Benchmarking and HPC systems

Aad van der Steen
NCF/HPC Research
The Netherlands

1. Why benchmarking?
2. Reasons for benchmarking
3. Benchmarking for knowledge
4. Work in progress ...
5. The End

Why benchmarking? — 1

Is there any sense in benchmarking?

Because:

- "One of the top capability machines in the world". (IBM, BG/P)
- "HP BladeSystem Servers lead with 35% of systems in the TOP500". (HP)
- "Appro is uniquely positioned to support High-Performance Computing markets". (Appro)
- "The Cray XT5™ system combines unprecedented sustained application performance...". (Cray)

They are ALL the best you can get. So why worry?

Why benchmarking? — 2

Well, maybe a little benchmarking...

Operation	Length	NEC SX-8 (Mflop/s)	Intel Clovertown (Mflop/s)
$x = x + \beta y$	1,000	3644	5102
	10,000	6976	2439
	10,000,000	7767	379
$x_i = y_i + x_{i-1}$	1,000	251	880
	10,000	246	885
	10,000,000	253	148

It seems that some computers are sometimes more best than others.

Reasons for benchmarking — 1

There appear to be three main reasons for benchmarking:

1. Benchmarking for selling systems.

2. Benchmarking for buying systems.

3. Benchmarking for knowledge.

The latter two reasons do not exclude each other.

Reasons for benchmarking – 2

Characteristics of benchmarking for selling systems:

- Tries to adhere to “standards”
- Event-driven

Advantages:

- “Standard” benchmarks can be easily compared with competition
- Benchmarking effort can be minimised (but see later)

Disadvantages:

- “Standard” benchmarks are often not very informative
- Much benchmarking has to be repeated again and again because projects are unrelated.

Reasons for benchmarking – 3

Characteristics of benchmarking for buying systems:

- Often occurs on an ad-hoc basis
- Event-driven

Advantages:

- Closely matches needs of the buyer (hopefully)
- Benchmarking effort can be minimised (but see later)

Disadvantages:

- Benchmark is often moving target
- Results often difficult to compare/interpret

Reasons for benchmarking – 4

Characteristics of benchmarking for knowledge:

- Tries to relate system properties and benchmark results
- Tries to interrelate benchmark performance results
- Tries to interrelate benchmark performance and performance of applications

Advantages:

- Systems can be compared with some justification
- Might be basis for performance predictions
- Generates knowledge about the system

Disadvantages:

- Gives no exact answers for specific applications
- Will always be in development

Benchmarking for knowledge – 1

Benchmarking for knowledge can minimise/rationalise the benchmark efforts of the buyer (and consequently, the vendor).

Over the years many attempts are made to devise a good synthetic benchmark that uncovers all or most important system properties.

Recent/current examples:

- EuroBen
- HPCC
- NAS Parallel Benchmarks (NPB)
- SPEC

Benchmarking for knowledge – 2

A good synthetic benchmark should adhere to the following properties:

- Programs should not be too low-level (e.g., measure cache latencies)
- Programs should not be too complicated (interpretation of results becomes impossible)
- Programs should yield results for fundamental operations
- Programs should span important basic algorithms (based on fundamental operations)
- There should be a certain amount of redundancy (because of operation context)
- Basic algorithms should span most of the important application areas

So a certain amount of hierarchy is involved.

Benchmarking for knowledge – 3

In how far do the synthetic benchmarks mentioned have the properties we need?

Euroben:

Pro:

- Simple to run; short duration
- Internal correctness check
- Measures basic operations and algorithms. Possesses the desired hierarchy
- Has implementations with OpenMP and MPI

Con:

- Lots of output; result is not one figure of merit
- Lacks good I/O operation measurements
- Lacks basic algorithms for Molecular Dynamics and Bio-Informatics

Benchmarking for knowledge – 4

HPCC:

Pro:

- Finds useful lower/upperbounds of some system metrics
- Simple to run; short duration
- Internal correctness check

Con:

- Not easy to relate all metrics to actual performance
- Result is not one figure of merit

Benchmarking for knowledge – 5

The NAS Parallel benchmark (NPB)

Pro:

- Reasonably easy to run
- Addresses the CFD area well
- MPI and hybrid implementation available

Con:

- Codes, even the kernels, are too complicated to interprete results
- Too much geared to one application area
- No single figure of merit

Benchmarking for knowledge – 6

The SPEC benchmark (but which?)

Pro:

- Very rigorous run rules
- Wide variety of programs
- Results widely distributed

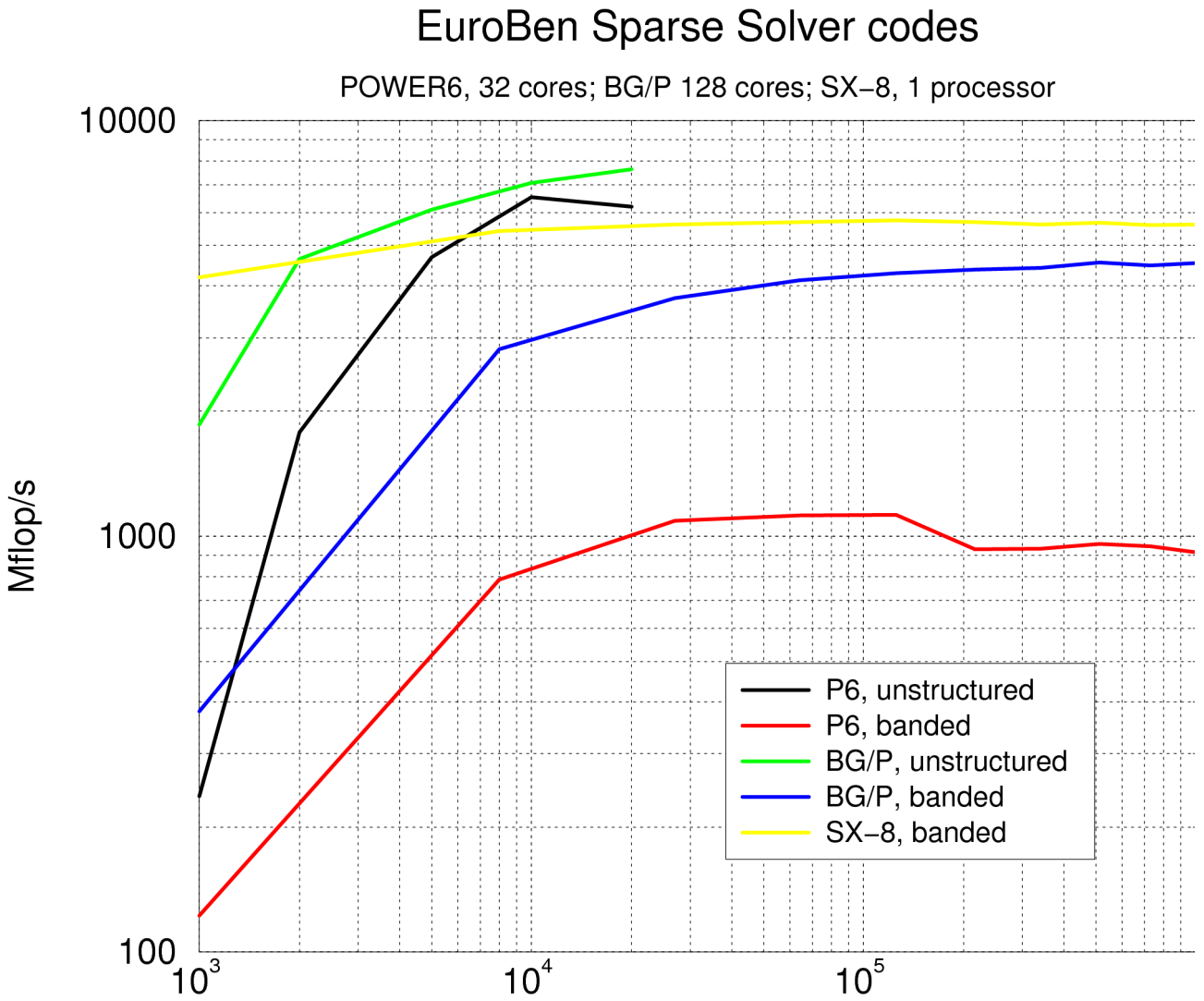
Con:

- Very rigorous run rules
- Wide variety of programs
- Codes much too complicated to interprete results
- HAS one figure of merit

Work in progress ... – 1

So what are we to do?

Synthetic benchmarks seem to have their use:



Work in progress ... – 3

In PRACE, Workpackage 6 is involved in the software aspects of Petascale systems — including benchmarking.

This entails:

- Selection of “example codes” from different application areas
- Porting, optimising, benchmarking these codes for prototype systems
- Devising a synthetic benchmark to complement these example codes

So, are we now in the same situation as with SPEC and the NPB?

Not entirely. But still much to do.

Work in progress ... – 4

Wish/work list for PRACE WP6 (and beyond):

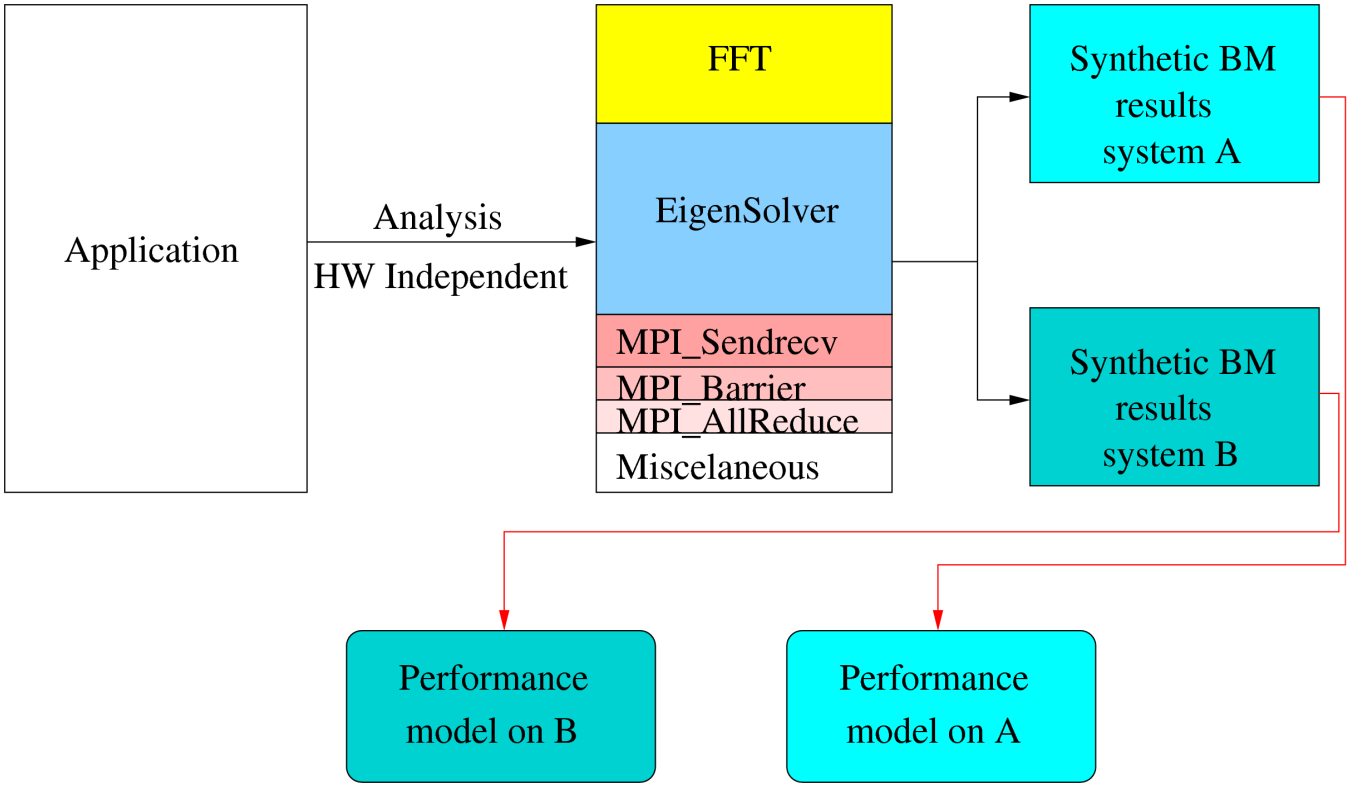
- Continuously adapt synthetic and application benchmarks (synthetic benchmark starts on basis of EuroBen benchmark, SkaMPI benchmark, and IOR benchmark)

- Extend the amount of platforms to port them to (e.g., computational accelerators, many-core processors)

Above all:
We need a tool to model application performance on the basis of the results from the synthetic PRACE benchmark.

Work in progress ... – 5

Such a tool will enable us to do more general performance predictions:



It will also enable us to pose “what if” questions.

NCF

PRACE

HPC Research

The End

Acknowlegdements:

– Arnold Meijster, Groningen University
For helping with running the EuroBen Benchmark on the BlueGene/P

– Rainer Keller, HLRS
For running the EuroBen Benchmark on the NEC SX-8

and

Thank you for your attention