

# Intro HPC: Blatt 11

26.1.2015

## **11.1 Reading**

### **11.1.1 On Achieving High Message Rates**

I

### **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 Address Space. The approach uses a shared memory engine to map some GPU registers to a cluster wide global memory. For testing, a custom 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, bandwidth 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 technical implementation could have been a bit more detailed, otherwise a good paper.