# NSF-supported Research on Sustainable Computing at *PittCS*

Panos K. Chrysanthis Kirk Pruhs
Department of Computer Science -- University of Pittsburgh
http://energy.cs.pitt.edu

"*What matters most to the computer designers at Google is not speed, but power, low power, because data centers can consume as much electricity as a city*." --- Eric Schmidt, CEO Google

As the above quote demonstrates, energy and power have recently become first order considerations for the IT industry. This is explained by several converging trends. The first trend is that the total energy used by computing technology, and the energy density within the technology, has been increasing exponentially for decades. This can be viewed as a corollary of Moore's law, which states that the number of transistors that can be placed inexpensively on an integrated circuit is doubling every two years. The second trend is that the number of computing devices, particularly mobile computing devices, has been also increasing exponentially for a similar time frame. As a result the total energy used by the IT industry is currently a few percent of the total energy usage in the United States, roughly comparable to the energy used by the airline industry. The final trend is that the effectiveness of battery technology has only been increasing linearly.

This rapid emergence of power and energy as a first order resource has spurred academic and industrial researchers to develop more sustainable computing technologies at all levels: *Hardware, Power Management Systems Software* and *Power-Aware Applications*. Thanks in part to the funding from *the National Science Foundation (NSF),* the Department of Computer Science at the University of Pittsburgh (PittCS) has been at the forefront of sustainable computing research. This funding supported 12 PhD graduates (9 men and 3 women), and several masters/undergraduate students. We now summarize sustainable computing research in general and NSF sponsored research within PittCS in particular.

| Summary of Recent NSF Support Related to Sustainable Computing within PittCS | | |
|---|---|---|
| IIS-9502091 | CAREER: Developing Pragmatic Mobile Data Management Systems | Chrysanthis |
| CNS-0123705 | Middleware Support for Multicast Data Dissemination | Chrysanthis, Pruhs |
| CNS-0325353 | ITR: Secure CITI: A Secure Critical Information Technology Infrastructure for Disaster Management | Mossé, Comfort, Melhem |
| CCF-0448196 | Algorithmic Support for Temperature Aware Computing & Networking | Pruhs |
| CCF-0514058 | Algorithmic Support for Power Aware Computing and Communication | Pruhs |
| IIS-0534531 | Algorithms and Metrics for New Generation Data Stream Management Systems | Chrysanthis, Labrinidis, Pruhs |
| IIS-0746696 | CAREER: User-Centric Data Management | Labrinidis |
| CCF-0830558 | Algorithmic Support for Power Management | Pruhs |
| CCF-0811295 | Tera-PCM: A Low Power Terabyte Main Memory using Phase Change Memory | Childers, Mossé, Melhem |
| CNS-0936386 | Science of Power Management | Pruhs |
| CNS-0952273 | CA-RAM: Enabling Fast and Versatile Packet Processing for Future Large-Scale Networks | Cho |

**Hardware:** Hardware is the lowest level of the IT technology stack, and the starting point for the technology revolution. Out of sheer necessity, some of these technology changes have already been adopted by industry. For example, approximately five years ago, chip makers such as Intel and AMD, hit a "thermal wall" caused it to radically redesign its chips to contain multiple lower power processors rather than one high power processor because they could no longer economically cool the chips. Many experts predict that the number of processors per machine will grow exponentially for the next decade or two. Other hardware technology revolutions are widely anticipated in the near future. For example, as industry anticipates machines with petabytes of memory, it seems necessary to develop more energy efficient memory technology.

**NSF-supported Phase Change Memory Research within PittCS**: New memory designs can save power by utilizing novel memory technology. While current memory technology can be used to build large-scale memories, this technology is impractical because it consumes vast amounts of power. Profs. Childers, Mossé and Melhem are working on new designs constructed with phase-change memory. Phase change memory (PCM) is a new technology that has good scalability, exceptionally low power consumption, and low susceptibility to faults. Despite these attractive characteristics, there are several challenges that must be overcome to use PCM as main memory. Two important limitations in comparison between current memory technology and PCM are relatively slow performance and poor write endurance. To help alleviate these limitations they designed a new memory architecture that employs several innovative architectural mechanisms and strategies to manage PCM read/write latency and to increase PCM longevity. Profs. Cho and Melhem also explore how PCM can be used to significantly lower the power consumption of the routers that form the backbone of the Internet.

**Power Management Software:** New hardware creates new management problems for systems software. For example, new processor designs can save power by scaling the speed of the processor. As dynamic power is roughly the cube of the speed, even small reductions in speed can save significant energy. For example the Intel i7 processor can be run at a top speed of 1.6 Gigahertz at a power of 25 Watts down to a speed of 600 Megahertz at a power of 6 Watts. There is a need to create power management systems software to most effectively use these hardware mechanisms. Some examples of commercial power management systems include AMD's PowerNow and Cool'n'Quiet technologies, and Intel's SpeedStep technology.

**NSF-supported Speed Scaling Research within PittCS:** Speed scaling, as a power management technique, has to date mostly been applied in a reactive fashion, largely decoupled from the scheduling policies. For example, if the system detects that a processor is in danger of exceeding its thermal threshold, then the system will react by throttling the processor's speed. It has been clear for some time that integrated proactive power management and scheduling policies will result in better power management than simple reactive strategies. For example, rather than reacting to a thermal emergency, a proactive integrated policy would ideally avoid a thermal emergency by intelligently scheduling the tasks and scaling the speed. Proactive power management is complicated by the fact that the dual objectives of power and quality of service are conflicting, that is, the more power that one uses, the better quality of service that can be provided. The algorithmic solutions to proactive power management problems involve increasing power when the improvement in the scheduling objective justifies the increased cost in the power management objective. Prof. Pruhs has developed proactive power management policies that are in some sense provably optimal, and has been developing the new mathematical techniques required to reason about energy, power and temperature as resources.

**Energy-Aware Systems and Applications**: To achieve maximum energy savings, systems and applications can be made power-aware so that they can collaborate with the power management system in setting the appropriate "energy use policy" or can adapt to the power state at any given point in time. For example, intelligent data processing can increase the idle time of resources such as the disk or the network card, hence enabling their powering down, which would result in significant energy savings.

**NSF-supported Mobile Data Management Reseach within PittCS**: Handheld computers and smart phones are pervasive, with about one active cell phone for every two people on the planet. Being battery-operated, minimizing the power consumption on these mobile devices has been a fundamental challenge. Prof. Chrysanthis has developed power-aware data management algorithms that intelligently use all the scarce resources of mobile devices. The basic principle of these algorithms is the support of disconnected operations that avoid expensive communication and selectively powering down the communication subsystem to save battery life and communication dollars. He has also developed energy-efficient wireless data dissemination schemes for query results and web documents that work in synergy with caching on mobile devices to achieve a better balance between response time and energy savings at the mobile devices.

**NSF-Supported Data Management for Cyber-Physical Systems within PittCS**: Sensor networks offer today an unprecedented level of interaction with the physical world and, as such, are key elements in bridging the physical world of the planet and the digital world of computers. Similar to mobile devices, a fundamental challenge in battery-operated wireless sensors is to reduce their energy consumption. Profs. Chrysanthis, Melhem, Mossé and Labrinidis have produced a wide range of data management algorithms and communication protocols (primarily funded through the S-CITI NSF ITR project), focusing in quality of data and quality of service, in the presence of resource constraints, such as those characterizing emergency response environments. For example, at the point of generation of data (e.g., from networks of sensors or mobile devices), they proposed energy-efficient data acquisition techniques, which combine in-network processing and in-network, data-centric storage in order to minimize communication and save energy, prolonging the life of the sensor network to fully cover the mission's critical time requirements. To effectively propagate data within such sensor networks, they also proposed semantic-based, multi-criteria, self-optimizing algorithms for constructing energy-efficient and robust routing topologies.

**Science of Power Management:** Recently there has been a realization that it is important to establish an underlying science of power management to support information technology engineering the way that the science of chemistry supports the engineering of pharmaceutical engineering or the science of statics supports civil engineering. Such a science would allow engineers to reason abstractly about energy, power and temperature when designing technology. Traditionally time and space have been the first order resources, but energy has very different physical characteristics; for example, while there are different physical limits on the time and space required for various computations, there are no physical limits on the minimum energy required for any computation. With NSF Support, Prof. Pruhs organized a workshop attended by a wide range of sustainable computing experts to construct a roadmap to the establishment of the science of power management.