Introdction

Background: HMC

Describe HMC and the Organizing Committee

Technical Areas

Technical areas provide the basic organizing structure for HMC and for this roadmap. The consortium and roadmap are organized in four technical areas: Applications and Libraries (AL), Programming Models (PM), Architecture and Metrics (AM), and Performance and Analysis (PA).

Applications and Libraries

The focus of the Applications and Libraries area is to identify the processes and libraries needed to make existing applications run effectively on large-scale accelerator based systems. The roadmap embraces all science and engineering applications that run at scale on current (non accelerator based) platforms, including the DOE INCITE applications, NNSA applications, and applications that have earned allocations on the largest NSF platforms. To avoid becoming mired in the plethora of applications, the Applications and Libraries area will define "Application Communities" that focus on issues common to a class of applications. The organizing principles for these communities might emphasize algorithmic strategies (CFD, coupled physics, UQ, etc) or some combination of the Berkeley Motifs. These communities will make process recommendations, e.g. whether it is better to accelerate small pieces of the application or to re-factor the entire application. They will also identify critical libraries and provide input to the Programming Models area to support productivity and platform independence. The long-term goal is to develop a strategy that supports the **co-design** of applications and architectures. The Applications and Libraries area will also develop a strategy, working with the Architecture and Metrics area, to address the need for application teams to have "early access" to new platforms.

Programming Models

The focus of the PM TC is to identify technologies able to improve programmer productivity and application performance portability. Productivity addresses: 1) support for data layout across intra- and inter-node memory hierarchies and 2) support for dynamic parallelism and the disjoint task hierarchies associated with the CPU and the accelerators. Performance portability is concerned with porting legacy codes and the need to ensure that the benefits of these efforts are not limited to the current platforms. There are several areas of interest: 1) code optimization techniques; 2) language/runtime interoperability (MPI, OpenMP, OpenCL); 3) automatic translation (e.g., OpenMP <-> OpenCL).

Architecture and Metrics

The focus of the AM TC is to identify and report on approaches for building large-scale accelerator-based hybrid computer systems in the near term and in the future. The time horizon of such a roadmap will span currently deployed or planned large-scale hybrid systems with node-attached accelerator devices (e.g. Cell, Tesla), the likely transition of

such systems to include tighter integration within cluster nodes, and future concepts which could include hybrid accelerator cores mixed with conventional general purpose cores on multi-chip carriers or on die. The technical committee will work to identify the types and degrees of parallelism provided by hybrid cores and to define key architectural metrics of this class of hybrid machine (e.g. connection latency and bandwidth, memory hierarchies, SIMD width, power efficiency) of importance to the targeted applications of the Consortium. They will work with the other TCs to support the overall co-design goal of the Consortium involving interplay of applications, architectures, programming, and performance and to build ties with and provide feedback to vendors. The committee will also develop strategies for early and broader access to these accelerator-based or future hybrid multicore systems.

Performance and Analysis

The focus of the PA TC is to apply techniques and tools for measurement, analysis, and modeling to performance engineer hybrid systems and applications, current and future. As the interface among the other areas of interest, the PA TC will work with the other TCs in order to characterize the performance of systems and applications, and guide the development of new hybrid and multicore architectures, as well as different mappings of applications on these architectures. The PTC will rely on state-of-the-art techniques, and will coordinate, in the short term, the development of tools for performance engineering of hybrid applications and systems. These tools aim to simplify application development by providing: insight for code optimization, performance observability in hardware and runtime systems, and automatic tuning.

Grading Criteria

Urgency **How soon is it needed?**	**Duration** **How long will it be useful?**	**Responsiveness** **How much will resources help?**	**Applicability** **How broadly can it be used?**	**Timeline** **How soon can we expect it?**
Critical needed now	Long Useful for the foreseeable future	High Resources enable significant progress	Broad Applicable beyond scientific computation	Immediate Results within 1-2 years
Important needs to be done within 3 years	Medium Useful for Exascale	Moderate Resouces enable progress	Science Applicable to scientific computing in general	Soon Results within 2-5 years
Useful needed after 3 years	Near Only useful for immediate systems	Low Resources have little affect on progress	Narrow Only applicable to HPC systems	Eventually Results after 5 years