ABHINAV BHATELE | TODD GAMBLIN | MARTIN SCHULZ | PEER-TIMO BREMER

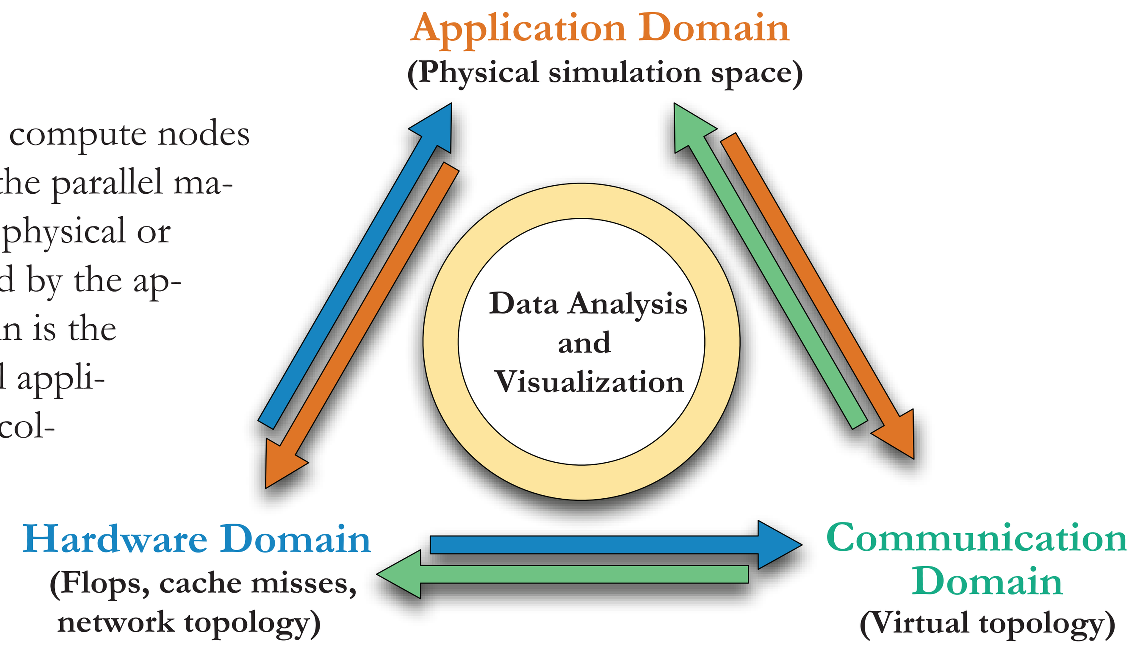
Collaborators: Brian T. N. Gunney, Katherine E. Isaacs, Aaditya G. Landge, Joshua A. Levine, Bernd Hamann and Valerio Pascucci

Intuitive Visualizations for Analyzing Exascale Workloads

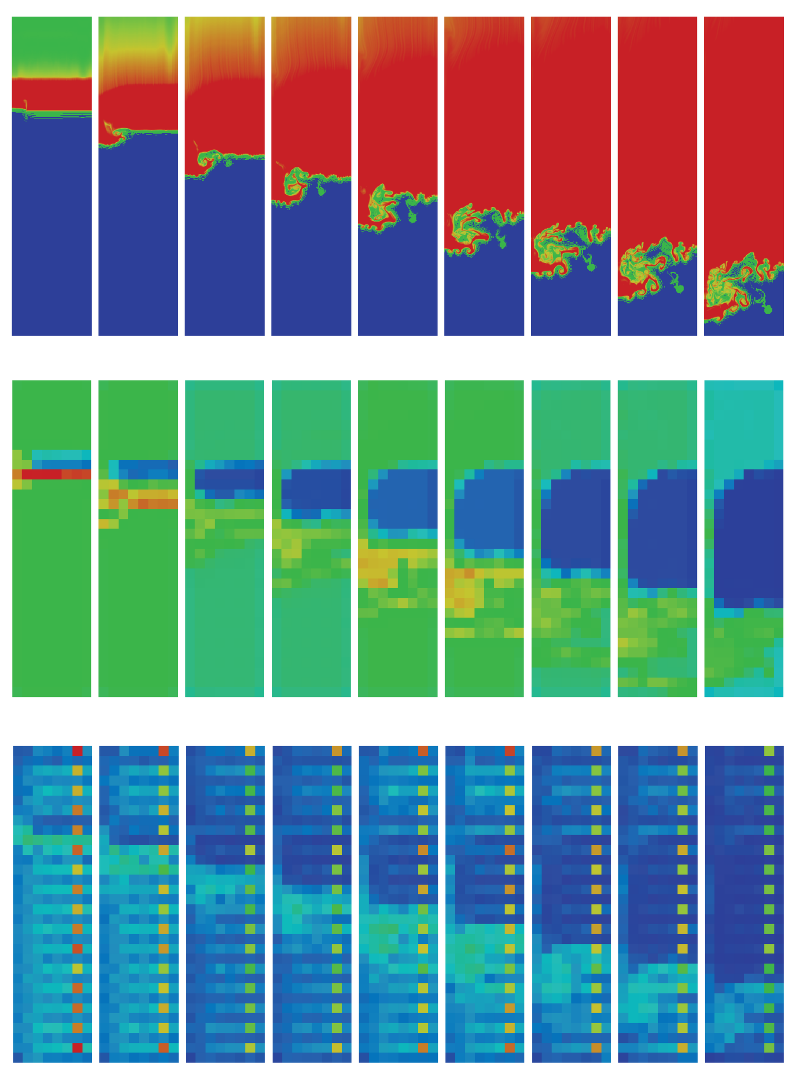
Performance analysis of parallel scientific codes is becoming increasingly difficult due to the rapidly growing complexity of applications and architectures. Existing tools fall short in providing intuitive views to reveal the root causes of performance problems. We have developed a new paradigm of projecting and visualizing performance data obtained from one domain onto other domains for faster, more intuitive analysis of applications. We gather performance data in three domains: hardware, application, and communication. For each domain, we define projections that allow the data to be viewed in the other domains. This framework is called the HAC model. The model not only allows us to directly compare the data across domains, but also allows us to use data visualization and analysis tools available in the other domains. Using these methodologies and visualization techniques, we demonstrate the careful unscrambling of otherwise tangled measurements caused by adaptive systems. By attributing performance anomalies directly to their causes, we are able to visualize performance measurements in the domains most intuitive to the user, which are not necessarily those in which the measurements are collected.

The HAC model

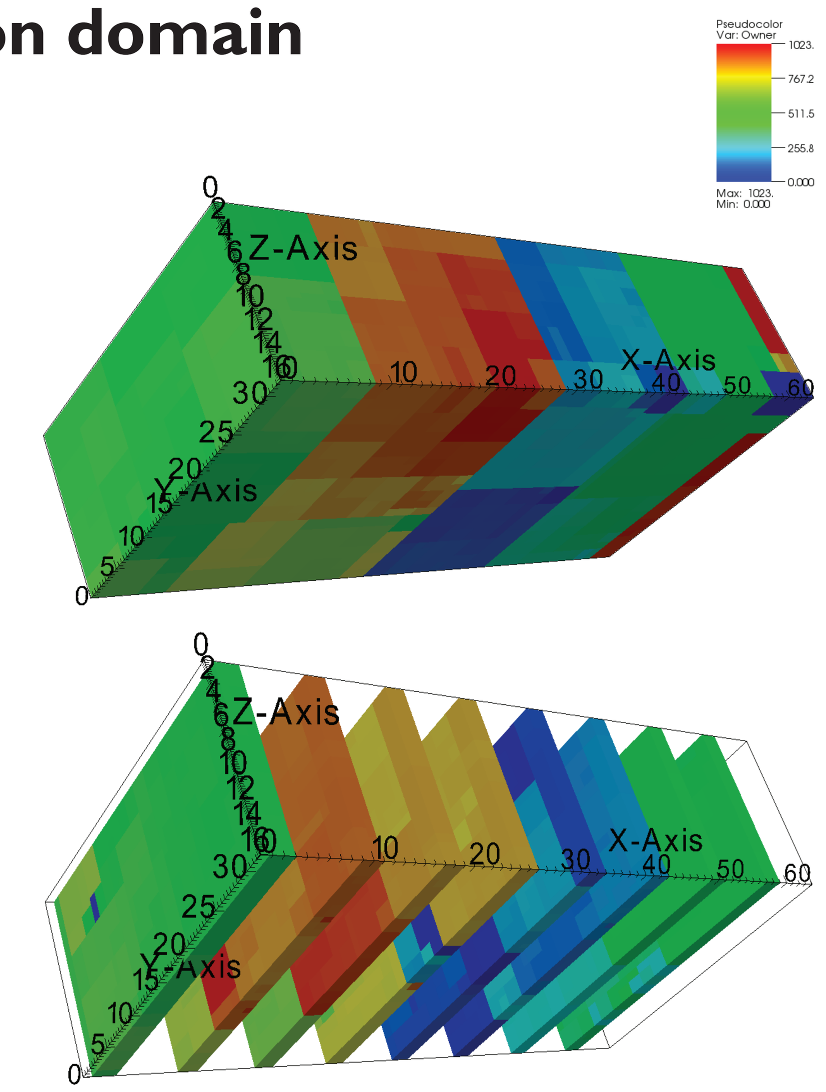
The hardware domain represents the compute nodes and physical interconnect that form the parallel machine. The application domain is the physical or other simulation space being modeled by the application. The communication domain is the virtual process topology for a parallel application. Performance data is typically collected in one of these domains. The HAC model projects data into domains where it is most intuitive to analyze.



Projections on the application domain



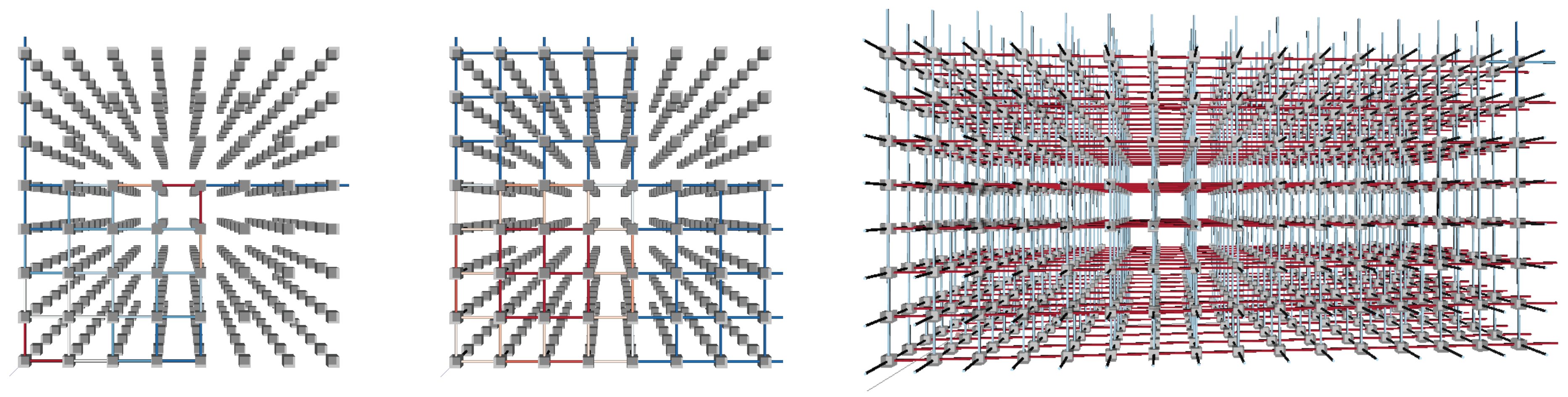
Floating point operations (middle) and L1 cache misses (bottom) mapped to the application domain (top) for multiple time steps of Miranda running on 256 cores of an Infiniband cluster.



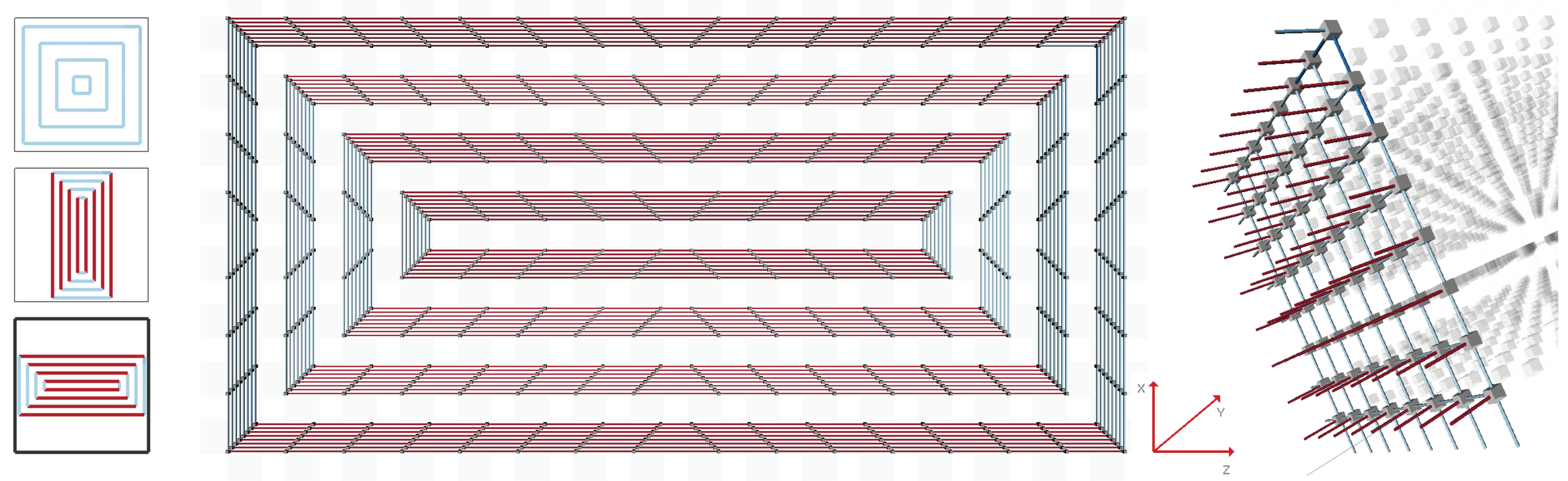
VisIt snapshots for a 1024-core run of SAMRAI on Blue Gene/P. Instead of coloring the patches by their physical properties, we color them by the MPI rank of the process that owns them. Similar colors on nearby patches indicate proximity on the physical interconnect.

Projections on the hardware domain

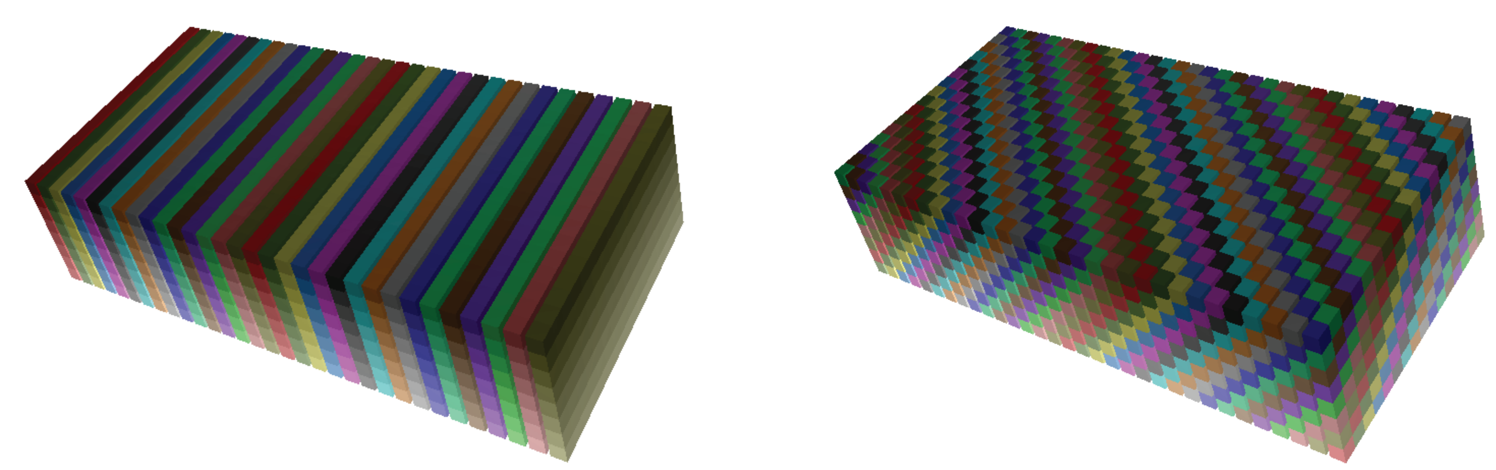
Communication is becoming the dominant performance bottleneck as we scale to a large number of cores. It becomes important to analyze communication in terms of contention on specific links (hot-spots) and distribution of network traffic on the links in various directions.



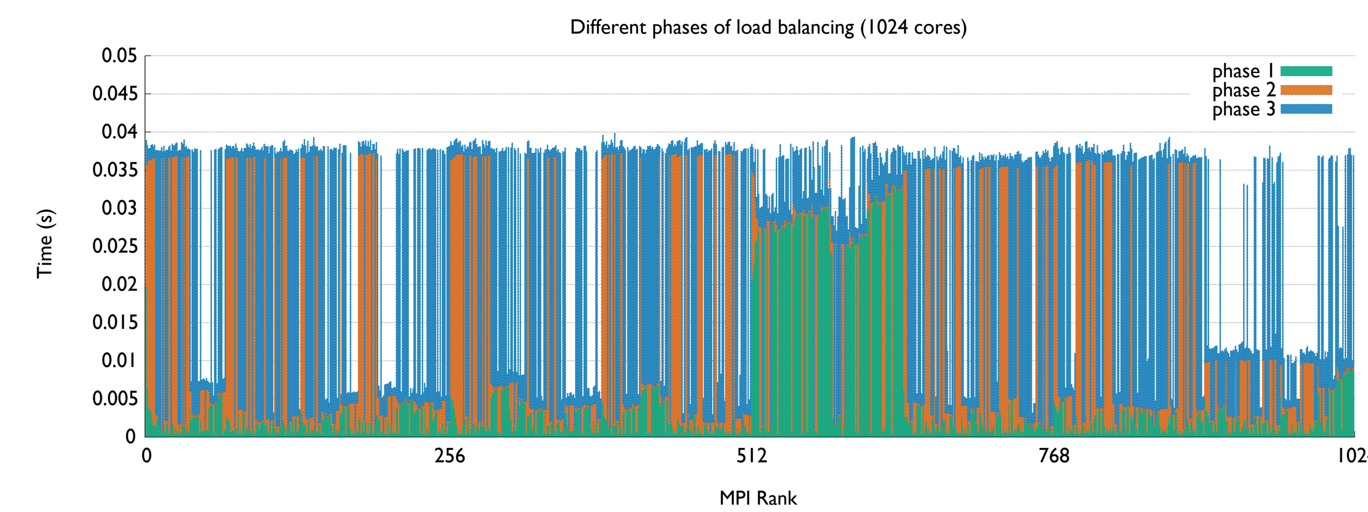
Blue Gene/P network counter data for a point-to-point message (above left), all-to-all over a subset (above center) and global all-to-all pattern (above right and below) visualized using Boxfish.



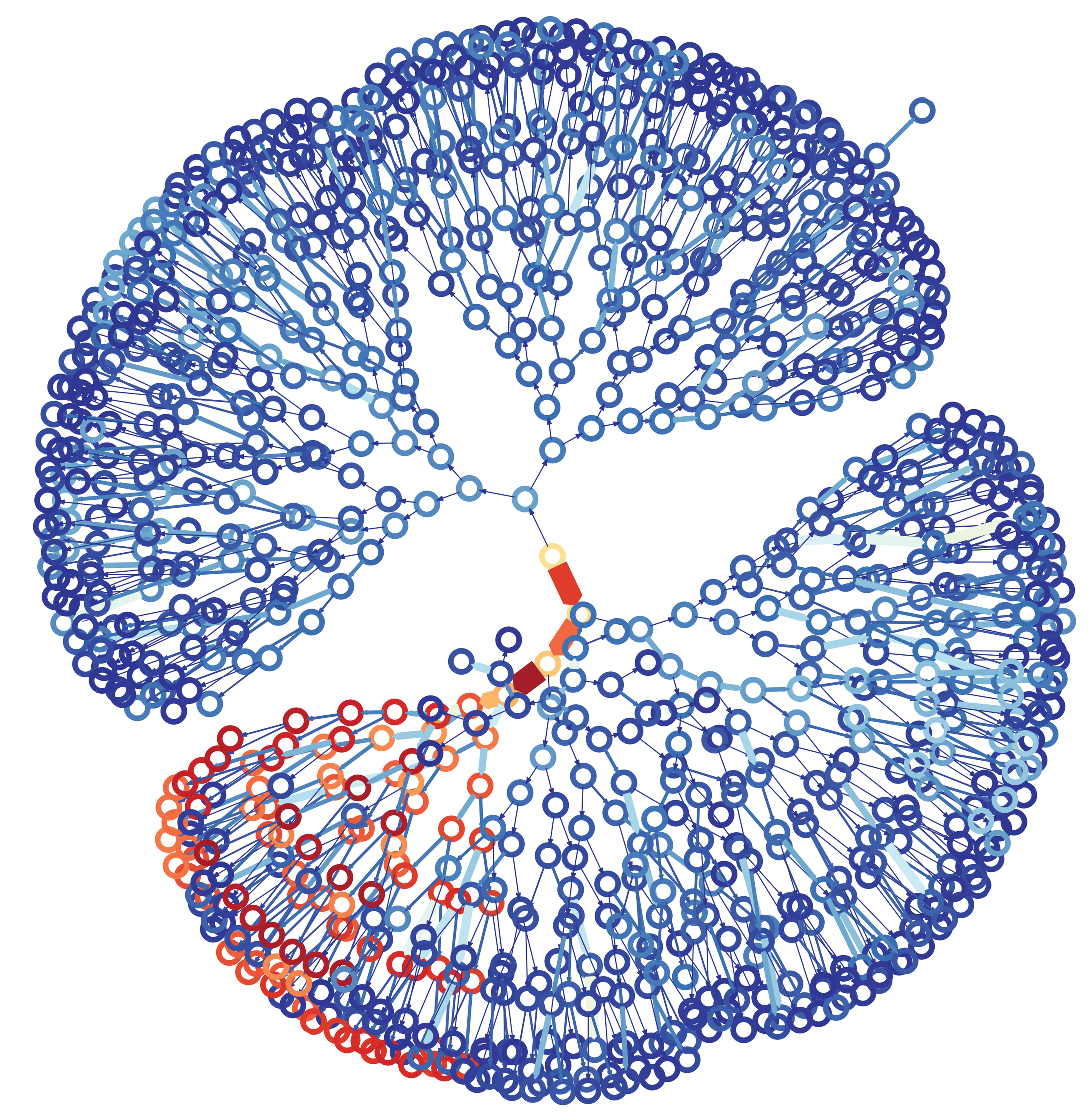
Default (MPI rank ordered, left) and a double tilted mapping (right) of a structured process grid of dimensions 32x8x16 to a torus of dimensions 8x16x32 generated using the Rubik mapping tool.



Projections on the communication domain



Times spent in the three load balancing sub-phases of SAMRAI plotted against the MPI ranks. Phase 1 *i.e.*, load distribution appears to lead to longer wait times in other phases. This MPI rank view does not reveal the cause of the performance problem.



The timing data for phase 1 projected to the communication domain which is a binary tree. Now, we can clearly see that a portion of the tree (in red) is significantly delayed. Coloring the edges by the number of boxes sent down, shows a flow problem arising from the movement of load from the rest of the tree to this sub-tree.

Lawrence Livermore National Laboratory | UC DAVIS UNIVERSITY OF CALIFORNIA

scalability.llnl.gov

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-POST-545412).

THE UNIVERSITY OF UTAH® | ADVANCED SIMULATION & COMPUTING