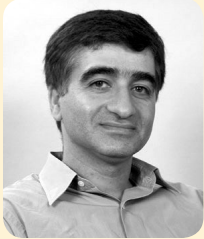


# Computer Engineering Newsletter

<http://ceng.usc.edu/>

August 2008



## Computer Engineering at USC

Massoud Pedram, Professor of Electrical Engineering and Chair of the Computer Engineering Division

Over the last few decades, a marvelous computer revolution has rapidly changed the way much of the world works. Developments in transistor manufacturing, electronic circuits, computers and information processors, and telecommunications have fundamentally altered human life. The field of Computer Engineering (CENG) is at the epicenter of this development. It encompasses a wide range of topics including operating systems, computer architecture, computer networks, system design, and computer-aided design.

The USC CENG is a division of the Electrical Engineering department at the Viterbi School of Engineering, which boasts of world-renowned faculty, high-profile and impacting research projects, and extremely motivated students and research staff. The USC's CENG program consistently ranks among the nation's best in the U.S. News and World Report analysis. Our faculty has received many prestigious awards, including many NSF Career awards and various IEEE/ACM recognitions.

The USC CENG seeks to play a major role and lead in the field of computer engineering through its research agenda and industrial collaborations. Additionally, it aims to train students to be industry, academic and government leaders, provide service to the society through our research and training, and create a supportive and stimulating intellectual environment wherein faculty, students, and staff can achieve their best. Our faculty believes in excellence in teaching and research, exposing students to fundamental knowledge and problem-solving techniques as well as practical and experimental skills that will enable them to excel in today's high-tech industry.

Our faculty specializes in a multitude of computer engineering fields, including computer architecture and chip multiprocessors, pervasive/grid computing, cyber-security, computer networks, wireless and sensor networking, embedded and reconfigurable systems, VLSI design, and electronic design automation. Cutting edge research projects span the gamut: from low power design of VLSI circuits to testing of circuit components to design and analysis of energy-efficient and robust mechanisms for querying, routing, and self-configuration in wireless sensor networks to modeling algorithms and architectures for reconfigurable computing, and more!

Many interdisciplinary research centers and laboratories have a home in CENG, including Cognitive Enabled ARCHitectures and the BioRC Biomimetic Real-Time Cortex. The CENG faculty members collaborate with partners from academia and industry both within the Viterbi School of Engineering at USC and worldwide. For example, the TransLight/PacificWave project is developing a distributed exchange facility on the West Coast (currently in Seattle, Sunnyvale, and Los Angeles) to allow interconnection of international research and education networks with US research networks.

CENG degree programs provide students with an intensive background in the function, simulation and analysis, architecture and organization, design and optimization of digital information processing systems. CENG has 19 faculty members, including research, teaching, and adjunct faculty. It serves about 120 undergraduate students through a joint degree program with the USC's Computer Science Department (CECS) and approximately 200 Master's level graduate students. The CENG faculty members supervise some 50 PhD students.

## New CENG PhD Students

The Division of Computer Engineering would like to extend a warm welcome to the new students who have joined our faculty research projects.

Bardia Zandian (Annaram)  
Carlo Torniai – Post Doc (Prasanna)  
Charalampos Chelmis (Prasanna)  
Chenqian Jiang (Prasanna)  
Chuanyou Li (Breuer/Gupta)  
Fatemeh Kashfi (Pedram)  
Georgios Konstantinidis (Raghavendra/Prasanna)  
Jonathan Joshi (Parker)  
Ma Nam (Prasanna)

Matthew Walker (Parker)  
Mohammad Mirza-Aghatabar (Breuer/Gupta)  
Na Chen (Prasanna)  
Naser Khosropour (Pedram)  
Sabyasachi Ghosh (Annaram)  
Suk Hun Kang (Annaram)  
Waleed Dweik (Annaram)  
Yi-Hua Edward Yang (Prasanna)  
Yoon Sik Cho (Parker)

## CENG News

Rahul Jain has joined the CENG faculty in August 2008 as an Assistant Professor

Murali Annavaram is a recipient of the 2008 IBM Faculty Award

Massoud Pedram has been assigned as the Editor-in-Chief of the ACM Transactions on Design Automation of Electronic Systems

Bhaskar Krishnamachari has been promoted to Associate Professor

Sandeep Gupta is assigned as the EFC Chair through September 2008

Bhaskar Krishnamachari has been honored with the Mellon Award for Excellence in Mentoring in the category "Faculty to Graduate Mentoring"

Kostantinos Psounis won an award from CISCO Systems for his work on multi-hop wireless transport protocols

John Silvester was elected as Secretary of the Corporation for Education Network Initiatives in California (CENIC)

## CENG PhD Student Council

Animesh Pathak, CENG PhD Student Council Member

The CENG PhD student council was created in Fall 2007 by the CENG faculty with the aim to provide the PhD students in Computer Engineering a platform to express their concerns and work towards improving their academic and social lives at USC. The council consists of nominated representatives of each research group in the CENG division, and is facilitated from the faculty side by Professor Bhaskar Krishnamachari.

As adopted by the council members, the objective of the CENG PhD Student Council is to: (i) Help CENG PhD students enrich their experience as graduate students at USC by organizing events and programs to aid in their professional development and help build a stronger community; (ii) Serve as a platform for CENG PhD students to discuss the issues they face in their capacity as students and as graduate assistants, and work with the CENG faculty and other departmental resources to identify solutions and implement them.

Over the past academic year, the council met monthly, and was successful in achieving two important goals. Firstly, it successfully petitioned for and obtained access to the 3rd floor copy room for PhD students after regular working hours. This effort, on which the council worked closely with the administration to come up with a set of usage policies, has led to the availability of hot water and printing/copying facilities at all times. Secondly, the council's request led to the procurement of new computers for the TA office of the digital design courses.

In addition to the above, the CENG PhD student council has also drawn up detailed plans for the activities to be targeted in Fall 2008. Included among them are a central website with information about the research tools and facilities available in the various CENG research groups, and a comprehensive survey about the state of affairs of the CENG PhD students. Additionally, the council plans to organize "CENG's Next Top Toastmaster", a seminar series where PhD students are encouraged to present their work in an easy to understand manner, and can win prizes based on audience votes.

Students interested in being a part of this representative body are encouraged to contact the council members. The election of the President, Vice President, Secretary and Treasurer of the council will be done in its first meeting of Fall 2008.

More information about the CENG student council can be found at <http://cengsc.wikidot.com/>

---

---

## The CENG Staff

Estela Lopez, Division Coordinator, CENG and Controls

The CENG staff consists of a Division Coordinator, Budget Business Analysts and a Budget Business Technician. Our staff is responsible for compiling and analyzing data to prepare budget proposals for the CENG faculty. We develop, analyze, and execute budgets, which are used to allocate current resources and estimate future financial needs.

Through collaborative teamwork our staff enables the Division of CENG to fulfill the University's mission by providing the highest level of quality support, service and resources to faculty, students and staff. We maintain a work environment – both internal and external to our department - that is conducive to mutual respect, professional growth, and productivity.

Estela Lopez: Division Coordinator, CENG and Controls

Annie Yu: Budget Business Analyst

Janice Thompson: Budget Business Technician

Shane Goodoff: Budget Business Analyst, CENG & Controls



## Error-tolerance: a.k.a. all computations need not be correct

Melvin A. Breuer, Professor of Electrical Engineering and Computer Science, VLSI/CAD

A circuit is **error-tolerant** (ET) with respect to an application/system, if (1) it contains defects that cause internal and may cause external errors and/or performance degradation, and (2) the application/system that incorporates this circuit produces *acceptable results*. An error-tolerant **system** is one where at least one of its circuits can be error-tolerant. As it turns out, many systems are error-tolerant.

### Why would anybody want to use a defective chip?

The notion of error tolerance is motivated by three important trends in information processing, namely changes in fabrication technology, changes in the mix of applications, and emergence of new paradigms of computation. **Fabrication technology:** As we get closer to what some call the “end of CMOS”, we see the emergence of highly unreliable and defect-prone technologies. This is accompanied by rapid development of new computing technologies such as bio, molecular, and quantum devices. Most of these new technologies are also extremely unreliable and defect-prone. However, these new technologies also provide the ability to carry out massive numbers of computations in parallel and at speeds that far exceed those currently achieved by CMOS devices.

**Applications:** Increasingly larger fractions of the total number of chips fabricated in any given year implement multi-media applications and process signals representing audio, speech, images, video and graphics. The outputs of such systems eventually become input signals to human users. There are several interesting aspects to the computational requirements for such systems. (i) The result of computation, i.e., the output data, is not measured in terms of being right or wrong, but rather on perceptual quality to its human users. (ii) Most such systems are by design *lossy*, in the sense that the outputs deviate from perfection due to sampling of input signals, conversion to digital, quantization, lossy encoding, decoding and conversion to analog signals. (iii) Many such applications require parallel architectures as they are *computationally intensive* and have *real-time performance constraints*.

**Emerging paradigms of computation:** Several new paradigms are emerging on how functions are computed and what requirements are placed on the “correctness” and “accuracy” of the results. These methodologies include: *evolutionary computation* attempts to solve a problem based on analogies made with biological systems; *neural nets* where damage to a few elements or links need not significantly impair the overall performance; *approximate computations* where a radical departure from discrete correct/incorrect computation is required; and *probabilistic computations* where accuracy is measured as the probability of a result being correct.

### We need new ways to specify, design and test systems!

Due to the abovementioned emerging trends, it is critical to develop new paradigms for digital system specification, design and test that deal efficiently with high levels of *imperfections*, i.e., process variations, defect densities, and noise sensitivity, inherent in modern CMOS and new technologies. Current techniques, such as fault-tolerance, defect-tolerance and design-for-manufacturability, are not sufficient to address these issues. In the past designers have focused on gate minimization, performance and low-power; now yield is the real issue.

**Errors:** At the system level, we have used *error metrics* such as PSNR, distortion, mean-opinion-score (MOS) and bit error rate to characterize the severities of errors. At the level of logic blocks, we have concentrated on two types of error metrics, namely error significance and error rate. An *error significance* metric quantifies the amount by which the response at a circuit's outputs, interpreted as a numeric or coded symbol, deviates from the corresponding error-free response. *Error rate* is defined as the asymptotic value of the fraction of a large number of functional responses from a circuit that are erroneous. The characteristics of the application determine what types and specific combinations of metrics are appropriate and need to be part of a product spec. Analysis of the entire application also determines the limits for acceptability, or *thresholds*, on each error severity metric by itself or for each appropriate combination.

### So, what are some error-tolerant systems?

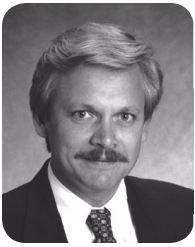
Our case studies on MPEG and JPEG encoders, answering machine, and decoders for turbo codes clearly support three main conclusions. (i) The notion of error tolerance can indeed provide significant benefits for a wide range of applications. (ii) The notion of error tolerance can be applied in different ways to provide benefits with respect to a wide range of criteria, including yield, power and delay. (iii) The characteristics of algorithm-, architecture-, and circuit-level designs have significant impact on the benefits that error tolerance can provide. In particular, as a rule-of-thumb, parallel architectures provide significantly higher benefits than lower-cost serial ones.

### So what areas are currently being researched?

**Test technology:** We have developed test techniques for estimating the error rate of a circuit as well as for binning circuits based on a range of error-rates. We have also developed test techniques for identifying the significance of an error, as well as the fraction of errors that exceed a given threshold value.

**Cost analysis:** We have developed a model for evaluating the cost effectiveness of marketing error-tolerant chips, based on many parameters such as yield, wafer costs, test costs and the fraction of defective chips that result in acceptable performance.

**Synthesis:** We are just beginning to develop synthesis and analysis techniques to guide the design of new systems that take advantage of the concept of error-tolerance and acceptable performance.



## Slack simulation of Chip MultiProcessors on Chip MultiProcessors

Michel Dubois, Professor of Electrical Engineering, Computer Architecture

### Chip MultiProcessors

Moore's law predicts that the computing power of microprocessors will double every 18-24 months at constant cost. Many see today strong evidence that the "end of the ride" is near, mostly because the miniaturization of CMOS technology is fast reaching its limit, the so-called CMOS endpoint.

In the past the only noteworthy trade-offs in architecture design were between cost, area and time (i.e., performance). Today, designs are challenged by several technological limits and the design problem is one of maximizing performance under multiple technological constraints. Power, wire delays, reliability, and the sheer complexity of designs must be taken into account during the design process. Although some of these problems can be addressed at the hardware (circuit) level or at the software level, architecture can play a significant role to maintain the viability of CMOS technology.

Because of these technological trends future micro-architectures will have to adopt some form of parallel execution model. Generically, this emerging form of micro-architecture is referred to as Chip MultiProcessors (or CMPs) and is one of the major focus of my research. One of the issue I address in my research is how to simulate future CMPs on current CMPs.

### Slack simulation

Traditionally, in a single-threaded simulation of a multiprocessor, the simulation of every core is interleaved on a clock by clock basis (cycle-by-cycle or cycle accurate simulation). Unfortunately this approach is very hard to parallelize. In this research we explore the simulation paradigm of simulating each core of a target CMP in one thread and then spreading the threads across the hardware thread contexts of a host CMP.

We relax the synchronization condition in various schemes which we call *slack simulations*. In slack simulations the various simulated cores do not necessarily synchronize after each simulated cycle, as in cycle accurate simulation, but rather there can be some slack between them. *Simulation Slack* is the cycle count difference between any two target cores in the simulation. Small slacks --such as a few cycles-- drastically reduces the amount of synchronization among simulation threads (as compared to cycle-by-cycle execution) and thus improves the efficiency of the simulation, at the cost of small or negligible simulation errors.

To enable the correct simulation of individual cores in a CMP so that each core simulation can be integrated in a simulation thread, we have made considerable modifications to SimpleScalar in order to simulate the micro-architecture of OoO cores with their cache hierarchies. Simulations are parallelized using POSIX Threads (Pthread).

The general framework of our CMP parallel simulation environment comprises two types of Pthreads: several core threads and one simulation manager thread. One core thread simulates a single target core of a CMP with its L1 cache. The simulation manager thread has two functions. Its first function is to simulate the on-chip lower level cache hierarchy including L2 cache banks and their interconnection with cores. The second is to orchestrate and pace the progress of the entire simulation.

### Simulation platform

The experimental platform is a Dell PC server powered by two Intel Quad-core Xeon processors (total of 8 cores). The operating system is a recent release of Ubuntu Linux Version 6.06. We have selected three benchmarks: *FFT*, *Jacobi*, and *SimpleSum*. The simulation is composed of 5 Pthreads. We use a maximum of 4 hardware thread contexts and therefore only one of the processors in the host.

We consider different slack simulation schemes: cycle-by-cycle or cycle accurate (CC); bounded slack simulation with 1-cycle slack (S1); bounded slack simulation with 5-cycle slack (S5); and slack simulation with unbounded slack (SU). We observe the following. Unbounded slack is at least 100% faster than cycle-by-cycle on 4 cores. More importantly, we see that even a small 5-cycle slack already yields a large improvement over cycle-by-cycle simulation. Relative simulation errors (with respect to cycle accurate simulation) are typically very small, even in the case of unbounded slack.

### Current and future work

One of the disadvantage of slack simulation with large slacks is that accuracy is sometimes compromised. Advanced techniques based on simulation violation detection are being explored. One technique is adaptive slack in which the slack is varied based on the frequency of detected violations. For future research, we also want to evaluate whether speculative simulation with rollback on detected violations will be feasible and useful.



## Opportunistic Spectrum Access in Cognitive Radio Wireless Networks

Bhaskar Krishnamachari, Professor of Electrical Engineering, Computer Networks

With the growing demand for high-rate wireless services and pervasive wireless embedded devices, it has become essential to make efficient use of the available wireless spectrum resources. Experimental studies suggest that on average, at any given place/time, only about 5% of the spectrum under 3GHz is being utilized today.

One solution approach that seems promising is a hierarchical overlay system, in which there are primary and secondary wireless users in some given band of spectrum. The primary users essentially “own” the spectrum and are free to use it as they desire, while the secondary users are permitted to use the spectrum in an opportunistic manner so long as they do not cause too much interference to the primary users. One indication of the popularity of this approach is its use in the DARPA XG project, and the IEEE 802.22 draft standard which aims to allow unlicensed radios to operate in TV bands.

In the past year, I’ve been exploring some interesting theoretical questions in this context in collaboration with colleagues at other institutions. The basic model is simple. There are  $N$  channels. Time is slotted. The occupancy behavior of the primary users on each of the channels over time is described by independent and identical 2-state discrete-time Markov chains. A single secondary user picks one of the channels at the beginning of each time slot and listens on it for a bit to see if it is occupied. If not, the secondary user is free to transmit a packet. The optimization problem, which can be formulated as a partially observable Markov decision process (POMDP), is to make channel switching decisions at each step based on past observations and knowledge of the Markov chain, in order to maximize the expected number of packets transmitted over a finite horizon. This basic formulation is originally due to Qing Zhao (UC Davis).

Qing Zhao and I explored together one possible solution - a myopic policy whereby at each time step the secondary user probes the channel that is most likely to be free. One fascinating result we proved is that for this problem, the myopic policy can be implemented in a uniquely simple semi-universal way that doesn’t require the user to know the exact statistics of the primary user behavior, only whether the primary users’ presence on each channel over consecutive time steps is positively correlated ( $P_{11} \geq P_{01}$ ) or negatively correlated ( $P_{01} < P_{11}$ ).

We conjectured based on simulations that the myopic policy is actually the optimal policy (if true, this would actually be quite a surprising result as in general POMDP’s are PSPACE hard and rarely admit such a simple solution). We were able to prove this result for  $N=2$  channels fairly easily.

Then it got interesting. We struggled to generalize it further. Several proof techniques work for  $N=2$  but collapse with 3 channels as the problem changes structurally. Months of joint effort with Tara Javidi (UCSD), Qing Zhao, and Mingyan Liu (U. Michigan) yielded the result that the optimality of myopic holds for all  $N$ , but only for positively correlated primary user behavior under somewhat restrictive additional conditions. Finally, with even more effort, we got an even stronger result (due primarily to Mingyan Liu and her student, Sahand Ahmad) - myopic is optimal for all  $N$  so long as the primary users’ behavior is positively correlated over time.

But wait: it gets even more interesting. For the case of  $N=3$  we also have that myopic strategy is optimal even for negatively correlated primary user behavior. However, for  $N=4$ , there is a counter-example! So where it stands now is that the general conjecture is not true after all, in the case of  $N > 3$  and negatively correlated primary user behavior. The extremely simple myopic solution is not always optimal for this problem. But it is for a large range of cases, and even when it is not, appears to be near-optimal. It is an open problem to determine tighter conditions when optimality holds for general  $N$  in the case of positively correlated primary user behavior, and bounds on the distance from optimal when it does deviate.

We have additional results that cover sensing error (when the secondary user’s detection of the primary user’s presence can be erroneous). The case of multiple secondary users appears even more challenging. Some preliminary results that my student Hua Liu, Qing Zhao and I have regarding multiple contending secondary users show that it can be challenging to use coordination and learning, but that they are essential for performance. It appears that ideas from distributed Artificial Intelligence and Game Theory will be essential in addressing these challenges.

In the long run, there is much ground to be covered in relaxing restrictive assumptions in the analyses for both single and multiple users so that we can get not only more general theoretical results, but also practically useful algorithms. I am also hoping to explore the experimental side of this field in the future with the help of the inexpensive software radio platforms that have started to become available.