XPRESS: eXascale Programming Environment and System Software

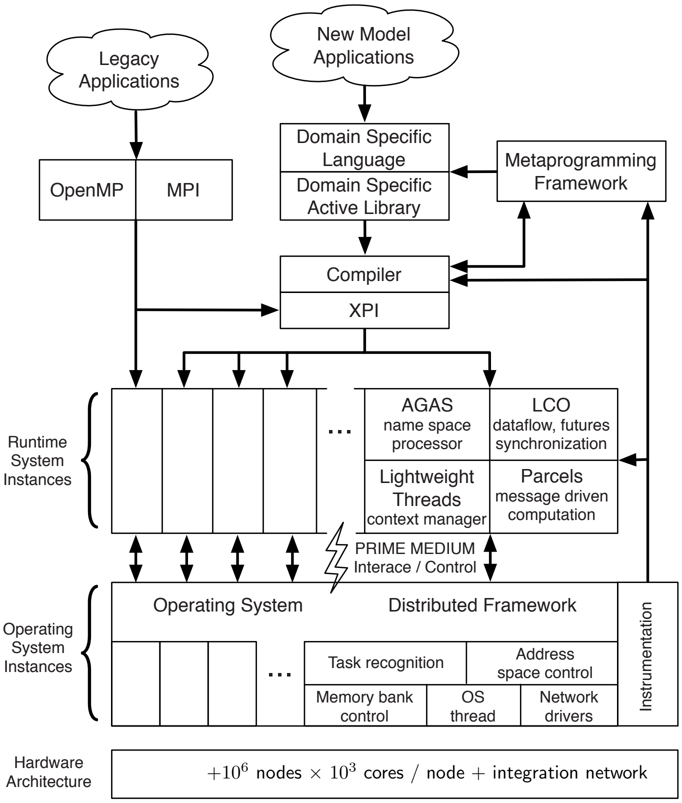
Goal

The XPRESS Project is one of four major projects of the DOE Office of Science ASCR X-stack Program initiated in September, 2012. The purpose of XPRESS is to devise an innovative system software stack to enable practical and useful exascale computing around the end of the decade with near-term contributions to efficient and scalable operation of trans-Petaflops performance systems in the next two to three years; both for DOE mission-critical applications. To this end, XPRESS directly addresses critical challenges in computing of efficiency, scalability, and programmability through introspective methods of dynamic adaptive resource management and task scheduling.

Objective

XPRESS will define a system-software architecture, OpenX, to represent the full functionality ultimately anticipated for an exascale computing system. While a paper specification, it will include important interfaces between the programming system and underlying runtime and OS called XPI and between the Runtime system and operating system called RIOS. Compliancy with these interface specifications will facilitate different X-stack configurations comprising components of different design, possibly by different development teams to accelerate progress towards advanced DOE systems. XPRESS will implement a critical subset of OpenX software modules to integrate and test key functionality to demonstrate and apply a working software system incorporating the defining innovative concepts upon which XPRESS is defined. These will include the LXK lightweight kernel OS and the HPX-4 runtime system interoperating through RIOS and driven by workloads via the XPI interface.

OpenX Software Architecture


Legacy Applications
New Model Applications
OpenMP
MPI
Domain Specific Language
Domain Specific Active Library
Metaprogramming Framework
Compiler
XPI
Runtime System Instances
...
AGAS name space processor
LCO dataflow, futures synchronization
Lightweight Threads context manager
Parcels message driven computation
PRIME MEDIUM Interace / Control
Operating System Instances
Operating System
Distributed Framework
Task recognition
Address space control
Memory bank control
OS thread
Network drivers
Instrumentation
Hardware Architecture
$+10^6$ nodes × 10^3 cores / node + integration network

Strategy

The technical strategy guiding the co-design of the constituents of the XPRESS software stack is conceived to directly address specific factors degrading performance efficiency and scalability according to the formula:

$$P = e * S * a * \eta$$

Where P is performance, e is efficiency, S is scaling, a is availability, and η is the single thread performance normalization factor. Four factors captured by the acronym, SLOW, are identified as principal contributors to performance degradation and are directly addressed by the XPRESS strategy. Starvation, or insufficiency of concurrent work, impacts both scalability and efficiency requiring more parallelism to be exposed and exploited. Latency effects need to be mitigated through a combination of locality management, reduction of messaging, and hiding. Overhead wastes time and energy but worse also limits the fineness of granularity that can be effectively exploited, reducing scalability. Waiting due to contention for shared physical or logical objects will further reduce efficiency. Another factor, resilience (which may make it SLOWR) impacts availability.

The XPRESS research is exploring means of dramatically improving efficiency and scalability through a set of guiding and interrelated operational and structural principles that together comprise an advanced execution model, ParalleX. ParalleX replaces the static Communicating Sequential Processes (CSP) execution model primarily familiar to all through MPI with the means of a fully dynamic and adaptive paradigm to exploit the capabilities of future generation performance oriented runtime systems and hardware architecture enhancements. The ParalleX model and manifesting runtime system software incorporate key semantic constructs and mechanisms for dynamic adaptive resource management, task scheduling, and parallelism discovery. It combines lightweight multi-threading with event-driven (sometimes message-driven) computation coordinated by powerful synchronization objects in the context of global address space. Together, these mutually supportive elements enable new implementation strategies to facilitate runtime techniques for dramatic performance improvement towards extreme scale computing.

Technical Approach

XPRESS is organized as a set of cooperative tasks to develop and test a software architecture based on the above concepts to deliver working scalable and efficient runtime environments for leading to exascale computing. These major tasks include:

- Performance models & metrics – provide parameters and their mutual sensitivities to guide co-design and quantify operational behavior
- ParalleX execution model – guiding principles for co-design of components of OpenX software stack
- OpenX software architecture – a conceptual framework for the co-design and interoperability of proof-of-concept XPRESS software stack including the RIOS interface protocol specification between the operating system and runtime system
- LXK operating system – Lightweight kernel operating system for order constant scalability and low/no noise to manage resources
- HPX runtime system – support of application dynamic adaptive resource management, task scheduling, and introspective control policies
- XPI advanced programming model – intermediate form and low-level (readable) programming interface reflecting the ParalleX model, providing a target for source-to-source high level parallel language translation, and supporting early direct programming experimentation and measurement
- Legacy application mitigation – ensuring seamless transition of legacy codes and programming methods to the future generation of ParalleX based exascale systems
- Experiments and evaluation – Critical to determining degree of effectiveness and likelihood of ultimate success as well as guiding corrective design changes to achieve DOE objectives
- Documentation – as well as reporting to DOE X-stack program management, to provide early adopters with sufficient information to apply prototype programming and execution environment

Team

The XPRESS Project is led by Sandia National Laboratories (SNL) and engages a team of 8 institutions including: Indiana University (IU), University of North Carolina (UNC/RENCI), Oregon University (OU), University of Houston (UH), Louisiana State University (LSU), Oak Ridge National Laboratory (ORNL), and Lawrence Berkeley National Laboratory (LBNL).

Sandia National Laboratories

Oak Ridge National Laboratory

LSU

Berkeley Lab

http://xpress.sandia.gov