The influence of two modern compiler infrastructures on energy consumption of the HPCG benchmark

TECHNISCHE UNIVERSITÄT DARMSTADT

EnA-HPC @ ISC 2017

Armin Jaeger
armin.jaeger@hrz.tu-darmstadt.de

Hochschulrechenzentrum, TU Darmstadt

joint work with
Jan-Patrick Lehr
and
Christian Bischof

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- The compiler is an integral part in every software production pipeline.
  - Traditional focus on “time to solution”.
- Our focus: To which degree the choice of the compiler influences a software’s energy demands.
- Tradeoffs concerning metrics like “energy to solution” and “time to solution”.
- HPCG benchmark is a good example of HPC workload.

Hardware and Software Platform

- Intel system based on four Xeon E7-4890v2 processors (*Ivy Bridge*) with a thermal design power (TDP) of 155 Watt and a main memory of 1024 GB.
- The processor's frequency spectrum ranges from 1.2 GHz to 2.8 GHz with a step size of 200 MHz, and in addition a turbo mode (TM).
- The memory bandwidth depends on the CPU frequency setting, therefore we are interested on the impact on the HPCG benchmark.

- "**LIKWID**" to access the Running Average Power Limit (**RAPL**) counter via model-specific registers (MSR).
  The framework was configured to collect the data for **PWR_PKG_ENERGY** (PWR0 Counter). Additionally we use it for monitoring the core binding and frequency settings.
- "**taskset**" to bind processes or threads to the appropriate cpu-cores.
- "**ipmi-tool**" to validate of the RAPL observations through IPMI facility.

Compiler infrastructures and software versions

TECHNISCHE UNIVERSITÄT DARMSTADT

- GNU Compiler Collection (**GCC**) in version 5.2.0.
- Intel Compiler Suite (**ICC**) in version 2016 (vanilla).

- The target application is compiled with the different available optimization flags $O\{0|1|2|3|\text{fast}\}$.
- All compiled binaries support MPI and OpenMP.
- "**LIKWID**" in version 4.0.1.
- The high-performance conjugate gradient benchmark ("**HPCG**") in version 2.4.

Baseline measurement and the application energy consumption

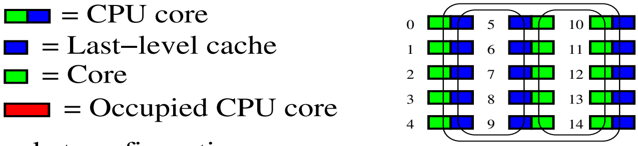
TECHNISCHE UNIVERSITÄT DARMSTADT

- We assume a significant service reduction allows a less perturbed approximation of the average application energy consumption.
- We strip down the production machine image by removing unnecessary services.
- For the idle-mode we measure the energy consumption of the socket for a timeframe of 300 seconds (10 runs per frequency).
  - With services the average is around 6000 J.
  - Stripped down the average is around 3000 J.
- This baseline measurement is subtracted from the measured energy consumption of the socket to derive the application energy consumption.
  - The measurement process is performed for every frequency.

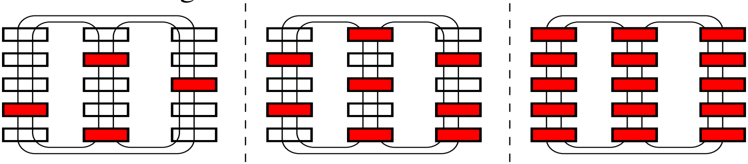
Benchmark setup I – Thread Distribution

TECHNISCHE
UNIVERSITÄT
DARMSTADT

CPU core 1–socket scheme



1–socket configurations:



4 , 8 , 15 threads

Benchmark setup II – measurement

TECHNISCHE UNIVERSITÄT DARMSTADT

A run must fulfill

three criteria to be considered as “valid” run:

1) No MSR register overflow occured.
2) Frequency setup and core-binding pattern are successfully validated.
3) HPCG’s internal validation tests were passed.

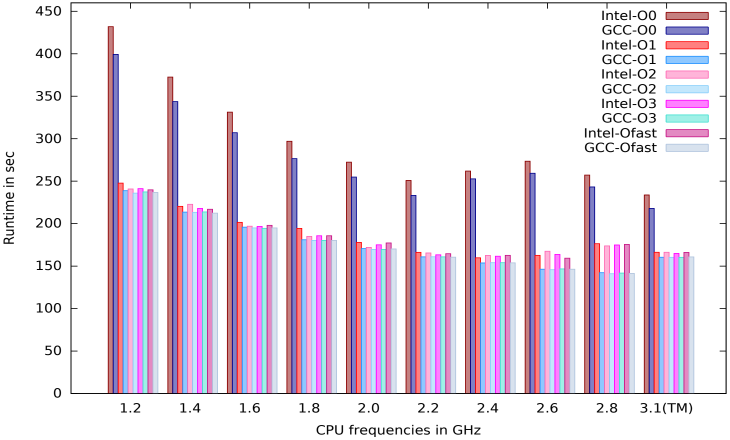
For every configuration,
we obtain at least 5 “valid” runs to determine the arithmetic mean value.
Each configuration consists of

a) selected processor frequency (10)
b) thread distribution (3)
c) compiler version (2)
d) and compiler optimization level (5).

Note: The corrected sample standard deviation is $\leq$ 3% for inbound and $\leq$ 5% for outbound measurements.

Application runtime
HPCG – 1 socket – Ivy Bridge – 4 cores

TECHNISCHE UNIVERSITÄT DARMSTADT



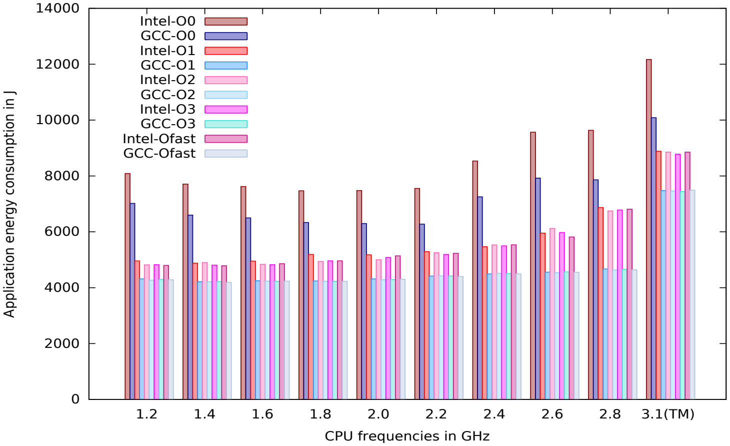
Application runtime
HPCG – 1 socket – Ivy Bridge – 4 cores

TECHNISCHE UNIVERSITÄT DARMSTADT

- In all cases the GCC binary runs faster than to the ICC binary.
- The shortest application runtime for the GCC is at 2.8 GHz.
- The runtime at turbo mode is at the same level as the 2.2 GHz configuration.
- At 2.8 GHz and O1-Ofast the GCC binary is around 18% faster than the ICC binary.

Application energy consumption
HPCG – 1 socket – Ivy Bridge – 4 cores

TECHNISCHE UNIVERSITÄT DARMSTADT



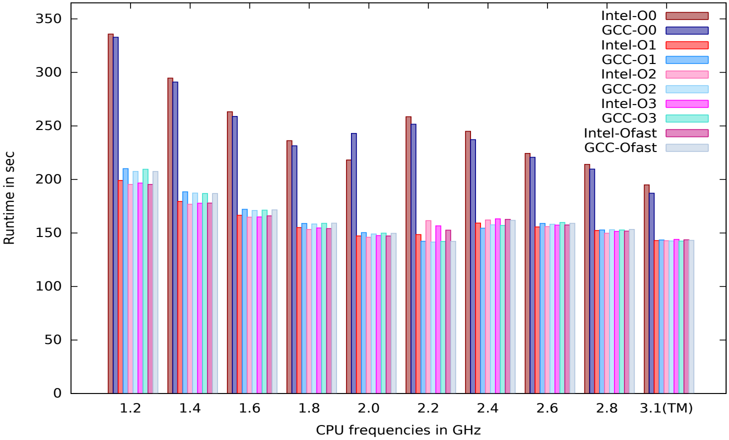
Application energy consumption
HPCG – 1 socket – Ivy Bridge – 4 cores

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Corrected sample standard deviation of 2.5%.
- The energy consumption stays practically at the same across the different frequencies for GCC with optimization level O1-Ofast up to 2.8 GHz, even though the runtime decreases with the optimization level.
- ICC shows increasing energy consumption with increasing frequency.
- For both compilers, Turbo Mode significantly increases energy consumption, but slows down the code compared to 2.8 GHz.
- At 2.8 GHz and O1-Ofast the GCC binary uses around 31% less energy than the ICC binary.

Application runtime
HPCG – 1 socket – Ivy Bridge – 15 cores

TECHNISCHE UNIVERSITÄT DARMSTADT



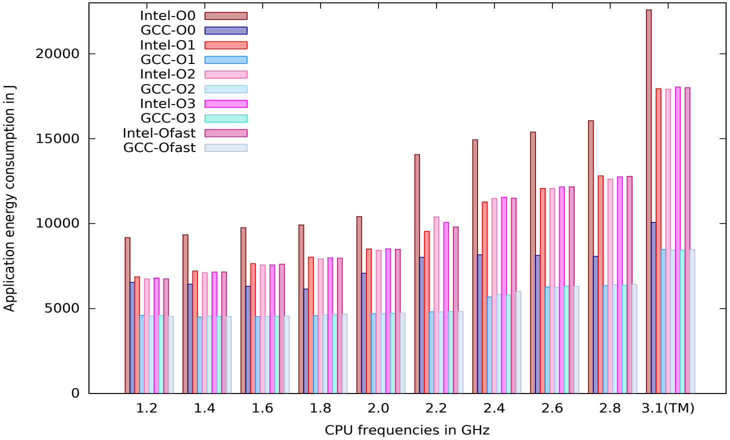
HRZ

Application runtime
HPCG – 1 socket – Ivy Bridge – 15 cores

- Up to 2.0 GHz the ICC binary is somewhat faster than the GCC binary.
- For 2.2 GHz and 2.4 GHz the GCC binary is a bit faster.
- For 2.6 GHz and above the two compilers binaries are nearly on par.

Application energy consumption
HPCG – 1 socket – ivy-bridge – 15 cores

TECHNISCHE UNIVERSITÄT DARMSTADT



HRZ

Application energy consumption
HPCG – 1 socket – Ivy Bridge – 15 cores

TECHNISCHE UNIVERSITÄT DARMSTADT

- Corrected sample standard deviation of 1.5%.
- Up to 2.2 GHz the energy consumption for GCC with O1-Ofast stays at same level.
- Increase of application energy consumption for GCC from 2.4 GHz up to 2.8 GHz (17-24%) and then an additional 25% for the Turbo Mode.
- The ICC shows increasing energy consumption with increasing frequency, a factor of around 2.2 between 1.2 GHz and Turbo Mode.
- The GCC requires less energy (up to 55%) than ICC in all configurations over the whole frequency spectrum
- The ICC binary consumes considerable more energy than the GCC binary - in the worst case up to twice the amount of energy.

{min, max}-values of runtime in seconds and energy consumption in KJoule for GCC 5.2.0 and ICC 2016 binaries for different thread counts.

NOTE:

- application energy consumption values
- values in both columns do not necessarily denote the same configuration!

#threads	*t* in *sec* {min, max}	Energy in *kJ* {min, max}	compiler
4	{141, 399}	{4.2, 10.1}	GCC
4	{159, 432}	{4.8, 12.2}	Intel
8	{139, 355}	{4.2, 9.7}	GCC
8	{144, 366}	{5.2, 15.4}	Intel
15	{142, 333}	{4.5, 10.1}	GCC
15	{143, 336}	{6.8, 22.6}	Intel

Validation of the current RAPL observation via IPMI

HPCG benchmark with 60 Threads and O2 optimization level:

- doubling the default input values
- measuring the whole node.

f	t	Energy in kJ	compiler
3.1(TM)	882	122	GCC
3.1(TM)	825	194	Intel
2.8	947	78	GCC
2.8	982	121	Intel
1.2	1576	81	GCC
1.2	1568	107	Intel

The difference in energy consumption is still significant, but because of the additional consumers (memory, fans, mainboard and others) not the same.

Summary and future work

TECHNISCHE UNIVERSITÄT DARMSTADT

- The compiler can have a strong influence on energy consumption.
- The **time to solution** does not differ much in most cases.
- The **energy to solution** changes significantly depending on the underlying compiler infrastructure.
- The benchmarking setup requires to be aware of other system services and counter overflows.
- In the future we are planing to investigate for application energy consumption of HPCG, HPL, STREAM for a higher core count (at least 60 with RAPL) and different CPU architectures.

Thank you for your attention!!!

Any questions?