Xen - Virtualisation, Performance, Grid-Computing?

Marcus Hardt
Forschungszentrum Karlsruhe GmbH

www.eu-egee.org

Information Society

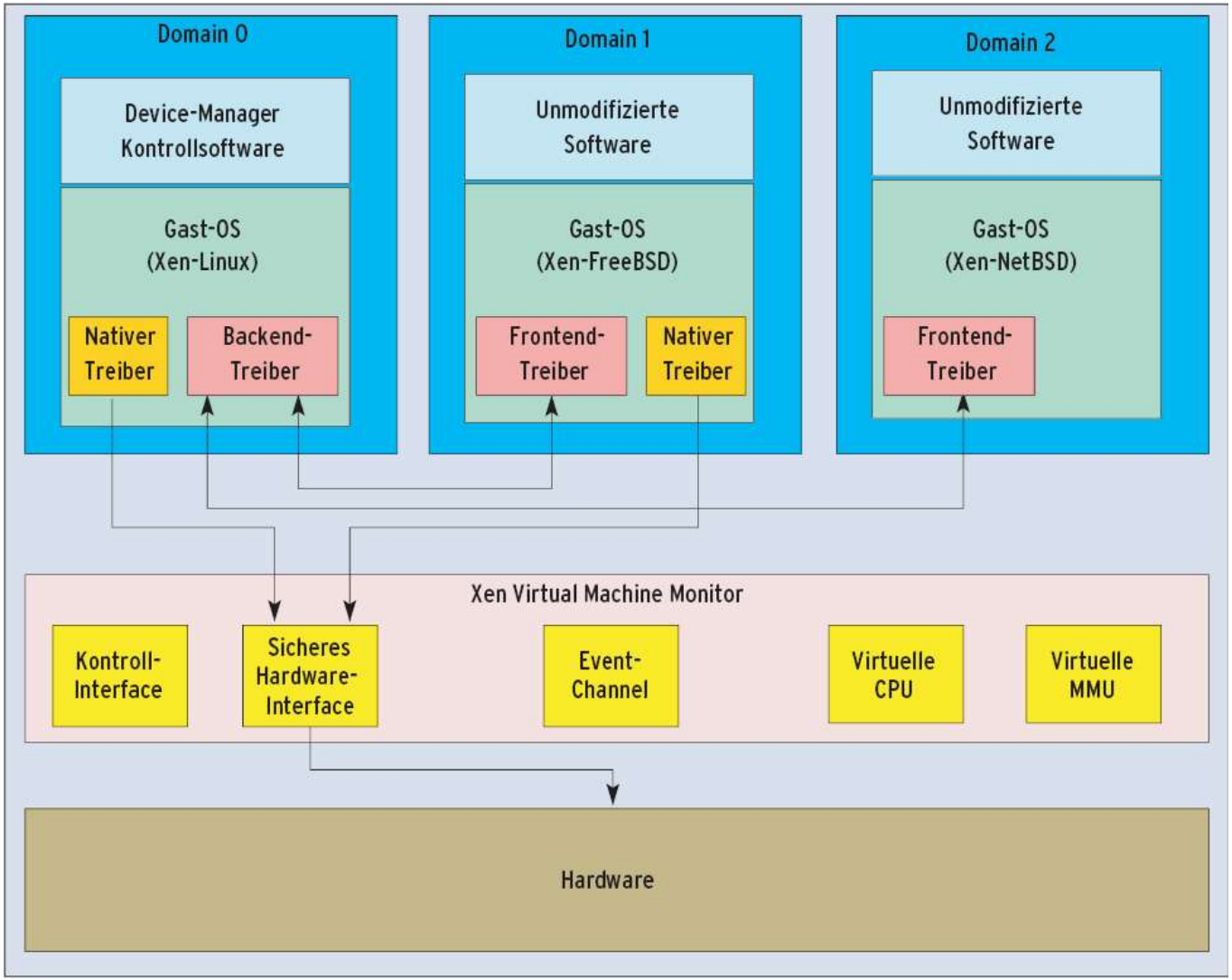
Overview

- **Xen**
  - What is Xen
  - Configuration / Features
- **Xen Performance**
  - What to measure
  - Results
- **Xen use in training courses**
  - Installation
  - Experience

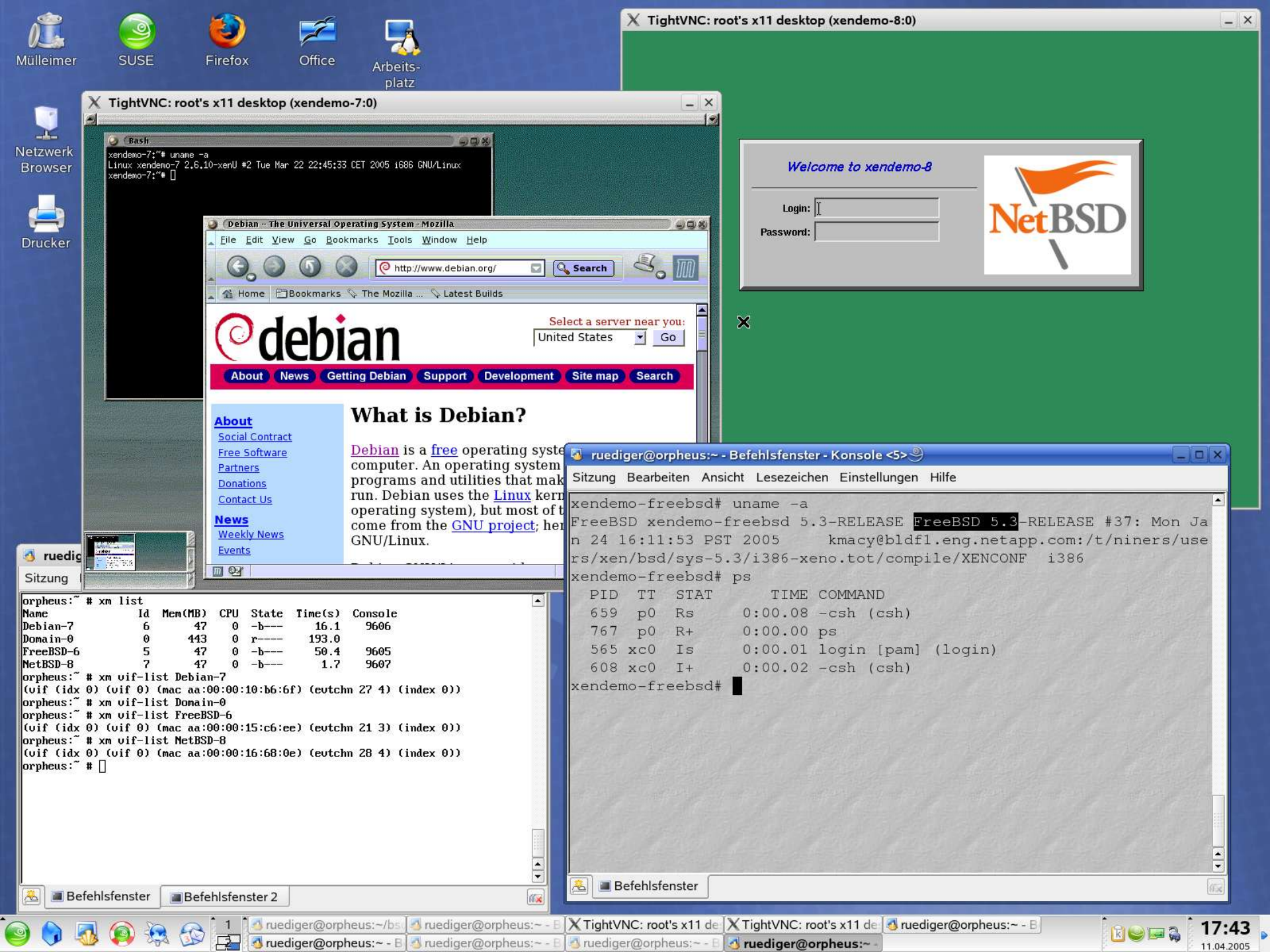
Xen: History

- Approx. 2 years old
- Started by the *Systems Research Group* of the University of Cambridge, U
- Originally part of the Xenoserver project
  - Idea: A distributed network of OS environments tailored to the user's needs
- **Xen is thus closely related to the ideas of Grid Computing !**
- Now available in Version 2.07
- Outlook: Native execution of arbitrary Intel-based OS feasible using hardware virtualisation features (Intel Vanderpool)
- Ports to 64 bit platforms underway (with the help of AMD, Intel, ...)

Xen: Paravirtualisation



- Priviledged calls are done through dedicated interface in domain 0
- Advantage: Very high per-formance (low overhead, very little emulation necessary)
- Disadvantage: Guest-OS must be ported to Xen (but not the applications !)
- But: very minor adaptations, in the range of *O*(3000 LOC)


Mülleimer
SUSE
Firefox
Office
Arbeits-platz
Netzwerk Browser
Drucker
TightVNC: root's x11 desktop (xendemo-8:0)
TightVNC: root's x11 desktop (xendemo-7:0)
Bash
xendemo-7:~# uname -a
Linux xendemo-7 2.6.10-xenU #2 Tue Mar 22 22:45:33 CET 2005 i686 GNU/Linux
xendemo-7:~#
Debian -- The Universal Operating System - Mozilla
File Edit View Go Bookmarks Tools Window Help
http://www.debian.org/
Search
Home Bookmarks The Mozilla ... Latest Builds
debian
Select a server near you:
United States
Go
About News Getting Debian Support Development Site map Search
About
Social Contract
Free Software
Partners
Donations
Contact Us
News
Weekly News
Events
What is Debian?
Debian is a free operating system
computer. An operating system
programs and utilities that make
run. Debian uses the Linux kernel
operating system), but most of the
come from the GNU project; her
GNU/Linux.
Welcome to xendemo-8
Login:
Password:
NetBSD
ruedig
Sitzung
orpheus:~ # xm list
Name Id Mem(MB) CPU State Time(s) Console
Debian-7 6 47 0 -b--- 16.1 9606
Domain-0 0 443 0 r----- 193.0
FreeBSD-6 5 47 0 -b--- 50.4 9605
NetBSD-8 7 47 0 -b--- 1.7 9607
orpheus:~ # xm vif-list Debian-7
(vif (idx 0) (vif 0) (mac aa:00:00:10:b6:6f) (evtchn 27 4) (index 0))
orpheus:~ # xm vif-list Domain-0
orpheus:~ # xm vif-list FreeBSD-6
(vif (idx 0) (vif 0) (mac aa:00:00:15:c6:ee) (evtchn 21 3) (index 0))
orpheus:~ # xm vif-list NetBSD-8
(vif (idx 0) (vif 0) (mac aa:00:00:16:68:0e) (evtchn 28 4) (index 0))
orpheus:~ #
Befehlsfenster
Befehlsfenster 2
ruediger@orpheus:~ - Befehlsfenster - Konsole <5>
Sitzung Bearbeiten Ansicht Lesezeichen Einstellungen Hilfe
xendemo-freebsd# uname -a
FreeBSD xendemo-freebsd 5.3-RELEASE FreeBSD 5.3-RELEASE #37: Mon Jan 24 16:11:53 PST 2005 kmacy@bldf1.eng.netapp.com:/t/niners/users/xen/bsd/sys-5.3/i386-xeno.tot/compile/XENCONF i386
xendemo-freebsd# ps
PID TT STAT TIME COMMAND
659 p0 Rs 0:00.08 -csh (csh)
767 p0 R+ 0:00.00 ps
565 xc0 Is 0:00.01 login [pam] (login)
608 xc0 I+ 0:00.02 -csh (csh)
xendemo-freebsd#
Befehlsfenster
17:43
11.04.2005

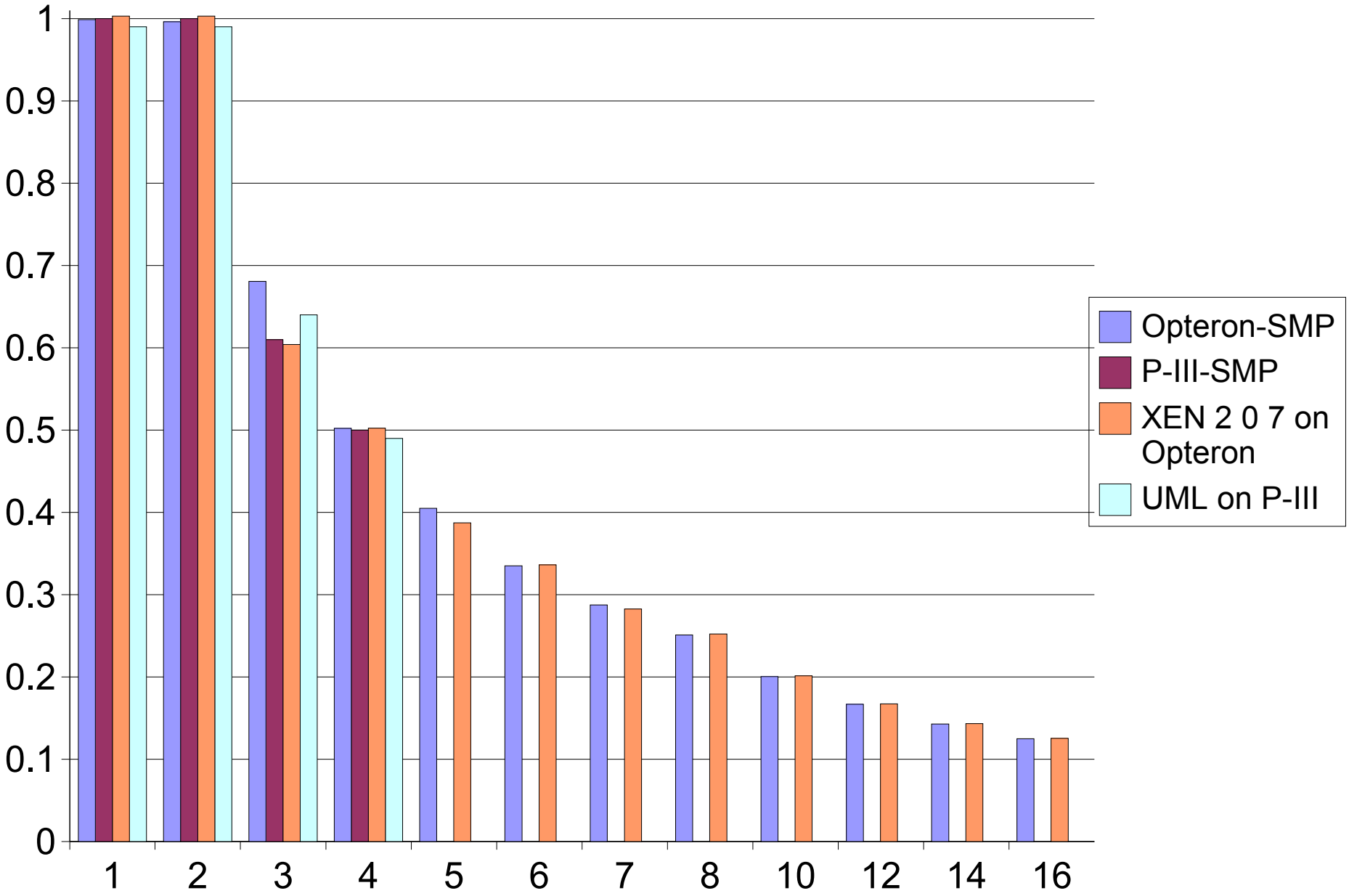
Xen: Features

- **Suspend & resume domains**
- **Destroy domains**
- **Migrate domains**
  - Migrate between different physical hosts
  - Life migration allows downtime in sub second range
  - Network connections are kept alive
- **Python scripted configuration files**

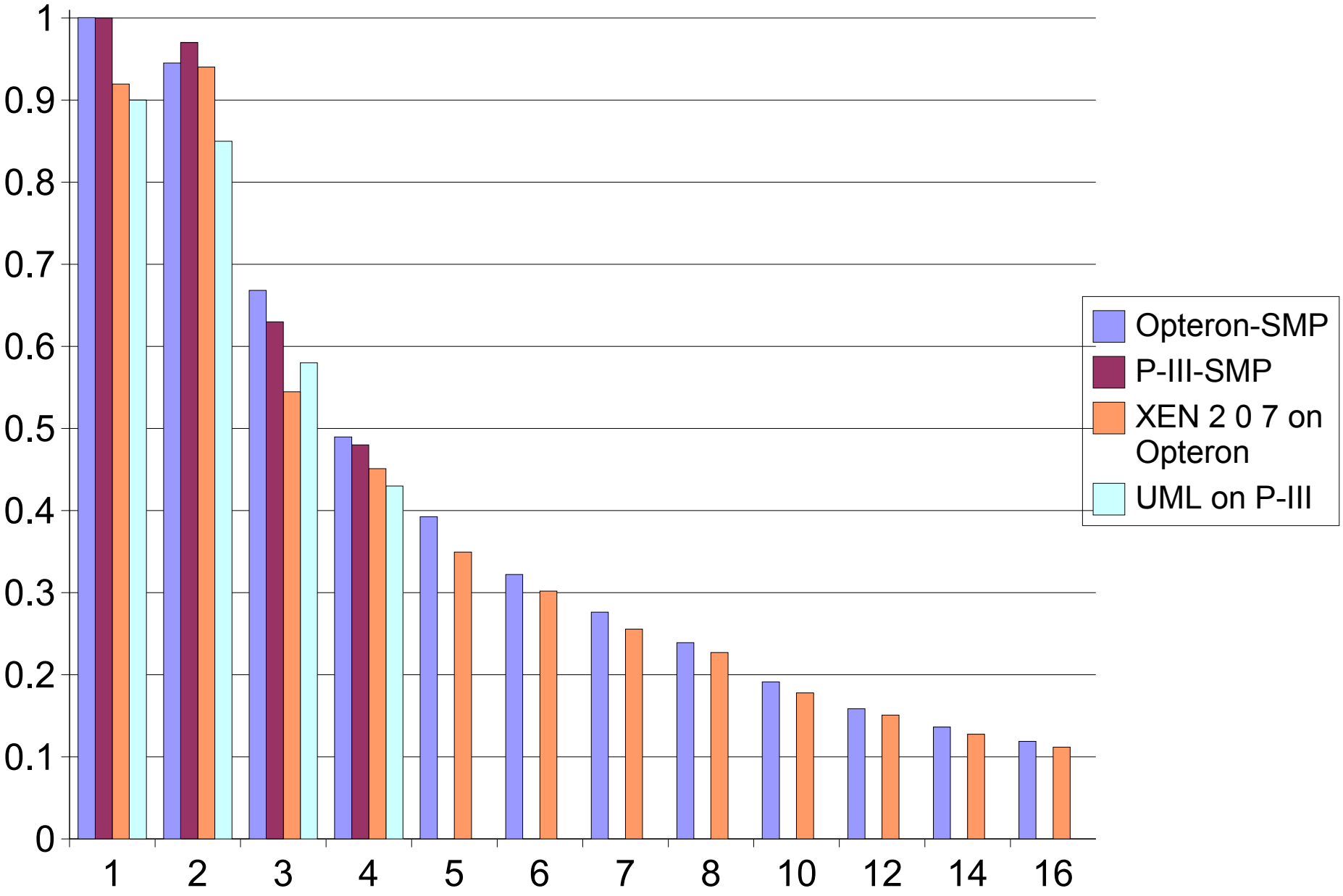
Xen Performance

- **How to measure Performance?**
  - Hardware:
    - Dual-PIII-700MHz / 1GB RAM / 40GB Disk / 100Mbit/s
  - Benchmarks
    - Covering the different system parameters
      - *CPU, MEM-IO, Disk-IO, kernel compilation*
    - Software set taken from freebench.org, samba.org, kernel.org
  - Reference Measurement 1-16 parallel runs on plain smp
  - Benchmark installation booted and run on 1-16 xen domains
  - Comparison Measurement on 1-4 UML instances
    - Only 1GB RAM on the UML test machines

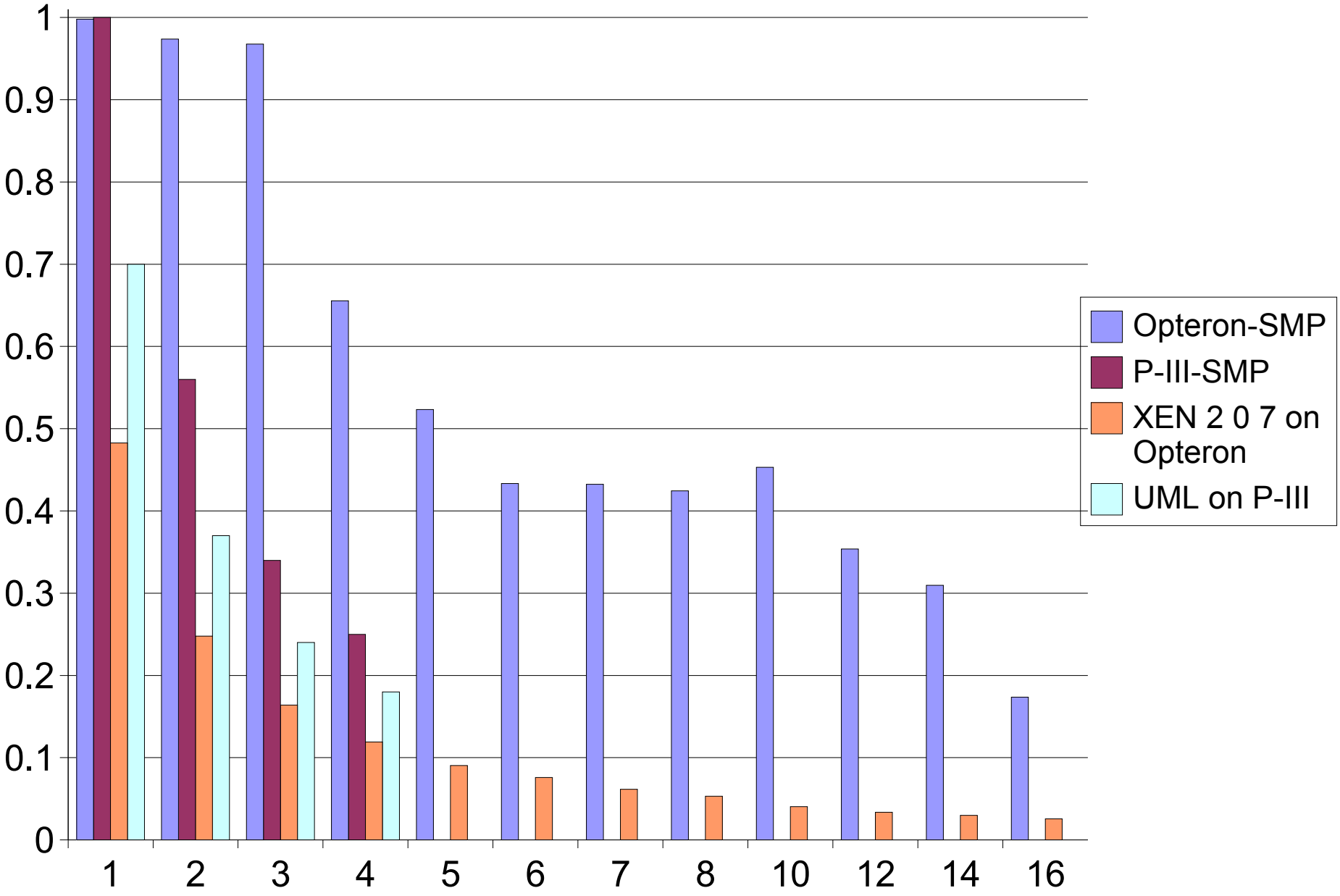
CPU


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
10
12
14
16
Opteron-SMP
P-III-SMP
XEN 2 0 7 on Opteron
UML on P-III

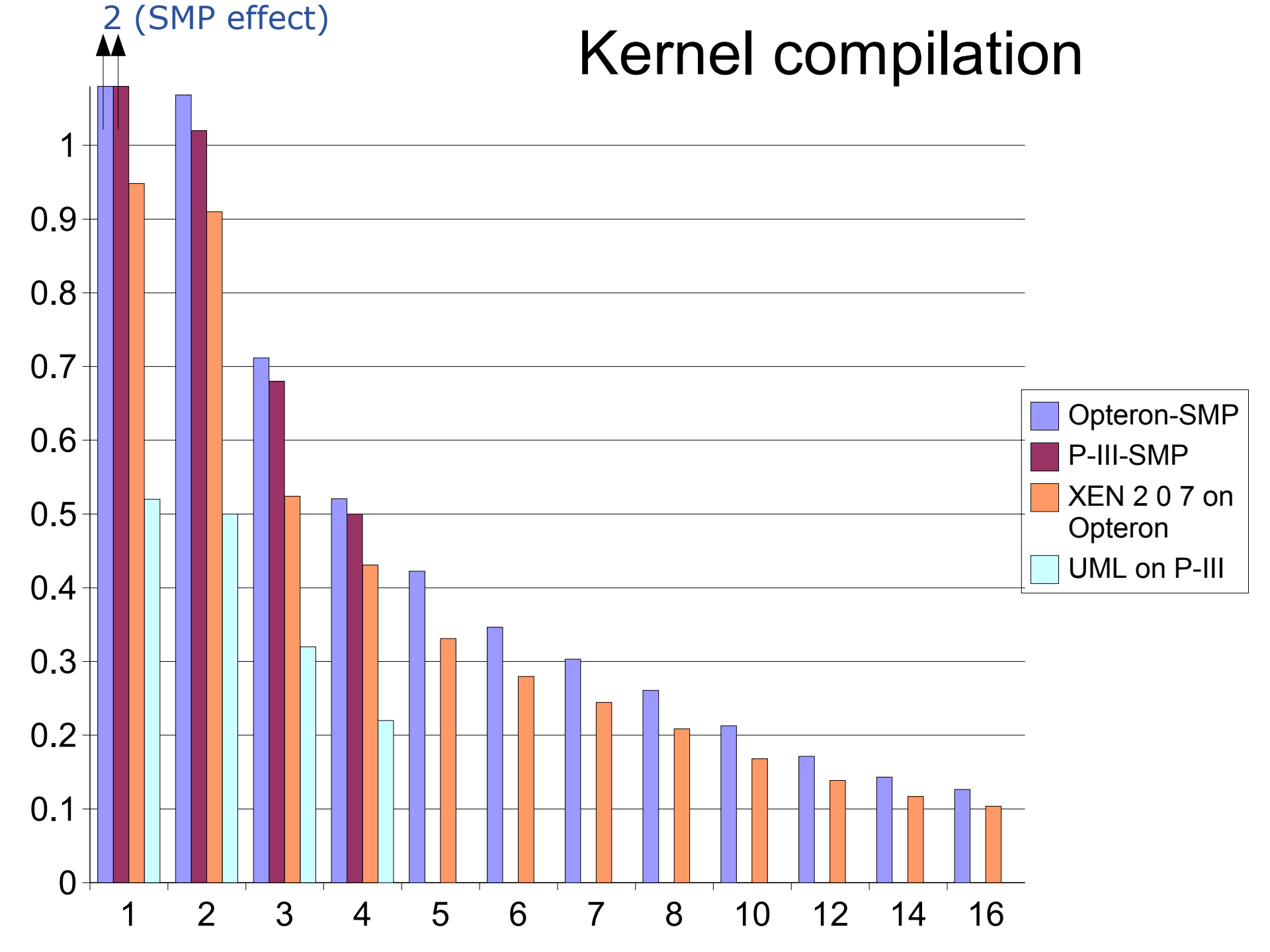
Memory


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
10
12
14
16
Opteron-SMP
P-III-SMP
XEN 2 0 7 on Opteron
UML on P-III

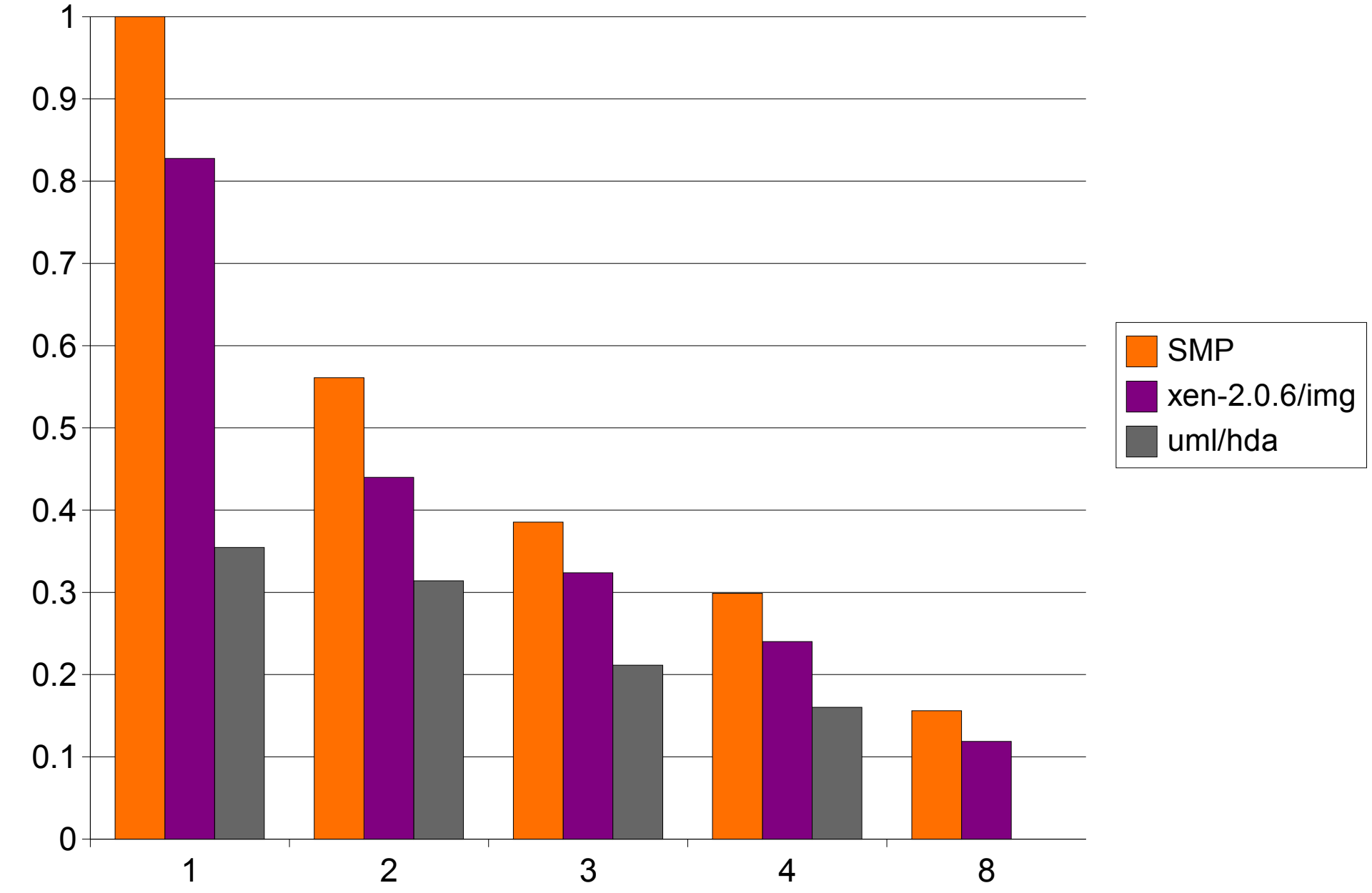
Disk


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
10
12
14
16
Opteron-SMP
P-III-SMP
XEN 2 0 7 on Opteron
UML on P-III

Kernel compilation



NET



Xen: Training

- **Installation Course on cluster/grid computing:**
  - Summer School on Gridcomputing at FZK
  - ~40 Students vs. 16 available PCs
  - PCs required for max 3 days

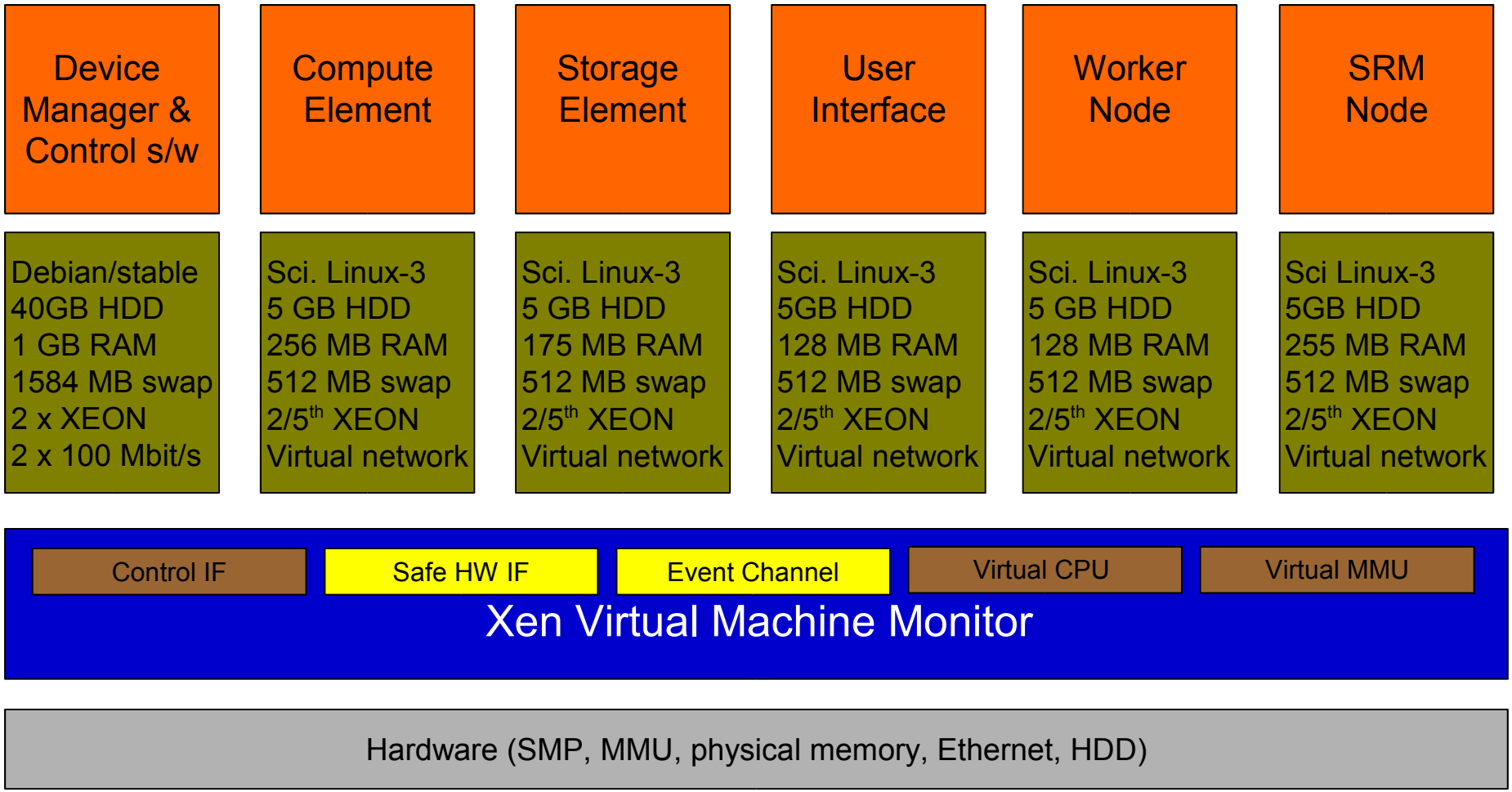
    => My boss won't buy the missing 60 PCs for that time
  - Virtualisation provides:
    - No need to buy additional 60 PCs (obvious)
    - No need to install 60 additional PCs
    - Students can check output of booted Xen domains via ssh
  - Last year we moved and installed 40 PCs (1.5 Racks) over to the office building....

Xen: Training

- **Simple installation of a virtual cluster:**
  - Linux installation:
    - `mount -o loop image mnt`
    - `ssh <installed machine> tar csp / | (cd mnt;tar xsp)`
    - Additional modifications:
      - *`/etc/fstab`*
      - *`/etc/passwd`*
      - *`/lib/tls`*
  - Image duplication
    - `for i in `seq 1 75`;do cp image image-$i; done`
  - Booting
    - `for i in `seq 1 75`;do xm create <conf> id=$i; done`

Xen: Training

- **Preparation:**
  - Image file with Scientific Linux
    => Image files can be cloned
    => 75 identical machines ready over lunchtime
- **The course itself:**
  - One PC per Group
  - 5 virtual machines per PC
    - CE, SE, UI, IO, SRM
  - Students logged into the virtual machines only
    => No notion of virtualisation
  - Access to Host systems possible
    - Observation of boot process
    - Network configuration of clients can be done
      => Remote installation trainings possible


Device Manager & Control s/w
Compute Element
Storage Element
User Interface
Worker Node
SRM Node
Debian/stable
40GB HDD
1 GB RAM
1584 MB swap
2 x XEON
2 x 100 Mbit/s
Sci. Linux-3
5 GB HDD
256 MB RAM
512 MB swap
2/5th XEON
Virtual network
Sci. Linux-3
5 GB HDD
175 MB RAM
512 MB swap
2/5th XEON
Virtual network
Sci. Linux-3
5GB HDD
128 MB RAM
512 MB swap
2/5th XEON
Virtual network
Sci. Linux-3
5 GB HDD
128 MB RAM
512 MB swap
2/5th XEON
Virtual network
Sci Linux-3
5GB HDD
255 MB RAM
512 MB swap
2/5th XEON
Virtual network
Control IF
Safe HW IF
Event Channel
Virtual CPU
Virtual MMU
Xen Virtual Machine Monitor
Hardware (SMP, MMU, physical memory, Ethernet, HDD)

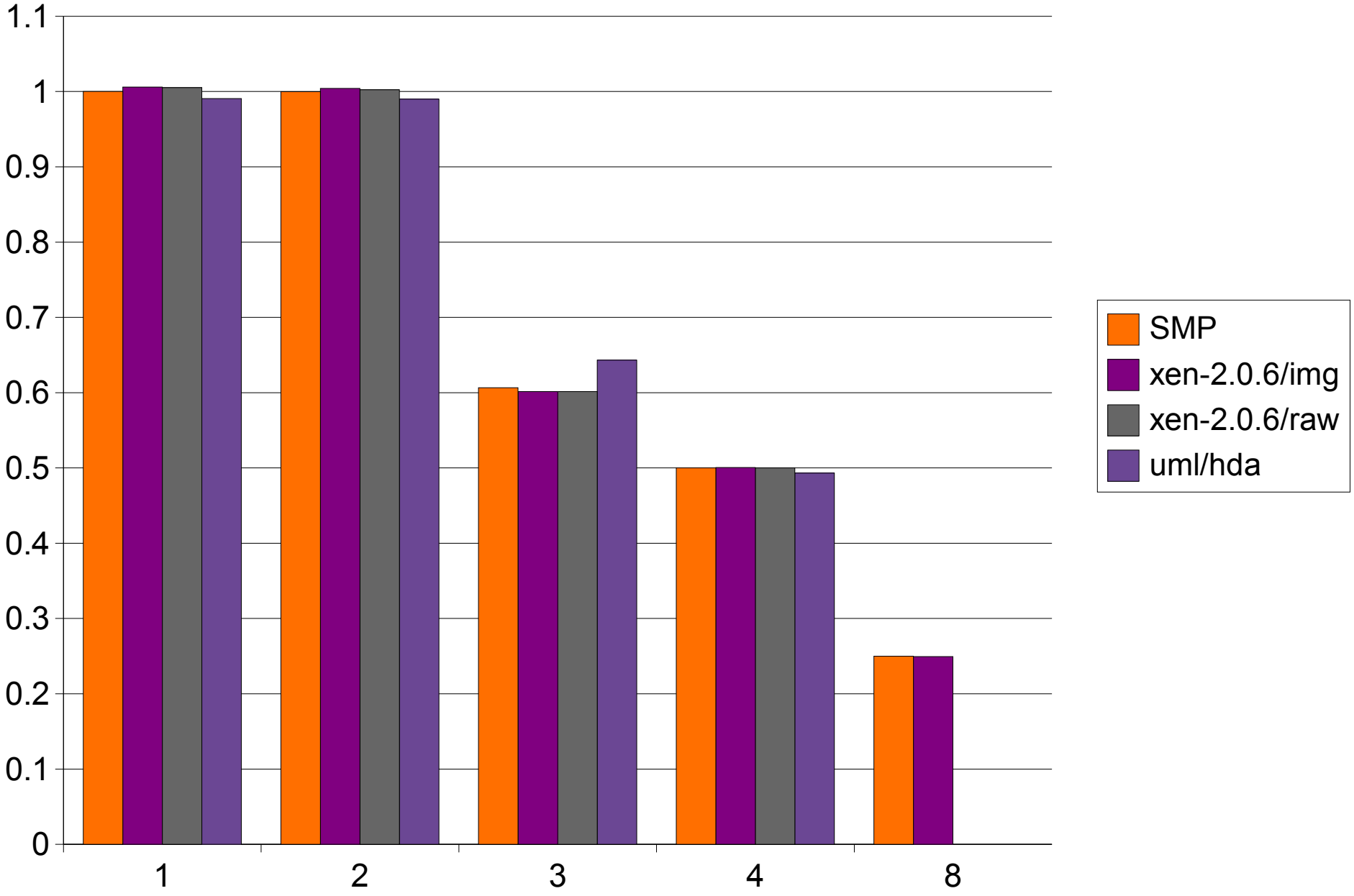
Xen: Issues

- **No Complaints about performance**
  - Even though oldish (P-III-700GHz 1GB-RAM) used
- **Linux cannot keep images on NFS**
  - Use SAN, GNBD or iSCSI instead
- **/lib/tls problem**
  - mv /lib/tls /lib/tls.disabled.for.xen
  - DB4 problems may still occur
- **Memory consumption**
  - Quite static memory configuration
  - Complete OS requires a lot of RAM
    => More resource-efficient sharing system on my agenda

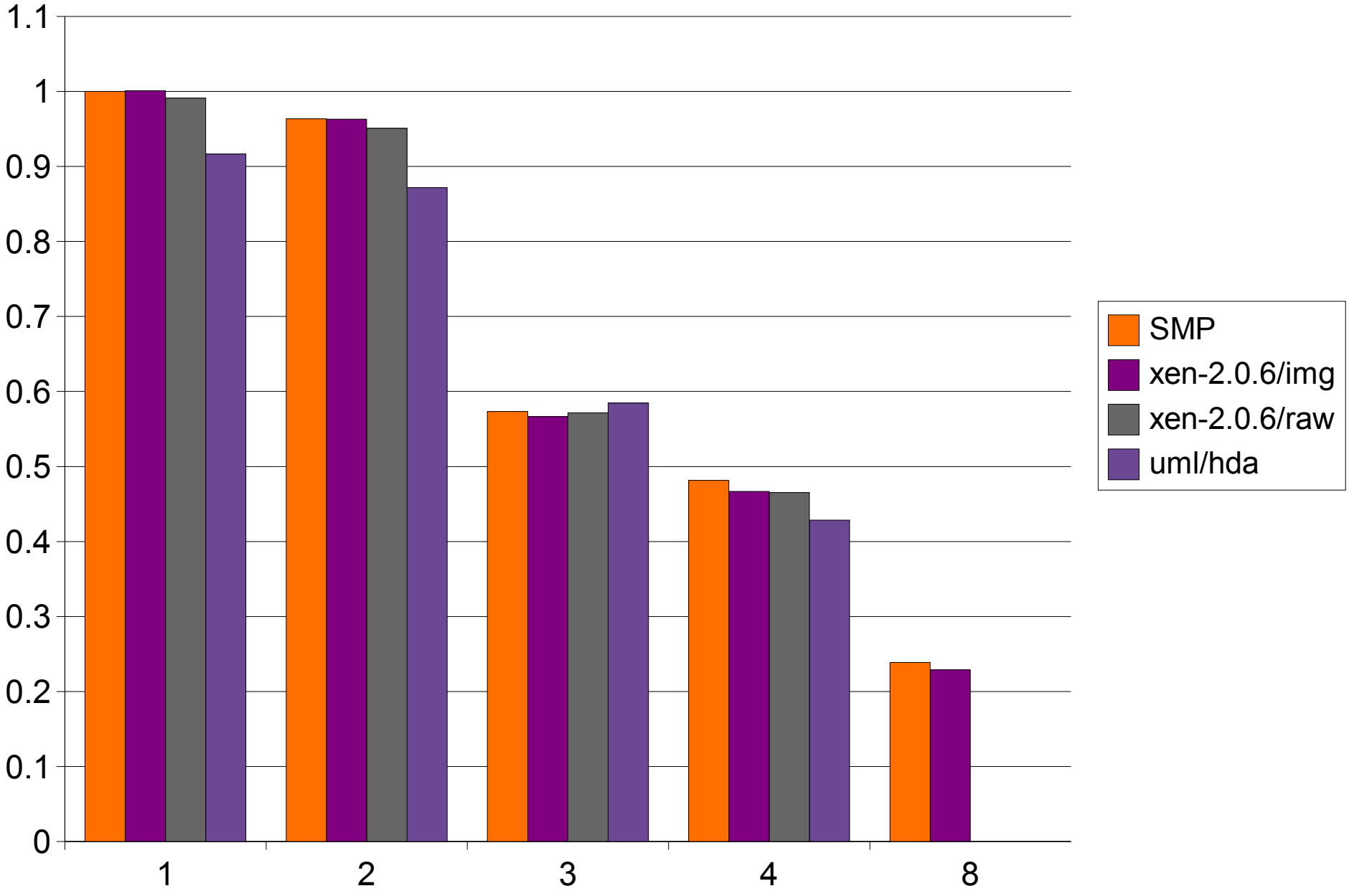
Conclusion

- **Valuable Virtualisation Environment**
  - “gridsite in a box” for test purposes
- **Good Performance**
  - Est 5% virtualisation cost (except disk/net i/o: 20%)
  - Better than userspace tools (UML, VMWare Workstation)
- **Easy to install and handle**
  - Image file contains (almost) the whole system
    => Allows for portable training environments
- **Very active user community**
  - Fast and good answers via mailinglist
- **Commercial Support available**
- **Supported by hardware manufacturers**

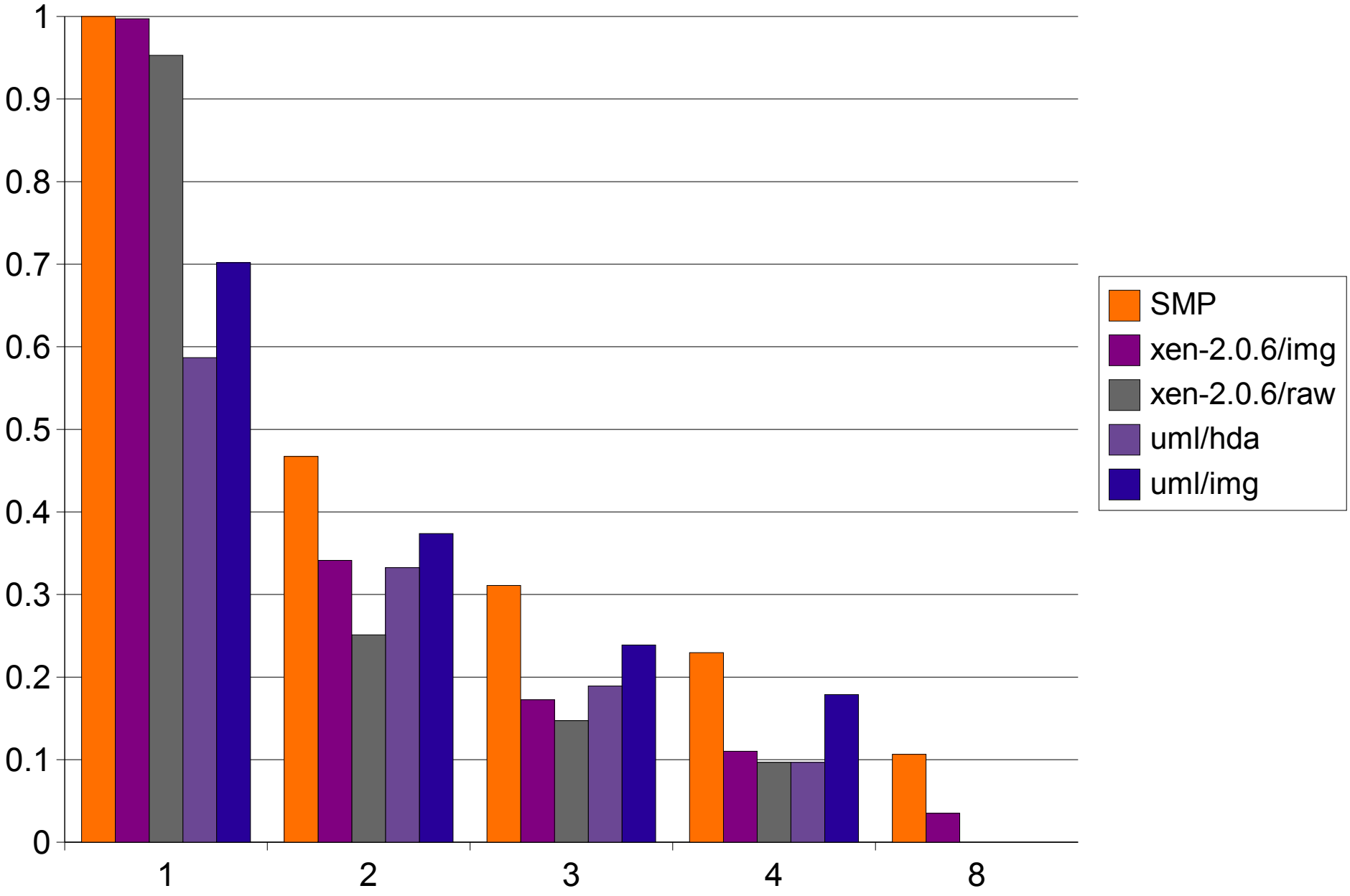
CPU


1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
8
SMP
xen-2.0.6/img
xen-2.0.6/raw
uml/hda

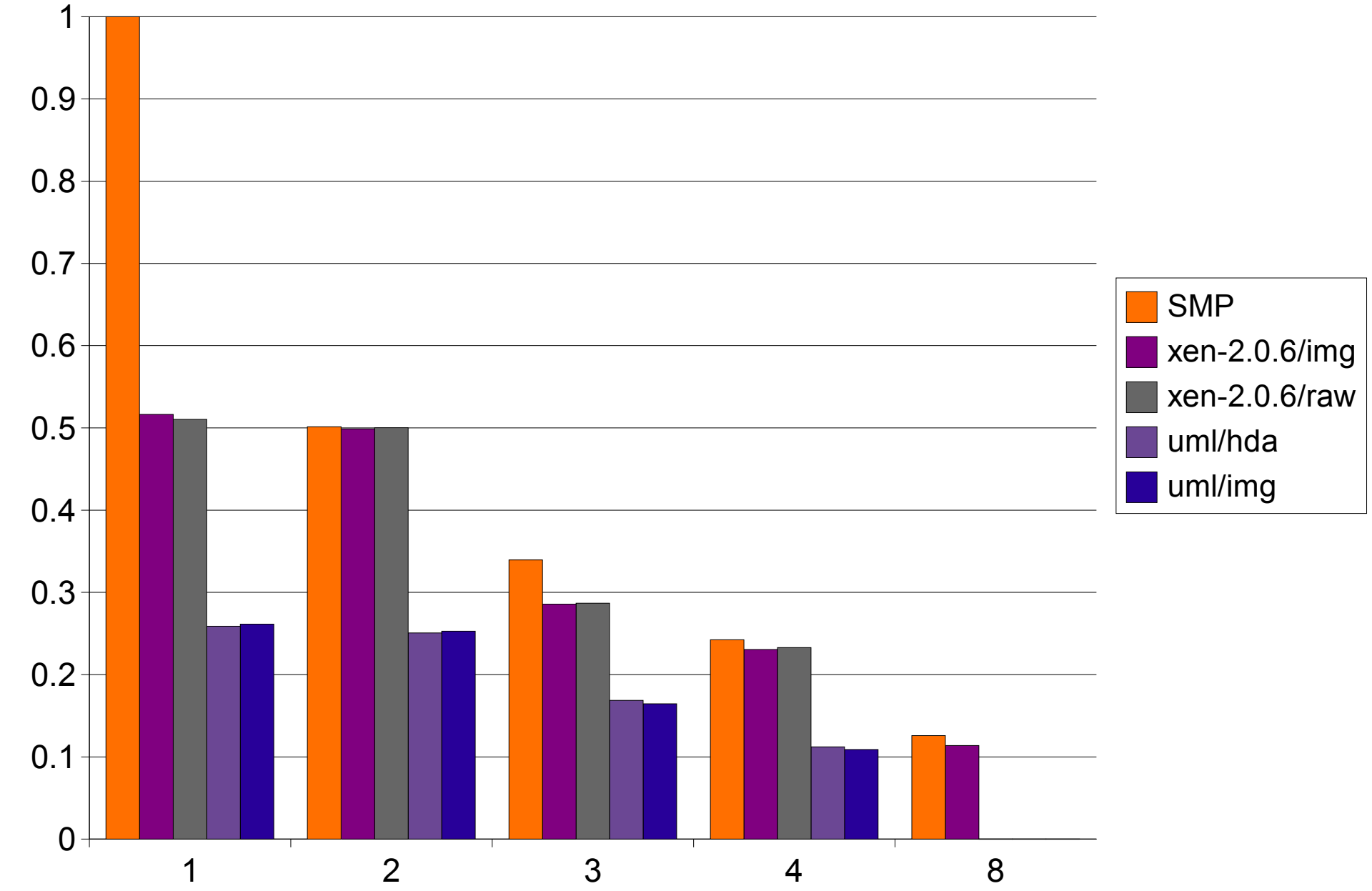
MEM


1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
8
SMP
xen-2.0.6/img
xen-2.0.6/raw
uml/hda

DD-2.0.6


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
8
SMP
xen-2.0.6/img
xen-2.0.6/raw
uml/hda
uml/img

Kernel


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
8
SMP
xen-2.0.6/img
xen-2.0.6/raw
uml/hda
uml/img

Xen: Training

No complaints on performance
(although oldish hardware used: P-III with 1GB RAM)