CPU Benchmarking at GridKa
– Update Nov. 2007 –

Manfred Alef

Grid Computing Centre Karlsruhe (GridKa)

Forschungszentrum Karlsruhe
Institut für Wisscnschaftliches Rechnen
Hermann-von-Helmholtz-Platz 1
D-76344 Eggenstein-Leopoldshafen

http://www.fzk.de, http://www.gridka.de

firstname.lastname@iwr.fzk.de

FALL HEPiX 2007

CPU Benchmark Talks to Last HEPiX Conferences:

➔ Measurements in GridKa environment:

- Operating system: Scientific Linux 3 or 4, i386
- Compiler: gcc-3.4.x

➔ Benchmark used: SPEC CPU2000

➔ Metric: SPECint_base2000

Run 1 benchmark per core in parallel –
the total box performance is the sum of all results.

New: Switched to SPEC CPU2006:

➔ The SPEC CPU2000 benchmark suite has been retired by SPEC in February 2007 and replaced by the new CPU2006 suite.

Results submissions to SPEC's web page (http://www.spec.org) are no longer being accepted.

(However, the SPEC CPU2000 benchmark is still used within the HEP community, e.g. for accounting purposes in the WLCG project.)

New: Switched to "Rate" Metric:

➔ The SPEC CPU benchmarks provide 2 metrics:

- Speed: Is used for comparing the ability of a computer to complete single tasks.
- Rate: Measures the <u>throughput</u> of a machine carrying out a number of jobs.
  Multiple copies of the benchmarks are run simultaneously. (Typically, the # of copies is the same as the # of CPU cores on the machine.)

New: Switched to "Rate" Metric:

➔ In SPEC CPU2006, there is no longer a difference between the measurements of a speed run and a "1 copy" rate run.

(In contrast to that, these numbers differ by a factor of 0.011...0.012 when the SPEC CPU2000 suite is used.)

New: Integer + Floating Point Measurements:

➔ The SPEC CPU benchmarks provide 2 sets of software packages:

- Integer benchmarks (e.g., CINT2000 or CINT2006),
  C, C++.
- Floating point benchmarks (e.g., CFP2006),
  C, C++, FORTRAN 77, Fortran 90, Fortran 95.
  Compiler issues ...

More info: http://www.spec.org/cpu2006/CINT2006/
http://www.spec.org/cpu2006/CFP2006/

New: Using GCC 4.x for SPEC CPU2006 Measurements:

➔ **Fortran compiler:**

- GCC 3.x (g77) doesn't compile source code written in Fortran 90 or Fortran 95.
- GCC 4.x (gfortran) does it!

- *Scientific Linux 4 comes with*
  GCC 3.4.x (default) + GCC 4.1.x (gcc4, g++4, gfortran).

Differences between SPEC CPU2000 and CPU2006:

➔ **Memory footprint:**

- Some of the packages coming with CPU2000 fit into the cache. They by-pass the memory bottleneck.
- The new benchmarks will use much more RAM and can't escape from a memory bottleneck.

  At least 1 GB (i386) or 2 GB RAM (x86_64) per benchmark copy are required.

Differences between SPEC CPU2000 and CPU2006:

➔ **Benchmark workload:**

The new benchmarks will use a much bigger workload.

- The run time of the CINT2000 integer benchmark suite was about 2 hours on recent cluster hardware.
- A CINT2006 run takes at least one day.

(Reportable runs, "size=ref".)

Benchmark Results:

➔ **Benchmark environment – worker nodes at GridKa:**

◆ Intel Pentium-3 1.26 GHz	*2001 – 2002*
Intel Xeon 3.06 GHz	*2004*
AMD Opteron 270 (2.00 GHz dual-core)	*2006*
Intel Xeon 5160 (3.00 GHz dual-core)	*2007*
Intel Xeon E5345 (2.33 GHz quad-core)	*2007 – 2008*

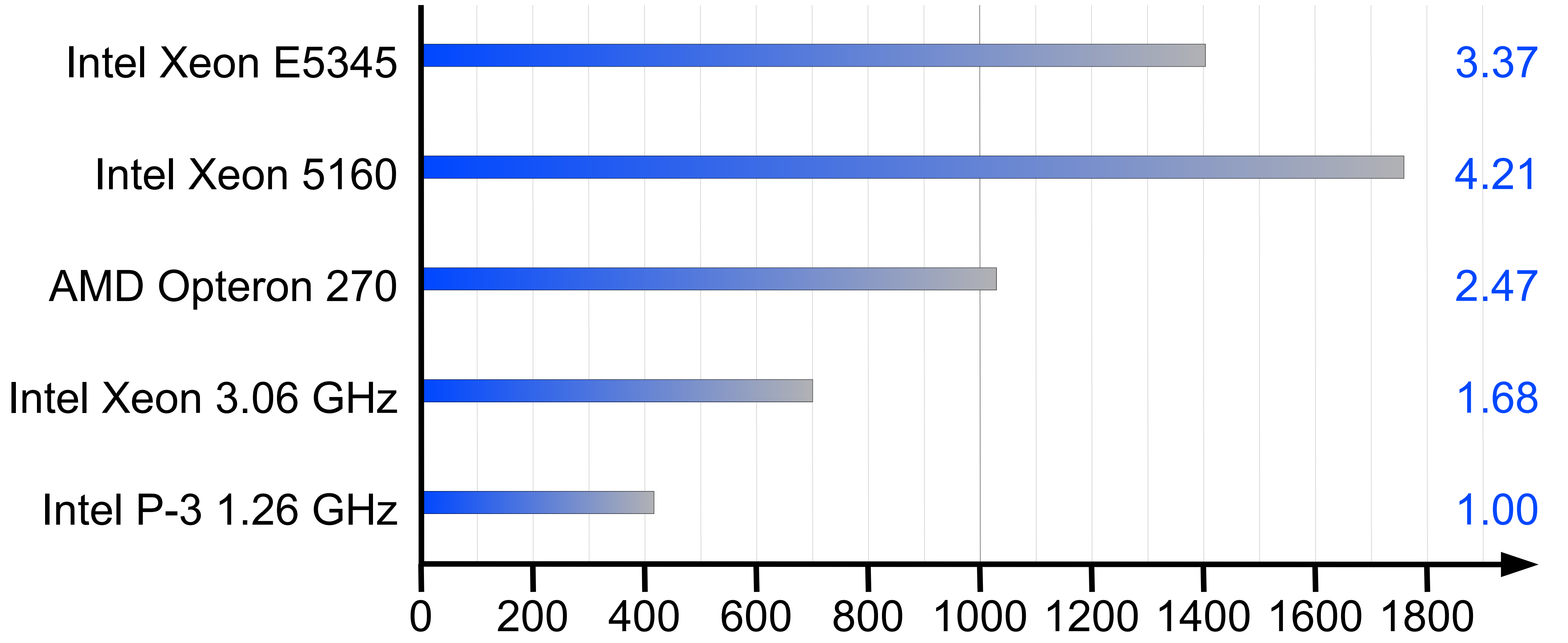
Dual CPU, 0.5-2 GB RAM per core, 1 hard disk (E5345: 2 disks)

Benchmark Results:

➔ **General remarks about optimizing flags:**

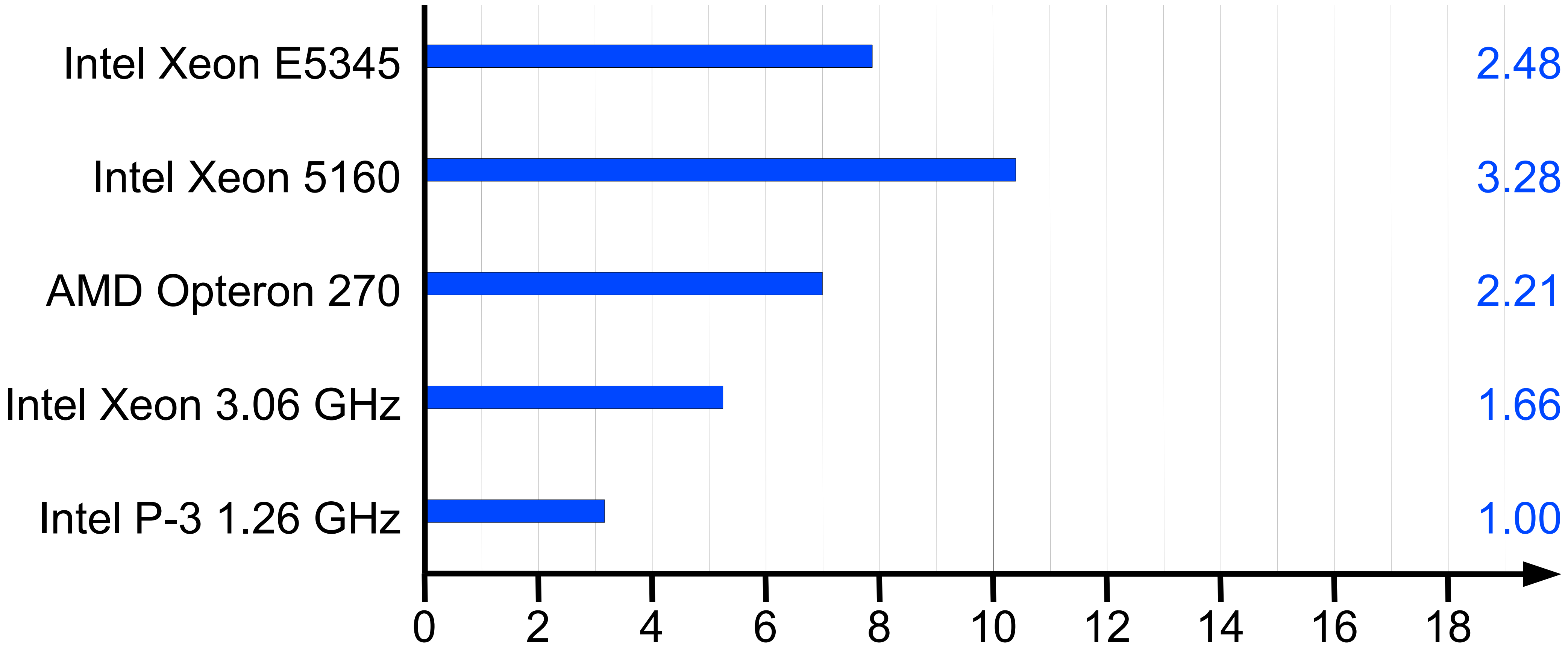
- “HIGH optimization” (measure “best” performance):
  -O3 -funroll-loops [-march]
- “LOW optimization” (flags used by LHC experiments):
  -O2 -pthread -fPIC

Average CPU Core Performance (SPECint_base2000, i386)



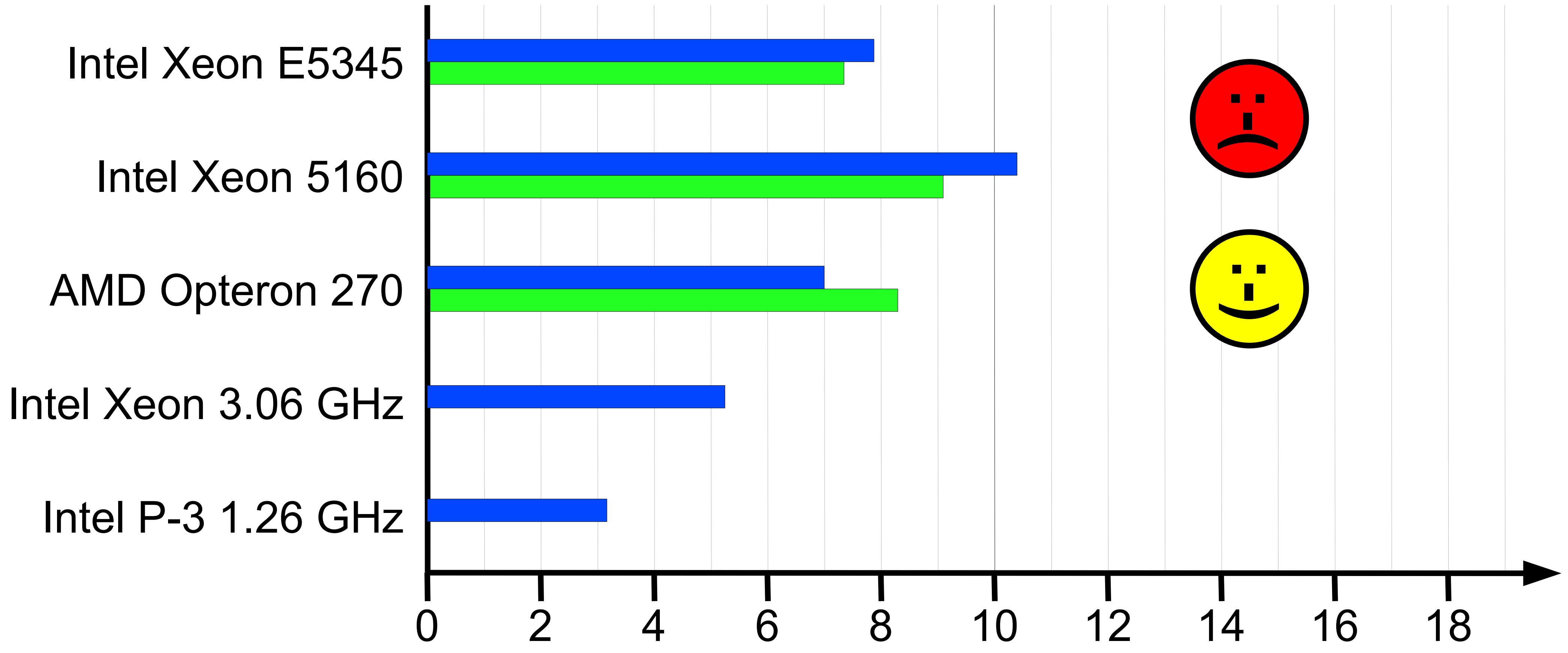
SL 3/4 i386, gcc-3.4.x ▨ HIGH optimization, 1 benchmark copy per core, average performance per core

Average CPU Core Performance (SPECint_rate_base2006, i386)



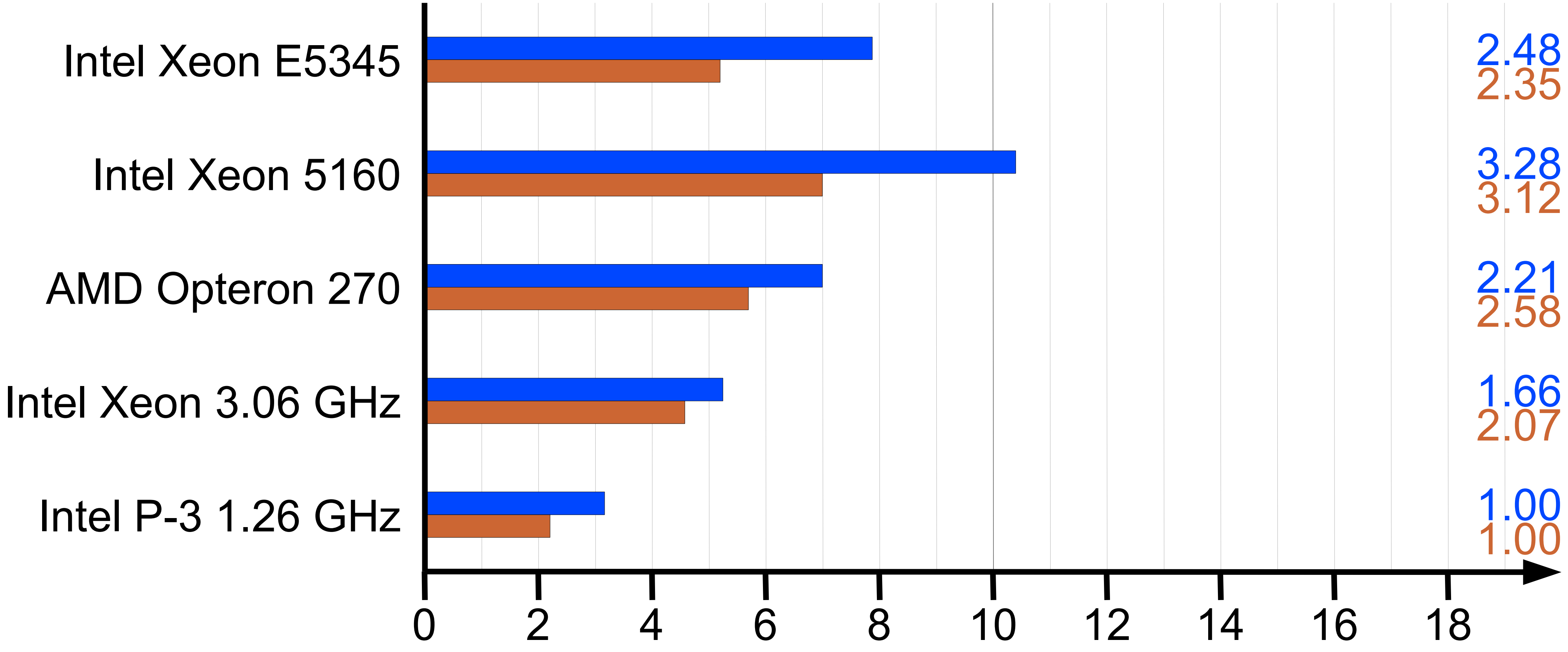
SL 3/4 i386, gcc-4.1.x 🟦 HIGH optimization, 1 benchmark copy per core, average performance per core

Average CPU Core Performance (SPECint_rate_base2006, i386 / x86_64)



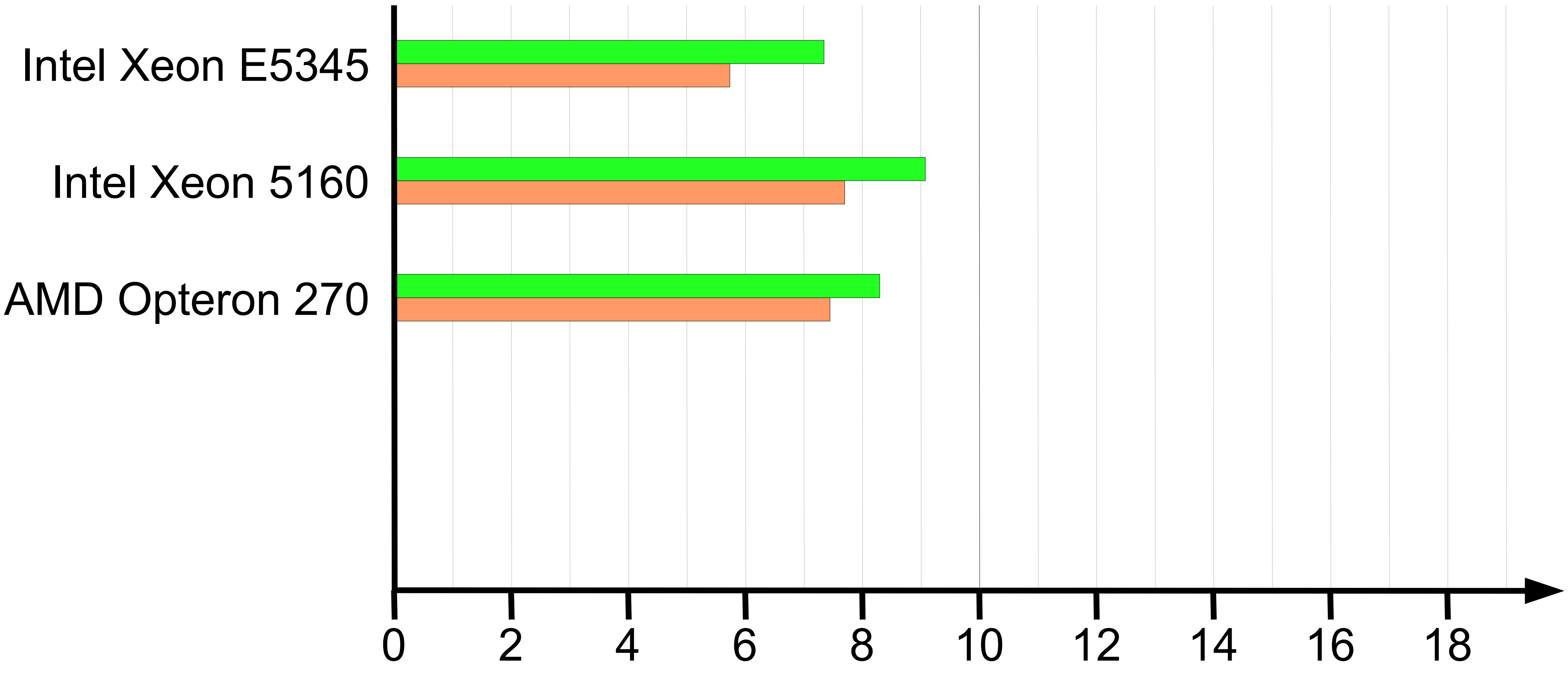
SL 4 **i386**, gcc-4.1.x 🟦 HIGH optimization
SL 4 **x86_64**, gcc-4.1.x 🟩 HIGH optimization
1 benchmark copy per core, average performance per core

Average CPU Core Performance (SPECint / fp_rate_base2006, i386)

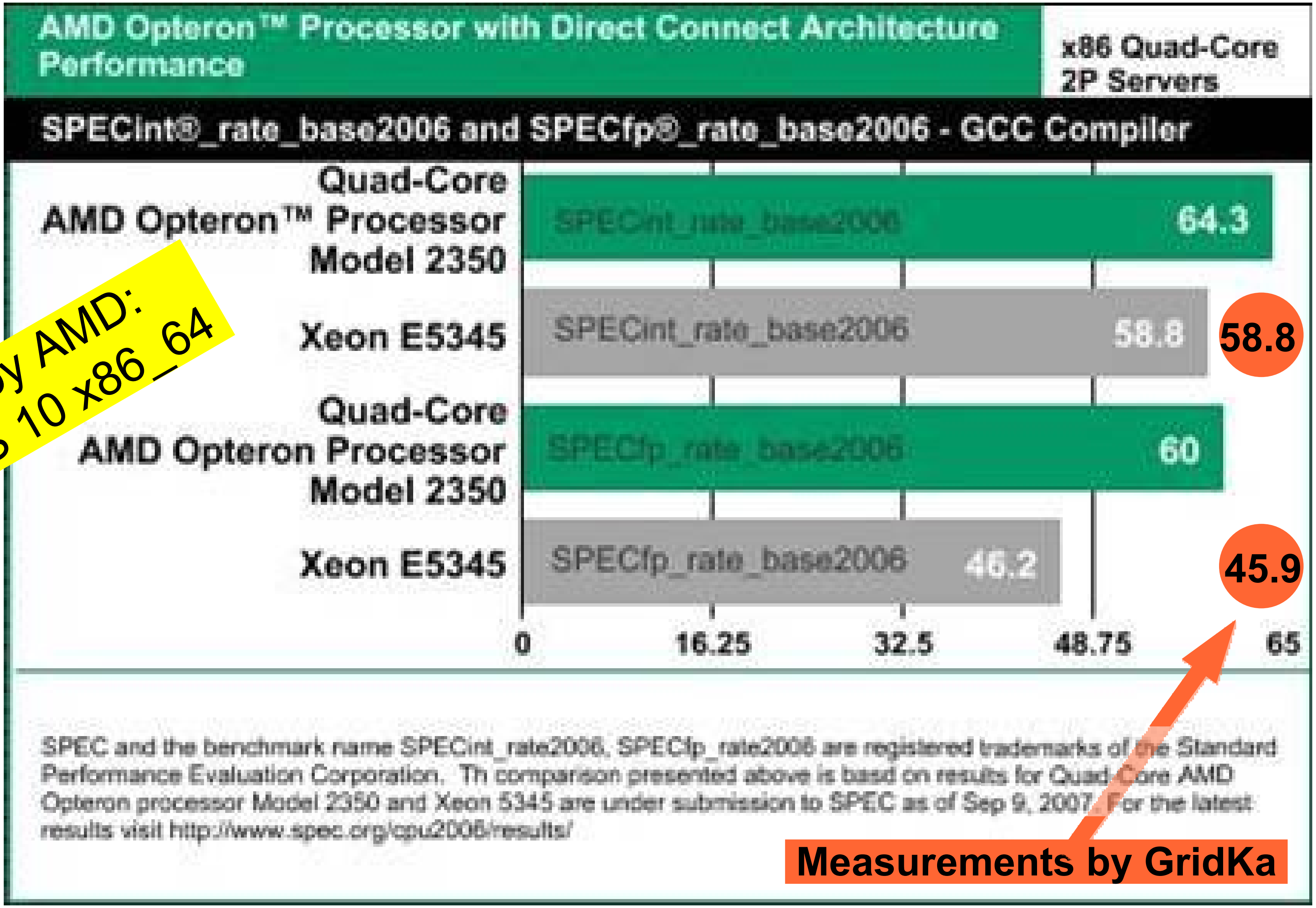


CINT2006: SL 4 i386, gcc-4.1.x 🟦 HIGH optimization
CFP2006: SL 4 i386, gcc-4.1.x 🟧 HIGH optimization
1 benchmark copy per core, average performance per core

Average CPU Core Performance (SPECint / fp_rate_base2006, x86_64)



CINT2006: SL 4 x86_64, gcc-4.1.x 🟩 HIGH optimization
CFP2006: SL 4 x86_64, gcc-4.1.x 🟧 HIGH optimization
1 benchmark copy per core, average performance per core



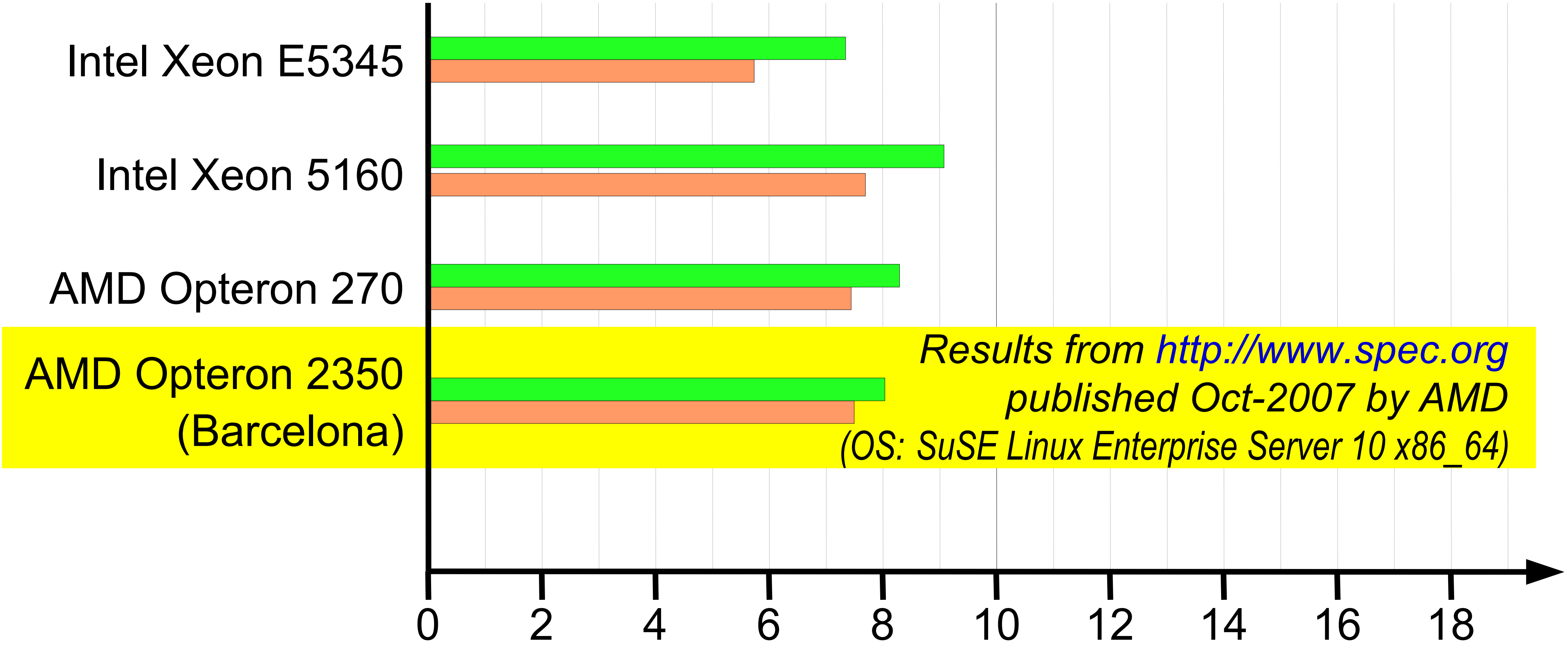
SPEC and the benchmark name SPECint_rate2006, SPECfp_rate2006 are registered trademarks of the Standard Performance Evaluation Corporation. Th comparison presented above is basd on results for Quad-Core AMD Opteron processor Model 2350 and Xeon 5345 are under submission to SPEC as of Sep 9, 2007. For the latest results visit http://www.spec.org/cpu2006/results/

All results by AMD:
OS=SLES 10 x86_64

Measurements by GridKa

http://developer.amd.com/content/images/misc/SPEC-gcc_2006-09-12.jpg

Average CPU Core Performance (SPECint / fp_rate_base2006, x86_64)



Results from http://www.spec.org
published Oct-2007 by AMD
(OS: SuSE Linux Enterprise Server 10 x86_64)

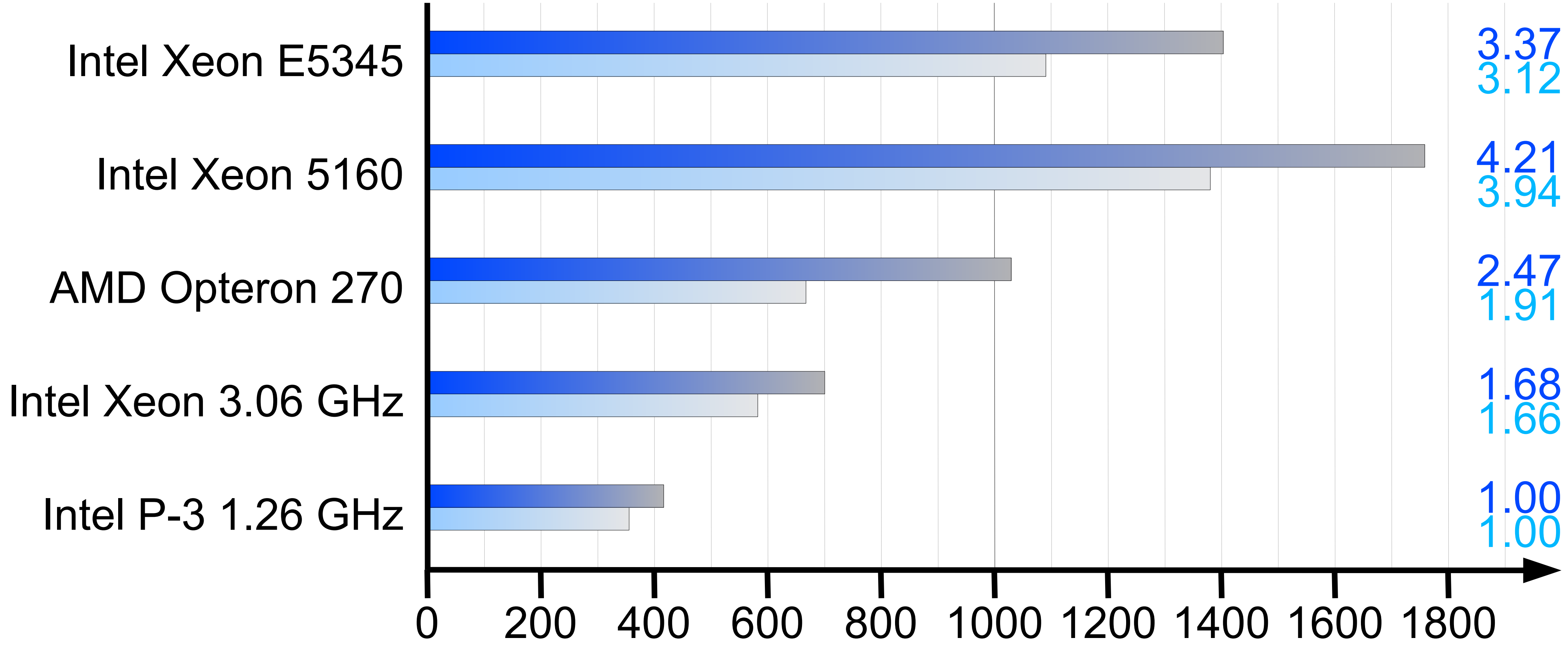
CINT2006: SL 4 x86_64, gcc-4.1.x 🟩 HIGH optimization
CFP2006: SL 4 x86_64, gcc-4.1.x 🟧 HIGH optimization
1 benchmark copy per core, average performance per core

Benchmark Results – “LOW” Tuning:

➔ Some HEP experiments are using another set of compiler flags:
-O2 -pthread -fPIC

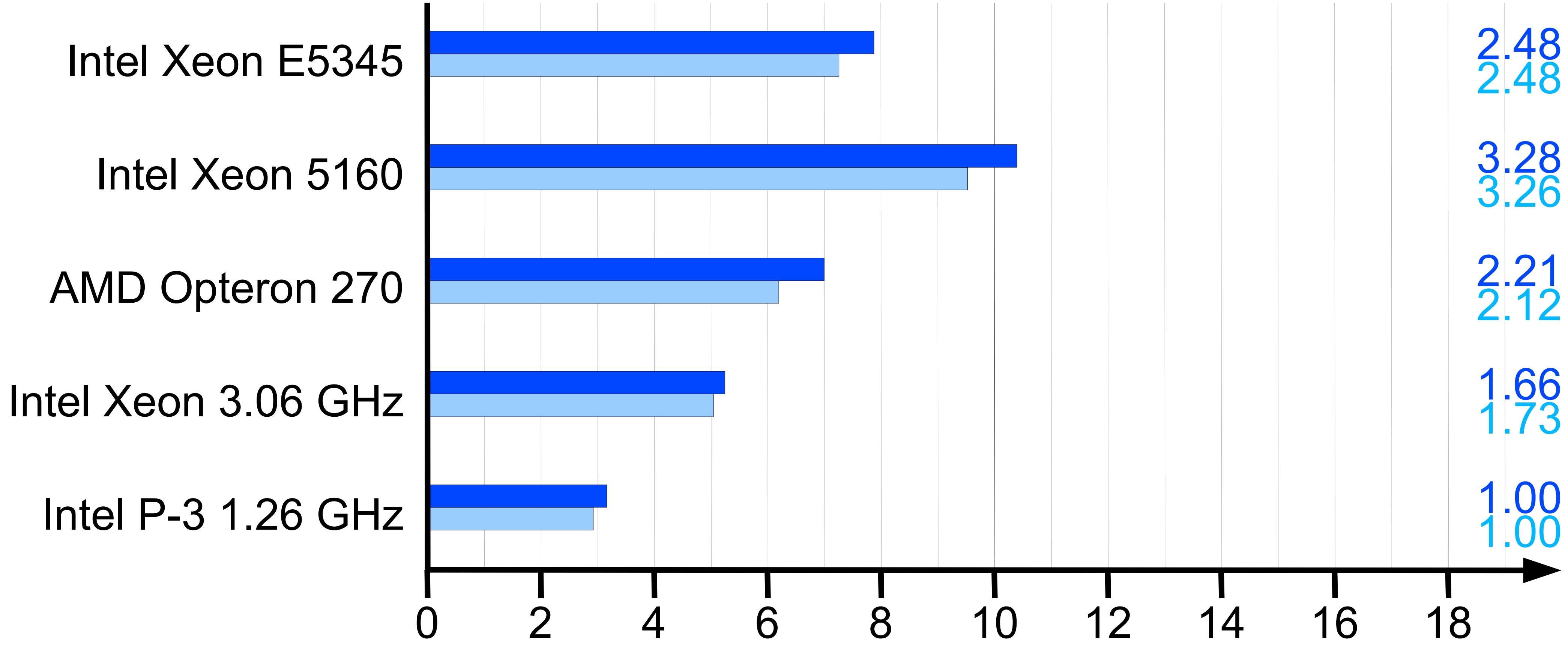
➔ How do the results differ?

Average CPU Core Performance (SPECint_base2000, i386)



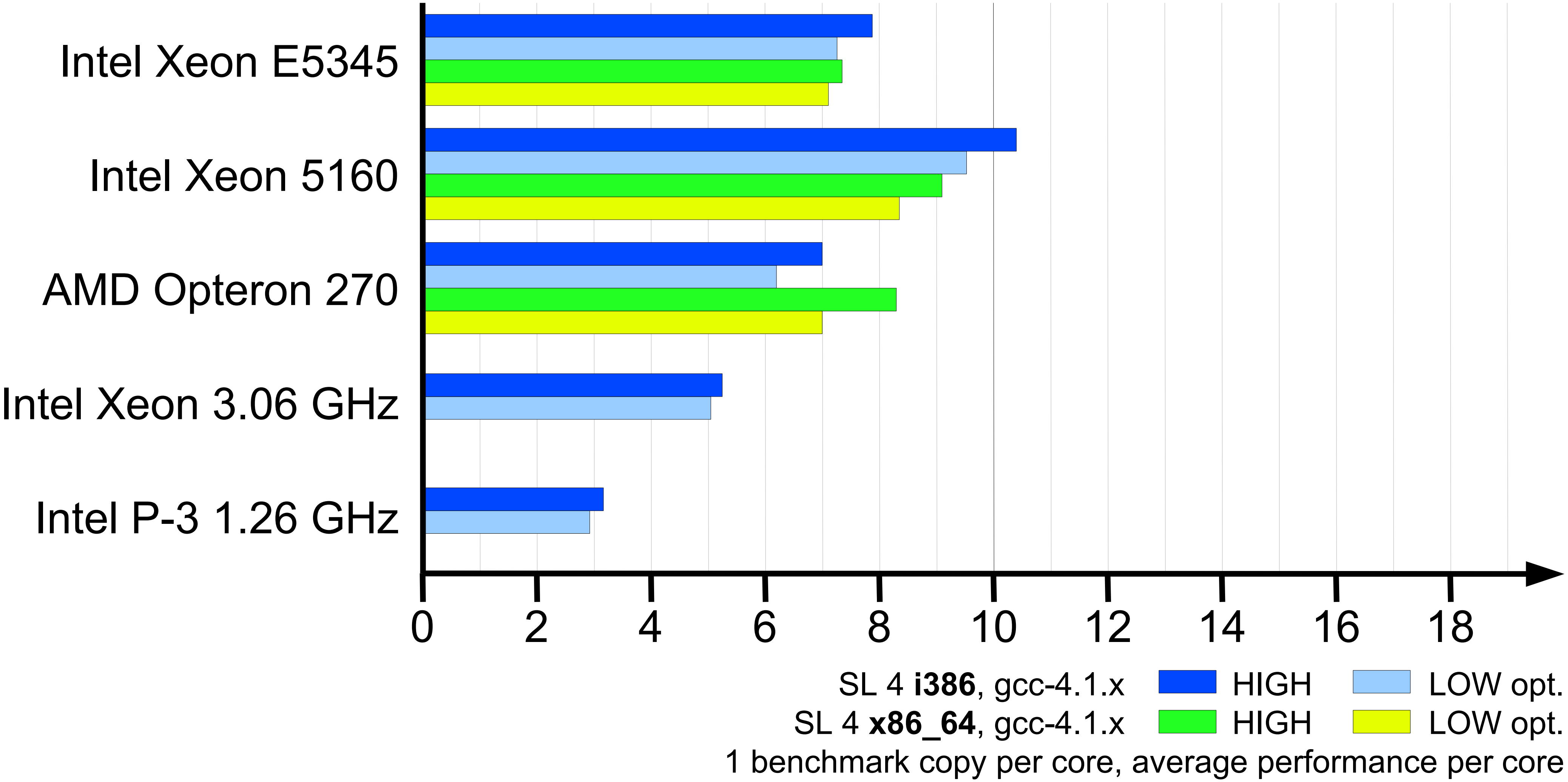
SL 3/4 i386, gcc-3.4.x **HIGH** **LOW** opt., 1 benchmark copy per core, average performance per core

Average CPU Core Performance (SPECint_rate_base2006, i386)

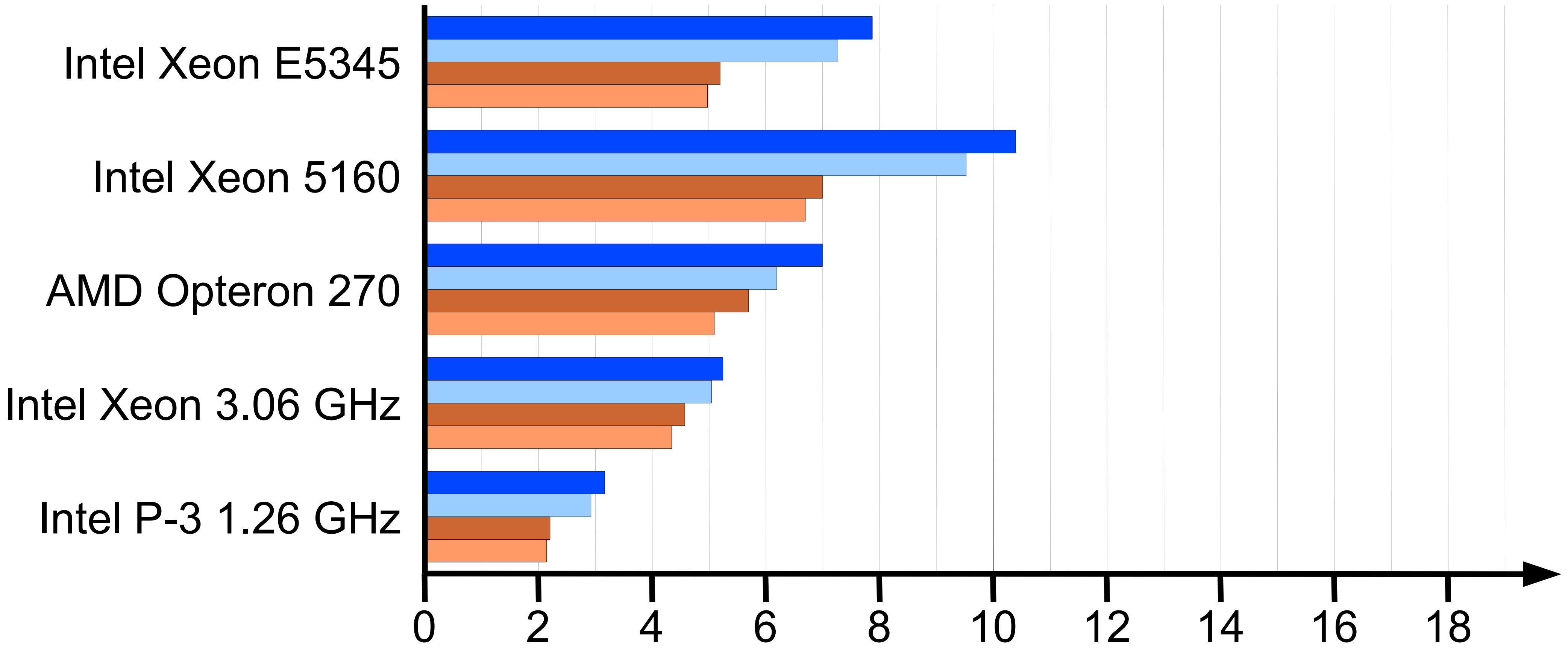


SL 3/4 i386, gcc-4.1.x ■ **HIGH** ■ **LOW** opt., 1 benchmark copy per core, average performance per core

Average CPU Core Performance (SPECint_rate_base2006, i386 / x86_64)



Average CPU Core Performance (SPECint / fp_rate_base2006, i386)



CINT2006: SL 4 i386, gcc-4.1.x 🟦 HIGH 🟦 LOW opt.
CFP2006: SL 4 i386, gcc-4.1.x 🟧 HIGH 🟧 LOW opt.
1 benchmark copy per core, average performance per core

Conclusions (1):

- ➔ SPEC CPU2000 measurements on recent hardware are somewhat too optimistic.
- ➔ SPEC CPU2000 has been retired and replaced by its successor CPU2006.

Conclusions (2):

➔ SPEC CPU2006 results:

- Larger memory footprint → cache issues removed.
- No difference in ramp-up 2001–2007 between high and low optimized measurements!
- Performance 32bit → 64bit OS:
  Opteron +10% :-) , Xeon –10% :-(.
- Floating point measurements:
  Opteron ≈ 10% faster than Xeon
  (compared with integer performance).

Comments, Questions?