Information Science and Technology Seminar Series

Uzi Vishkin
University of Maryland Institute for Advanced Computer Studies (UMIACS)

"Current Mainstream Computing: If in a hole, should we keep digging?"

Wednesday, January 11, 2012
3:00 - 4:00 PM
TA-3, Bldg. 1690, Room 102 (CNLS Conference Room)

Abstract: Following the considerable slowdown in the improvement of single processor cores since 2004, all mainstream machines are now parallel. The central problem for the field is how to build them for cost-effective programming and good speedups. Alas, only heroic programmers can exploit the vast parallelism in today's mainstream (e.g., PCs) computers. Rejection of their parallel programming by most programmers is all but certain. The tendency to import old parallel computing ideas does not work, as parallel computing has been long plagued with programming difficulties; an example of a popular introduction to parallel programming that is unnecessarily traumatic will be discussed. But, we can draw a lesson from what got parallel computing started on the wrong foot. The 'build-first figure-out-how-to-program-later' approach to the first parallel machines was followed by fitting parallel languages to these arbitrary architectures and standardization of these language fits doomed later parallel architectures.

The starting point for our holistic explicit multi-threaded (XMT) on-chip platform had been the Parallel PRAM algorithmic theory, which is second in magnitude only to the serial algorithmic theory, won the "battle of ideas" in the 1980s, and has been challenged repeatedly without success. Taking advantage of the increasing chip real estate, the XMT PRAM-like programming incorporates programming for locality and reduced synchrony as 2nd order considerations. An on-chip interconnection network provides high bandwidth, and along with the memory architecture, low latencies for regular and irregular memory access. A patented stored-program and program-counters apparatus spontaneously starts as many threads as processors; using a hardware scheduler based on a prefix-sum primitive it also handles effectively very short threads. The XMT architecture scales up to at least 1024 processors on chip, and down, providing backwards compatibility on serial code. Implementations include hardware (a 64-processor machine), and downloadable compiler and simulator. Order-of-magnitude improvements over vendors' many-cores have been demonstrated for ease-of-programming and speedups. The former includes teaching more advanced parallel algorithms at earlier developmental stages (e.g., high-school vs. graduate school), report by an NSA employee of minimal effort for XMT programming beyond a serial version for several parallel graph algorithms, and a joint UIUC/UMD course in which no student was able to get speedups over serial on an OpenMP/SMP, while their speedups on XMT were in the range 7X to 25X. The latter include speedups over serial by order of magnitude beyond state-of-the-art CPUs and GPUs for simple irregular fine-grained problems, and up to 43X beyond such CPUs and GPUs for more advanced ones; these results account for silicon area and power.

Biography: Uzi Vishkin has been Professor at the University of Maryland Institute for Advanced Computer Studies (UMIACS) since 1988. Per his ACM Fellow citation, he "played a leading role in forming and shaping what thinking in parallel has come to mean in the fundamental theory of Computer Science". The presentation framework in several parallel PRAM algorithms textbooks, which also include quite a few parallel algorithms he co-authored, is the 1982 Shiloach-Vishkin work-depth methodology. His research team's recent work on his explicit multi-threaded (XMT) many-core 'PRAM-On-Chip' architecture refuted the common wisdom that PRAM algorithms are irrelevant for practice. He is an ISI-Thompson Highly Cited Researcher, was named a Maryland Innovator of the Year for his PRAM-On-Chip venture whose main patent was cited in quite a few patents of major vendors, and his proposal for Reinvention of Computing for Parallelism was ranked first among 49 proposals in a University System of Maryland competition for Maryland Research Centers of Excellence.