Operating two InfiniBand grid clusters over 28 km distance

Sabine Richling, Steffen Hau, Heinz Kredel, Hans-Günther Kruse

IT-Center University of Heidelberg, Germany
IT-Center University of Mannheim, Germany

3PGCIC-2010, Fukuoka, 4. November 2010

Motivation

Circumstances in Baden-Württemberg (BW)

- Increasing demand for high-performance computing capacities from scientific communities
- Demands are not high enough to qualify for the top German HPC centers in Jülich, Munich and Stuttgart

⇒ Grid infrastructure concept for the Universities in Baden-Württemberg

Motivation

Special Circumstances in Heidelberg/Mannheim

- Both IT-centers have a long record of cooperations
- Both IT-centers are connected by a 10 Gbit dark fibre connection of 28 km (two color lines already used for backup and other services)

$\Rightarrow$ Connection of the clusters in Heidelberg and Mannheim to ease operation and to enhance utilization

Outline

bwGRiD cooperation

D-Grid

- German Grid Initiative (www.d-grid.de)
- Start: September 2005
- Aim: Development and establishment of a reliable and sustainable Grid infrastructure for e-science in Germany
- Funded by the Federal Ministry of Education and Research (BMBF) with $\sim$ 50 Million Euro

bwGRiD

- Community project of the Universities of BW (www.bw-grid.de)
- Compute clusters at 8 locations: Stuttgart, Ulm (Konstanz), Karlsruhe, Tübingen, Freiburg, Mannheim/Heidelberg, Esslingen
- Central storage unit in Karlsruhe
- Distributed system with local administration
- Access for all D-Grid virtual organizations via at least one middleware supported by D-Grid

bwGRiD

bwGRiD – Objectives

- Verifying the functionality and the benefit of Grid concepts for the HPC community in BW
- Managing organisational and security problems
- Development of new cluster and Grid applications
- Solving license difficulties
- Enabling the computing centers to specialize

bwGRiD – Access Possibilities

Access with local university accounts (via `ssh`):
→ Access to a local bwGRiD cluster only

Access with Grid Certificate and VO membership using a Grid middleware (e.g. Globus Toolkit: `gsissh`, `GridFTP` or Webservices):
→ Access to all bwGRiD resources

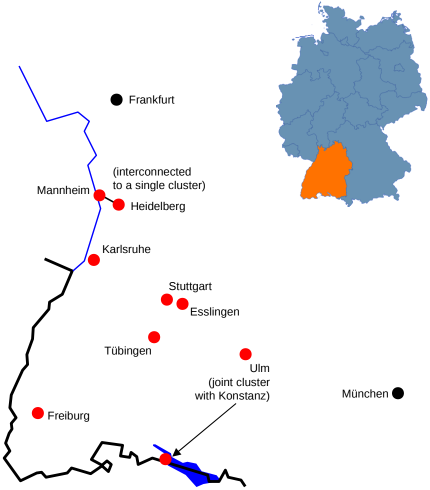
bwGRiD – Resources

Compute cluster:

- Mannheim/Heidelberg: 280 nodes
  Direct Interconnection
- Karlsruhe: 140 nodes
- Stuttgart: 420 nodes
- Tübingen: 140 nodes
- Ulm (Konstanz): 280 nodes
  Hardware in Ulm
- Freiburg: 140 nodes
- Esslingen: 180 nodes
  more recent Hardware

Central storage:

- Karlsruhe:
  128 TB (with Backup)
  256 TB (without Backup)



bwGRiD – Software

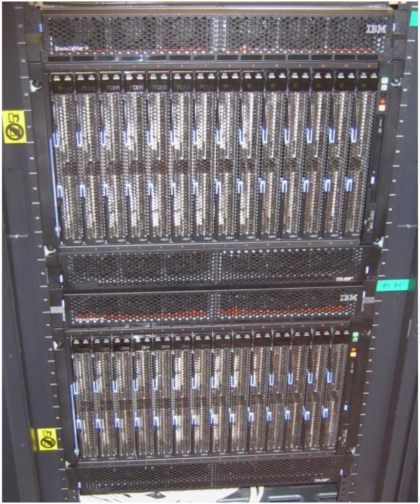
- Common Software:
  - Scientific Linux, Torque/Moab batch system, GNU and Intel compiler suite
  - Central repository for software modules (MPI versions, mathematical libraries, various free software, application software from each bwGRiD site)
- Application areas of bwGRiD sites:
  - Freiburg: System Technology, Fluid Mechanics
  - Karlsruhe: Engineering, Compiler & Tools
  - Heidelberg: Mathematics, Neuroscience
  - Mannheim: Business Administration, Economics, Computer Algebra
  - Stuttgart: Automotive simulations, Particle simulations
  - Tübingen: Astrophysics, Bioinformatics
  - Ulm: Chemistry, Molecular Dynamics
  - Konstanz: Biochemistry, Theoretical Physics

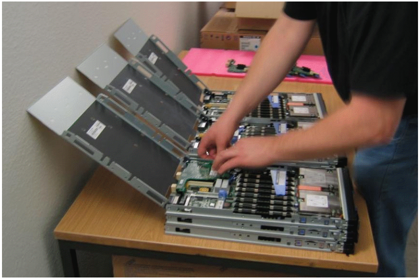
Interconnection of two bwGRiD clusters

Hardware before Interconnection

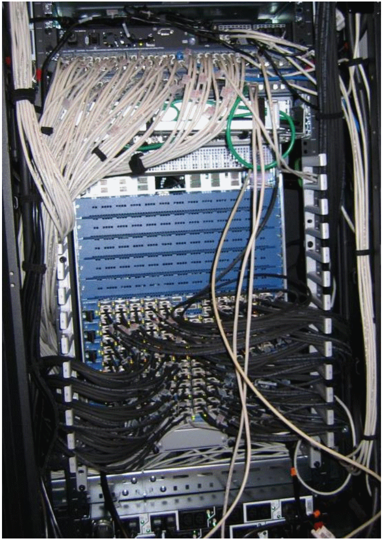
- 10 Blade-Center in Heidelberg and 10 Blade-Center in Mannheim
- Each Blade-Center contains 14 IBM HS21 XM Blades
- Each Blade contains
  - 2 Intel Xeon CPUs, 2.8 GHz (each CPU with 4 Cores)
  - 16 GB Memory
  - 140 GB Hard Drive (since January 2009)
  - Gigabit-Ethernet (1 Gbit)
  - Infiniband Network (20 Gbit)
- $\Rightarrow$ 1120 Cores in Heidelberg and 1120 Cores in Mannheim

Hardware – Bladecenter



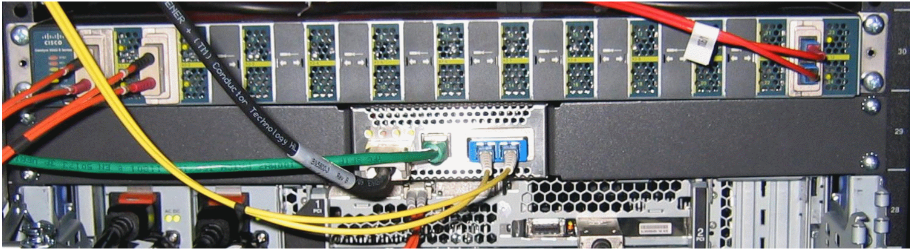


Hardware – Infiniband



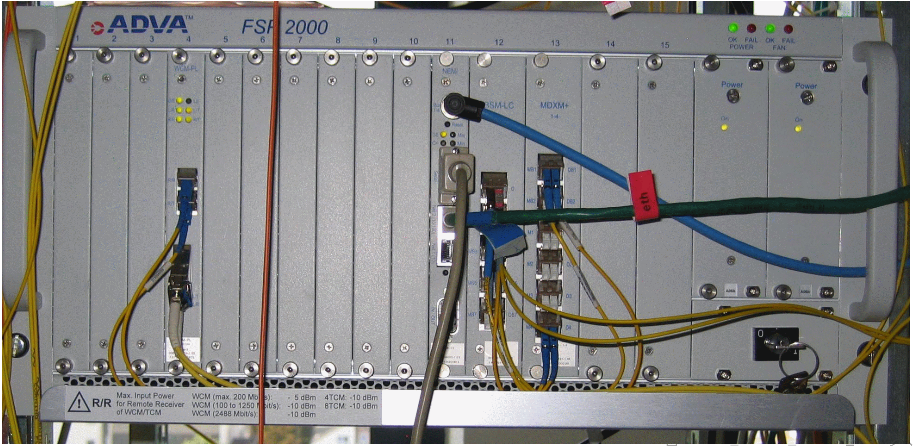
Interconnection of the bwGRiD clusters

- Proposal in 2008
- Acquisition and Assembly until May 2009
- Running since July 2009
- Infiniband over Ethernet over fibre optics: Longbow adaptor from Obsidian
  - InfiniBand connector (black cable)
  - fibre optic connector (yellow cable)



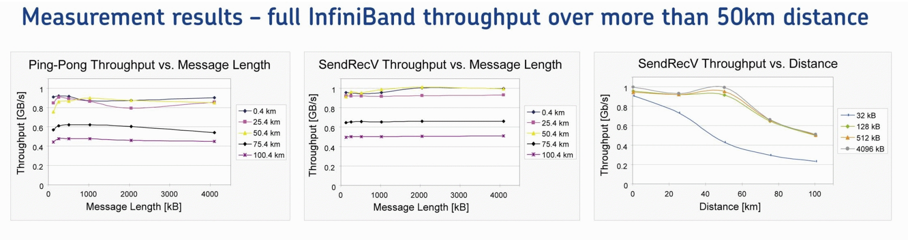
Interconnection of the bwGRiD clusters

- ADVA component: Transformation of white light from Longbow to one color light for the dark fibre connection between IT centers



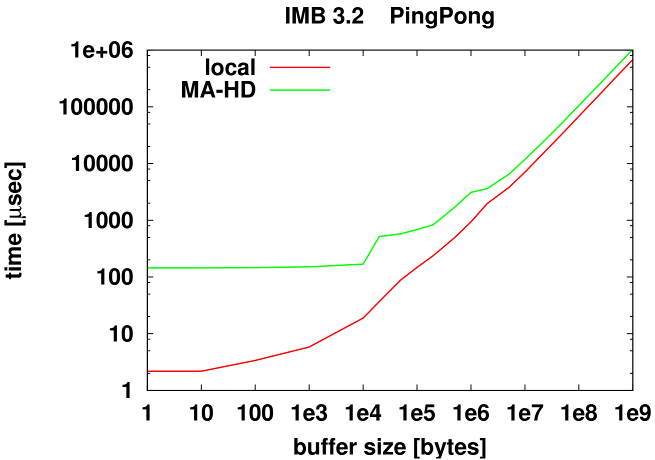
MPI Performance – Prospects

- Measurements for different distances (HLRS, Stuttgart, Germany)
- Bandwidth 900-1000 MB/sec for up to 50-60 km
- Latency is not published



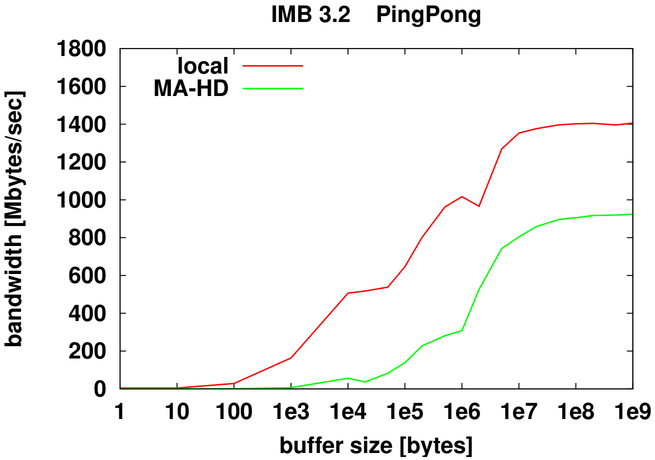
MPI Performance – Latency

Local: $\sim 2\ \mu$sec
Interconnection: $145\ \mu$sec



MPI Performance – Bandwidth

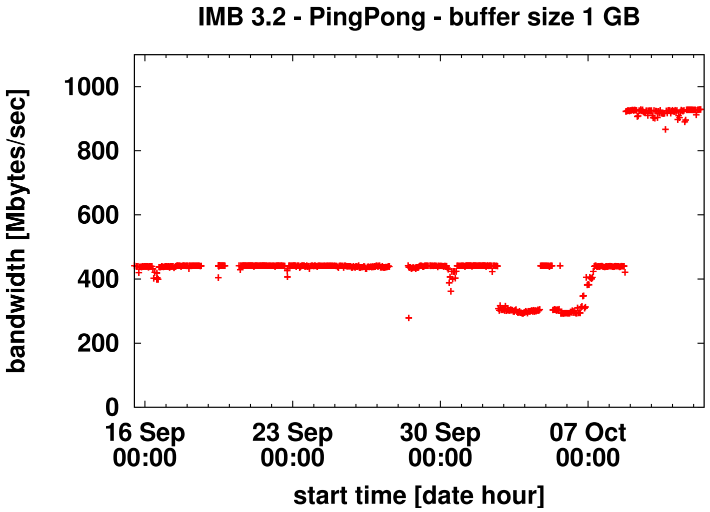
Local: 1400 MB/sec
Interconnection: 930 MB/sec



Experiences with Interconnection Network

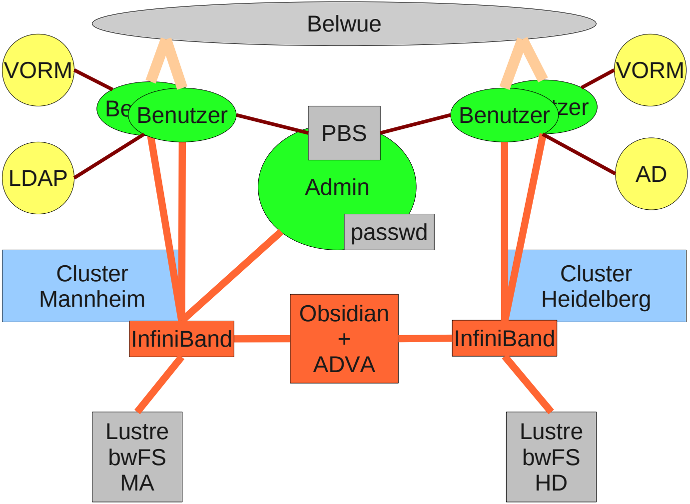
- Cable distance MA-HD is 28 km (18 km linear distance in air)
  $\Rightarrow$ Light needs 143 μsec for this distance
- Latency is high:
  145 μsec $=$ Light transit time $+$ 2 μsec local latency
- Bandwidth is as expected: about 930 MB/sec
  Local bandwidth 1200-1400 MB/sec
- Obsidian needs a license for 40 km
  - Obsidian has buffers for larger distances
  - Activation of buffers with license
  - License for 10 km is not sufficient

MPI Bandwidth – Influence of the Obsidian License



Cluster operation

bwGRiD Cluster Mannheim/Heidelberg



bwGRiD Cluster Mannheim/Heidelberg – Overview

- Two clusters (blue boxes) are connected by InfiniBand (orange lines)
- "Obsidian and ADVA" (orange box) represents the 28 km fibre connection
- bwGRiD storage systems (grey boxes) are also connected by Infiniband
- Access nodes ("Benutzer") are connected with 10 GBit (light orange lines) to the outside Internet "Belwue" (BW science net)
    - Access with local accounts from Mannheim ("LDAP")
    - Access with local accounts from Heidelberg ("AD")
    - Access with Grid certificates ("VORM")
- Ethernet connection between all components is not shown

Node Management

- Compute nodes are booted via PXE and use NFS read-only export as root file system
- Administration server provides
  - DHCP service for the nodes (MAC-to-IP address configuration file)
  - NFS export for root file system
  - NFS directory for software packages accessible via module utilities
  - queuing and scheduling system
- Node administration (power on/off, execute commands, BIOS update, etc.) with
  - adjusted shell scripts originally developed by HLRS
  - IBM management module (command line interface and Web-GUI)

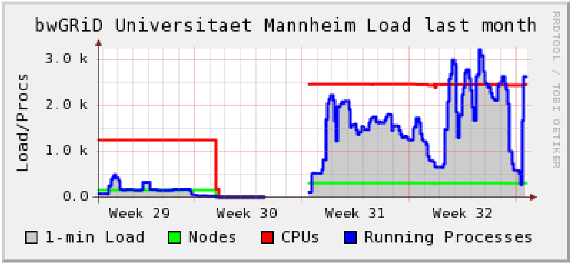
User Management

- Users should have exclusive access to compute nodes
  - user names and user-ids must be unique
  - replacing `passwd` with reduced `passwd` proofed unreliable
  - better is a direct connection to PBS for user authorization via PAM module
- Authentication at the access nodes
  - directly against directory services: LDAP (MA) and AD (HD)
  - or with D-Grid certificate
- Combining information from directory services from both universities
  - Prefix "ma", "hd" or "mh" for group names
  - Adding offsets to group-ids
  - Adding offsets to user-ids
  - Activated user names from MA and HD must be different
- Activation process
  - Adding a special attribute for the user in the directory service (for authentication)
  - Updating the user database of the cluster (for authorization)

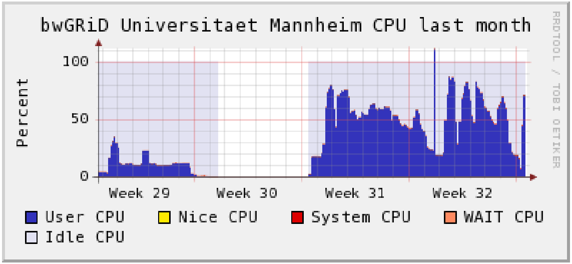
Job Management

- Interconnection (high latency, limited bandwidth) provides
  - enough bandwidth for I/O operations
  - not sufficient for all kinds of MPI jobs
- Jobs only run on nodes located either in HD or in MA (realized with attributes provides by the queuing system)
- Before interconnection
  - In Mannheim: mostly single node jobs → free nodes
  - In Heidelberg: many MPI jobs → long waiting times
- With interconnection better resource utilization (see Ganglia report)

Monitoring Report during activation of the interconnection



Number of processes



Percent CPU Usage

Performance modeling

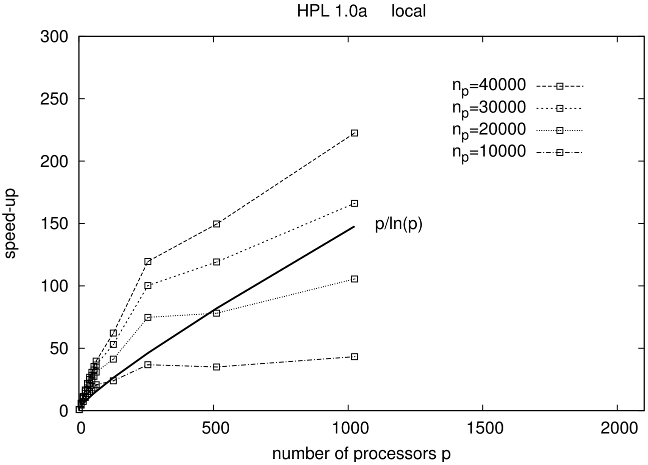
MPI Jobs running across the interconnection

- How does the interconnection influence the performance?
- How much bandwidth would be necessary to the improve the performance?
- How much would such an upgrade cost?

Performance modeling

- Numerical model
  - High-Performance Linpack (HPL) benchmark
  - OpenMPI
  - Intel MKL
- Model variants
  - Calculations on a single cluster with up to 1025 CPU cores
  - Calculations on the coupled cluster with up to 2048 CPU cores symmetrically distributed
- Analytical model for the speed-up to analyze the characteristics of the interconnection
  - high latency of 145 μsec
  - limited bandwidth of 930 MB/sec

Results for a single cluster

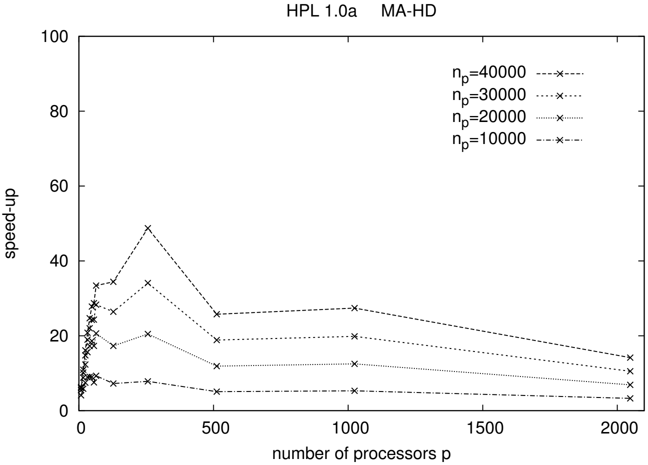


n_p load parameter (matrix size for HPL)

ideal speed-up for perfect parallel programs
$S_{\mathrm{ideal}}(p) = p$

speed-up for a simple model "all CPU configurations have equal probability"
$S_{\mathrm{simple}}(p) = p/\ln p$

Results for coupled cluster

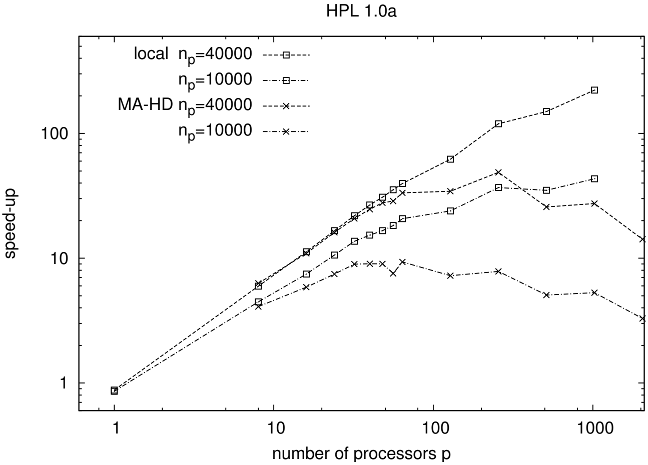


n_p load parameter (matrix size for HPL)

for $p > 256$ reduced speed-up by a factor of ~ 4 compared to single cluster

for $p > 500$ constant (decreasing) speed-up

Direct comparison of the two cases



n_p load parameter (matrix size for HPL)

for $p < 50$
speed-up for coupled cluster is acceptable, applications could run across interconnection effectively (in the case of exclusive usage)

Performance modeling

Following a performance model developed by Kruse (2009):
$t_{\mathrm{c}}(p)$: communication time
$t_{\mathrm{B}}(1)$: processing time for $p = 1$

Speed-up $$S_{\mathrm{c}}(p) \leq \frac{p}{\ln p + \frac{t_{\mathrm{c}}(p)}{t_{\mathrm{B}}(1)}}$$

For $t_{\mathrm{c}}(p) = 0$, we receive the result of the simple model:

$$S_{\mathrm{simple}}(p) = p/\ln p$$

Performance model for the high latency

Modeling $t_\mathrm{c}(p)$ as a function of the typical communication time between 2 processes $t_c^{(2)}$ an the communication topology $c(p)$:

$$t_\mathrm{c}(p) = t_c^{(2)} c(p)$$

Defining a rate $r = t_\mathrm{c}^{(2)}/t_\mathrm{A}$ between $t_\mathrm{c}^{(2)}$ and the computation time for a typical instruction $t_\mathrm{A} = t_\mathrm{B}(1)/n$:

Speed-up $$S_\mathrm{c}(p) \leq \frac{p}{\ln p + \frac{r}{n} c(p)}$$

Analysis for HPL ($n = \frac{2}{3} n_p^3$):

- for $n_p = 1000$: $\sim p/\ln p$ for small p, decrease for $p \geq 30$
- for $n_p = 10\,000$: $\sim p/\ln p$ for $p \leq 10\,000$, decrease for $c(p) > 10^6$

Analysis does not explain the numerical results.
Decrease of speed-up already for smaller p.

Performance model including a limited bandwidth

Modeling the interconnection as a shared medium for the communication of p processes with a given bandwidth B and average message length $< m >$:

$t_c^{(2)} = t_{\mathrm{L}} + \frac{\langle m \rangle}{(B/p)}$

$r(p) = \frac{t_{\mathrm{L}}}{t_{\mathrm{A}}} + \frac{\langle m \rangle}{t_{\mathrm{A}} B} p$

With the measured bandwidth $B = 1.5 \cdot 10^6$ and $\langle m \rangle = 10^6$:

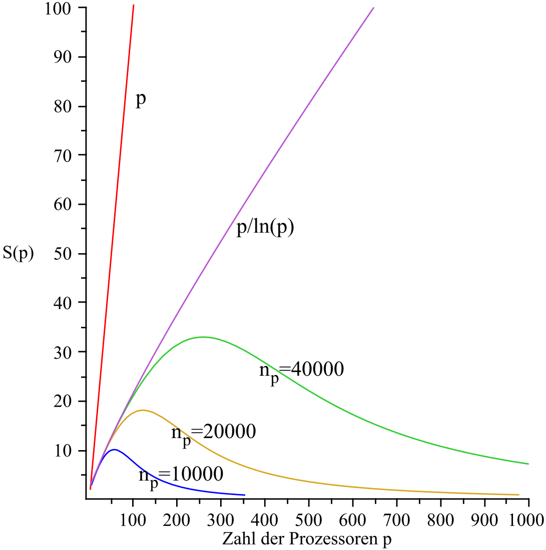
Speed-up $$S_{\mathrm{c}}(p) \leq \frac{p}{\ln p + \frac{3}{4}\left(\frac{100}{n_p}\right)^3 (1 + 4p) c(p)}$$

With assumption $c(p) = \frac{1}{2} p^2$:

- for $n_p = 10\,000$: $\sim p/\ln p$, decrease for $p \geq 50$
- for $n_p = 40\,000$: $\sim p/\ln p$, decrease for $p \geq 250$

Speed-up reproduces the measurements.

Speed-up of the model including limited bandwidth



n_p load parameter (matrix size for HPL)

⇒ limited bandwidth is the performance bottleneck for shared connection between the clusters

⇒ Doubling the bandwidth: 25 % improvement for $n_p = 40\,000$

⇒ 100 % improvement with a ten-fold bandwidth (in the case of exclusive usage)

Summary and Conclusions

InfiniBand connection of two compute clusters

- Network (Obsidian, ADVA and Infiniband switches) is stable and reliable
- Latency of 145 μsec is very high
- Bandwidth of 930 MB/sec is as expected
- Jobs are limited to one site, because MPI jobs would be slow (Interconnection is a "shared medium")
- Performance model predicts the cost for an improvement of the interconnection
- Bandwidth sufficient for cluster administration and file I/O on Lustre file systems
- Interconnection is useful and stable for a "Single System Cluster" administration
- Better load balance at both sites due to common PBS
- Solving organizational issues between two universities is a great challenge