Our research focuses on how to design multicore architectures in the presence of severe physical constraints, especially thermal, power delivery, process variations, and wear-out. We are chiefly focusing on these issues in the context of symmetric and heterogeneous designs, which provide the best balance between high single-thread performance and high throughput for parallel tasks. Support for asymmetry is also becoming essential as process variations create performance and power asymmetry even in organizations that were originally designed to be symmetric. We are one of the first groups to explore the use of graphics processors (GPUs) for general-purpose computing and the first to develop an architectural simulation infrastructure-Qsilver-for performance, power, and thermal studies. Additionally, we are exploring how the massive parallelism of the GPU and its novel SIMD and memory organization can most effectively be used. Most of this work is done in collaboration with Mircea Stan, John Lach, and Ben Calhoun.
**CPU/GPU Convergence**
Computer systems are increasingly exposing a heterogeneous programming model consisting of conventional CPUs and various accelerators, such as GPUs and reconfigurable hardware (FPGAs). This trend is driven both by the speedups these platforms afford as well as energy-efficiency and cooling considerations that limit scalability. However, suitable programming models, effective compilation, decisions on which tasks to run where, and associated hardware design decisions for heterogeneous systems are poorly understood. We are studying all these issues. We have also developed a suite of applications that support research and development in heterogeneous systems.

**Architectures for Managing Power, Temperature, and Reliability**
High power densities and their effects in silicon semiconductors today already pose significant challenges to circuit and system designers. As semiconductor manufacturing processes scale to smaller transistor sizes, circuit operating temperatures will continue to rise. Thermal management and improved on-chip power delivery have become vital to continued innovation in IT, energy efficiency, and economic health. Since its initial development and release, HotSpot, the temperature-aware design tool developed through our work has played a critical role in assisting researchers worldwide with the design and validation of thermal management techniques. Our work on HotSpot is expanding to provide significant extensions that will support the next generation of thermal modeling and other related research.

**Architectural Modeling and Simulation Methodology**
We are working to develop new capabilities for accelerating simulation in order to achieve quantum leaps in our understanding of complexity in complex systems (such as the heart and other organs) by automatically partitioning the work among the heterogeneous processing units that are now becoming commonplace in commodity computer systems, in particular GPUs and FPGAs. We use a combination of CPUs, GPUs, and FPGAs because they each target different bottlenecks.

**RECENT RESEARCH DEVELOPMENTS**
- A new extension to our HotSpot framework, VoltSpot, allows modeling of a chip's power distribution network. Our preliminary analysis suggests that power stability is just as much a threat to future technology scaling as total power and thermal limits.
- In collaboration with Brett Meyer at McGill, we have developed a novel architecture for safety-critical systems, in which cores can switch in and out of redundant execution modes in order to maximize throughput for non-critical tasks.

**RECENT GRANTS**
- NSF-Accelerating Simulations Using CPU+FPGA Heterogeneous Processing
- SRC and NSF-Heterogeneous Multi-core Architecture from Homogeneous Arrays using Configurable Interconnect

**SEAS Research Information**
Pamela M. Norris, Associate Dean
University of Virginia
Box 400242
Charlottesville, VA 22903
pamela@virginia.edu
434.243.7683