Abstract – Because it no longer is possible to improve computing capability by simply increasing clock frequencies, we have spent the better part of a decade in a new parallel computing era. Recently, as energy efficiency and power consumption have become increasingly important for modern parallel architecture designers, hardware resources for parallelism are shifting from general-purpose, multi-core designs to throughput-oriented computing with graphics processing units (GPUs), accelerators, and increasingly wide single instruction multiple data (SIMD) extensions on commodity processors that provide efficient, vector-based parallel computation. Compared to other hardware, SIMD extensions require less extra hardware, and SIMD instruction execution is essentially “free” from a power perspective, making vectorization an attractive option.

However, there are many obstacles to leveraging SIMD extensions. First, many algorithms exhibit concurrency in the form of divide-and-conquer, recursive “task parallelism”. Without enough data parallelism, it seems these algorithms are not well suited to SIMD extensions. Second, even with “obvious” data parallelism, many applications, particularly ones traversing irregular data structures, still cannot be mapped onto SIMD extensions straightforwardly because of the mismatch between the strict, lockstep behavior of SIMD parallelism and the dynamic, data-driven behavior of the programs that manipulate irregular data structures. This talk will introduce my research efforts addressing these challenges, including a novel transformation framework to expose data parallelism for task-parallel algorithms and a novel non-traditional solution consisting of an intermediate language and a runtime scheduler to efficiently vectorize applications traversing irregular data structures. In addition, this talk will cover other recent progress and exciting opportunities in using compiler techniques to leverage modern parallel architectures.

Bio - Bin Ren currently is a post-doctoral research associate in the High Performance Computing group at Pacific Northwest National Laboratory. He received his Ph.D. from the Department of Computer Science and Engineering at The Ohio State University. His primary research interest involves software systems, specifically programming systems and compiler support for parallel computing. His research has encompassed parallel architectures and hardware, static and dynamic compiler analysis, high-level parallel programming models, and various applications. He has been part of close collaborations with Microsoft Research, NEC Laboratories, Cray Inc., Purdue University, Washington University in St. Louis, and Washington State University. Results from his research have been published in leading computer systems and parallel programming venues, including PLDI, CGO, PACT, TACO, and ICS. His CGO’13 paper earned a Best Paper award, was featured as a SIGPLAN research highlight, and nominated as a CACM research highlight. Ren earned his bachelor’s and master’s degrees from Beihang University (China) in 2006 and 2008, respectively.