GASNet-EX:
Communication Support for Legion

Paul H. Hargrove

gasnet-staff@lbl.gov    gasnet.lbl.gov

Joint work with Dan Bonachea
and the LBNL Pagoda Project (CRD/CLaSS)
Support for lightweight communication for exascale applications, frameworks and runtimes

- **GASNet-EX** middleware layer providing a network-independent interface suitable for Partitioned Global Address Space (PGAS) runtime developers
- **UPC++** C++ PGAS library for application, framework and library developers, a productivity layer over GASNet-EX

---

**Diagram:**

- **DOE Scientific Applications:**
  - Arkouda
  - FLeCSI
  - FlexFlow
  - ExaBiome
  - ExaGraph
  - NWChemEx
  - AMReX
  - ...  

- **GASNet-EX**
  - One-sided Get/Put RMA
  - Atomics
  - Active Messages
  - Memory Technologies (Host memory, GPUs, ...)
  - Network Hardware (InfiniBand, Cray Aries, HPE Slingshot, Ethernet, Intel Omni-Path, ...)

- **Legion**
- **Berkeley UPC**
- **UPC++**
  - Fortran coarrays
  - SHMEM
  - ...  

**ECP Exascale Computing Project**
- Closely co-developed ECP projects
- Closely co-developed non-ECP projects
- Other interacting components

**Website:** [https://go.lbl.gov/pagoda](https://go.lbl.gov/pagoda)
GASNet-EX in the Legion Ecosystem
GASNet History

• Began in 2002 to provide a portable network runtime for three Partitioned Global Address Space (PGAS) languages: UPC, Titanium and CAF
  – Chosen over then-current alternatives: MPI-2, ARMCI
• Provides Remote Memory Access (RMA) and Active Message (AM) interfaces for implementing Partitioned Global Address Space (PGAS) programming models
• GASNet-EX is the next generation of GASNet
  – Updates GASNet-1 design to address the needs of newer programming models such as UPC++, Legion and Chapel
  – Incorporates 20 years of lessons-learned and focuses on the challenges of emerging exascale systems
  – Provides backward compatibility for GASNet-1 clients
GASNet…

- is “Global Address Space Networking”
- is an AM and RMA API for implementing PGAS models
- is designed for compilers and authors of low-level code
- is MPI-interoperable on most platforms
- performs comparably to (and often better than) MPI
- has influenced design of RMA in MPI-3 and later
- is the primary networking library for distributed execution of Legion
A comparison of uni-directional point-to-point host-memory flood bandwidth benchmarks, run March 2022 on OLCF’s Crusher system.

Shows the performance of RMA (Put and Get) operations using GASNet-EX and both RMA and message-passing (Isend/Irecv) using HPE Cray MPI.

Results were obtained using current GASNet tests and Intel MPI Benchmarks, respectively.

**UP IS GOOD**

- Puts peak at 4x smaller transfer
- Gets reach an 11% higher peak
GASNet-EX and Legion: GPU RMA

Realm is Legion’s low-level runtime, providing comm. services
- Originally implemented over GASNet-1
- Still works using legacy API support in current GASNet-EX
Realm introduced a new GASNet-EX backend (Dec 2020)
- Embraces EX-specific capabilities
- Leverages Immediate, NPAM, and LC handles for AM
- Most notable new capability is “memory kinds”: support for offloaded GPU xfers (GDR and ROCmRDMA)

Figures illustrate some performance benefits of memory kinds:
- Large GPU memory xfers: same bandwidth as host mem
- Small GPU memory xfers: 2.2x to 3.0x latency improvement

Multi-endpoint allows RDMA for all Realm-allocated host buffers
- Avoids copies needed with the GASNet-1 API
THANK YOU!

gasnet-staff@lbl.gov

gasnet.lbl.gov

GASNet

20th Anniversary

est 2002