a dataset representation that both offers a good space-time tradeoff for fast support (i.e., frequency) counting, and also automatically identifies transactions that contain the query itemset. We compare FP Trees, and Compressed Patricia Tries against several novel variants of vertical bit vectors. We compress vertical bit vectors using WAH encoding and show that simple lexicographic ordering may outperform the Gray code rank-based transaction reordering scheme in terms of RLE compression. These observations lead us to propose HDO, a novel Hamming-distance-based greedy transaction reordering scheme, and aHDO, a linear-time approximation to HDO. We present results of experiments performed on 15 common datasets with varying degrees of sparseness, and show that HDO- reordered, WAH encoded bit vectors may take as little as 5% of the uncompressed space, while aHDO achieves similar compression on sparse datasets. With results from over 109 database and data mining style frequency query executions, we show that bitmap-based approaches result in up to 102 times faster support counting, and that HDO-WAH encoded bitmaps offer the best space-time tradeoff.

Cristian Soviani

High-Level Synthesis for Packet Processing Pipelines

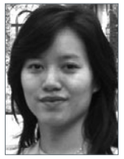
Packet processing pipelines are essential components of state-of-the-art network routers and switches. Their design methodology is a delicate trade-off between, on one hand, the desire of abstract specifications, short development time, and design maintainability and, on the other hand, a very aggressive performance requirement.

The present thesis proposes a coherent design methodology for packet processing pipelines. Following systematically the design flow, it starts with the high level specification of the pipeline, where a domain specific language is introduced. The next step is the RTL-synthesis of the high level model and the worst case throughput analysis of the synthesized pipeline, as well as a FIFO sizing technique. Finally, I address some characteristic RTL-level optimization issues.

I claim, based on experimental results, that my proposed techniques can dramatically improve the design process of these pipelines, while the resulting performance matches the expectations of a hand-crafted design.

The considered pipelines exhibit a pseudo-linear topology, which can be too restrictive in the general case. However, especially due to its high performance, such architecture may be suitable to applications outside packet processing, in which case some proposed techniques can be easily adapted.

As I ran my experiments on FPGA platforms, this work has an inherent bias towards this technology; however, most results are technology independent.



Jia Zeng

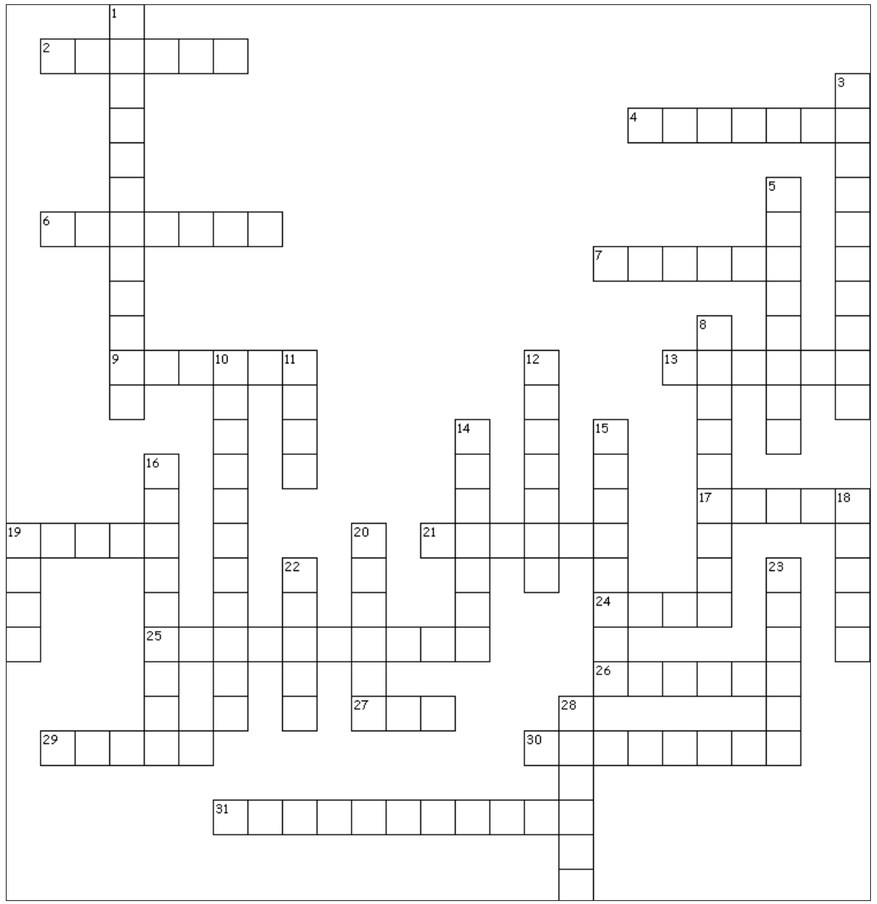
Partial Evaluation for Code Generation from Domain-Specific Languages

Partial evaluation has been applied to compiler optimization and generation for decades. Most of the successful partial evaluators have been designed for general-purpose languages. Our observation is that domain-specific languages are also suitable targets for partial evaluation. The unusual computational models in many DSLs bring challenges as well as optimization opportunities to the compiler.

To achieve an interesting level of optimization, partial evaluation has to be specialized to fit the specific paradigm. In this dissertation, we present three such specialized partial evaluation techniques designed for specific languages that address a variety of compilation concerns. The first algorithm provides a low-cost solution for simulating concurrency on a single-threaded processor. The second enables a compiler to compile modest-sized synchronous programs in pieces that involve communication cycles. The third statically elaborates recursive function calls that enable programmers to dynamically create a system's concurrent components in a convenient and algorithmic way. Our goal is to demonstrate the potential of partial evaluation to solve challenging issues in code generation for domain-specific languages.

Naturally, we do not cover all DSL compilation issues. We hope our work will enlighten and encourage future research on the application of partial evaluation to this area.

CS **Professors**



ACROSS

2 Detects intruders
4 Architect who reuses components
6 Can summarize the news into a few sentences
7 Augments reality
9 Director of the programming systems laboratory
13 Distinguished faculty teacher
17 Founded cricket website
19 Can see through raindrops
21 Our main weapon on the front line in 1004
24 Uses CS to study disease
25 Knuth prize winner
26 Working with NASA to send chips into space
27 Wrote the book on compilers
29 Wrote the book on Combinatorial Methods with Computer Applications
30 Wrote the book on firewalls
31 Young investigator award winner

DOWN

1 He's a quantum computer
3 A good listener
5 Computationally learns, at least theoretically
8 Can teach you to measure biometrics
10 Internet phone pioneer
11 Makes your queries fast
12 Web searcher
14 Likes to be embedded
15 Physical simulator
16 A strong man
18 Lends a hand to surgeons
19 Advised on Microsoft antitrust case
20 Is discriminative and generative
22 Wrote the book on complexity and information
23 Advanced cryptographer
28 Has SMARTS

Caption **Contest**

The **winner of Spring 2007's caption contest** for the picture of Steve Bellovin and Angelos Keromytis *(shown left)* was **Michael Rabin**, with the caption:

"Have you heard,
Zvi Galil gave up Tel Aviv
and is returning
to the CS department!"

Two new photos are included for this issue's caption contest. Please send suggestions to *newsletter-editors@lists.cs.columbia.edu*. The winners will be announced in the spring newsletter. In case you have trouble making out the faces, the second image includes Professors Malkin, Misra, Schulzrinne, Nayar, and Kender.

PHOTO 1

PHOTO 2