Scaling Usable Computing Capability

Tor M. Aamodt

University of British Columbia, Canada

Abstract. Programmers tackling the complexity of large software systems have long favored computing system flexibility over peak computing capability. The growing popularity of scripting languages suggests this tendency is increasing. On the other hand, current trends in semiconductor scaling are raising concerns over "dark silicon" and suggest hardware accelerators are necessary to improve raw computing capability. Accelerators tend to give up flexibility to obtain efficiency and performance. This talk argues a consequence of these trends is that there is an urgent need to develop accelerators that are both energy efficient and easy to program. Such devices are required if usable computing capability is to improve in the face of expected challenges in semiconductor scaling. While efficiency is easier to measure than programmability this talk will argue both can be optimized simultaneously. Examples of how this can be achieved on GPU computing architectures will be described including locality-aware thread scheduling and efficient GPU transactional memory.

Biography

Tor Aamodt is an Associate Professor in the ECE Department at the University of British Columbia. His recent research has focused primarily on GPU computing architectures. Several of his papers have been selected as "Top Picks" by IEEE Micro Magazine. He is in the ACM/IEEE MICRO Conference "Hall of Fame" and served as Program Chair and General Chair for ISPASS 2013 and 2014, respectively. He received his BASc, MASc and PhD from the University of Toronto. Before UBC, he worked at NVIDIA on the memory system of the first GPU supporting CUDA. More recently he was a visiting faculty at Stanford.

978-1-4799-3770-7/14/$31.00