Alexander Schnapp
Max Menges
introhpc02

Intro HPC: Blatt 11

26.1.2015

11.1 Reading

11.1.1 On Achieving High Message Rates

In this paper the authors are pesenting a network architecture based on EXTOLL optimized for high message rates, which is the main influecning factor for sending small messenges, together with the start-up latency. Therefore the system uses a Virtualized Engine for Low Overhead (VELO) for small messenges, which is using PIO to reduce the injection latency as much as possible. Also a Remote Memory Access (RMA) unit is employed, which uses DMA to handle large messages.

On software level they also differentiate between eager and rendevouz protocol In this way they achieve a messege rate of 9 million messeges per second. Which is beating Infiniband and of course the not compateble 10G Ethernet. In the current system fpgas are used but out of these result they want to predict the performence, changing to ASICs, which shoud allow for both higher clock frequencies and wider data paths.

I think it is a pretty good result at very small messege sizes, but since Infiniband is much fast at already 64 B messeges and larger the question is wether the measuerd advantage is also transferable to real workloads.

11.1.2 Global GPU Address Spaces for Efficient Communication in Heterogeneous Clusters

In this paper, the authors propose and implement a model for direct GPU to GPU message passing, by-passing the CPU, called GGAS – Global GPU Adress Space. The approach uses a shared memory engine to map some GPU registers to a cluster wide global memory. For testing, acustom network device was implemented on an FPGA.

The CPU is now no longer required to initiate communication actions and can be utilized for other actions. On a test implementation using two nodes, the authors ran several benchmarks including latency, bandwith and running a stencil code and compared their results to the performance of an Infiniband network with traditional communication methods. First results show a speedup in various performed tasks.

The paper describes what seems to be a novel idea for GPU communications. A follow up paper for extended measurements (better network device, scalability, etc.) would be interesting. Measurements seem a bit preliminary and the the technical implementation could have been a bit more detailed, otherwise a good paper.