Virtualisation: Use Cases and Performance

Marcus Hardt
Forschungszentrum Karlsruhe GmbH

www.eu-egee.org

Overview

- **Virtualisation Approaches**
- **Usecases for Virtualisation**
  - Realistic
  - Futuristic
- **Performance Measurements**

Approaches 1/2

- **zSeries**
  - Hardware supported + specialised OS provide VM to guest OSes (e.g. Linux, UNIX)
- **UserModeLinux**
  - Linux-only emulation
  - Still: large virtualisation overhead
  - Feature: Designed to run without root privileges
- **QEMU & some commercial systems**
  - Full system emulation
    - Emulate the full system including processor and peripherials => guest OS can not see the difference
    - Large virtualisation overhead expected

Approaches 2/2

- **Other Commercial server Virtualisation system:**
  - Full system virtualisation
    - Virtualise the host system
    - Redirect
    - Features: Run onmodified guest OSes (e.g. WinXP)
      Available for Linux + Windows hosts
- **Xen**
  - Designed for x86 architecture to overcome its lack of virtualisation
    - No hardware support for trapping direct access

  => Para-virtualisation
    - Requires cooperation (modification) of guest OS
      => Major free OSes are supported. (Free- Net- OpenBSD, Linux, Plan9). Windows XP was demonstrated by MS-Research
    - Features: suspend/resume, migrate

Use Cases / What to do with Virtualisation?

Use Case: 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....

Use Case: 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
  - No Complaints about performance
    - Even though oldish (P-III-700MHz 1GB-RAM) used

Virtualisation: Use Cases

- **IT Consolidation**
  - Start/stop servers on demand
    - maybe based on monitoring information (load, response time, ...)
  - Cheap and transparent high availability solutions possible
    - One standby server per IT department vs. one per service
  - Easy provisioning of machines
    - `cp Debian-stable.img webserver-cern.chimg`
    - `xm create ...`
  - Concentration of rarely used machines to one
    - est. 100-200 EUR per machine per year. (1 EUR/W/a)
  - Migration of domains may be helpful for administration

Virtualisation: Use Cases

- **Using Windows Desktops**
  - We have est 4000 Windows Desktops
    - Idle 66% of their time
    - Doesn't even require air-condition
  - With cross platform virtualisation:

    (VMWare, Virtual Server, Xen since 2006)
    - Run two different machines on every Desktop:
      - *Windows Desktop*
      - *Cluster Node*
        - Optional: Image supplied by customer
    - When Desktop is used, workernode can be suspended or migrated elsewhere

Virtualisation: Possible Use Cases

- **Submitting a job to "the grid"**
  - The grid =
    Scattered heterogenous resources with different admins
    - App.-developers, MW-developers and Site-admins prone to conflict
  - Virtualisation allows:
    - Cleaner separation of different interests:
      - *Application Developer is given (can modify) an OS image*
        - Image is transported to resource
        - ... booted ... processing ... executed ... results returned
      - *Application Developer can choose the MW he requires*
      - *Site Admins provide a run environment based on their favourite OS*

Virtualisation: Performance

- **How to measure Performance?**
  - Hardware reference (=1):
    - Dual Opteron 2.2GHz / 4GB RAM / 80GB SCSI Disk 1Gbit/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 ran on 1-16 xen domains
  - “Scheduler darlings”
    - some VMs finish four benchmarks while others only finish two
=> Measured time is time to finish three benches on every VM
=> pay tribute to unequal load distribution

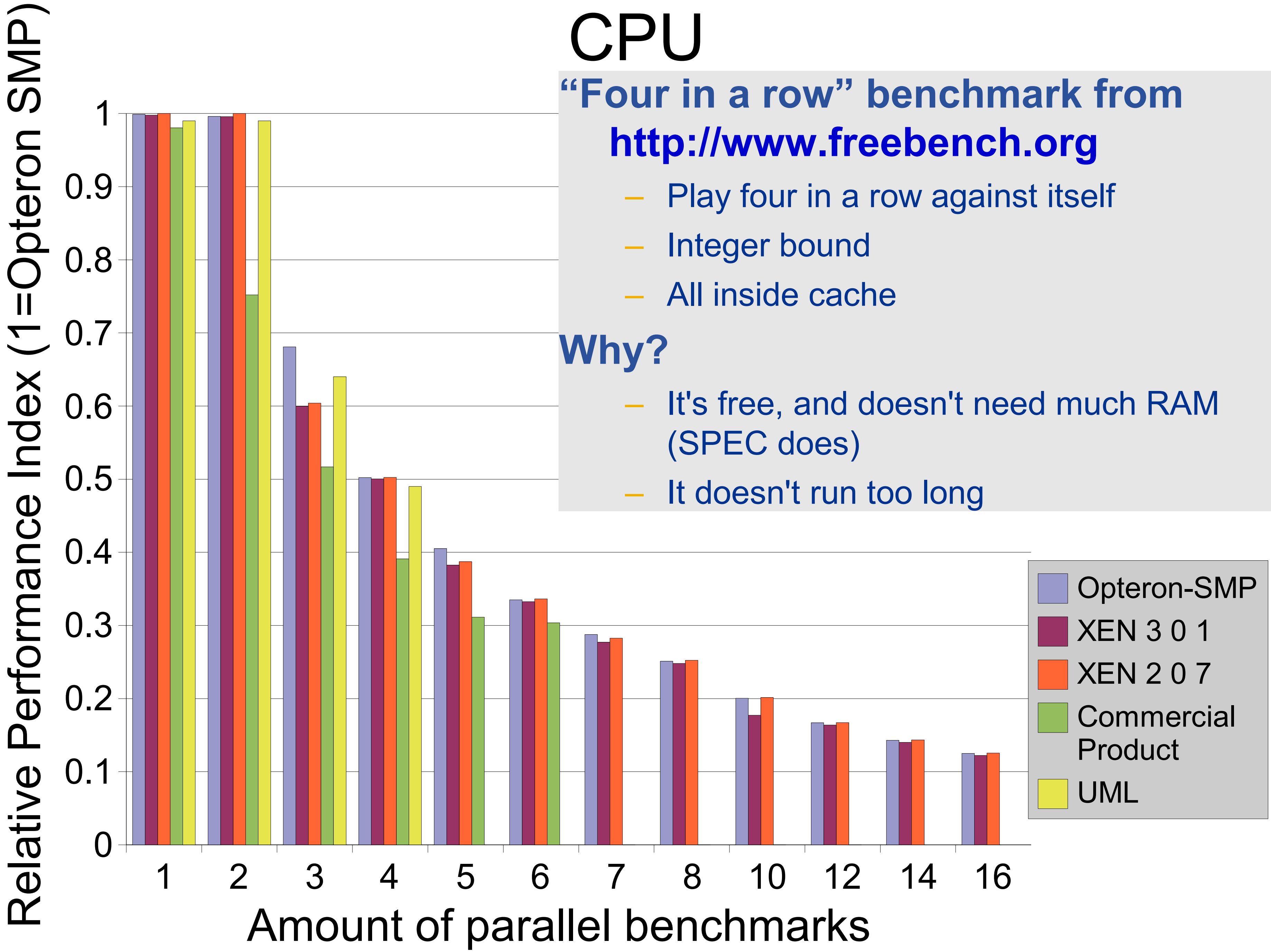
CPU

“Four in a row” benchmark from http://www.freebench.org

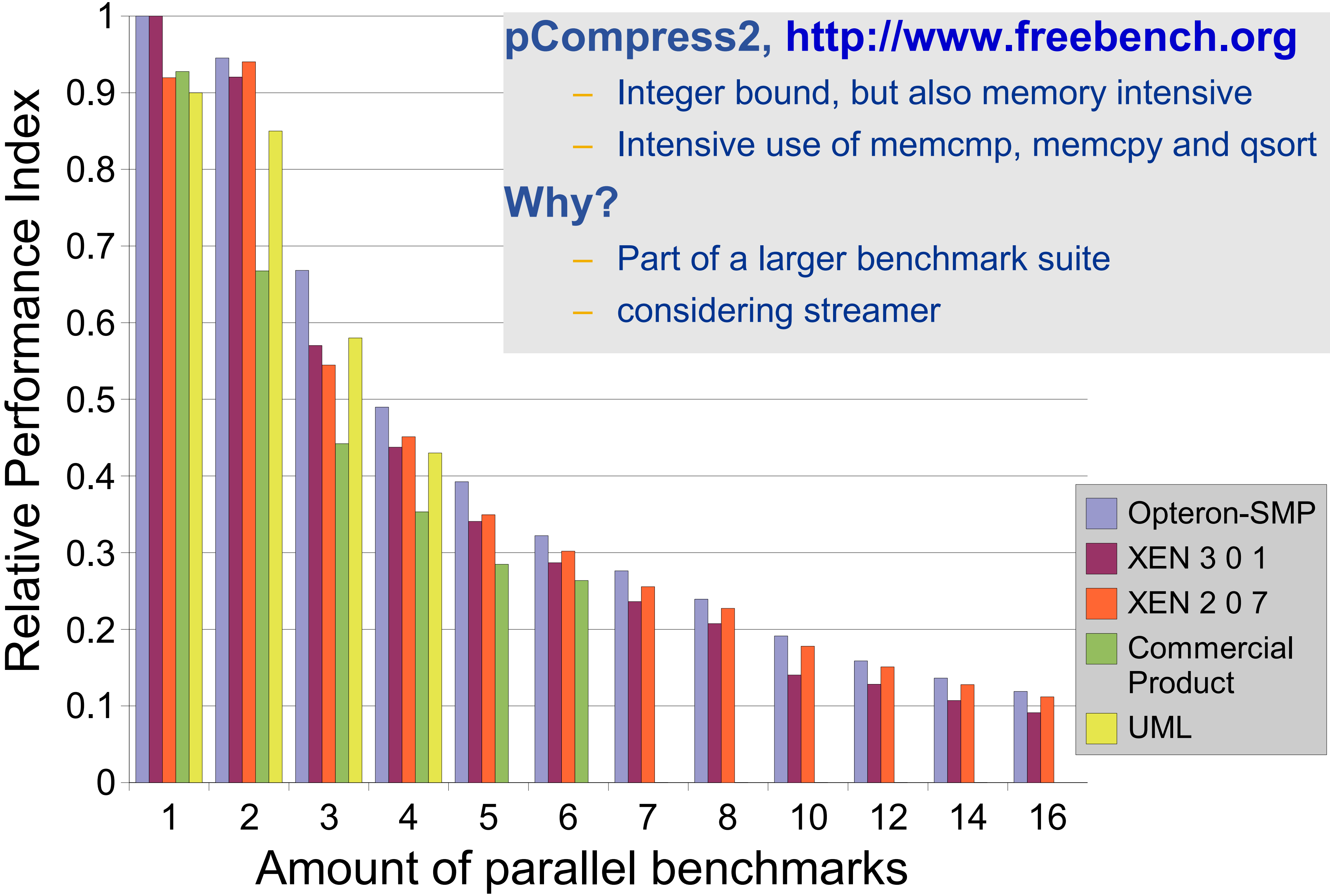
- Play four in a row against itself
- Integer bound
- All inside cache

Why?

- It's free, and doesn't need much RAM (SPEC does)
- It doesn't run too long



Memory



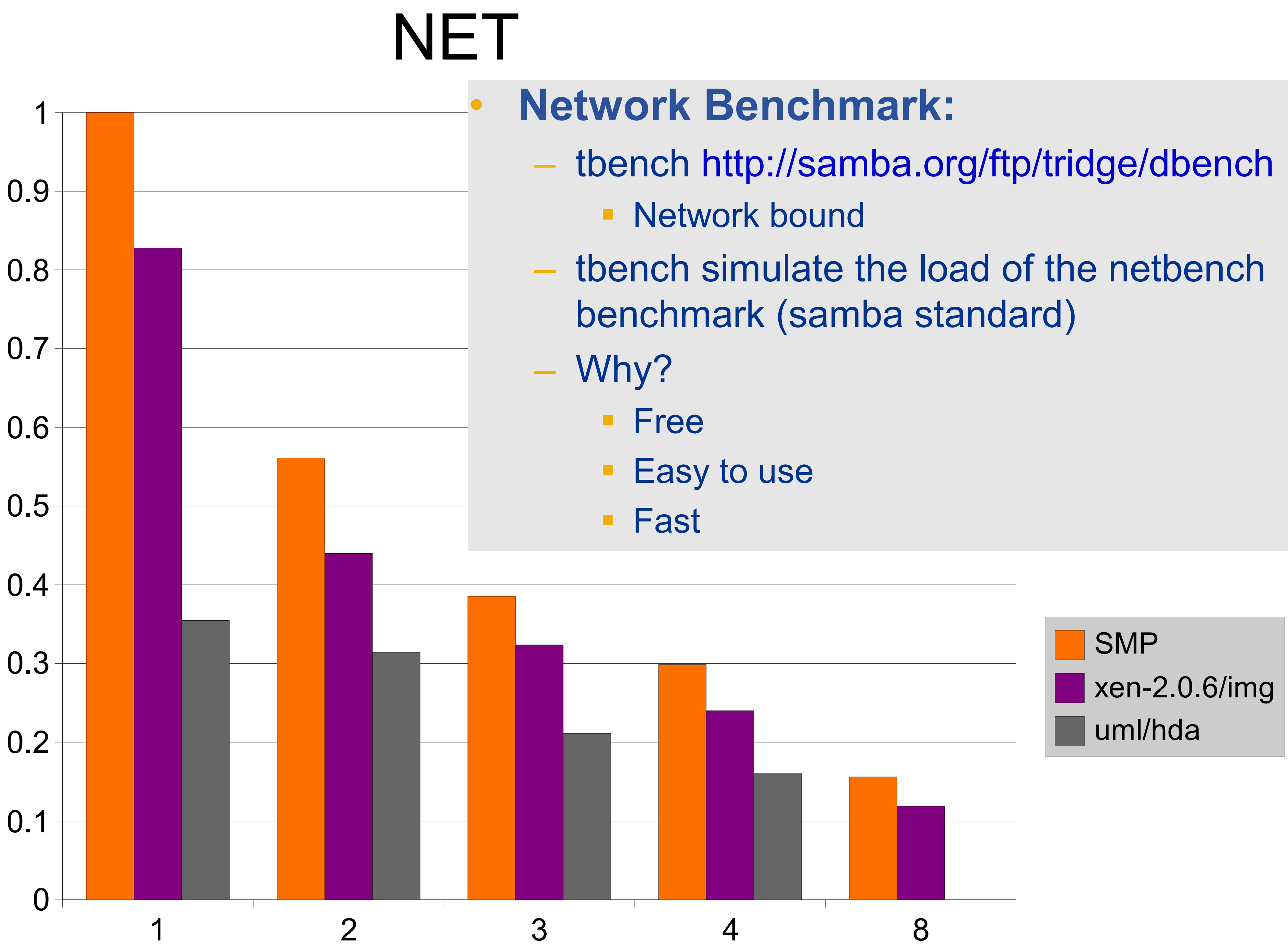
pCompress2, http://www.freebench.org

- Integer bound, but also memory intensive
- Intensive use of memcmp, memcpy and qsort

Why?

- Part of a larger benchmark suite
- considering streamer

- **Network Benchmark:**
  - tbench http://samba.org/ftp/tridge/dbench
    - Network bound
  - tbench simulate the load of the netbench benchmark (samba standard)
  - Why?
    - Free
    - Easy to use
    - Fast

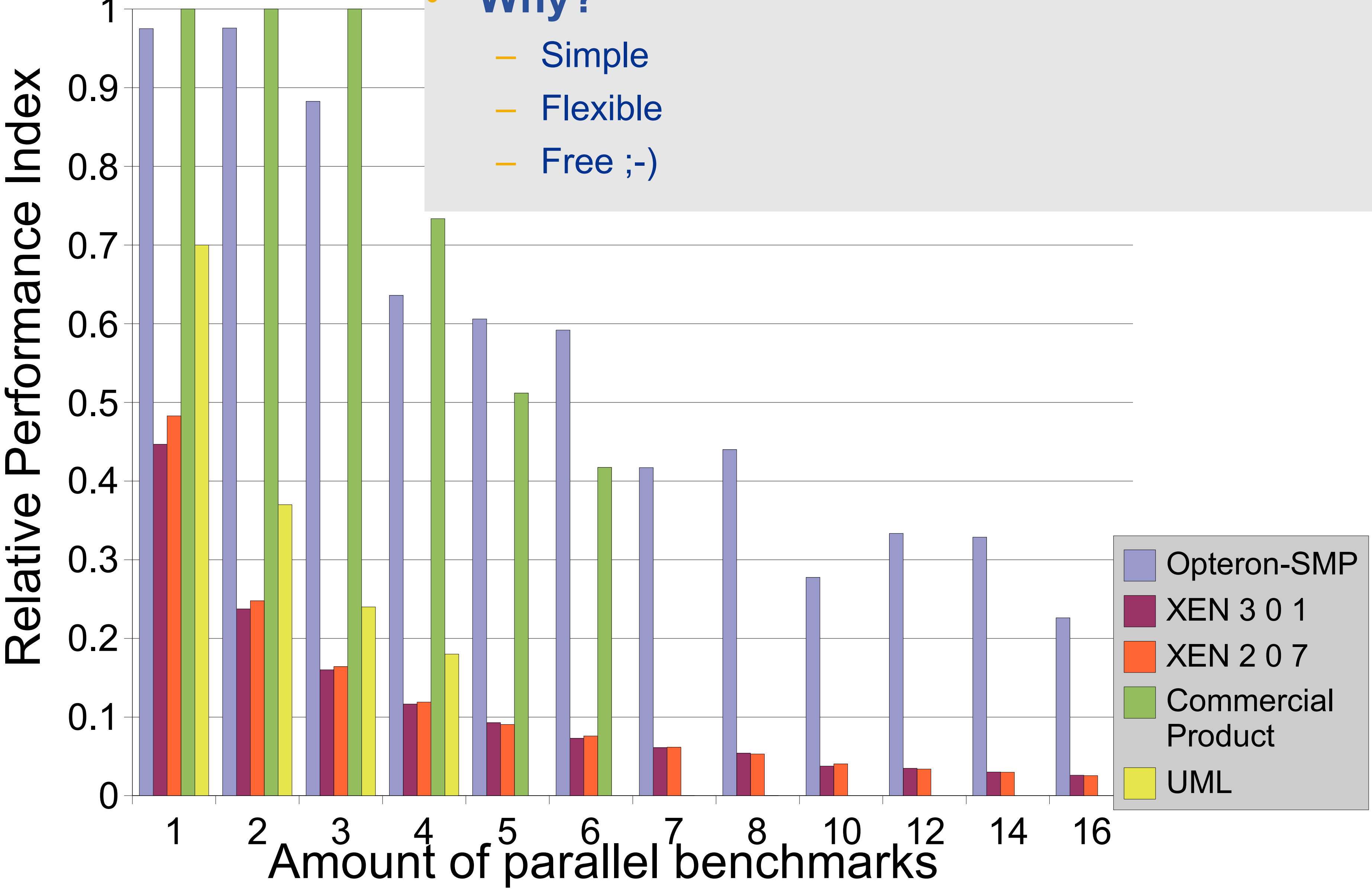


dd

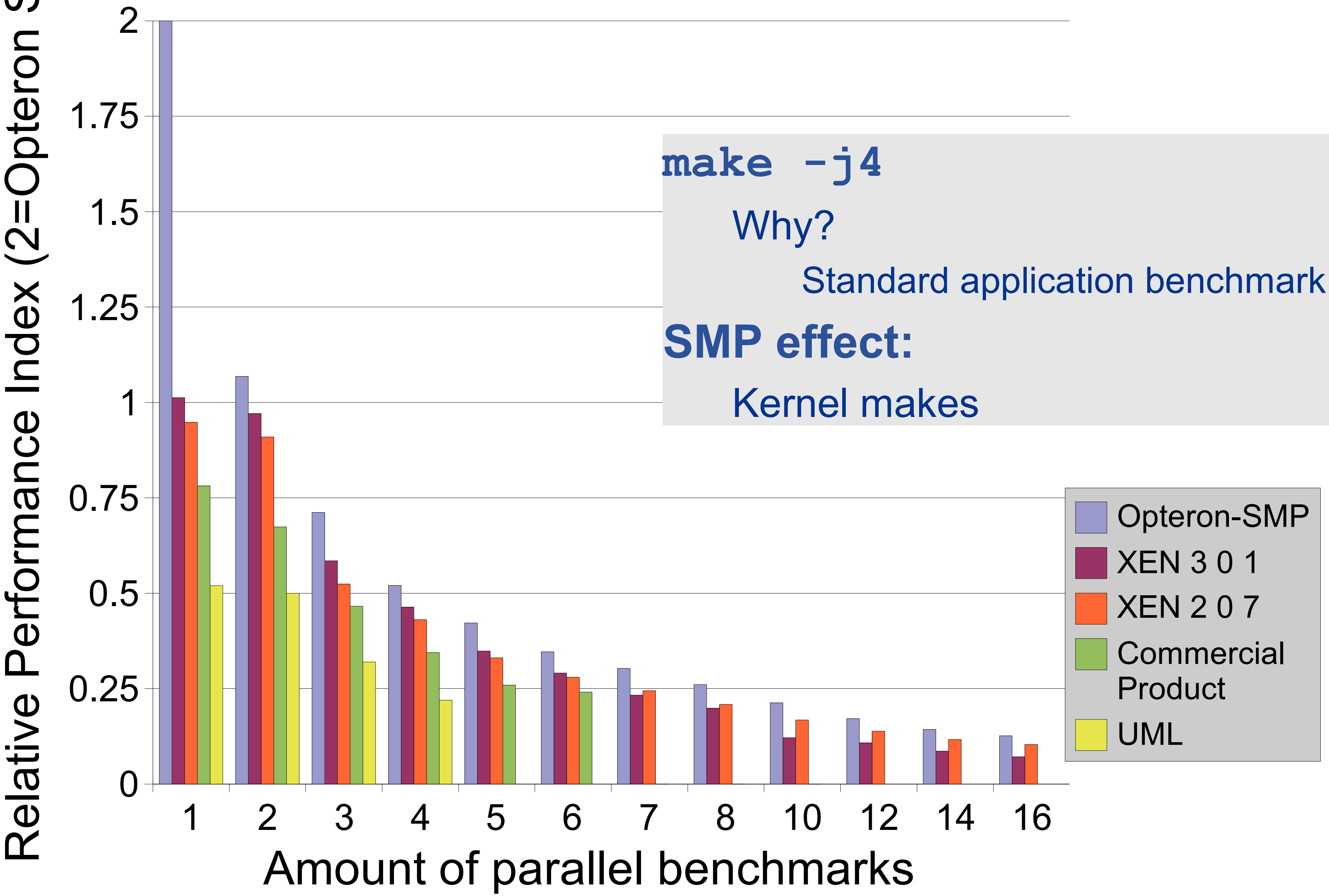
```
dd if=/dev/hda1 of=/dev/null bs=32k count=32k
```

Image backed vs. Partition backed

- **Why?**
  - Simple
  - Flexible
  - Free ;-)



Kernel



`make -j4`

- Why?
  - Standard application benchmark

SMP effect:

- Kernel makes

Conclusion 1/2

- **Xen**
  - Linux cannot keep images on NFS
    - Use of SAN, GNBD or iSCSI is recommended
  - Stability:
    - /lib/tls problem
      - ***`mv /lib/tls /lib/tls.disabled`*** *(careful when updating!)*
    - DB4 problems may occur
    - Stable enough for: Installation courses
      3 private webservers
  - Life migration capability
  - Support questions are answered within a few hours
- **User Mode Linux**
  - Does not scale well
  - Linux only

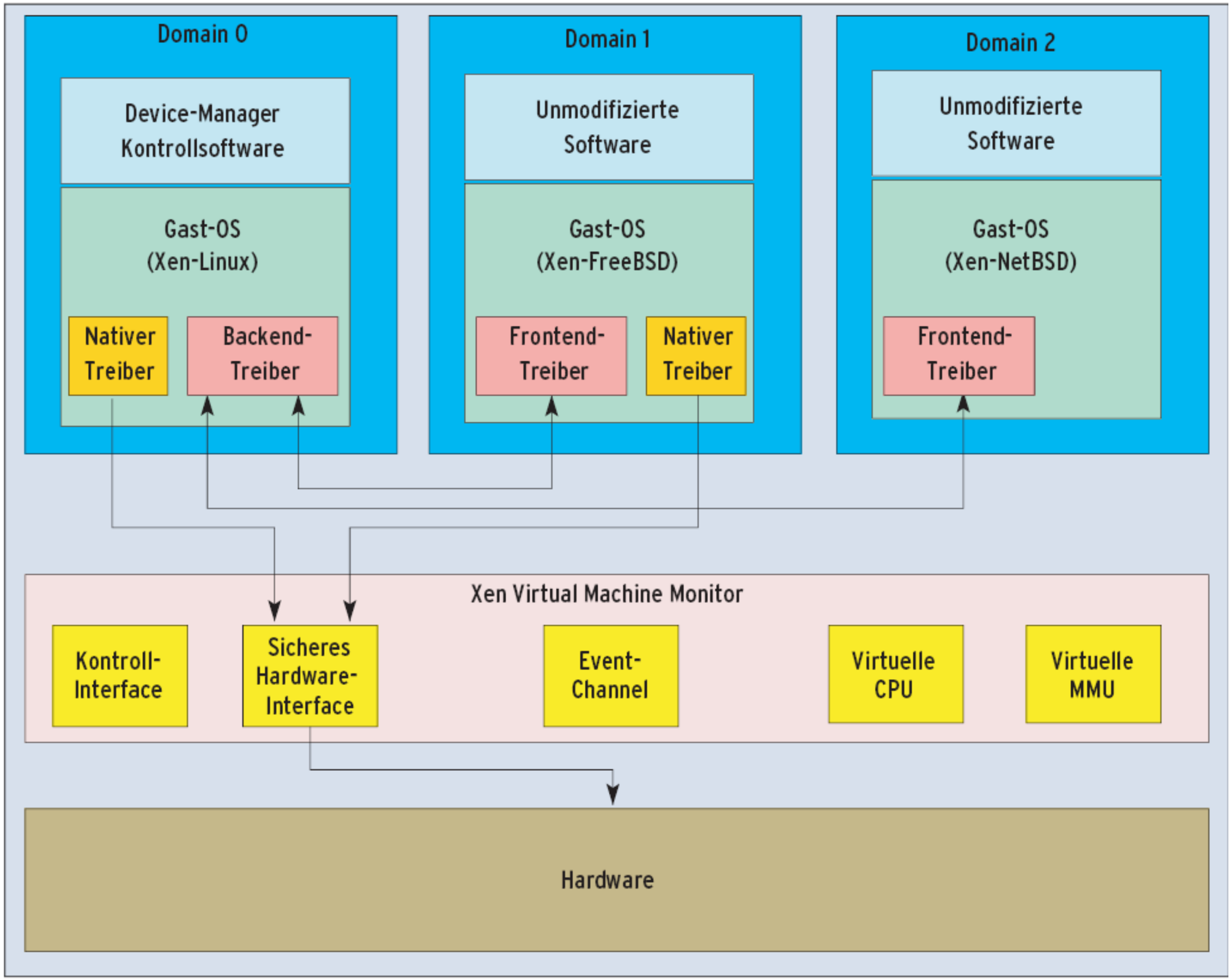
Conclusion 2/2

- **Commercial Product**
  - Requires GUI for running
  - Strong load-inequality (more than 10 parallell runs are very difficult)
  - Support is rather slow: 1-2 days to answer a ticket
- **Performance**
  - CPU: less than 10% virtualisation cost
  - Network I/O: 20% loss
  - Disk I/O: 50% loss on disk images
  - Xen slightly better than commercial products
- **Complete OS requires a lot of RAM**

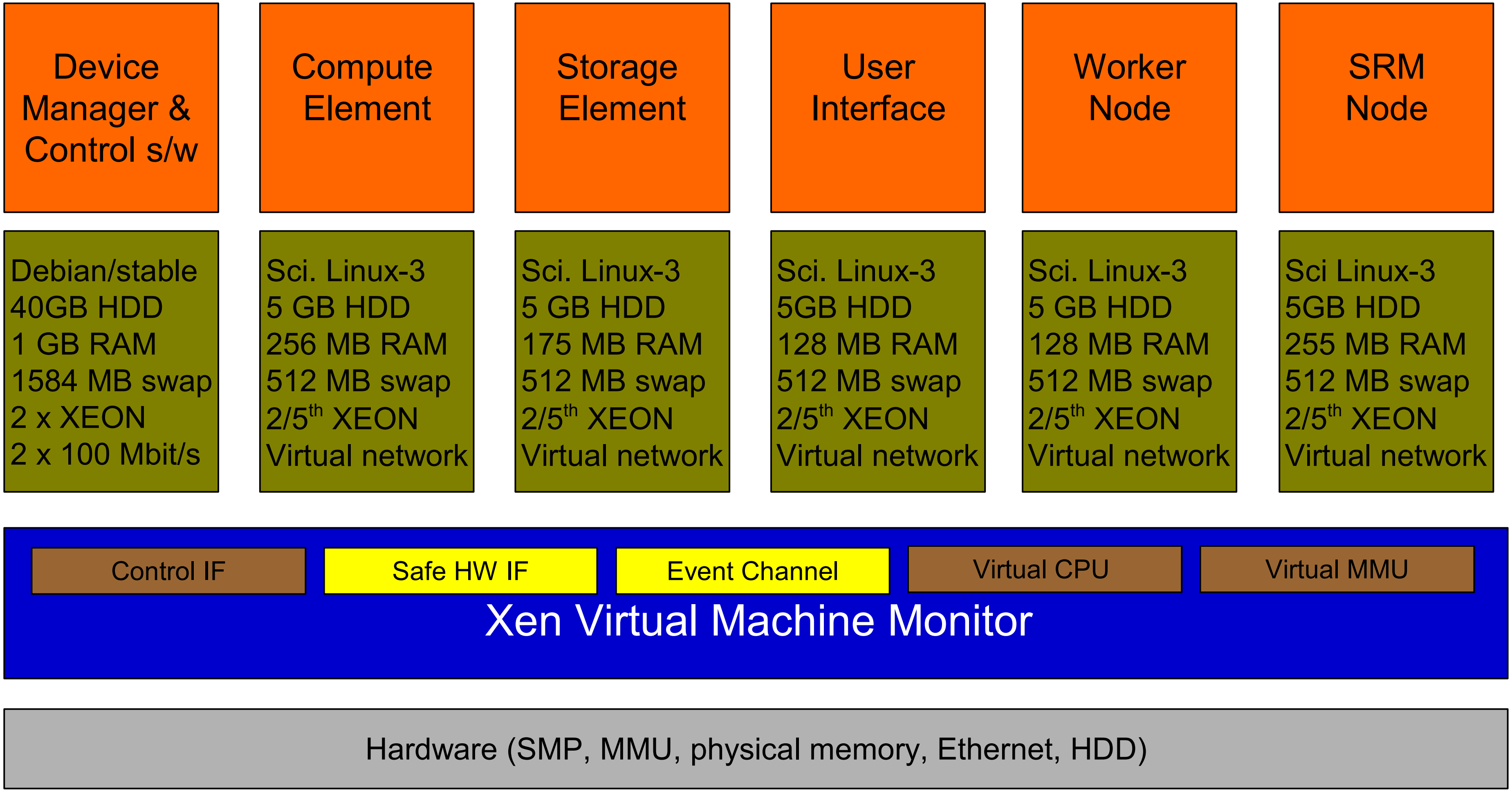
  => More resource-efficient virtualisation environments to be evaluated

Thank you for your time!

Xen: Paravirtualisation



- Priviledged calls are done through dedicated interface in domain 0
- Advantage: Very high performance (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)



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)

Virtualisation: Possible Use Cases

- **Load Balancing in Cluster Systems**
  - Oversubscription of the cluster
    - Some jobs do I/O, while others compute
  - Individual Operating Systems provided
  - Easier administration, especially of SMP machines
  - Migration helps administration
  - Phython based configuration increases flexibility
- **Flexible node allocation with SAN backed VMs:**
  - gpfs client on host-machine could provide FS to VMs

  => nfs4 over gpfs? (on IBM roadmap only for 2007)

Use Case: Training

- **Simple installation of a virtual cluster:**
  - Linux installation:

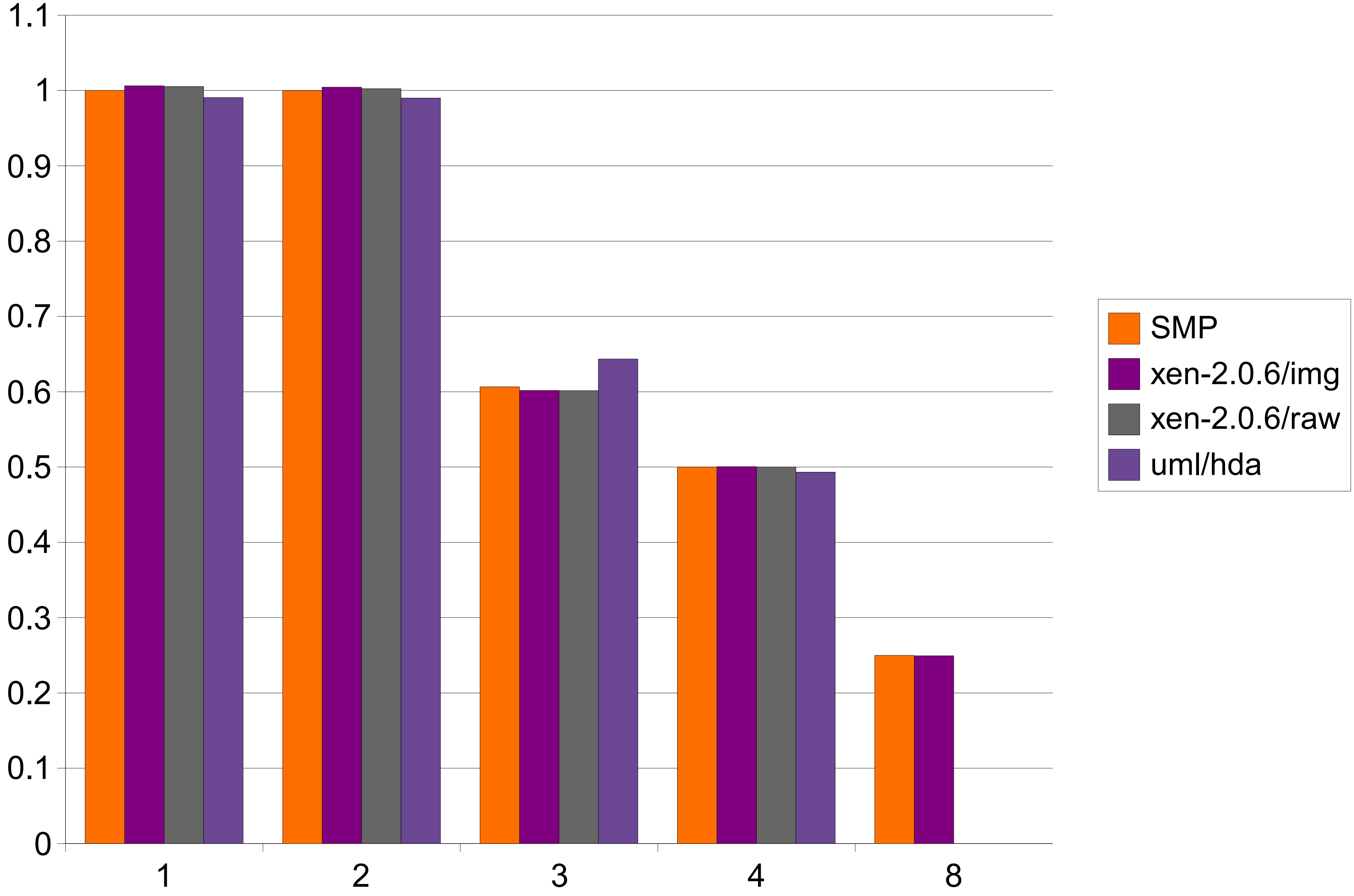
    -
      ```
      mount -o loop image mnt
      ```

    -
      ```
      ssh <installed machine> tar csp / | (cd mnt;tar xsp)
      ```
    - Additional modifications:
      - *`/etc/fstab`*
      - *`/etc/passwd`*
      - *`/lib/tls`*
  - Image duplication

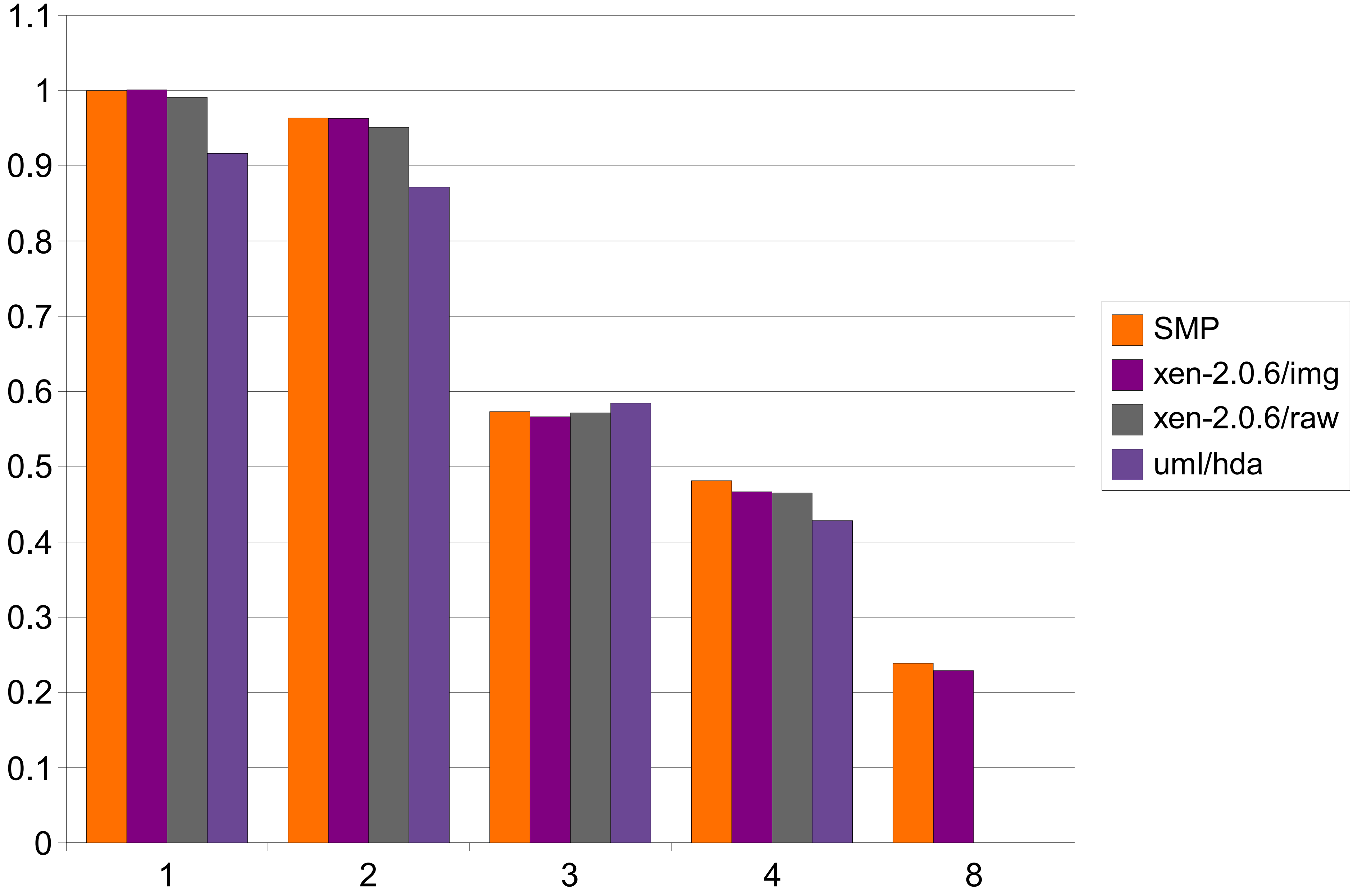
    -
      ```
      for i in `seq 1 75`;do cp image image-$i; done
      ```
  - Booting

    -
      ```
      for i in `seq 1 75`;do xm create <conf> id=$i; done
      ```

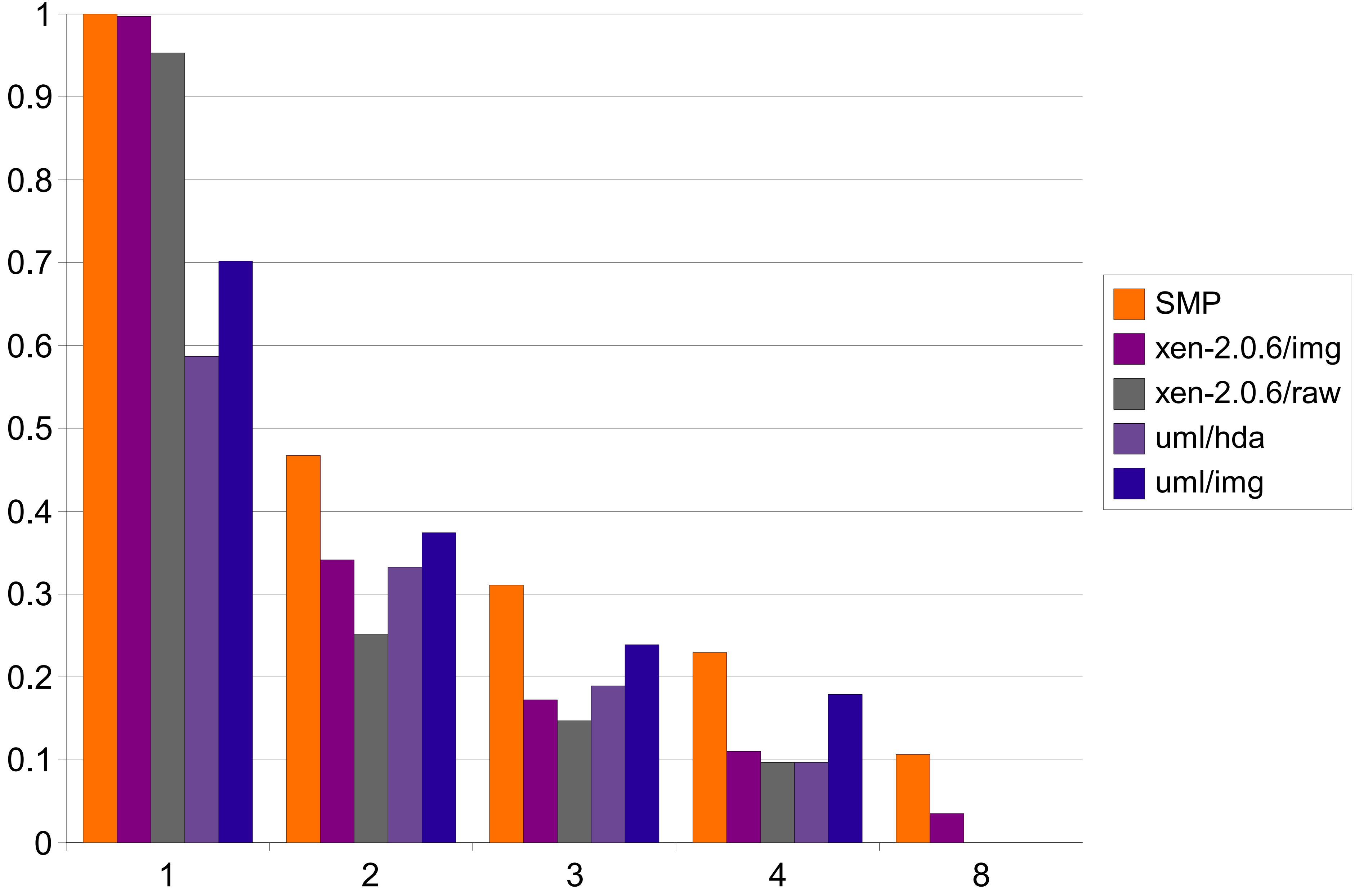
CPU


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

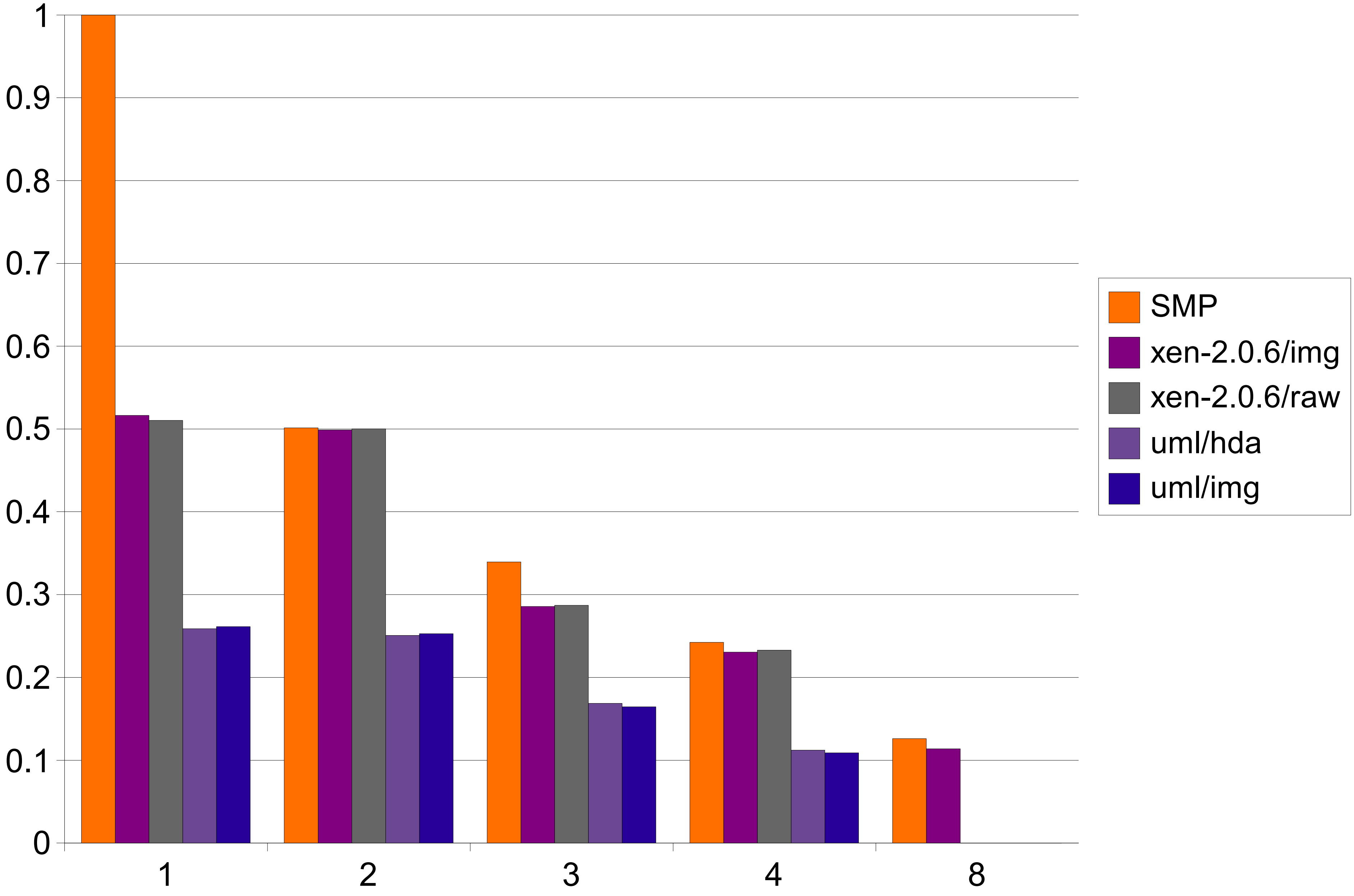
MEM


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