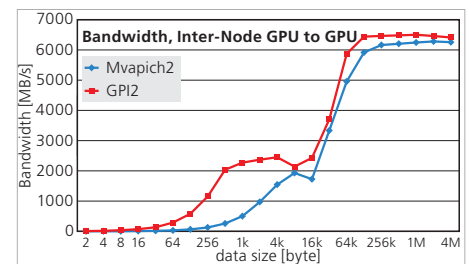
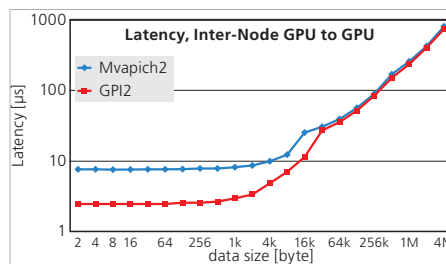


## GPI – SUPPORT FOR GPU<sub>s</sub>

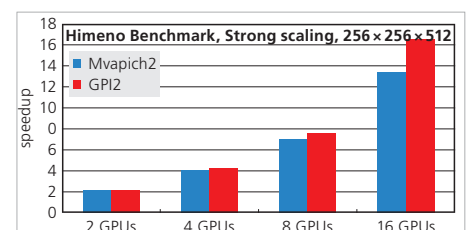
- GPI2: Programming interface to create a partitioned heterogeneous global address space
- Global memory areas can be located in host memory or in GPU memory
- Uses GPUDirect RDMA support from Mellanox and NVIDIA
- Direct remote read/write of GPU remote memory
- Minimal overhead for transferring data between GPUs
- Low latency for small data transfers
- Weak (point to point synchronization) for GPU memory segments

### Micro-Benchmarks



### Himeno

- Using Kepler K40GPUs
- Mellanox FDR Infiniband
- Overlapping of communication and computation
- Linear scaling
- Compared with Mvapih2, GPUDirect RDMA enabled



### Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM

Fraunhofer-Platz 1  
67663 Kaiserslautern  
Germany

#### Contact

Dr. Franz-Josef Pfreundt  
Phone +49 631 31600-44 59  
pfreundt@itwm.fraunhofer.de

Dipl.-Ing. Lena Oden  
Phone +49 631 31600-45 62  
lena.oden@itwm.fraunhofer.de

www.itwm.fraunhofer.de