Message-Passing Software Status and Plans

Karl Feind

Parallel Communication Engineering

SGI

Outline

Message-passing software strategy

Recent software enhancements

Future plans

MPT Themes

- Performance
- Platforms and Interconnects
- Standards

MPT Supports Fast Interconnects

Fast MPI-1

- ➢Fast send-receive of all message lengths
- ➢High message queuing rate
- ➢Fast MPI collectives

Fast extensions to MPI-1

- ➢SHMEM put/get
- ➢MPI-2 put/get
- ➢SHMEM global pointer

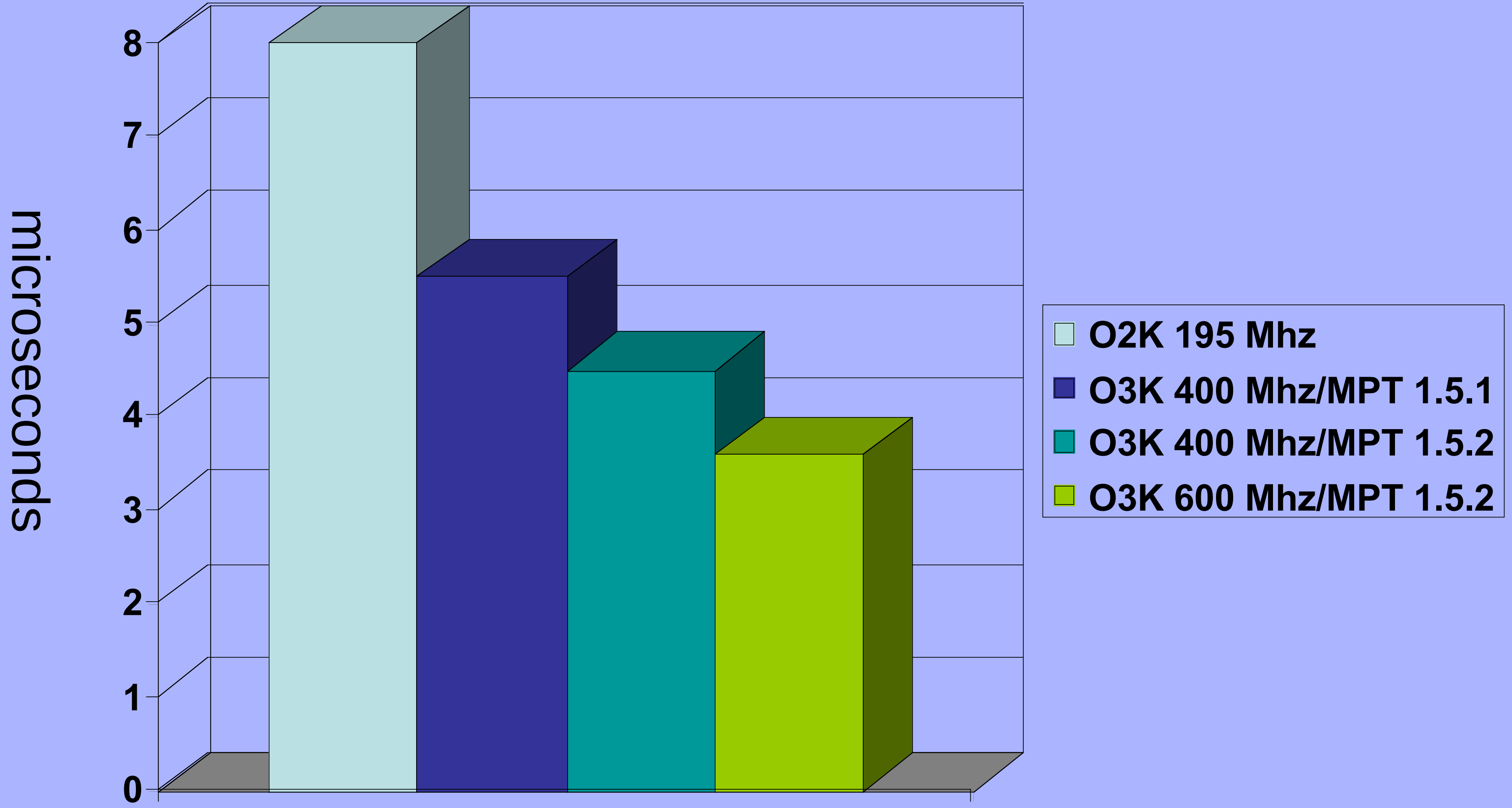
New Licensing

- MPT 1.6 is available
    - Available for no fee downloads
    - Will accompany IRIX 6.5.17 media
- PVM 3.3 is unbundled from MPT 1.6
    - Available at http://www.sgi.com/products/evaluation
- Message Passing Helpers available soon
    - See http://freeware.sgi.com
    - perfcatcher and default64 MPI wrappers

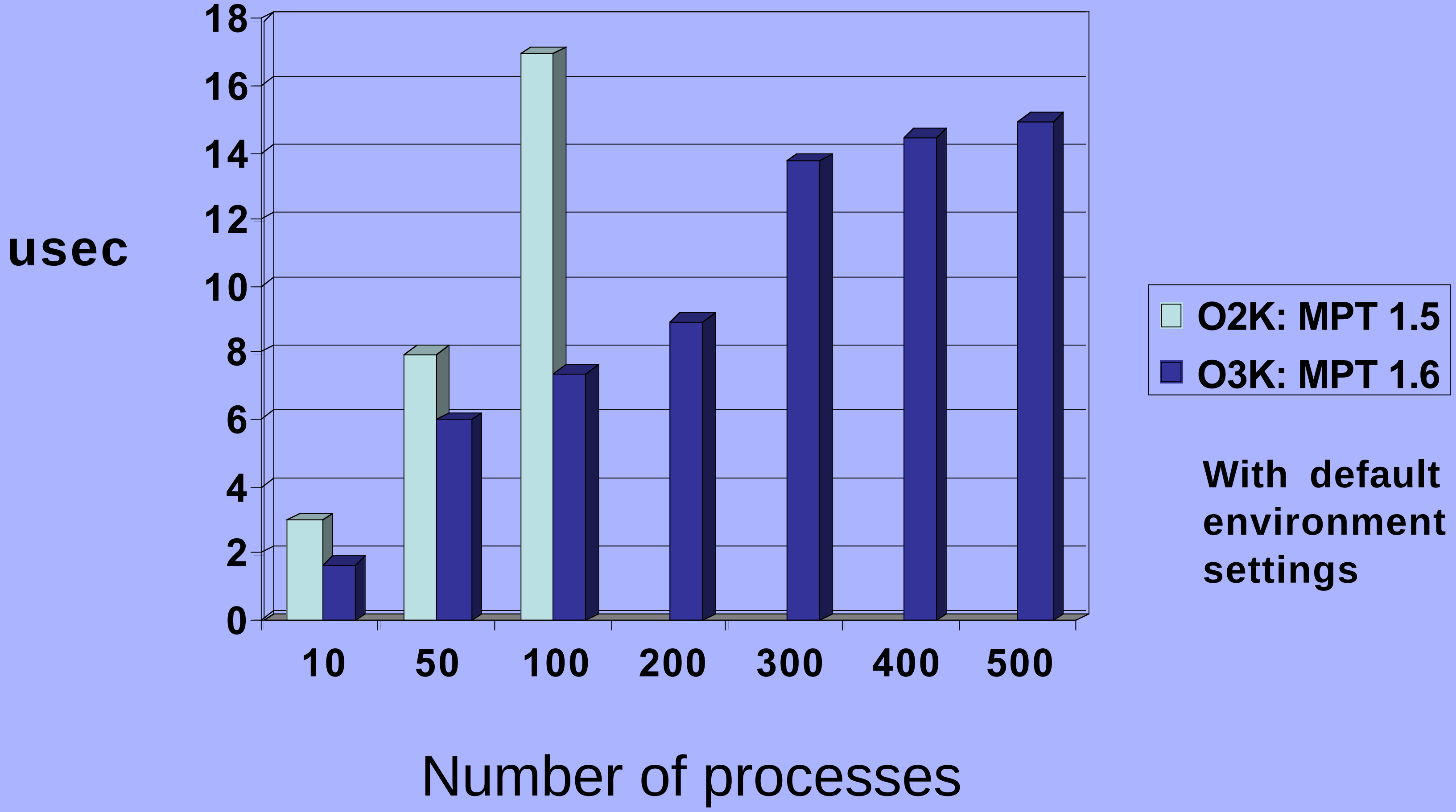
Reducing Communication Latency

- MPI send and receive
  - 1 microsecond better on Origin 3000
  - We tweaked the message-queue algorithm
- Barrier Sync time improved
  - Activated tree barrier by default

Improving Send/Receive Latency



Improving Barrier Sync Scalability



With default environment settings

More Single-Copy Send/Receive

Traditional single-copy

  ＝common block or symmetric heap data or MPI_Alloc_mem

  ＋set MPI_BUFFER_MAX

XPMEM-based single-copy (Origin 300/3000)

  ＝set MPI_XPMEM_ON

  ＋set MPI_BUFFER_MAX variable

See *MPI Programmer’s Manual*: Optimization and Tuning

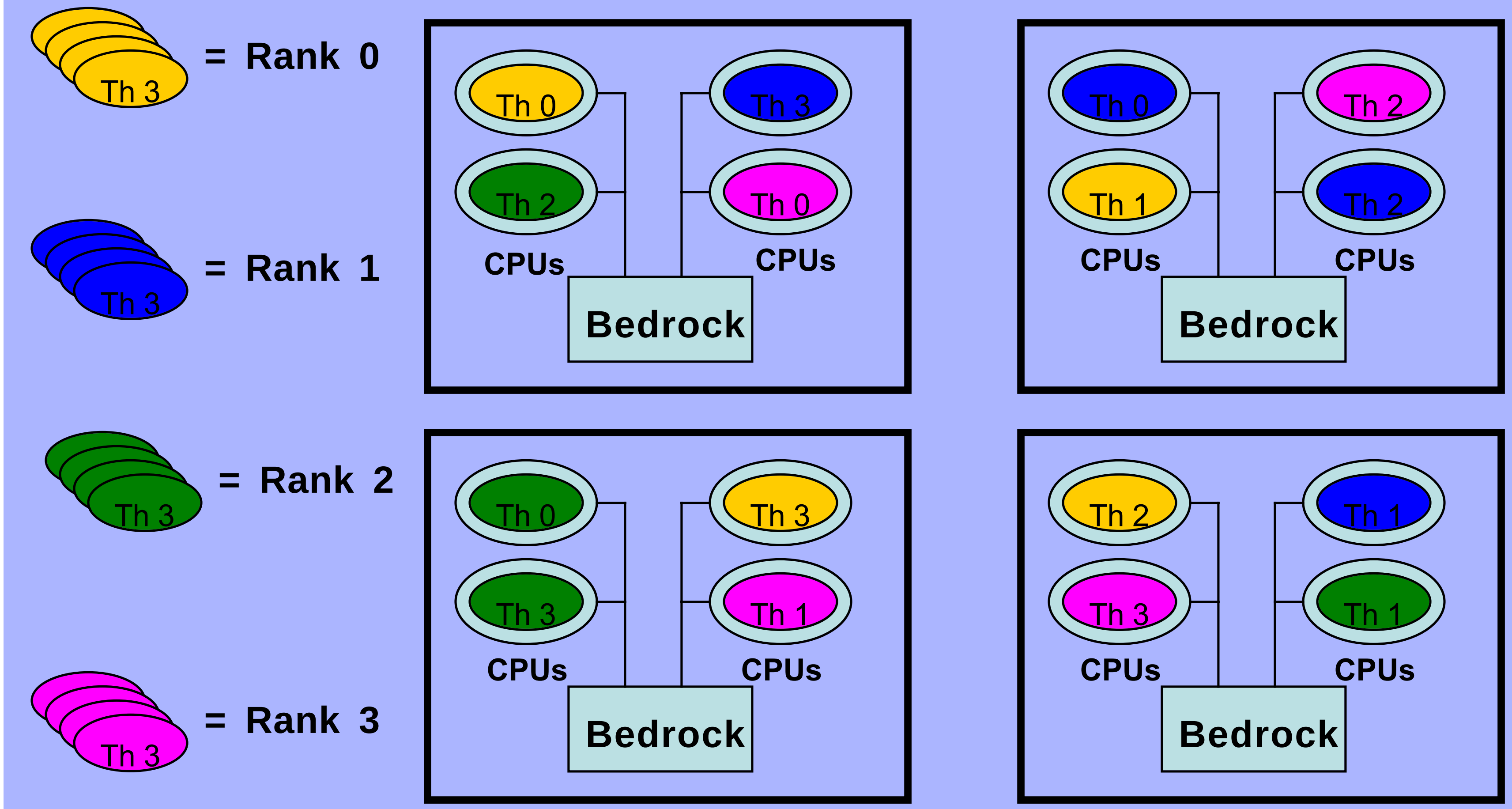
Coordinated MPI/OpenMP Hybrid Launch

MPI and OpenMP in the same application

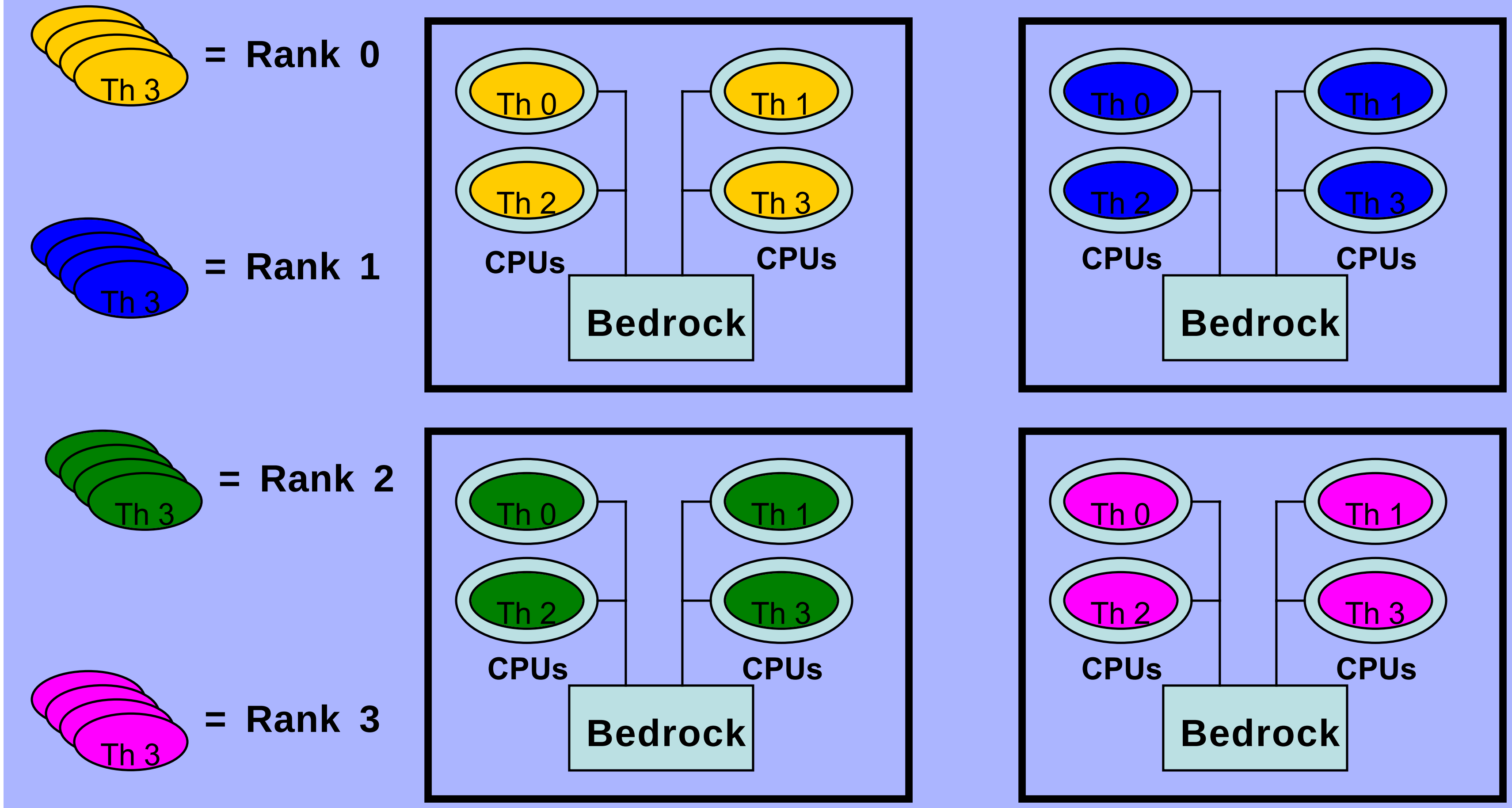
New interoperation controls in MPT 1.6

- Only on Origin 300 and Origin 3000
- MPI_OPENMP_INTEROP variable
- See mpi(1) man page

MPI and OpenMP: Random Placement



MPI and OpenMP: Coordinated Placement



MPI-2 Spawn

SGI users running coupled MPI models

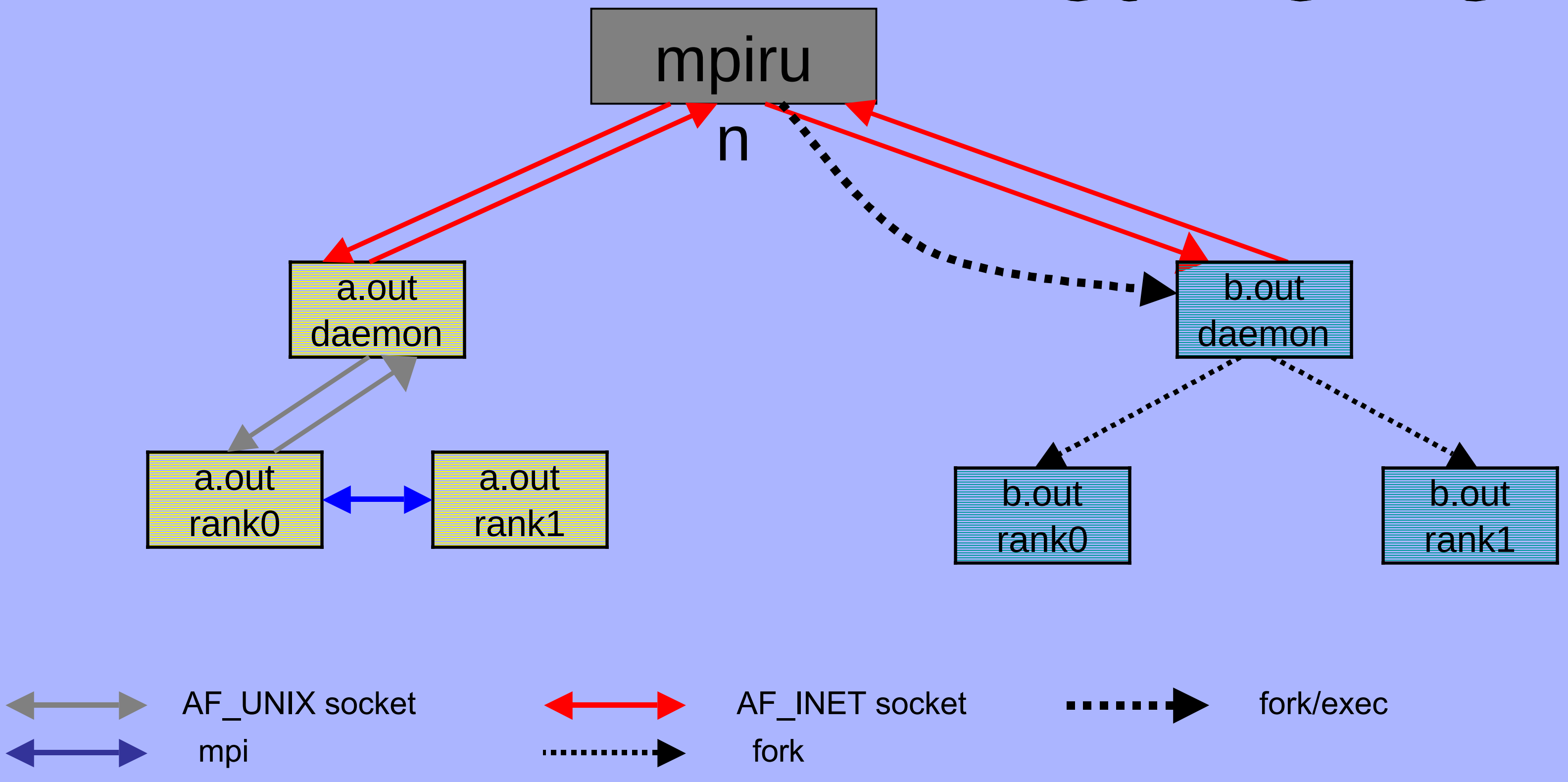
Support restricted to single host for now

See mpi_comm_spawn(3) man page

- MPI spawn operation

```
mpirun -up 4 -np 2 a.out
```

same ASH



Other Recent Additions

MPI Optimization Chapter added in *MPI Programmer's Manual*

GSN support for 512 P hosts

Improved MPI startup time on large systems

MPI-2 Features:

- MPI_Alloc_mem
- Fortran/C transfer of MPI handles
- Replacements for deprecated datatypes

Future Plans for Message-Passing Software

Large System Improvements

- Page table space sharing
- Program startup-time

More MPI-2

- Generalized requests
- MPI I/O refresh and MPI_Wait support
- Spawn and OpenMP ineterop refinements

Porting to SN McKinley