Parallel Technologies ParTec

ParaStation

The new paradigm of Cluster Computing

Ferdinand Geier
ParTec AG

Company background

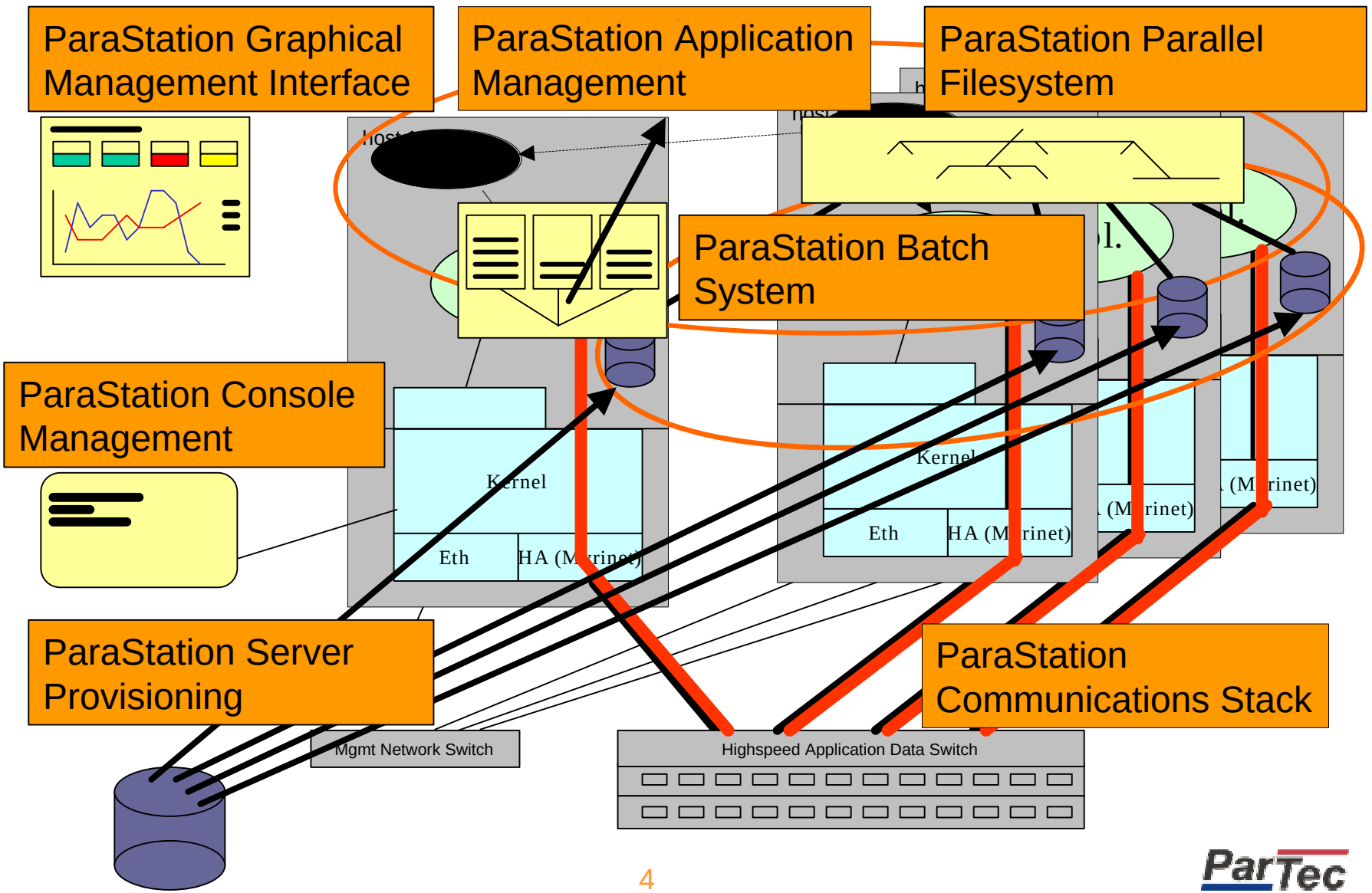
- 1995 Presentation of the ParaStation Hardware.
  - Development of the ParaStation Software.
- 1998 Presentation of ParaStation2 on top of MyriNet.
- 1999 Founding of ParTec AG.
  - Spin-Off from the Karlsruhe University.
- 2001 Presentation of the ParaStation3 software.
- 2003 April: Presentation of ParaStation 4.

ParaStation: Product overview

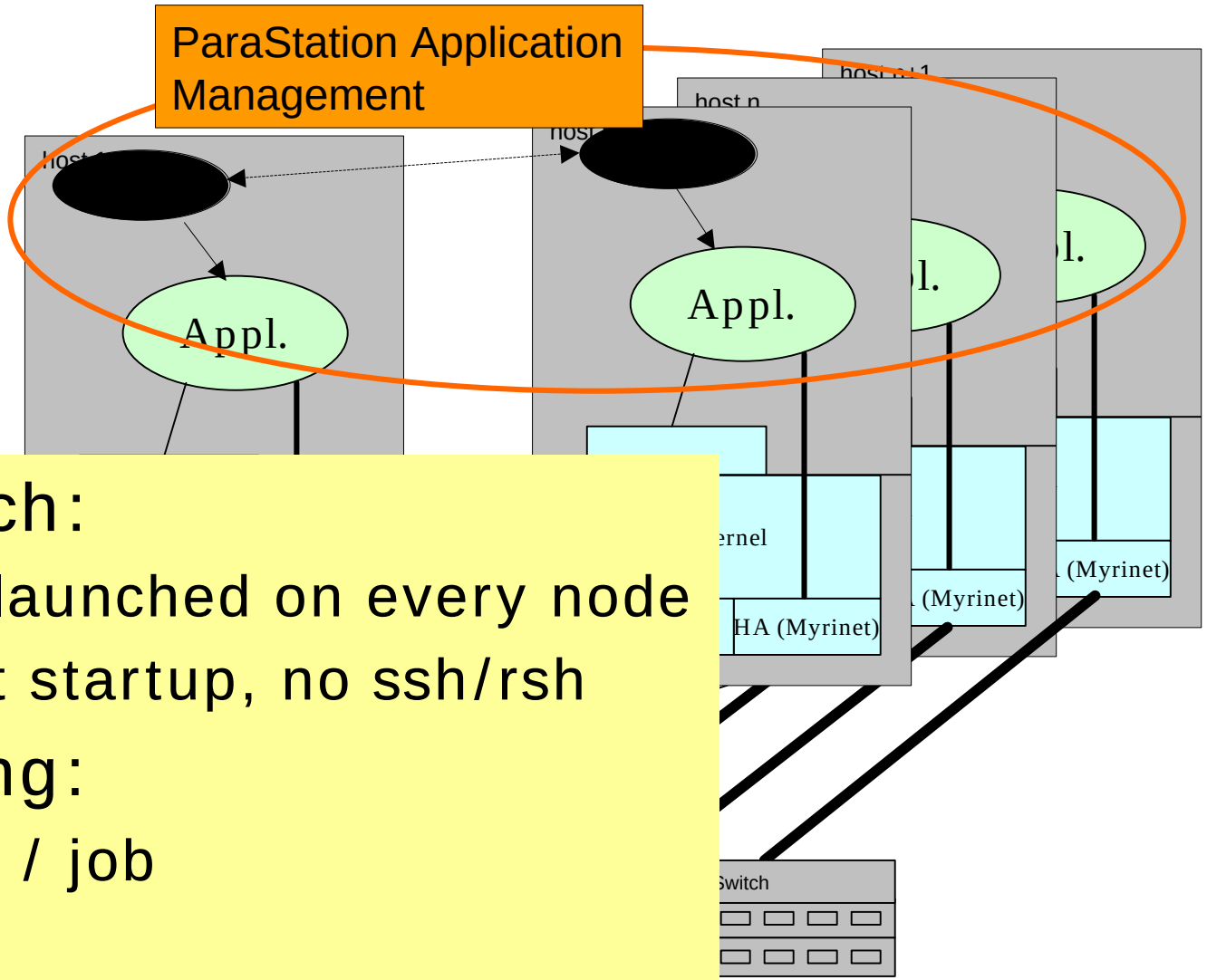
- ParaStation 3
  - based on Myrinet
  - User Level communication
- ParaStation FE
  - based on TCP/IP
- New: ParaStation 4
  - enhanced functionality
  - „hybrid communication“
- Expertise: **C**luster **C**ompentence **C**enter
  - Interconnects
  - Software
  - Benchmarks

ParTec

ParaStation: Overview



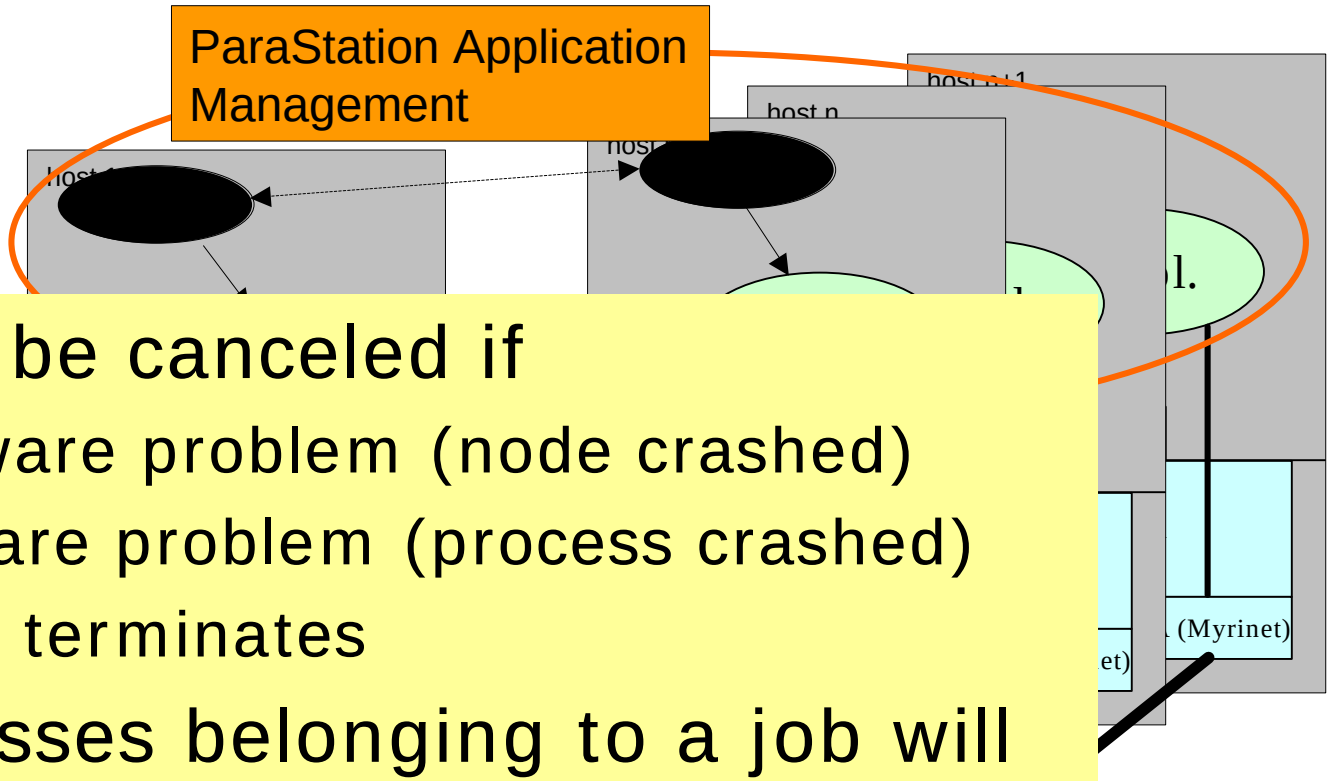
ParaStation: Management



- Job launch:
  - can be launched on every node
  - efficient startup, no ssh/rsh
- Monitoring:
  - process / job
  - node

ParTec

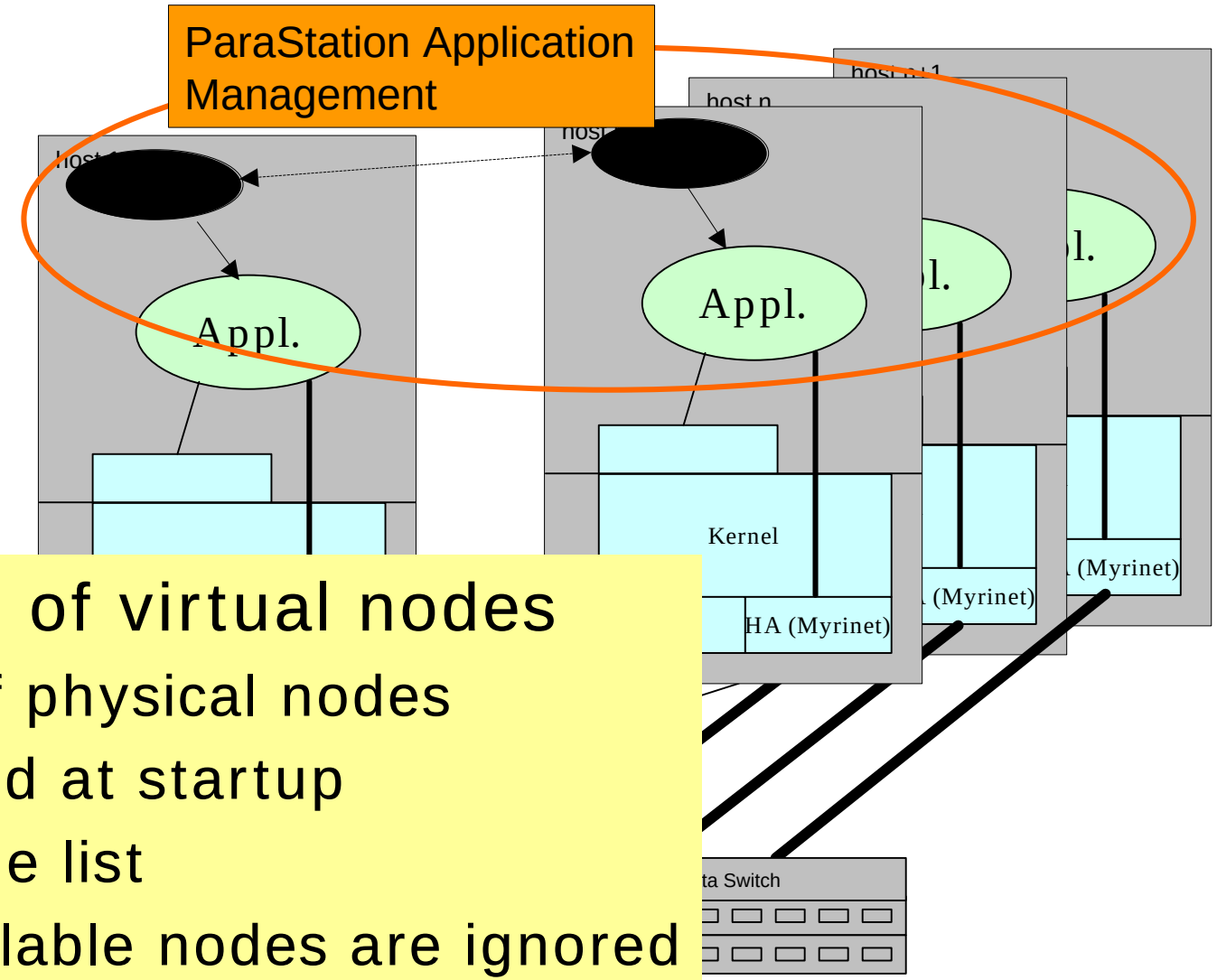
ParaStation: Management



- Jobs will be canceled if
  - a hardware problem (node crashed)
  - a software problem (process crashed)
  - process terminates
- All processes belonging to a job will be terminated!
- No orphaned processes left eating up CPU cycles!

ParTec

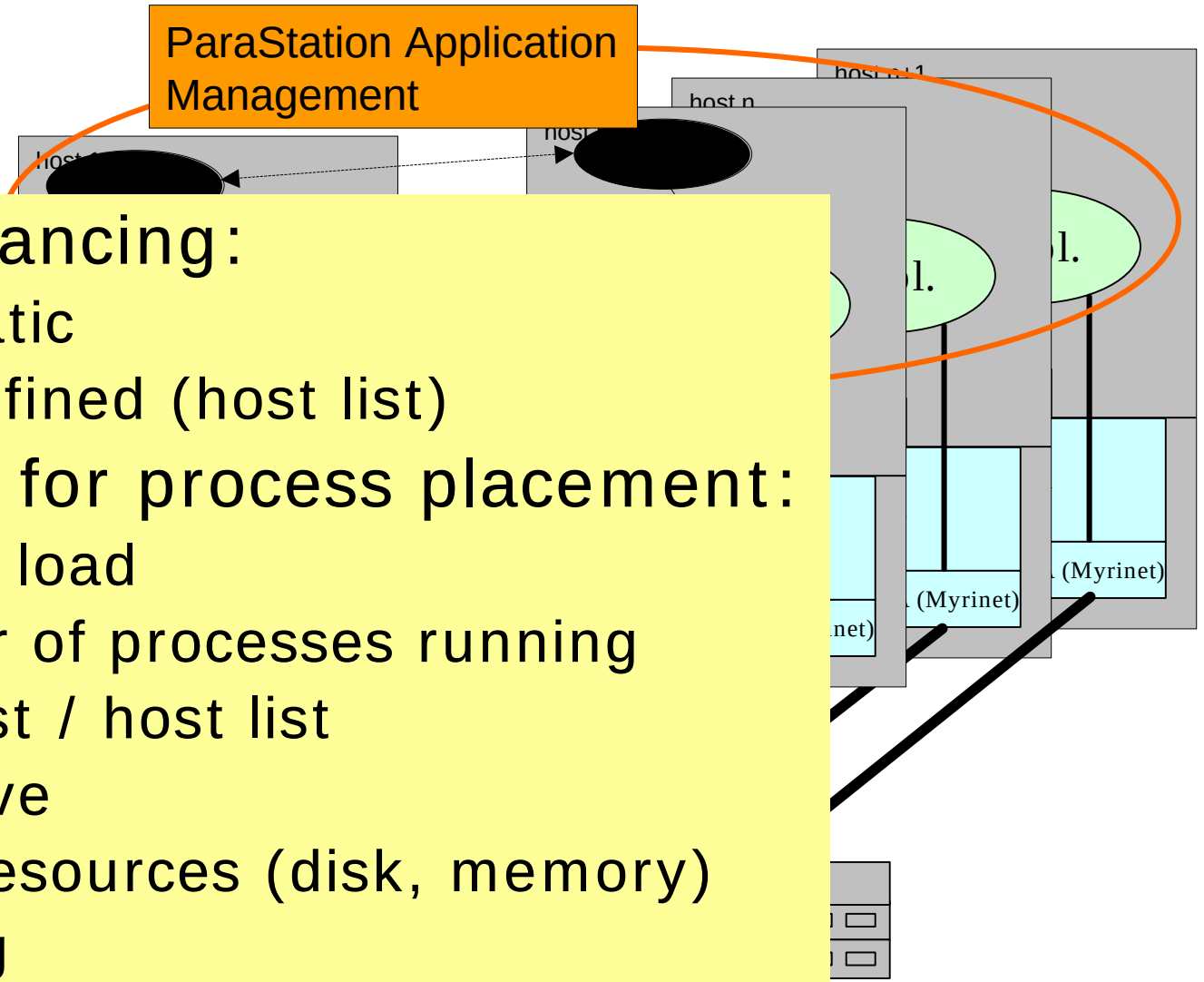
ParaStation: Management



- Concept of virtual nodes
  - pool of physical nodes
  - mapped at startup
  - no node list
  - unavailable nodes are ignored

ParTec

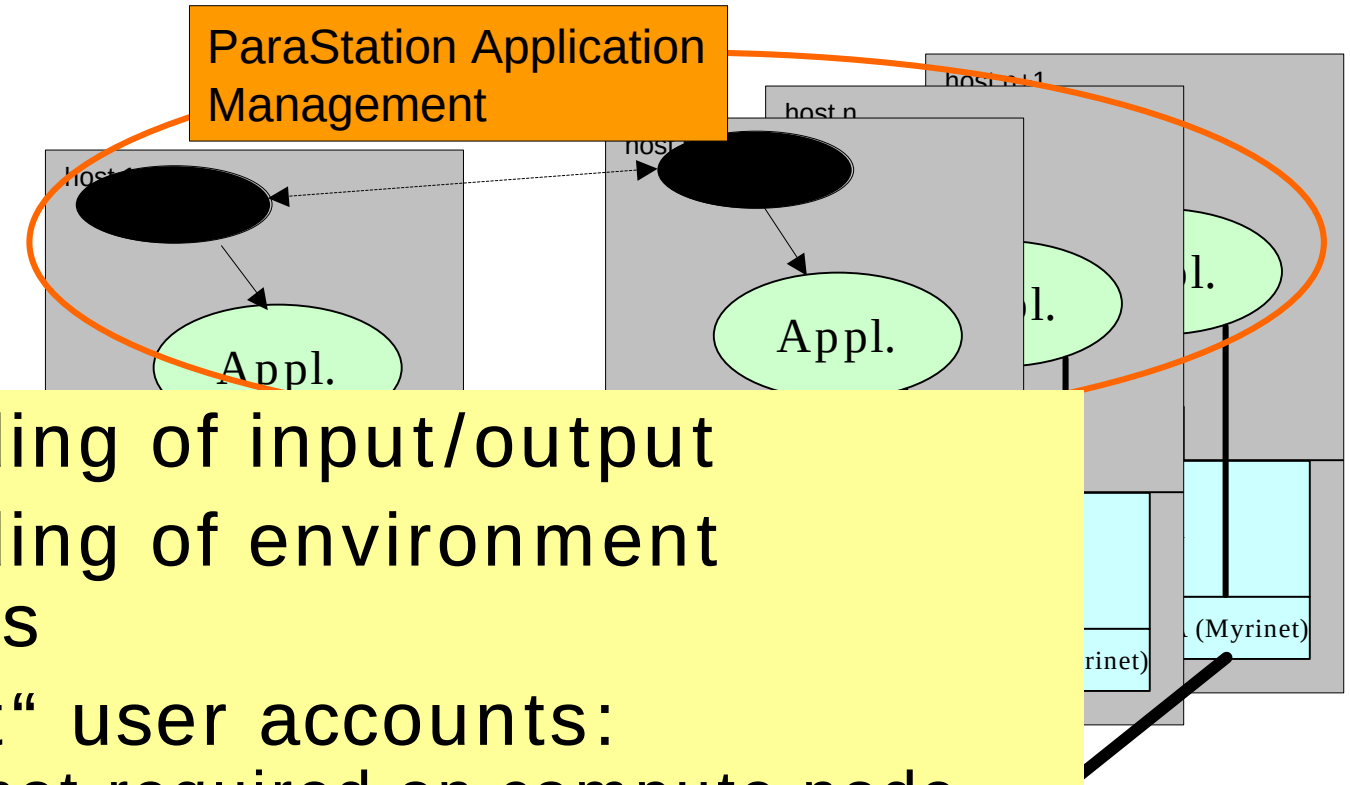
ParaStation: Management



- Load balancing:
  - automatic
  - user defined (host list)
- Criterias for process placement:
  - current load
  - number of processes running
  - node list / host list
  - exclusive
  - other resources (disk, memory)
  - farming

ParTec

ParaStation: Management

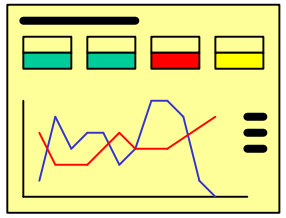


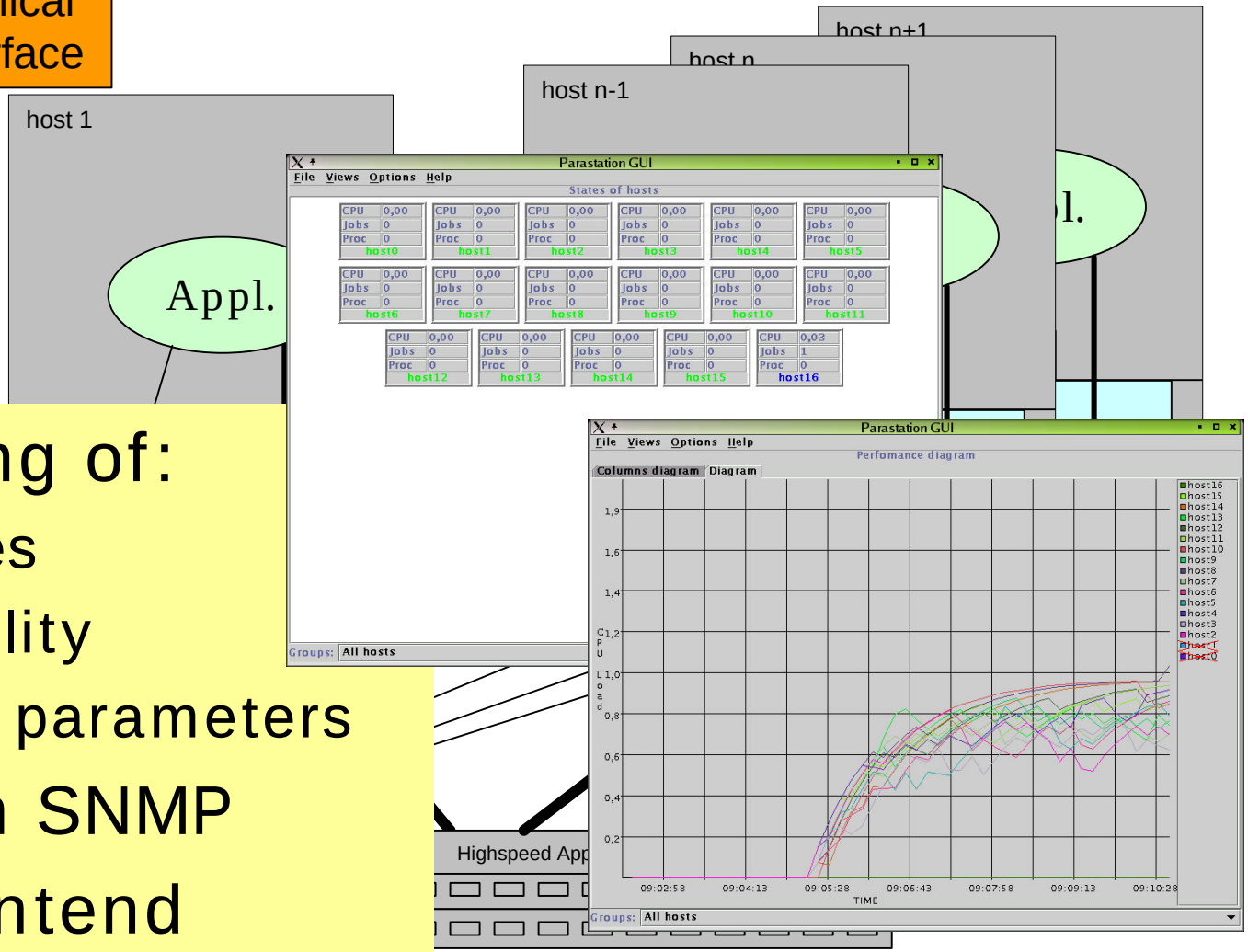
- Forwarding of input/output
- Forwarding of environment variables
- „implicit“ user accounts:
  account not required on compute node, only on frontend
- Launching of jobs can be restricted to dedicated users

ParTec

ParaStation: Graphical User IF

ParaStation Graphical Management Interface

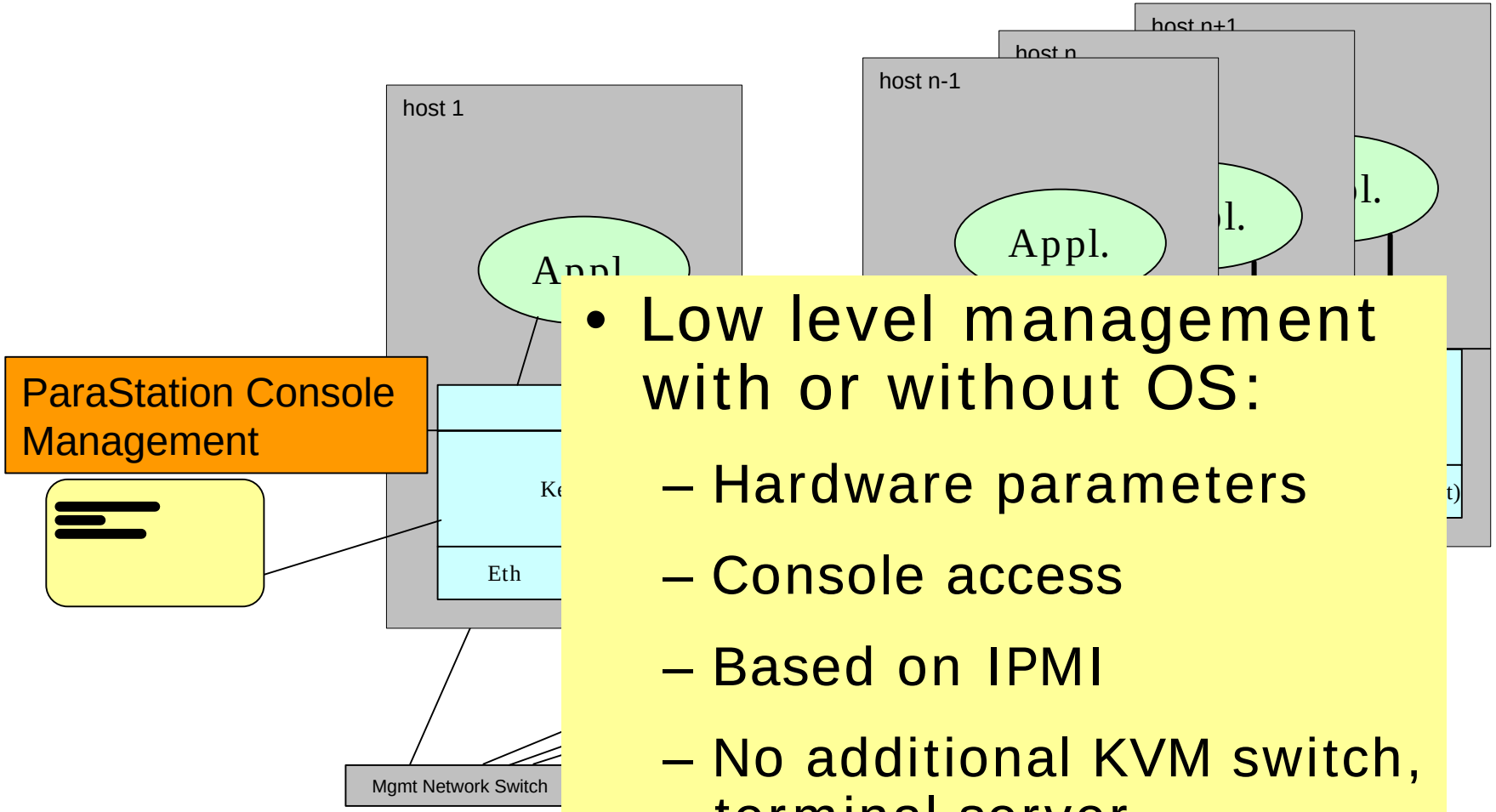




- Monitoring of:
  - Activities
  - Availability
  - System parameters
- Based on SNMP
- Web-Frontend

ParTec

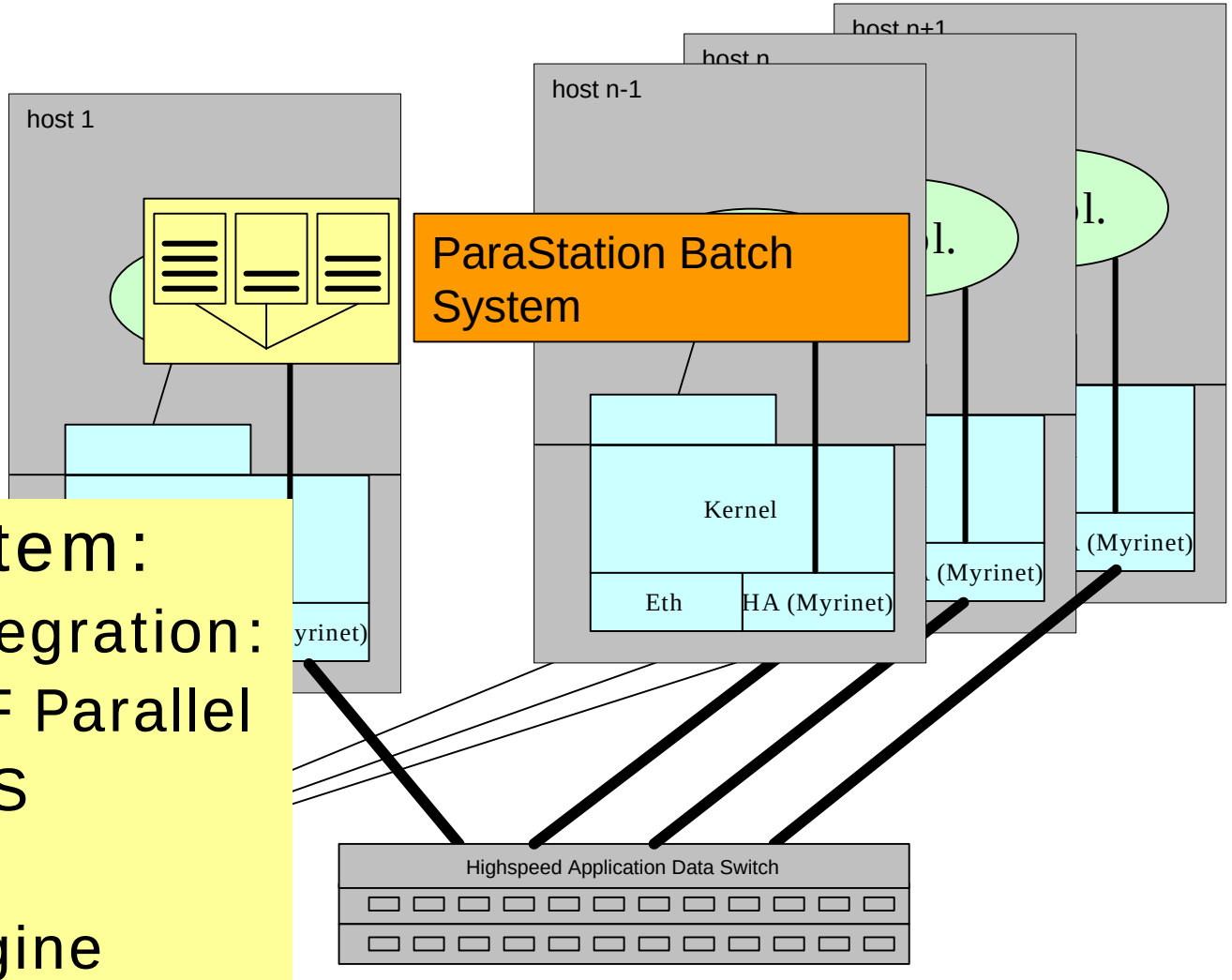
ParaStation: Lowlevel Monitor



- Low level management with or without OS:
  - Hardware parameters
  - Console access
  - Based on IPMI
  - No additional KVM switch, terminal server, ...

ParTec

ParaStation: Batchsystem

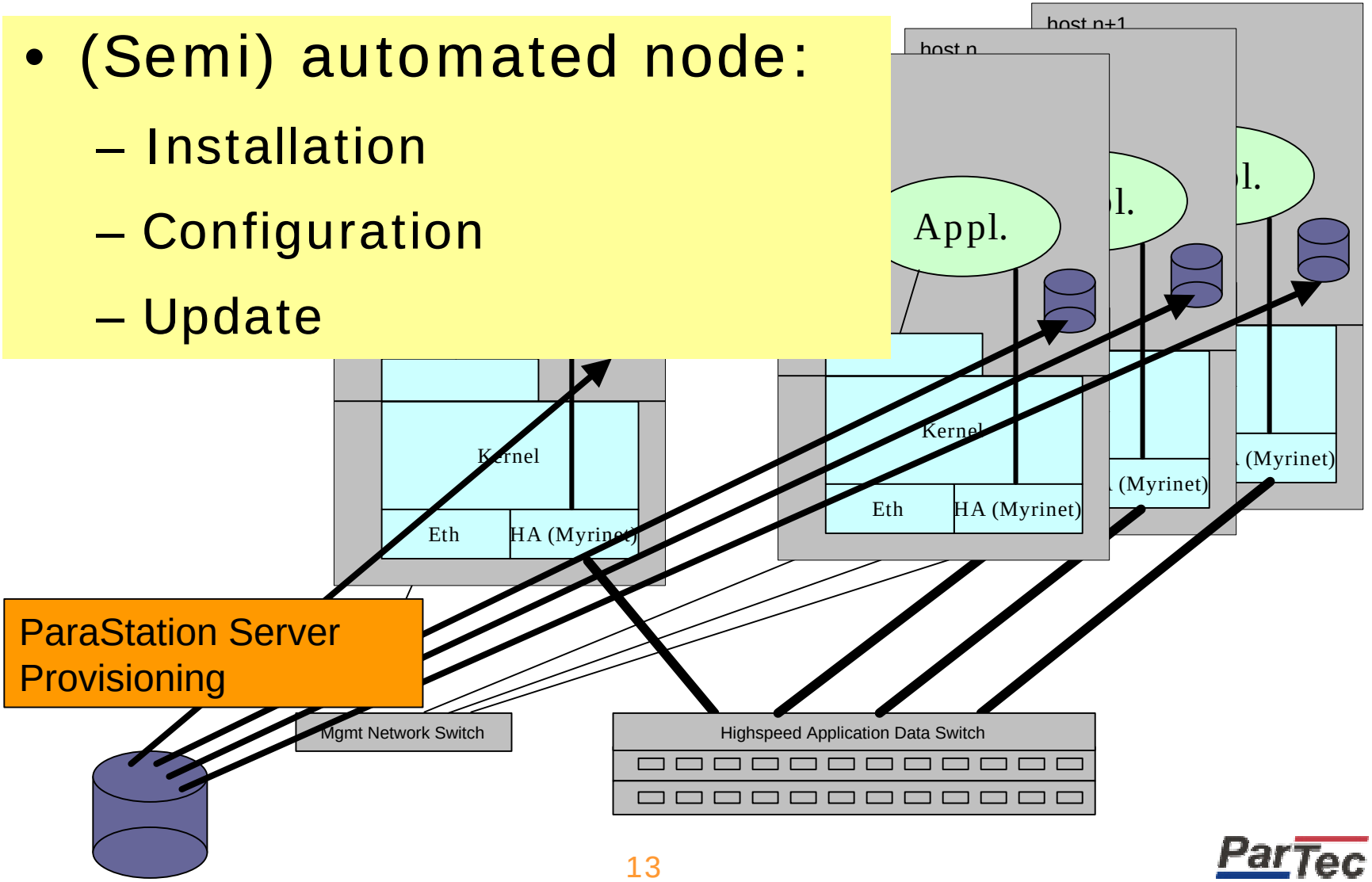


- Batchsystem:
  - Easy integration:
  - LSF, LSF Parallel
  - OpenPBS
  - PBS-Pro
  - Grid Engine

ParTec

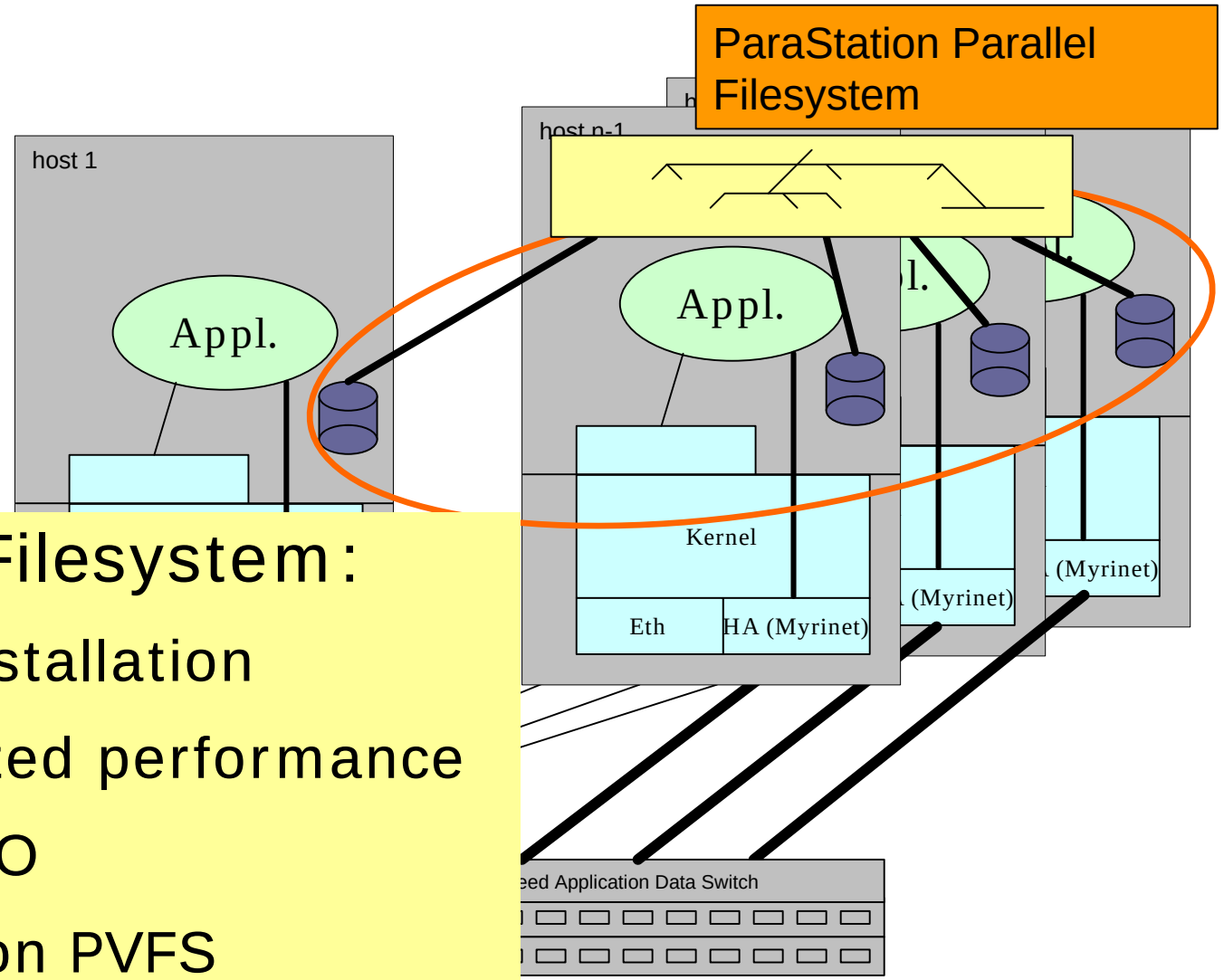
ParaStation: Installation

- (Semi) automated node:
  - Installation
  - Configuration
  - Update



ParTec

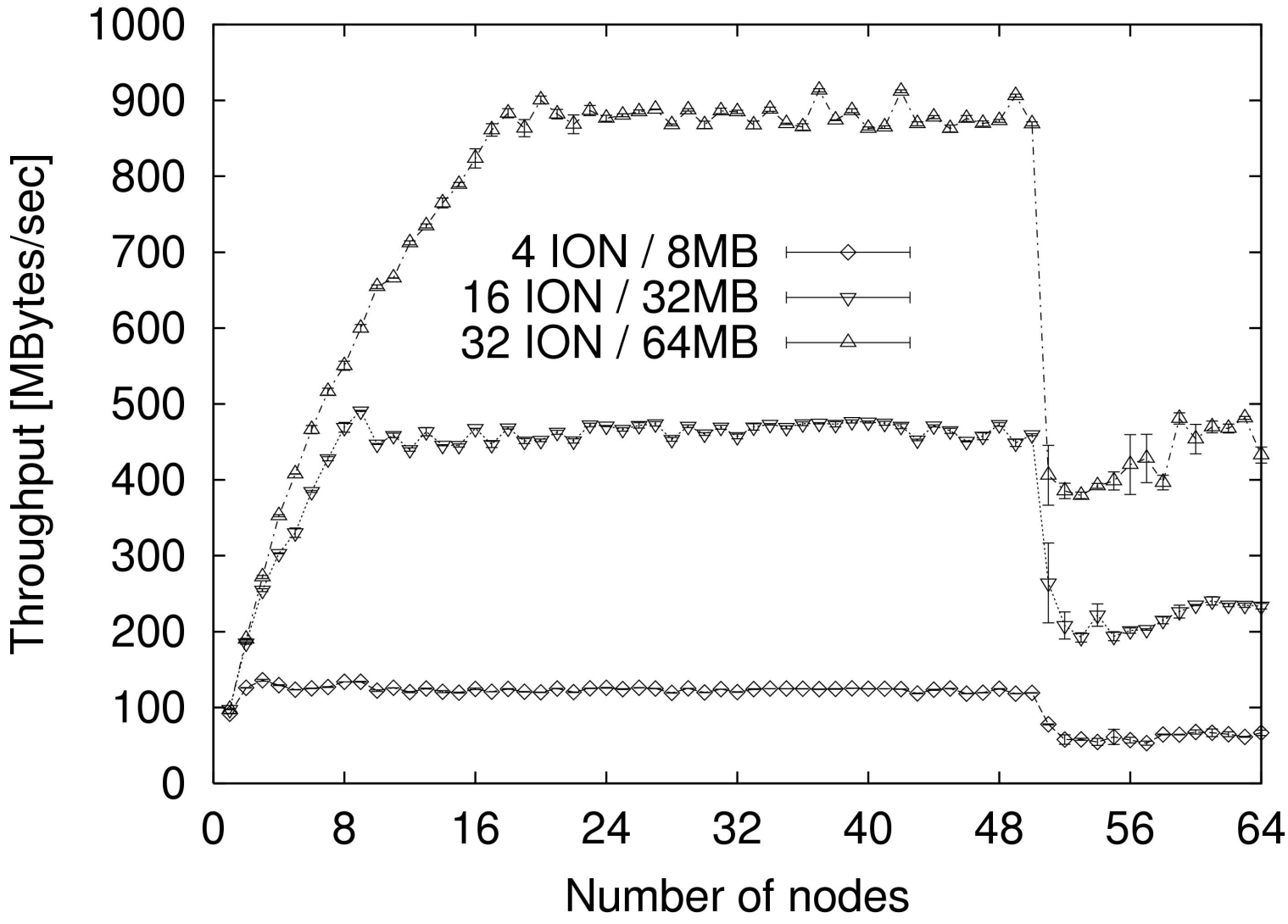
ParaStation: Parallel Filesystem



- Parallel Filesystem:
  - Easy installation
  - Optimized performance
  - ROM-I/O
  - Based on PVFS

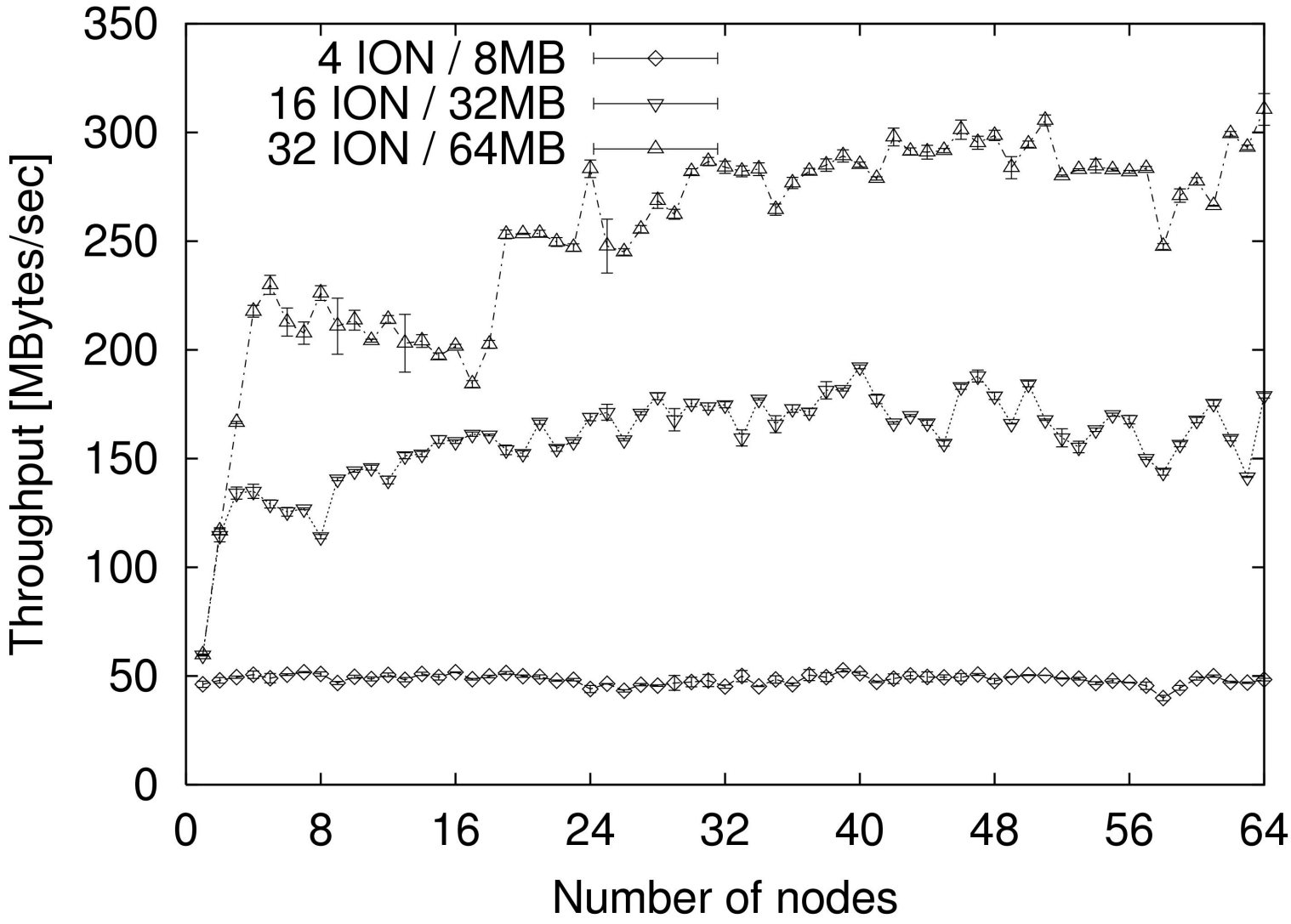
ParTec

PVFS - Write Performance



ParTec

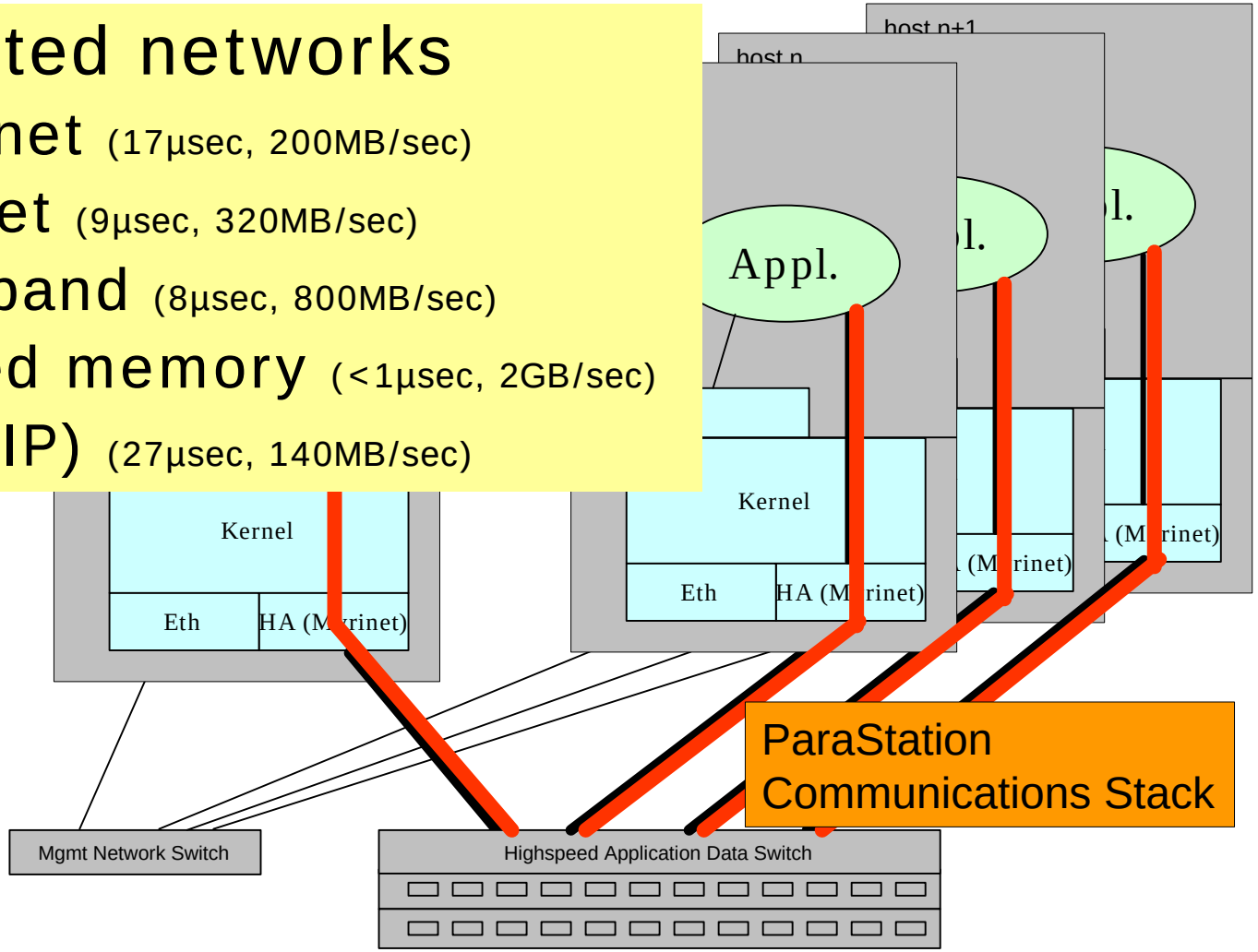
PVFS - Read Performance



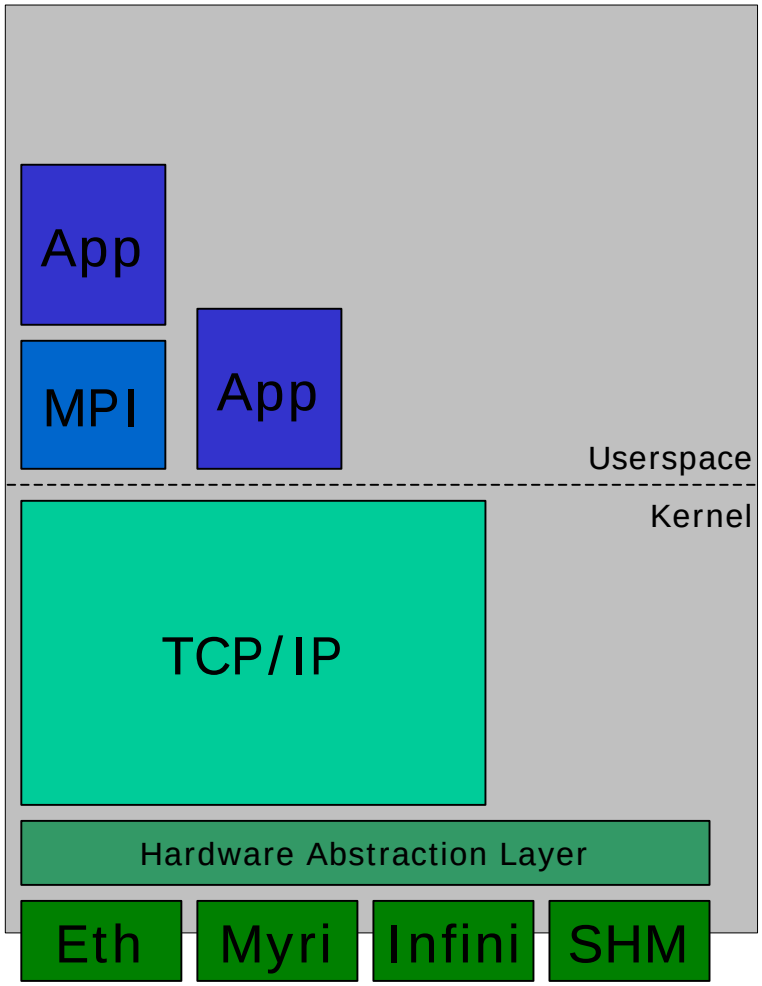
ParTec

ParaStation: Communication

- Supported networks
  - Ethernet (17µsec, 200MB/sec)
  - Myrinet (9µsec, 320MB/sec)
  - Infiniband (8µsec, 800MB/sec)
  - Shared memory (<1µsec, 2GB/sec)
  - (TCP/IP) (27µsec, 140MB/sec)



Kernel Communication



- Transparent to the application
- Multi hardware
- Effective local communication

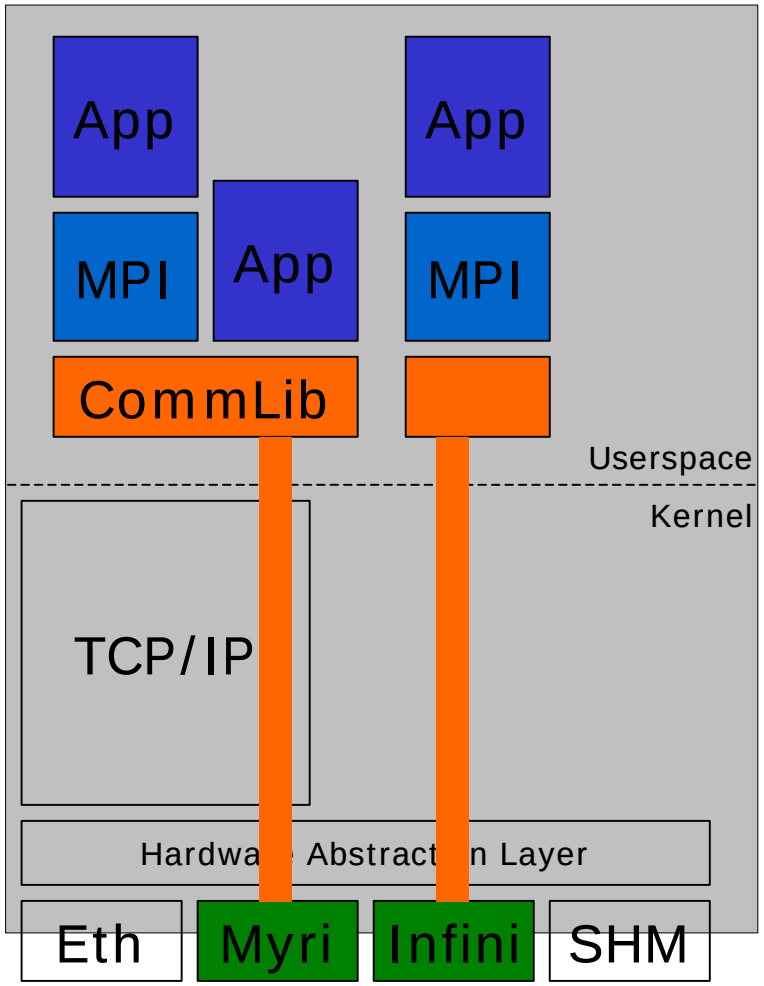
- Protocol overhead
- Expensive switch from Userspace to Kernelspace

ParTec

Userspace Communication

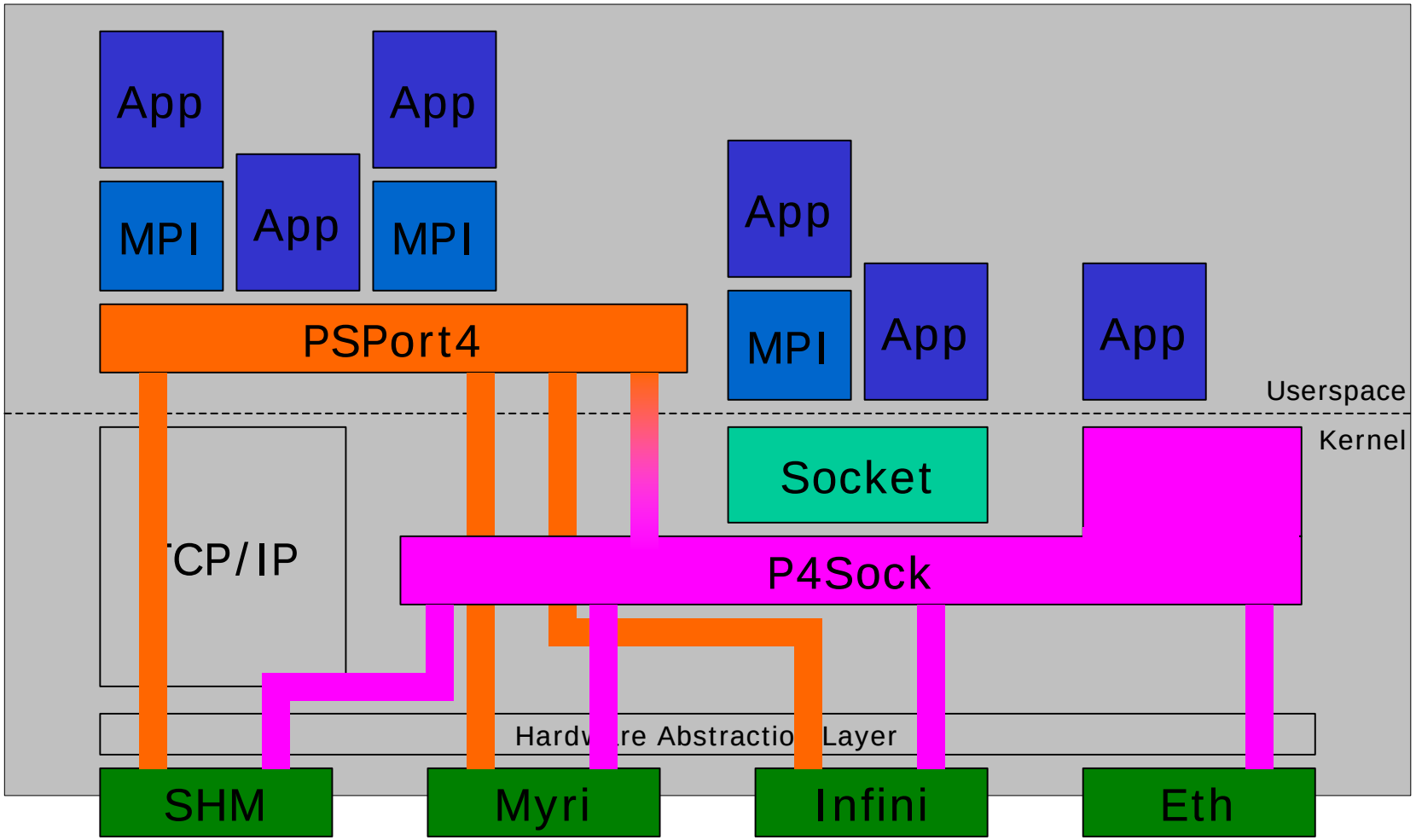
- Slim Protocol
- Direct hardware access
- Fast communication

- Multi hardware support harder to implement
- Security problem possible



ParTec

ParaStation: Architecture

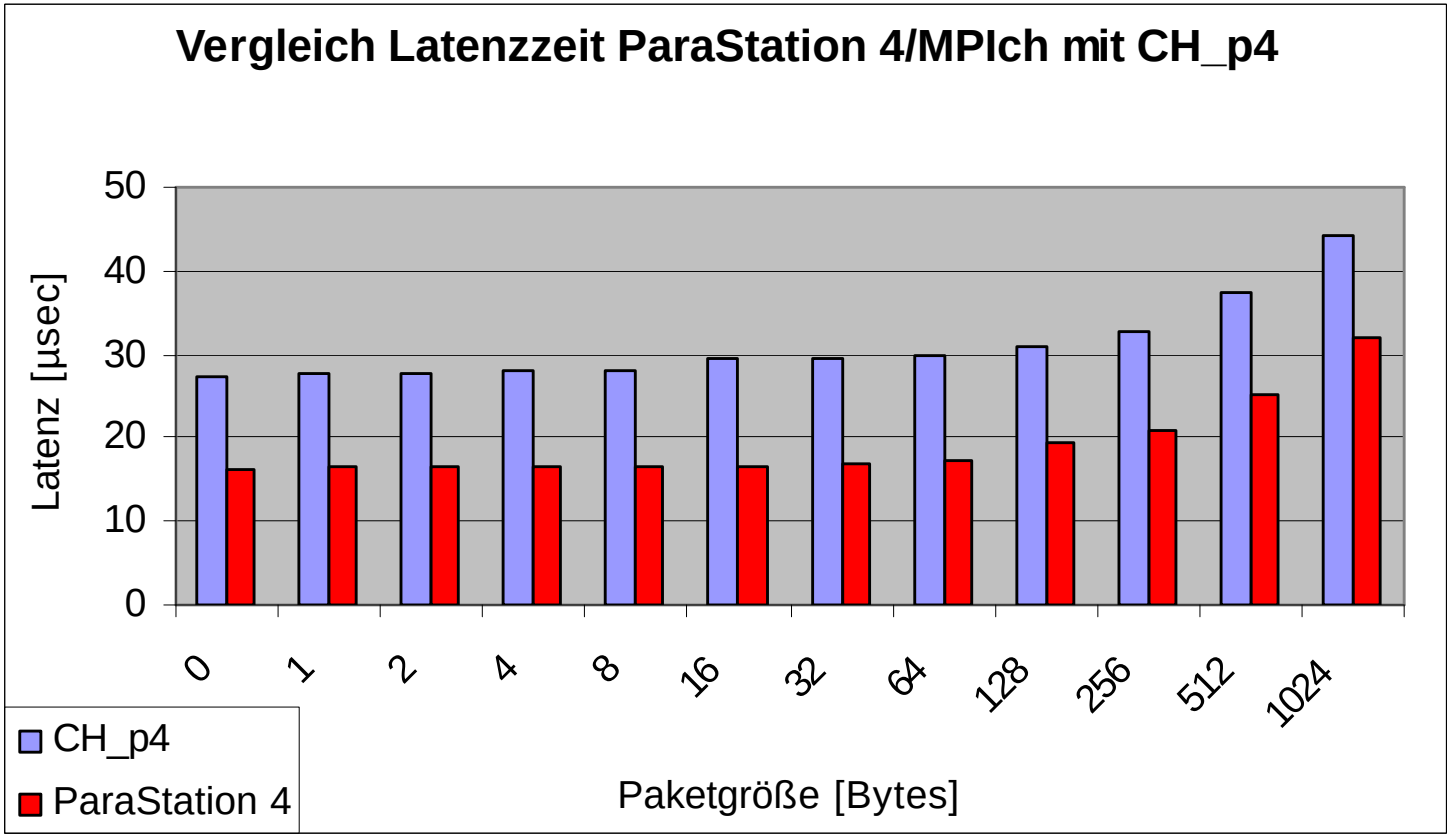


ParTec

ParaStation: Communication

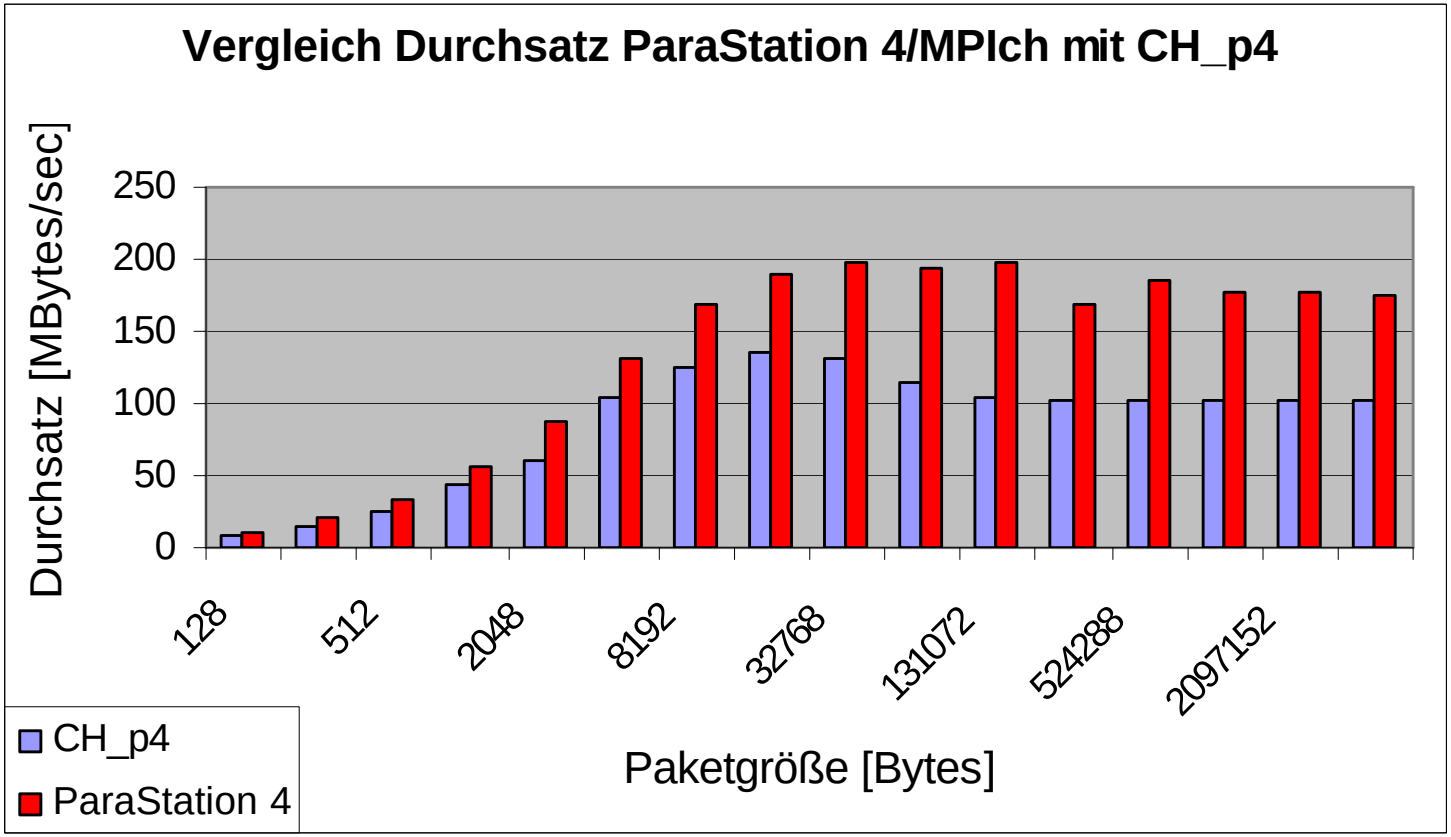
- Communication Benchmark:
  - Pallas MPI Benchmark PMB2.2
- Test System:
  - Dual XEON System 2.6 GHz
  - 2 GB Memory
  - SuperMicro P4DPE-G2 (E7500)
  - Intel E1000 (82540) on board
  - Broadcom NetXtrem BCM5701
  - All numbers without switch

ParaStation: Communication



ParTec

ParaStation: Communication



ParTec

ParaStation: Summary

- Modular, high speed, robust, easy to use compute cluster environment
- Linux
  - all major distributions
  - all kernel versions
- IA32, (IA64 soon)
- Fully supported (support@par-tec.com)

Contact

Parallel Technologies

http://www.par-tec.com

info@par-tec.com