Report 3: Early Evaluation of IBM BlueGen/P

Tyler Allen

1 Summary

BlueGene/P (BG/P) is a super computer architecture developed by IBM. It is the second generation architecture in the BlueGene series. This paper details the design of BG/P supercomputers. Several benchmarks were performed on an implementation of the BG/P supercomputer to overview its performance.

Each node in the cluster is built with a System-on-Chip (SoC) design containing four PowerPC-brand processors. The nodes each have 2GB memory shared among the processors. The SoC acts as a routing unit, and contains multiple network connections as part of the design. The architecture specifies a 3D torus network topology. There is a network external to the torus konwn as the "global collective network," which handles global operations or IO access.

The system supports several uses for each node, Symmetric Multiprocessor mode (SMP), Dual Node mode (DUAL), and Virtual Node mode (VN). These modes specify the number of MPI tasks that can be handled by each node, and as a result, the number of threads available to each MPI task.

The remainder compares the BG/P supercomputer at Argonne National Laboratory (Intrepid) to the Cray XT system at Oak Ridge National Laboratory (Jaguar). In general, Jaguar outperforms Intrepid due to a number of factors, but this was expected. The Intrepid system performs at a lower clock rate than Jaguar. The results also showed that Intrepid performs best under low-latency conditions. The Intrepid shows consistent results under different network communication protocols. Intrepid is also shown to scale well with different network message sizes. For scientific data, the Intrepid is shown to have varying results. In some cases, such as the Parallel Ocean Program, BG/P

scales and performs well. In other cases, such as the Community Atmospheric Model, BG/P scales poorly, and with fusion simulations such as GYRO, BG/P is shown to perform in line with the first generation BG model, BG/L. BG/P also shows a good performance/power ratio, except when considering certain scientific computation tasks.

2 Evaluation

This paper describes the BG/P architecture in detail, but is mainly focused on performance metrics. The goal of the paper seems to be to characterize the performance of the new BG/P architecture from bottom up. It first analyzes lower level issues, such as latency and bandwidth, before moving up to general benchmarks and then specific scientific application domains. This would probably be most useful to someone, at the time, interested in constructing a supercomputer, or to see the performance benefits of adding the global network to the 3D torus topology.

Using the Cray XT System is an interesting comparison tool. The paper does not compare the two in order to show that one outperforms the other. It is known that Intrepid is not as powerful as Jaguar. The goal is to compare scaling trends between the two. Unfortunately, the Cray system configuration seems to change later in the paper, causing variability in the type of nodes used in this comparison. This seems like it could be misleading, since not all the comparisons are necessarily against the same system. The issues possibly introduced by this are not discussed.

It is also noted several times that, since the Jaguar and Intrepid are not 1-to-1 comparable with hardware specifications, problems have to be altered in order to work within the constraints of Intrepid. The details of these changes are mentioned, but it seems that this could cause a difference in the shown trends. The paper also goes into detail about the different configuration modes for each node on the Intrepid system (SMP, DUAL, VN), but primarily uses VN or VN and SMP for the benchmarks. However, in at least one case, DUAL was the only feasible mode due to memory constraints.

3 Synthesis

It may have been useful to introduce some performance metrics from other systems for reference. The Cray XT is used as the golden standard for scalability in this paper, but prior results from similar systems may give a more grounded result.

It also would have been good to have a section analyzing the expected differences caused by parameter changes. Since parameters were changed to fit Intrepid system specifications, some forecast of expected issues would have been helpful. This is especially important when Cray XT3 nodes are used, since those nodes only perform 2 FLOP per cycle, 2 less than Intrepid and the standard Jaguar nodes. Additionally, explanation of why different configuration modes (SMP, DUAL, VN) were not used for each problem would be helpful. Sometimes VN and SMP are listed, but the reasoning behind that decision does not seem to be explicit. DUAL is chosen once due to memory constraints, so the logical assumption is that the others use more memory. What else goes into this decision?

References

[1] S. Alam, R. Barrett, M. Bast, M.R. Fahey, J. Kuehn, C. McCurdy, J. Rogers, P. Roth, R. Sankaran, J.S. Vetter, P. Worley, and W. Yu, *Early evaluation of ibm bluegene/p*, High performance computing, networking, storage and analysis, 2008. sc 2008. international conference for, 2008Nov, pp. 1–12.