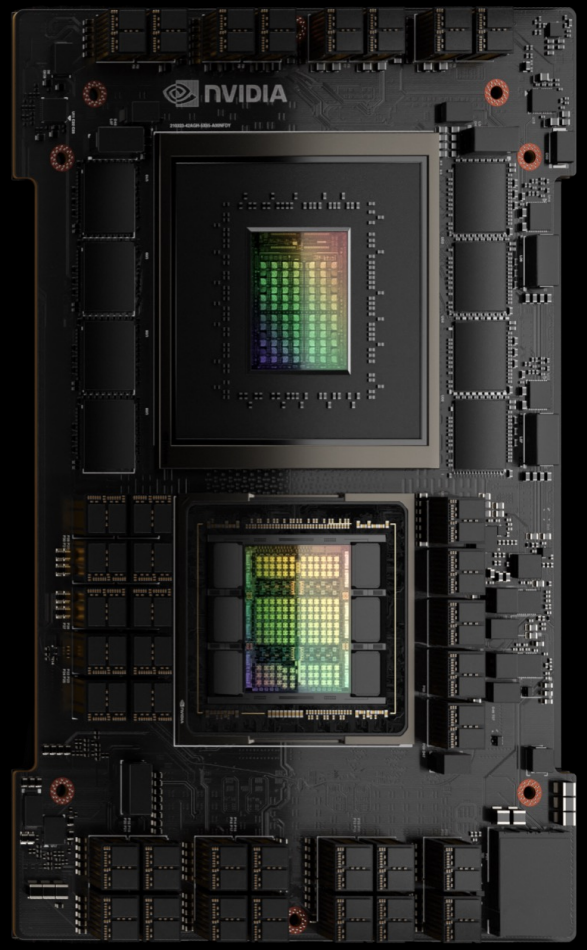
# UCX Update

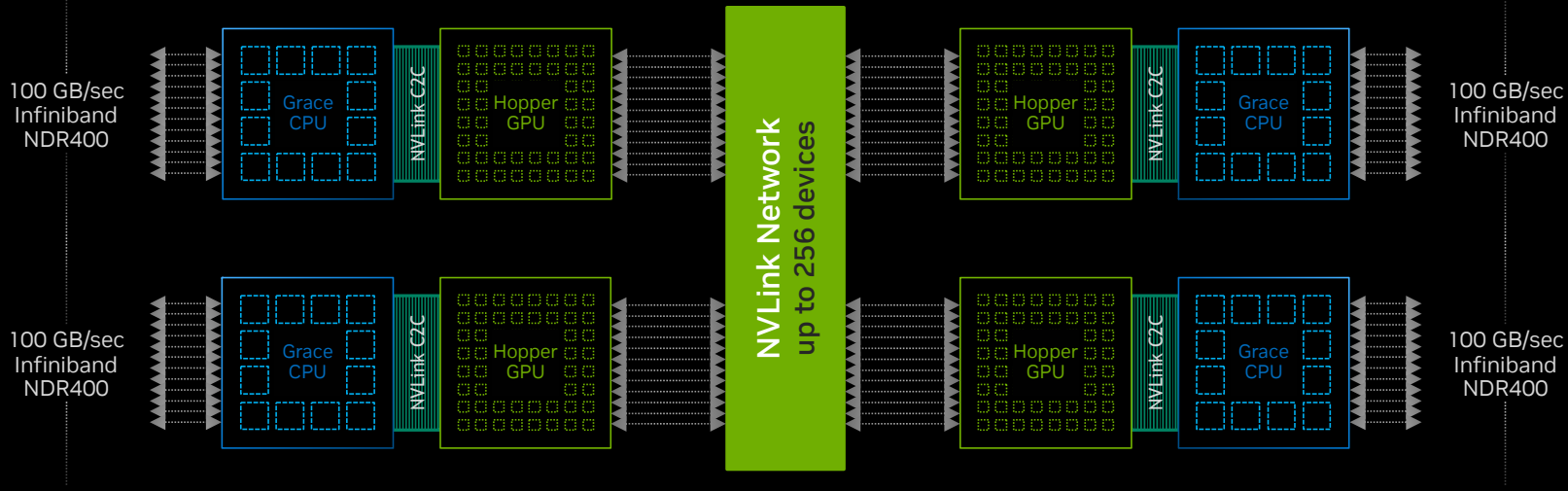
Jim Dinan, SC’23 MPICH BoF

# New Features in UCX




| Feature | Description |
|---|---|
| UCP Protocols v2 | Overhaul of protocol selection logic for memory-type and location-based transport and protocol optimizations |
| DMABUF support | Allow registration of pinned GPU memory using DMABUF file descriptors. Removes dependence on nvidia-peermem. |
| On-Demand Paging | Use ODP registration for system memory transfers on Grace-Hopper. Support CPU and GPU page migration via high-bandwidth C2C links. Opt-in. |
| cudaMallocAsync support | Allow passing async allocated memory to UCP operations |
| Hopper support | Enable CPU/GPU transfers for Hopper GPUs |

# Upcoming UCX Features




| Feature | Description |
|---|---|
| On-Demand Paging | Automatically use ODP for CUDA Managed Memory |
| Device pipeline protocols | Pinned GPU memory bounce buffers for managed/sysmem transfers<br>Using location hints to decide bounce buffer location for pipeline protocols |
| Multi-node NVLINK | Use NVLINK across OS instances for pinned GPU transfers |
| EGM | Use Extended GPU Memory for pipeline protocols |
| Multiple device contexts | Allow a single process/single UCP context to own multiple device contexts |
| DGX cloud | Virtual topology injection infrastructure and rail-optimized protocols |