Career: Data Locality Enhancement of Dynamic Simulations for Exascale Computing

Transformations

data reorder.   hybrid   job swap.

(data reloc.   ref. redirect.)

guidance for transformations

overhead hiding & minimization

Optimality & approximation

Complexity analysis

Approximating optimal layouts and mappings

Efficiency control

Adaptive CPU-GPU pipelining

Two-level efficiency-driven adaptation
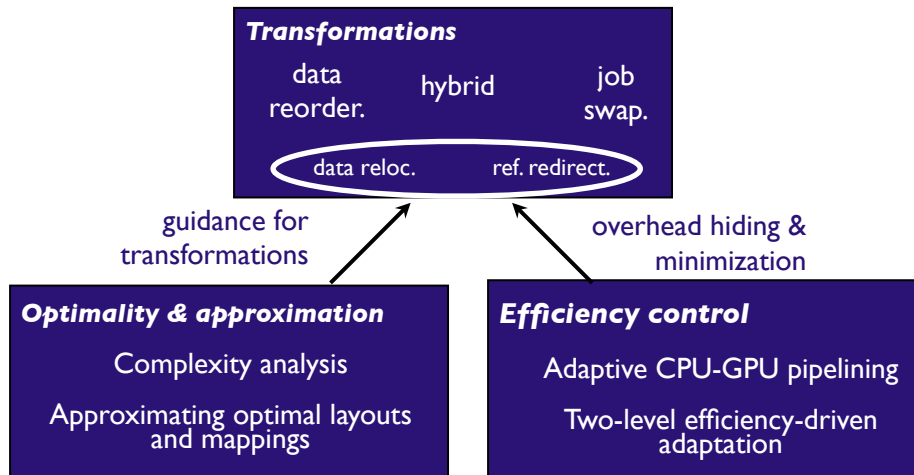
Novel Ideas

- Asynchronous streamlining
  - a) Adaptive CPU-Accelerator pipelining
  - b) Eliminating thread divergences on-the-fly
  - c) Runtime removal of irregular mem. accesses
- Neighborhood-aware locality optimizations
  - a) Exploiting non-uniformity of comp. elements
  - b) Resource-sharing-aware data transformation
  - c) Cross-thread locality analysis
  - d) Neighborhood-aware thread/job scheduling

Impact and Champions

- The research will provide a set of techniques and tools for scientific users to enhance program locality on multi- and many-core systems that is not possible to achieve with existing tools, helping alleviate the memory bottleneck in exascale computing.

- It will enhance the understanding in the relations between dynamic simulations and exascale computing systems, especially in the aspects of memory performance and concurrency exploitation.

- It will help advance computational sciences and promote academic research and education in the challenging field of scientific computing

PI: Xipeng Shen, The College of William & Mary

Milestones/Dates/Status

	Scheduled	Actual
streamlining alg. design	FEB 2012	FEB 2012
concurrent reuse distance	AUG 2012	AUG 2012
interplay of transformations	FEB 2013	FEB 2013
co-run perf. modeling	AUG 2013	AUG 2013
library develop.	FEB 2014	
co-sch. alg. design	AUG 2014	
two-level adaptation	FEB 2015	
other scheduling criteria	AUG 2015	
System integration	LOE through 2016	