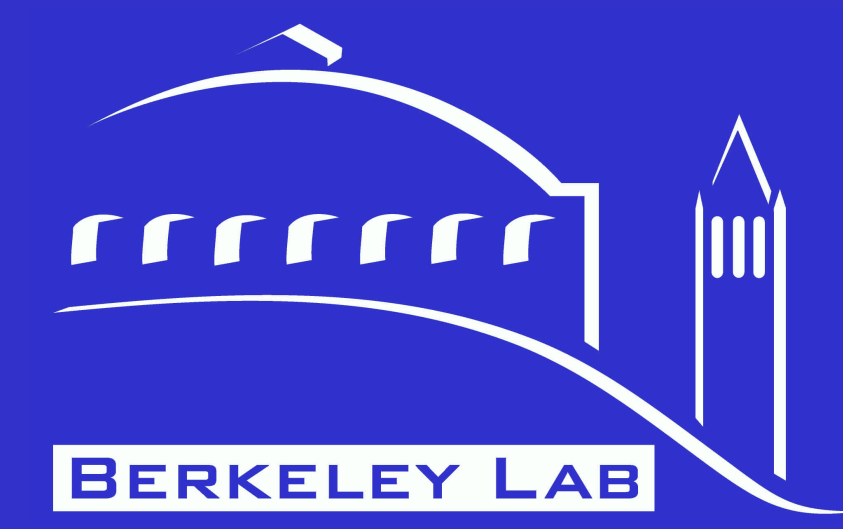




# GASNet



<http://upc.nerdc.gov>

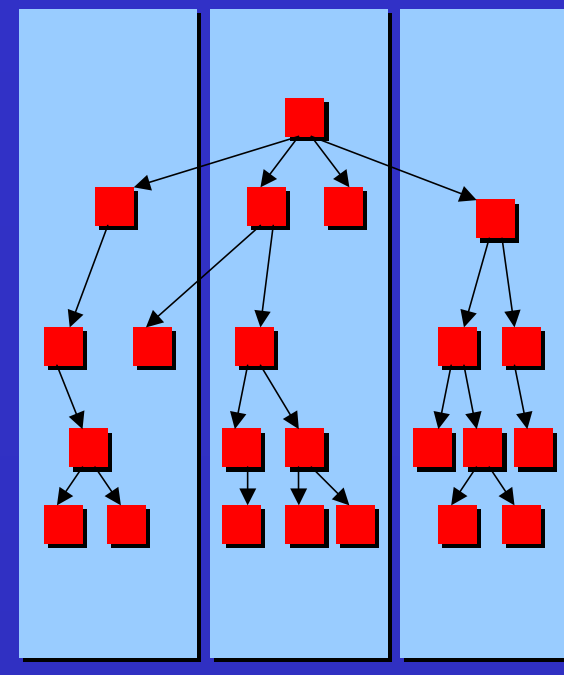
**U.C. Berkeley and LBNL**

[upc@lbl.gov](mailto:upc@lbl.gov)

**Christian Bell, Dan Bonachea, Wei Chen, Jason Duell, Paul Hargrove,  
Parry Husbands, Costin Iancu, Mike Welcome, Kathy Yelick**

## Global Address Space Languages

- Global address space languages support
  - Global pointers and distributed arrays
  - User controls layout of data across nodes
  - Direct read and write to remote memory
- Single Program Multiple Data (SPMD) control
  - Similar to using threads, but with remote accesses
  - Global synchronization, barriers
- Languages: UPC, Co-Array Fortran, Titanium
- GASNet - A common communication system tailored for global address space languages



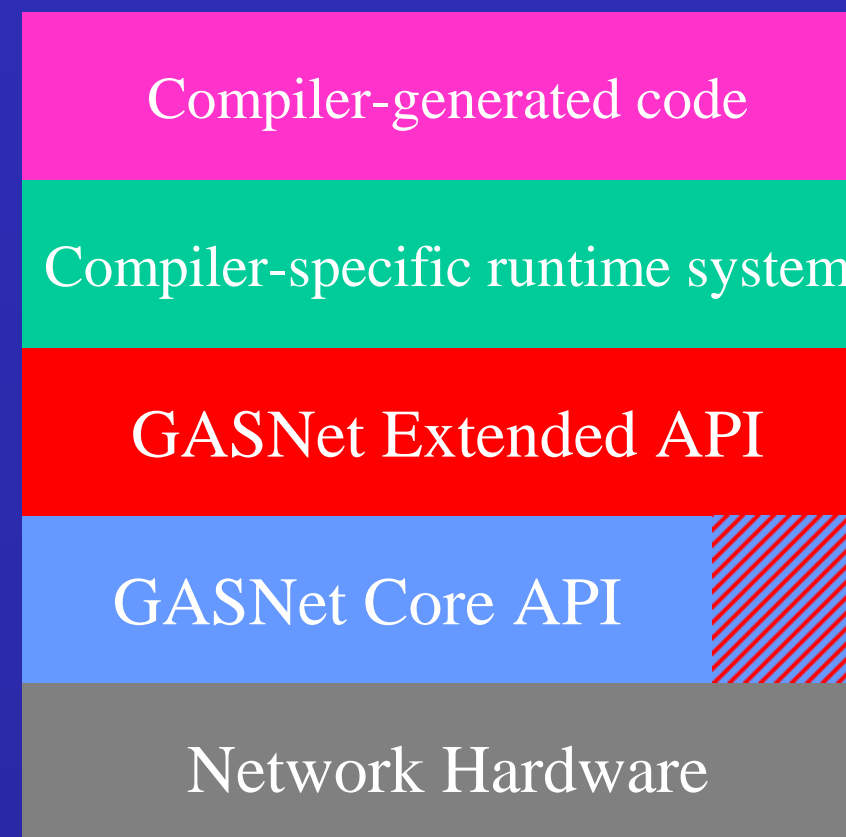
Distributed Data Structures

## GASNet Goals

- Language-independence: Compatibility with several global-address space languages and compilers
  - UPC, Titanium, Co-array Fortran, possibly others..
  - Hide language- or compiler-specific details, such as shared-pointer representation
- Hardware-independence: variety of parallel architectures & OS's
  - SMP: Linux/UNIX SMP's, Origin 2000, etc.
  - Clusters of uniprocessors or SMP's: IBM SP, Compaq AlphaServer, Linux/UNIX clusters, etc.
  - Support many high-performance networks: Infiniband, Myrinet/GM, Quadrics/elan, IBM/LAPI, Dolphin, MPI
- Ease of implementation on new hardware
  - Allow quick prototype implementations
  - Implementations can leverage performance features of hardware
- Provide both portability & performance

## GASNet Core API

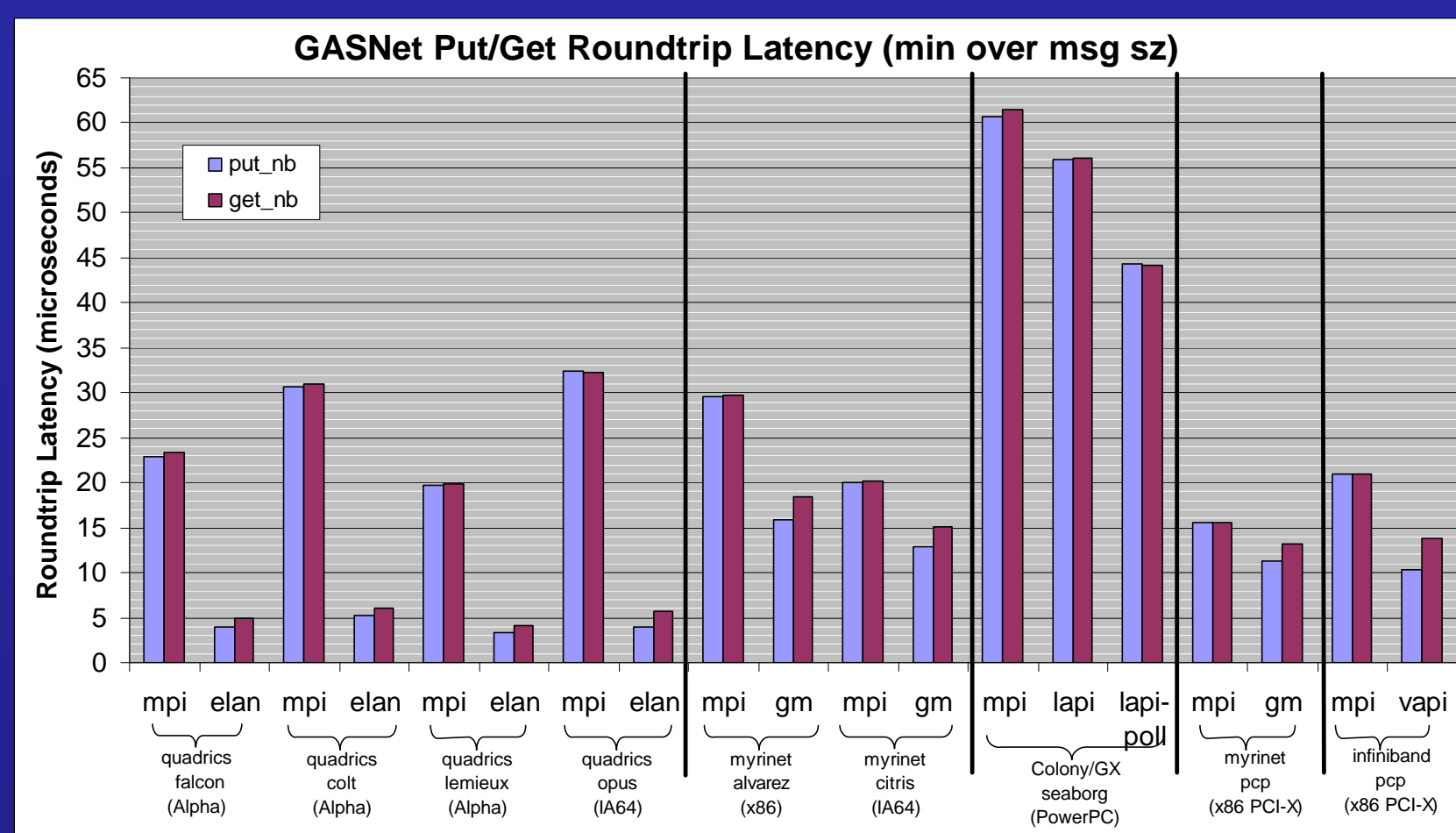
- Most basic required network primitives
- Implemented directly on each platform
  - Minimal set of network functions needed to support a working implementation
  - General enough to implement everything else
- Based heavily on active messages paradigm
  - Provides powerful extensibility mechanism



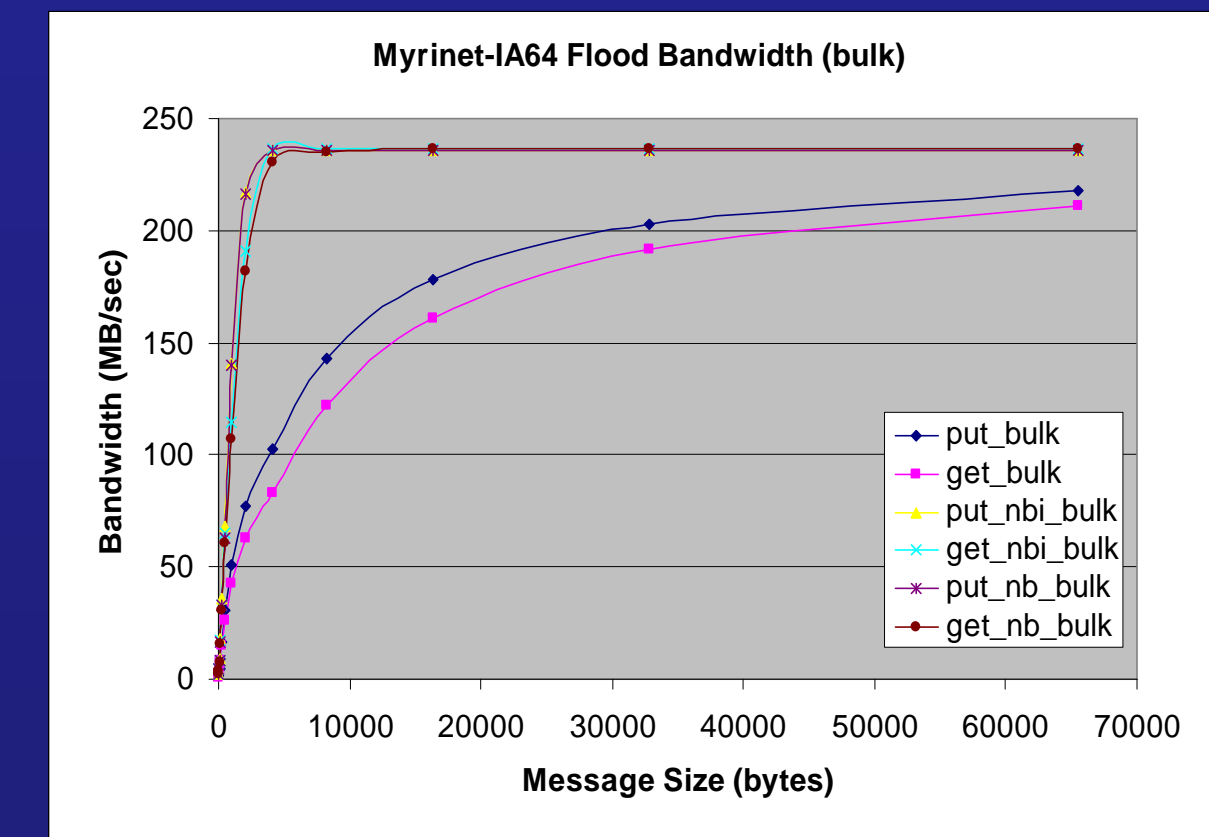
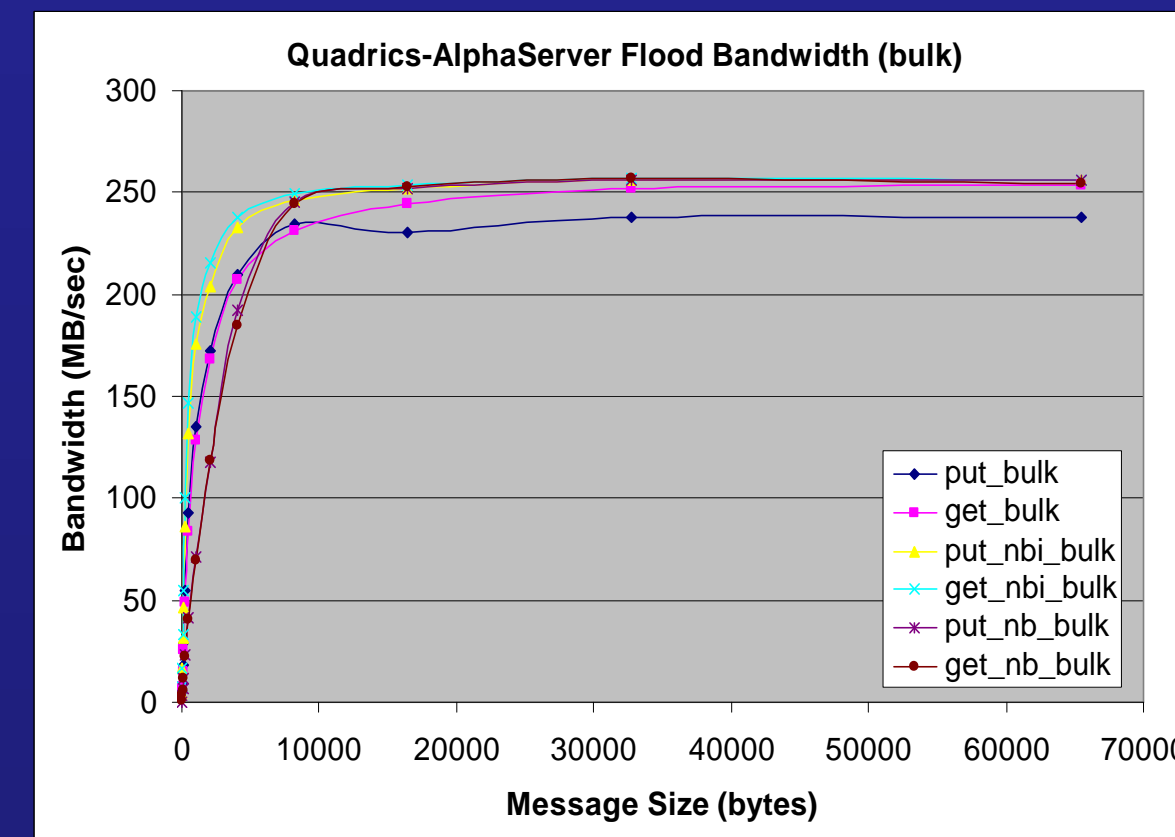
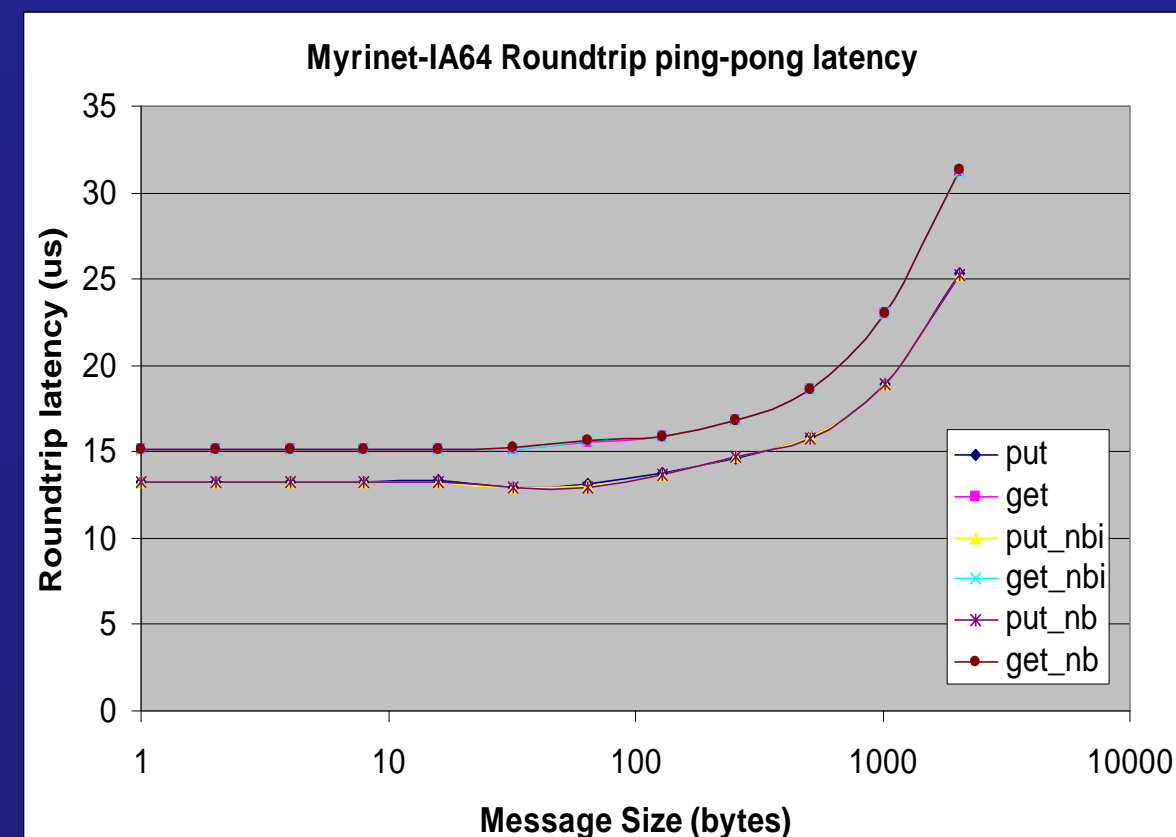
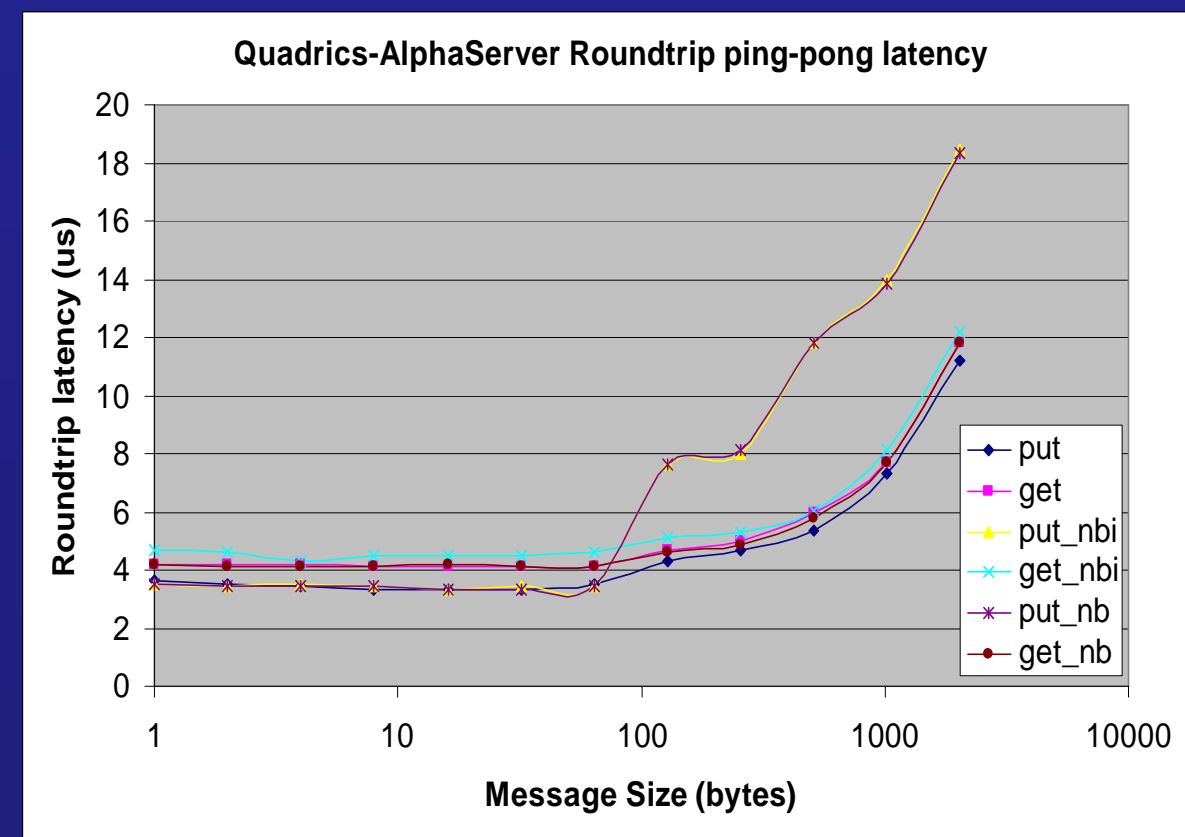
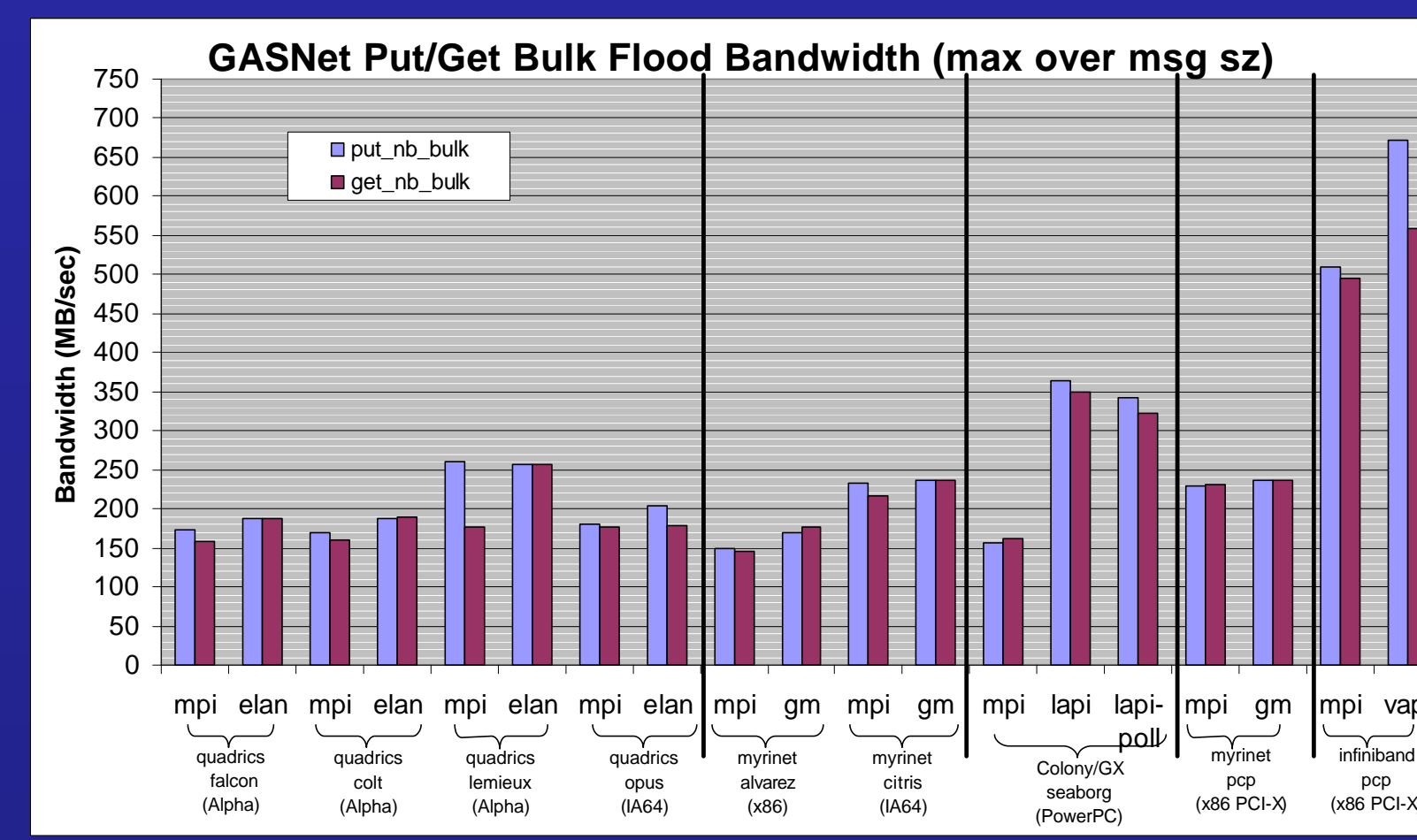
## GASNet Extended API

- Wider interface that includes more complicated operations
- We provide a reference implementation of the extended API in terms of the core API
- Implementors can choose to directly implement any subset for performance - leverage hardware support for higher-level operations

## Latency Performance

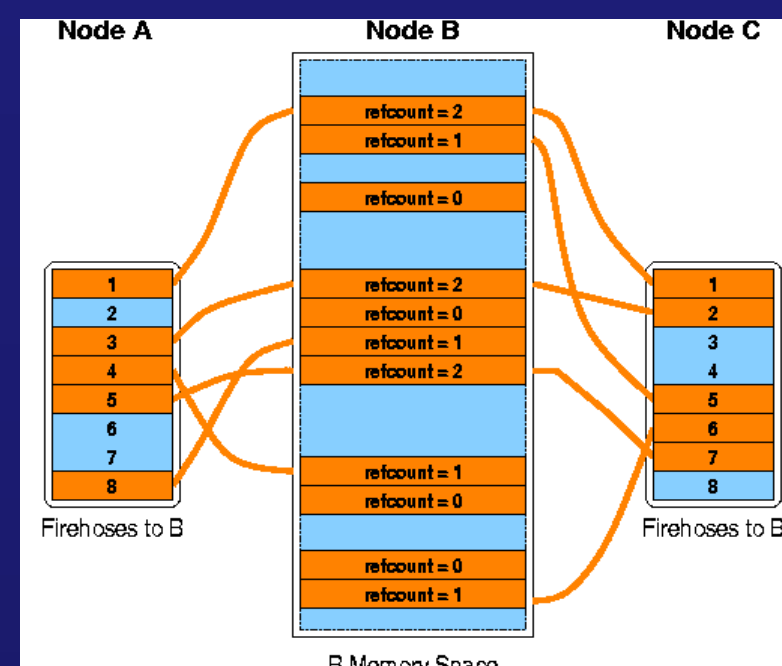


## Bandwidth Performance



## Implementing GASNet on InfiniBand

- Using Mellanox VAPI interface
  - Vendor implementation of the InfiniBand Verbs
  - Useful minor extensions beyond Verbs
- GASNet Core API: Active Messages
  - Based on Send/Recv operations
  - Simple flow control
  - Uses an additional thread for improved responsiveness
- GASNet Extended API: RDMA
  - Thin layer over InfiniBand RDMA (puts and gets)
  - Simple record attached to each CQE for completion
  - No dynamic memory registration yet (bounce-buffers used for out-of-segment references)



Firehose Algorithm for distributed management of DMA registration

## Future Work

### VAPI conduit:

- Dynamic memory registration (work in progress)
    - Extension of "firehose" algorithm for region-oriented registration
    - C. Bell and D. Bonachea. "A New DMA Registration Strategy for Pinned-Based High Performance Networks" CAC 2003
  - RDMA-based barrier
  - General performance tuning
- ### General GASNet:
- Collective Communication
  - Point-to-point Scatter/Gather and Strided put/get operations
  - New GASNet conduit implementations:
    - Florida State University: SCI/Dolphin
    - UC Berkeley: shmemp (Cray, SGI and Quadrics), and Cray X-1

