IA^3 2012 – SC12 Workshop on Irregular Applications: Architectures and Algorithms

Final Program

9:00 – 9:05 Welcome and Introduction

9:05 – 10:05 Session 1: Architectures for Irregular algorithms

CHOMP: A Framework and Instruction Set for Latency Tolerant, Massively Multithreaded Processors
John Leidel, Kevin Wadleigh, Joe Bolding, Tony Brewer and Dean Walker.

Abstract: Given the recent advent of the multicore era [1], we find that parallel application performance is no longer solely gated by an architecture's core arithmetic unit performance. Memory bandwidth has failed to grow at the same rate as effective core density. This paper presents a framework for constructing tightly coupled, chip-multithreading [CMT] processors that contain specific features well-suited to hiding latency to main memory and executing highly concurrent applications. This framework, deemed the "Convey Hybrid OpenMP" or CHOMP architecture, is built around a RISC instruction set that permits the hardware and software runtime mechanisms to participate in efficient scheduling of concurrent application workloads regardless of the distribution and type of instructions utilized. In this manner, all instructions in CHOMP have the ability to participate in the concurrency algorithms present in the hardware scheduler that drive context switch events. This, coupled with a set of hardware supported extended memory semantic instructions, means that the CHOMP architecture is well suited to executing applications that access memory using non-unit stride or irregular access patterns. Furthermore, the CHOMP architecture and framework contains specific logic and instruction set support that allows application-level, dynamic power gating of individual register files and function pipes.

Exploiting Coarse-Grained Parallelism in B+ Tree Searches on an APU
Mayank Daga and Mark Nutter.

Abstract: B+ Tree structured index searches are one of the most fundamental database operations and hence, accelerating them is exigent. GPUs provide a compelling mix of performance-per-watt and performance- per-dollar and thus, are an attractive platform for accelerating B+ Tree searches. However, tree search on discrete GPUs presents significant challenges for acceleration due to (i) the irregular representation in memory, and (ii) the requirement to transfer the tree to the GPU over the PCIe bus. In this paper, we present the acceleration of B+ Tree searches on a fused CPU+GPU processor (or APU). We counter the aforementioned issues by reorganizing the B+ Tree in memory and utilizing the novel heterogeneous system architecture which eliminates - (i) the need to transfer the tree to the GPU, and (ii) the limitation on the size of the tree than can be accelerated. Our approach exploits the coarse-grained parallelism in tree search wherein we execute multiple searches in parallel to optimize for the SIMD width, without modifying the inherent B+ Tree data structure. Our results illustrate that the APU implementation can perform up to 70M1 queries per second and is 4.9-fold (best case) and 2.5-fold (on average) faster than the hand-tuned SSE optimized, multi-threaded CPU implementation (6-cores), for varying orders of the B+ Tree with 4M keys. We also present an analysis on the effect of caches on performance, and on the efficacy of the APU to eliminate data-copies.

10:05 – 10:30 Coffee break

10:30 – 11:30 Session 2: Using GPUs for Irregular Applications

Breadth first search on APEnet
Mauro Bisson, Massimo Bernaschi, Enrico Mastrostefano and Davide Rossetti.

Abstract: We present preliminary results of a multi-GPU code for exploring large graphs (hundreds of millions vertices and billions of edges) by using the Breadth First Search algorithm. The GPU are connected by APEnet, a custom technology that allows for direct data exchange among GPUs.

An Irregular Approach to Large-Scale Computed Tomography on Multiple Graphics Processors Improves Voxel Processing Throughput
Edward Jimenez, Laurel Orr and Kyle Thompson.

Abstract: While much work has been done on applying GPU technology to computed tomography (CT) reconstruction algorithms, many of these implementations focus on smaller datasets that are better suited for medical applications. This paper proposes an irregular approach to the algorithm design which utilizes the GPU hardware's unique cache structure and employs small x-ray image data prefetches on the host to upload to the GPUs while the devices are operating on large contiguous subvolumes of the reconstruction.

This approach will improve the overall cache hit-rates and thus improve the performance of the massively multi-threaded environment of the GPU. Overall, utilizing small prefetches of x-ray image data improved the volumetric pixel (voxel) processing rate when compared to utilizing large data prefetches which would minimize data transfers and kernel launches. Additionally, this approach does not sacrifice performance on small datasets and is thus suitable for medical and industrial applications. This work utilizes the CUDA programming environment and Nvidia's Tesla GPUs.

11:30 – 12:30 Session 3: Programming models for Irregular Applications

Executing Optimized Irregular Applications Using Task Graphs Within Existing Parallel Models
Christopher Krieger and Michelle Strout.

Abstract: Many sparse or irregular scientific computations are memory bound and benefit from locality improving optimizations such as blocking or tiling. These optimizations result in asynchronous parallelism that can be represented by arbitrary task graphs. Unfortunately, most parallel popular programming models with the exception of Threading Building Blocks (TBB) do not directly execute arbitrary task graphs.
In this paper, we compare the programming and execution of arbitrary task graphs qualitatively and quantitatively in TBB, the OpenMP doall model, the OpenMP 3.0 task model, and Cilk Plus. We present performance and scalability results for 8 and 40 core shared memory systems on a sparse matrix iterative solver and a molecular dynamics benchmark.

Logic Programming for Parallel Irregular Applications
Jeremiah Willcock and Andrew Lumsdaine

Abstract: Parallel computers are becoming increasing heterogeneous and correspondingly more

difficult to program as a result. Irregular applications exacerbate this problem further, given that this class of applications is more diverse and uses different performance features of a computer system than more common application classes. Thus, new approaches are necessary to achieve performance and productivity simultaneously. Domain-specific languages are becoming increasingly popular for high-performance computing, both in the domains of regular (e.g., SPIRAL) and irregular (e.g., Green-Marl) applications. However, past languages for graph computations tend to be too limited to efficiently express the wide range of irregular algorithms needed in applications. Instead, this position paper advocates adapting a language from the database community, Datalog, to the domain of high performance irregular applications. Although the plain Datalog language is also insufficient for the class of applications targeted, extensions can be added to increase its expressiveness. Starting with a standard language also enables taking advantage of the literature on Datalog implementations, including in the contexts of parallelism and incremental execution of algorithms. Thus, this approach promises to be a good way to implement irregular applications with both productivity and performance.

2:15 – 3:00 Keynote 1

Big Data - Ops vs. Flops
Steve Wallach, Convey Computer

Abstract: Today's HPC systems are primarily designed for floating Point Calculations on 3D data structures. Big Data applications operate on: Graphs, Bytes strings, adjacency matrices, etc. We will examine the hardware micro-architectures and software paradigms of both of these environments for similarities and differences.

3:00 - 3:30 Coffee break

3:30 – 4:15 Keynote 2

Irregular Applications and their Architectural Challenges
Pradeep Dubey, Intel

4:15– 5:30 Panel

Next generation systems for irregular application: directions, challenges and opportunities.

Panelists:
Umit Catalyurek, Ohio State University
Pradeep Dubey, Intel
David Gleich, Purdue University
Alex Ramirez, Barcelona Supercomputing Center
Vinod Tipparaju, AMD
Steve Wallach, Convey Computer