Aalto University
School of Science
Degree Programme in Computer Science and Engineering

Gonçalo Marques Pestana

Energy Efficiency in High Throughput Computing

Tools, techniques and experiments

Master's Thesis
Espoo, 1 December, 2014

DRAFT! — October 28, 2014 — DRAFT!

Supervisors:	Professor Jukka K. Nurminen
Advisor:	Zhonghong Ou (Post-Doc.)

Aalto University
School of Science

Aalto University
School of Science
Degree Programme in Computer Science and Engineering

ABSTRACT OF
MASTER'S THESIS

Author:	Gonçalo Marques Pestana		
Title: Energy Efficiency in High Throughput Computing Tools, techniques and experiments			
Date:	1 December, 2014	**Pages:**	28
Major:	Data Communication Software	**Code:**	T-110
Supervisors:	Professor Jukka K. Nurminen		
Advisor:	Zhonghong Ou (Post-Doc.)		
abstract			
Keywords:	energy efficiency, scientific computing, ARM, Intel, RAPL, tools, techiques		
Language:	English		

Acknowledgements

I wish to thank all students who use LaTeX for formatting their theses, because theses formatted with LaTeX are just so nice.

Thank you, and keep up the good work!

Espoo, 1 December, 2014

Gonçalo Marques Pestana

Abbreviations and Acronyms

2k/4k/8k mode	COFDM operation modes
3GPP	3rd Generation Partnership Project
ESP	Encapsulating Security Payload; An IPsec security protocol
FLUTE	The File Delivery over Unidirectional Transport protocol
e.g.	for example (do not list here this kind of common acronymbs or abbreviations, but only those that are essential for understanding the content of your thesis.
note	Note also, that this list is not compulsory, and should be omitted if you have only few abbreviations

Contents

Chapter 1

Introduction

Chapter 2

Background

2.1 Energy consumption in Scientific Computing

2.1.1 Literature review

According to [2], the computing requirements for HPC have increased particularly in recent years. Projects of the magnitude and complexity of the Large Hadron Collider are overwhelming examples of that fact. To achieve results like the discovering of the Higgs boson and other significant scientific advances, it was necessary a to distribute the processing tasks across several partners and institutions through the WLCG. The equivalent capacity of such distributed systems was between 80,000 and 100,000 x86-64 cores in 2012. Further projects and discoveries will demand even more processing capacity from the WLCG. For example, as stated by [2], to upgrade the LHC detectors luminosity to its full power the datasets will increase sizes by 2-3 orders of magnitude and processing power will have to increase in proportion.

2.2 High Throughput Computing

2.2.1 Literature review

2.3 CERN and the LHC experiment

2.3.1 Literature review

2.4 Energy performance and measurement

2.4.1 Literature review

on importance of energy consumption for engineers and scientists The study conducted by [3], shows that engineers have been considering energy consumption as an important factor when developing software. It consists on an empirical study that aims to understand the opinions and problems of software developers about energy efficiency. The data that sustain the conclusions are mined from a well-known technical forum (*StackOverflow* [1]). Although the study is focused in an application-level energy efficiency, it shows that developers are aware of the importance of energy efficiency in computational systems. When trying to understand in depth what questions arise more frequently, it is shown that measurement techniques is amongst the most asked questions by developers. In addition, the study ascertains that the *"lack of tool support"* is an important handicap for the development of energy efficient software.

2.5 ARM architecture

2.5.1 Literature review

In [2]:

- After 2015, processors have hit scaling limits. Two different paths started to be taken on the processor industry: development of multiprocessor architectures that allow to run parallel tasks and the time clock frequency - which have been increasing throughout the years - stabilized.

- Most High Physics Computing systems run in clusters of several cores. Additional cores are parallelized and can run at the same time, which allows the system to scale. However, also commodities such memory, I/O streams and energy scale proportionally in such architectures.

Chapter 3

Tools and techniques for measuring energy efficiency of scientific software applications

Chapter 4

Experiments

4.1 Experiments methodology

The experiments were performed in different sets. Whereas the first two sets of experiments aim to provide a straightforward comparison between ARM and Intel technologies, the third set of experiments aims to study the influence of a NUMA environment in hight performance computing from an energy consumption perspective. In each set, we used different techniques and tools to perform the energy measurements. The techniques and tools used to perform the measurements are described and analyzed in depth in Section 3.

The first part of this chapter outlines the scope, methodology and measurement tools used for each set. The latest part shows the results of the experiments, which are analyzed in the next chapter.

Throughout the rest of the document, the different experiments will be termed as first (FSE), second (SSE) and third set of experiments (TSE).

4.1.1 First Set of Experiments

Done at Aalto. Explain methodology and scope.

4.1.2 Second Set of Experiments

Done at CERN. Explain methodology and scope.

4.1.3 Third Set of Experiments: RAPL in NUMA environment

Done at CERN. Explain methodology and scope.

4.2 Results

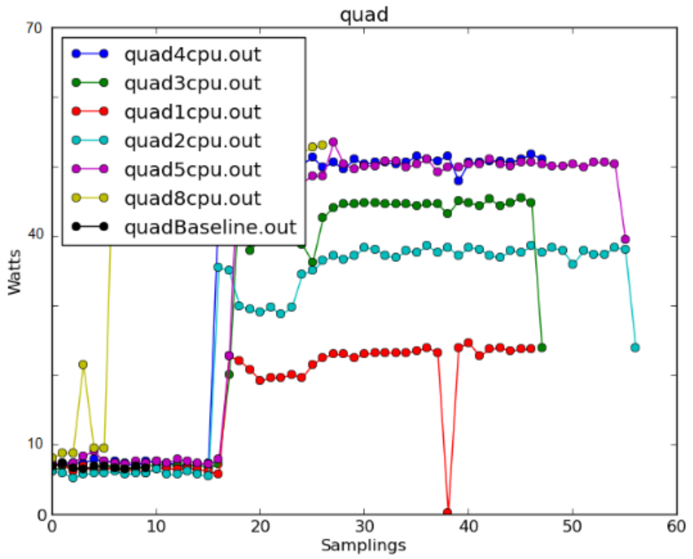


Figure 4.1: Full single threading CMS experiments on Intel Quad

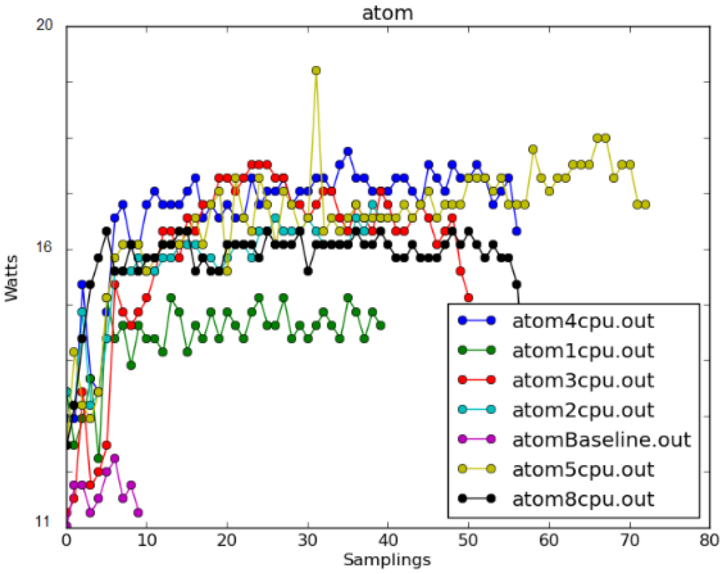


Figure 4.2: Full single threading CMS experiments on Intel Atom

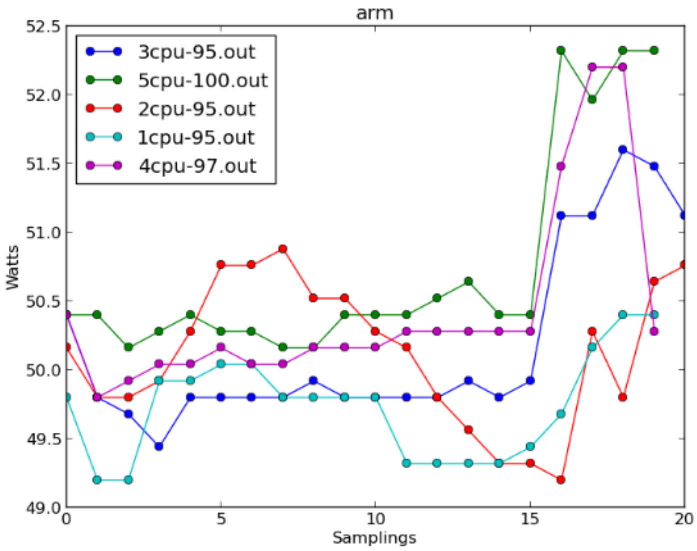


Figure 4.3: Full single threading CMS experiments on ARMv7 server

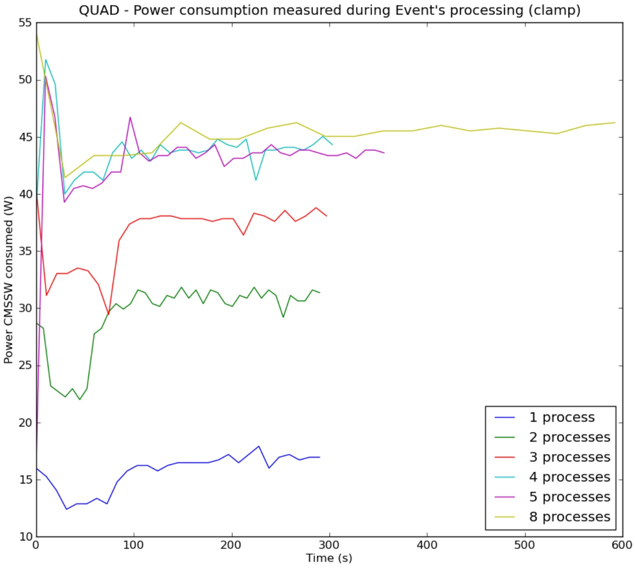


Figure 4.4: Full single threading CMS experiments on Intel Quad - event processing only

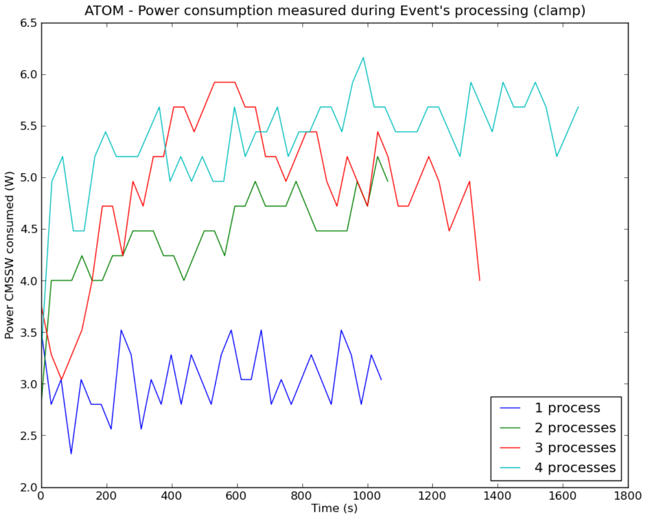


Figure 4.5: Full single threading CMS experiments on Intel Atom - event processing only

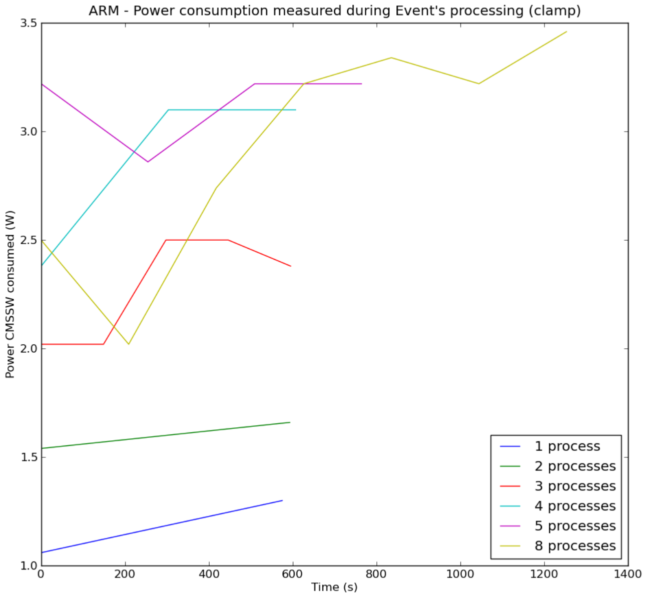


Figure 4.6: Full single threading CMS experiments on ARMv7 server - event processing only

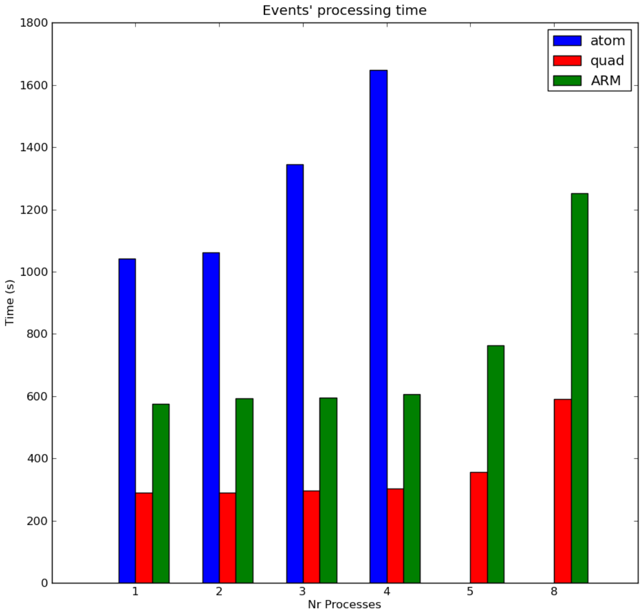


Figure 4.7: Processing time comparison

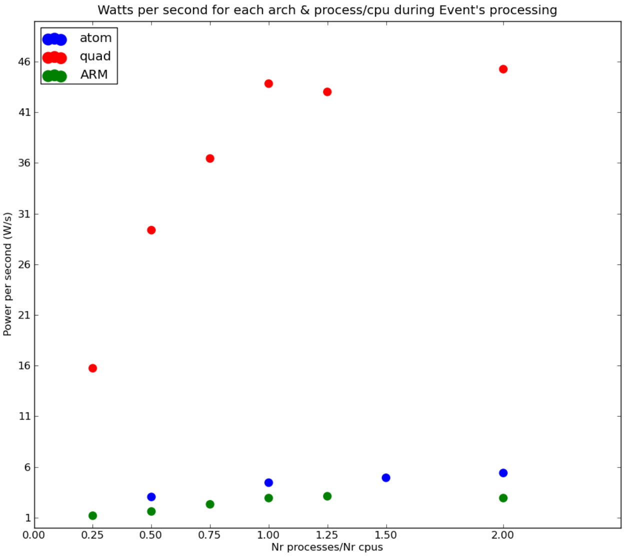


Figure 4.8: Processing stage comparison between architectures

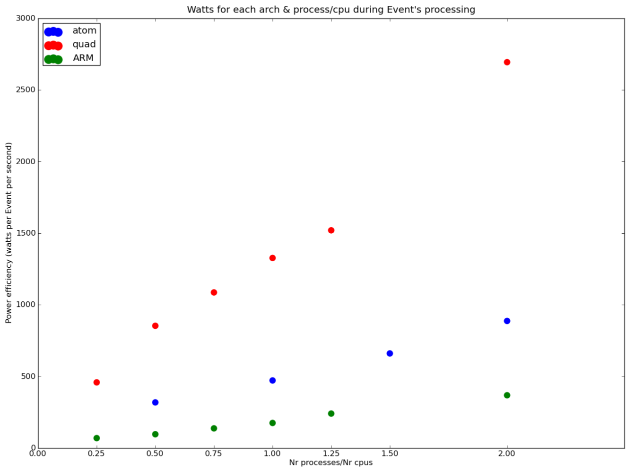


Figure 4.9: Processing stage comparison between architecturesi - 2

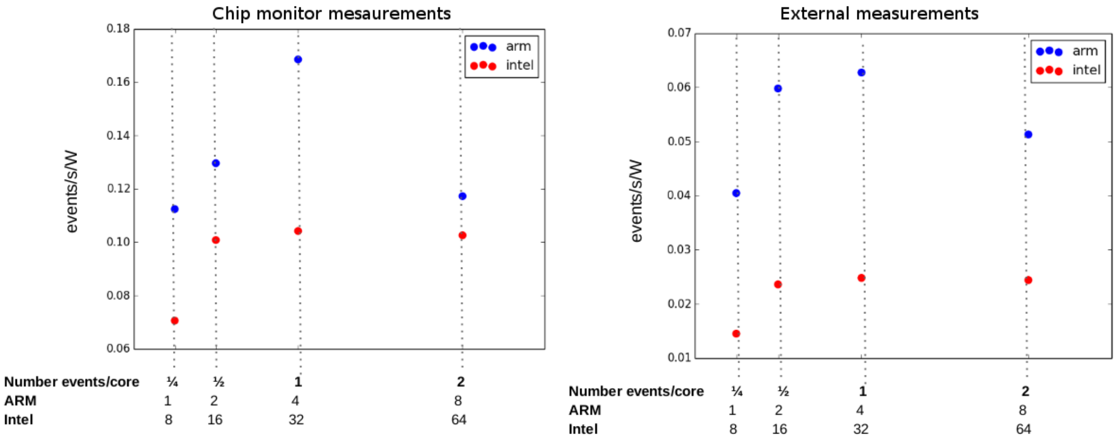
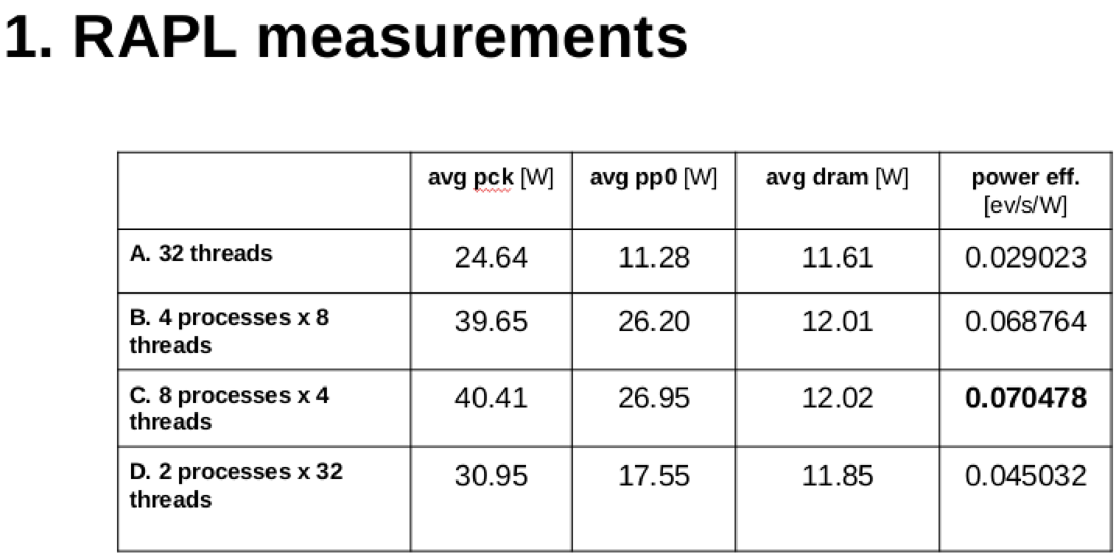


Figure 4.10: Multithreaded ParFullCMS comparison Intel Xeon vs ODROID ARMv7



1. RAPL measurements

	avg pck [W]	**avg pp0** [W]	**avg dram** [W]	**power eff.** [ev/s/W]
A. 32 threads	24.64	11.28	11.61	0.029023
B. 4 processes x 8 threads	39.65	26.20	12.01	0.068764
C. 8 processes x 4 threads	40.41	26.95	12.02	**0.070478**
D. 2 processes x 32 threads	30.95	17.55	11.85	0.045032

Figure 4.11: RAPL measurements with different load combinations

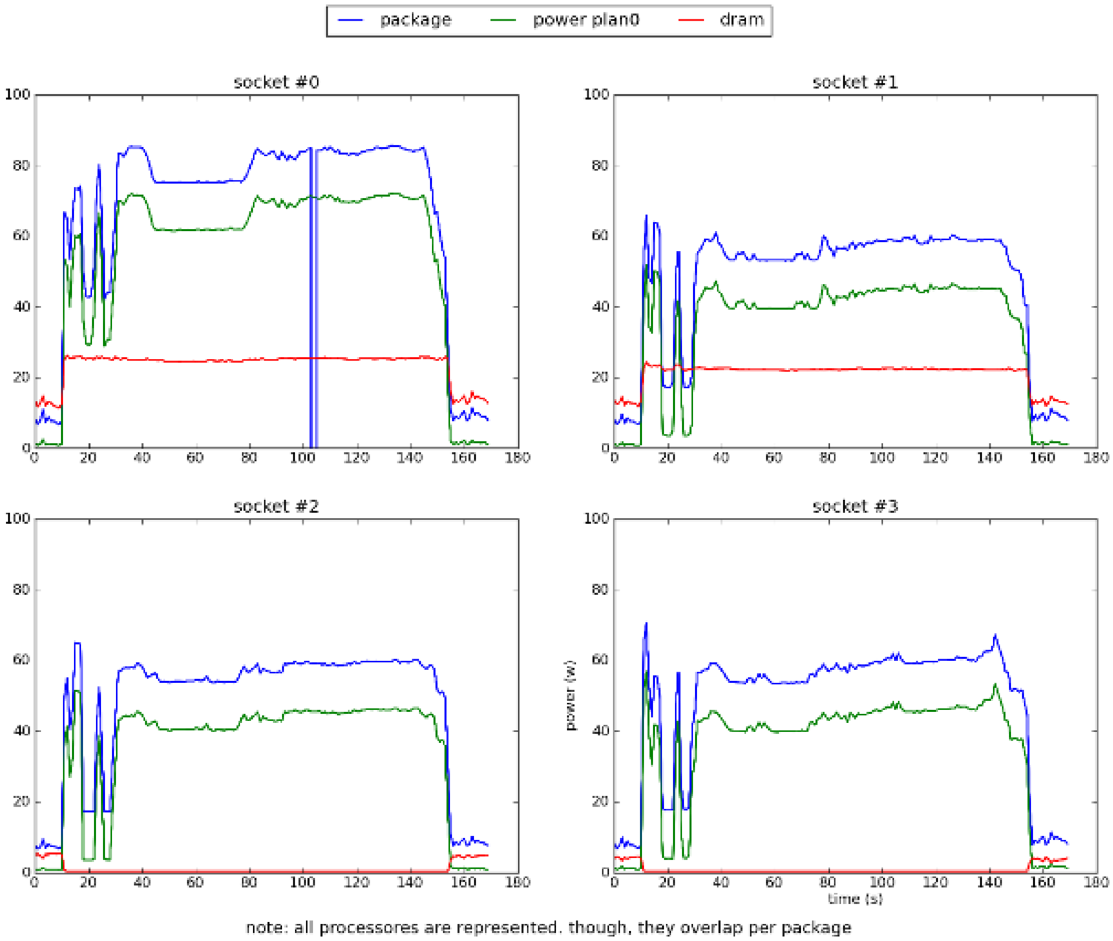


Figure 4.12: RAPL measurements of NUMA nodes - 16 processes with no explicit binding

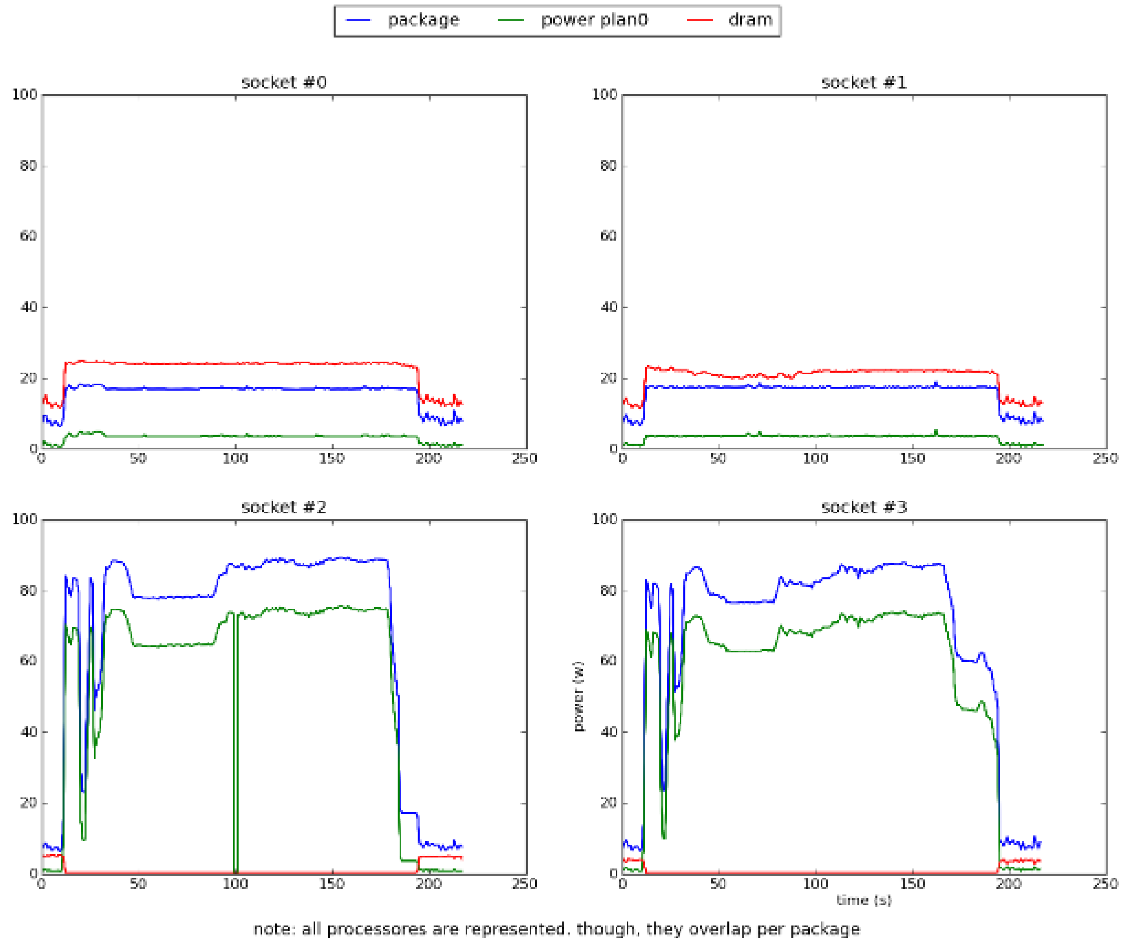


Figure 4.13: RAPL measurements of NUMA nodes - 16 processes. Explicit binding on node #2 and node #3 binding

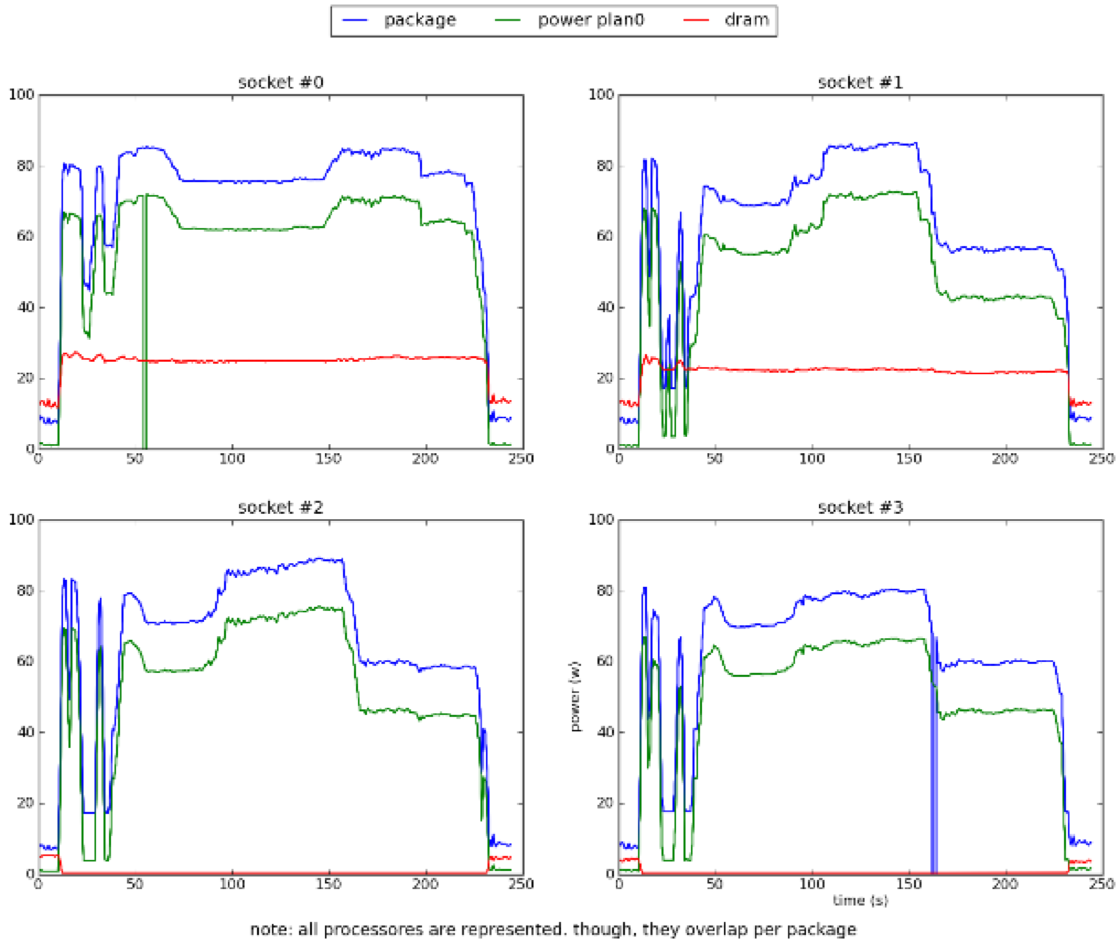


Figure 4.14: RAPL measurements of NUMA nodes - 32 processes with no explicit binding

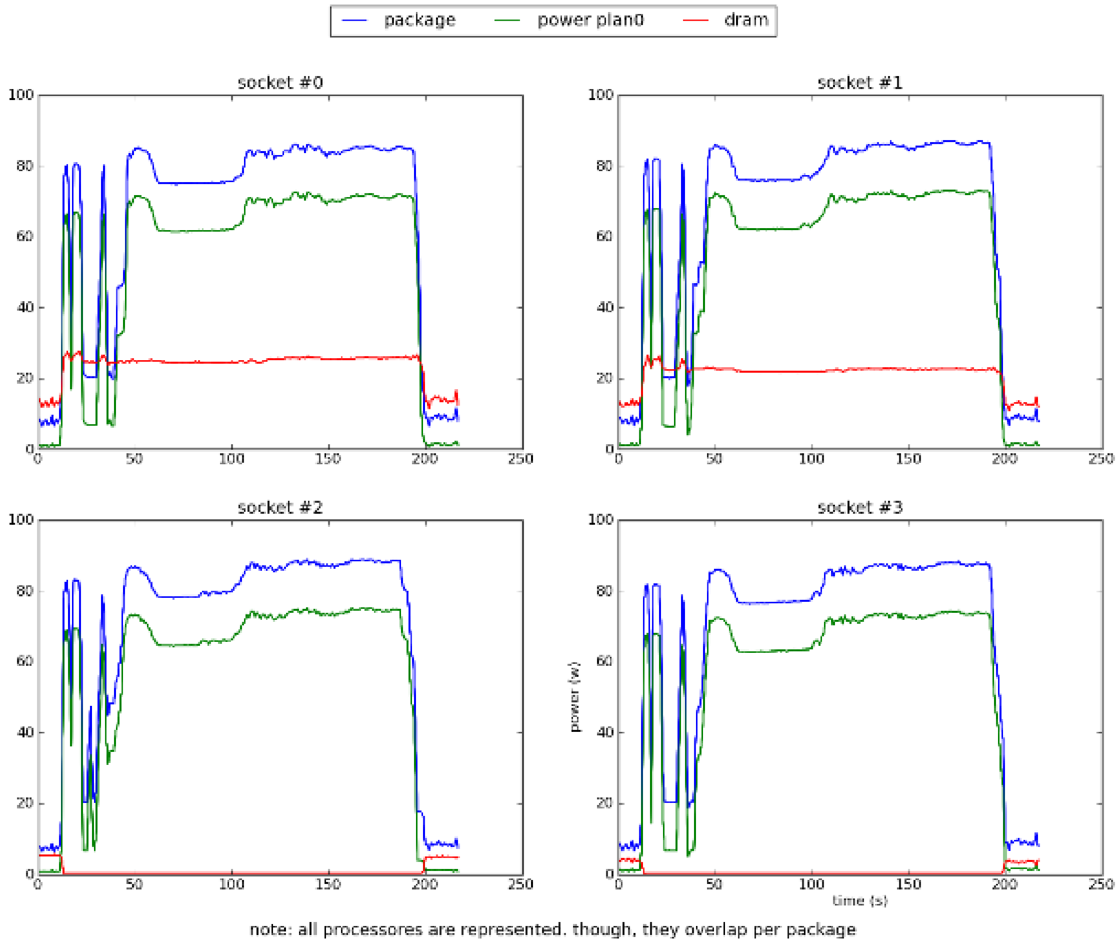


Figure 4.15: RAPL measurements of NUMA nodes - 32 processes. Processes distributed evenly explicitly - 8 processes per node.

Chapter 5

Analysis

The scope of the analysis presented in this section is twofold: to compare the platforms from an energy efficiency perspective and analyze the tools and techniques used on the different experiment sets.

Whereas the first two sections analyze the energy efficiency of the platforms studied and the particularities of the tools and techniques used, the last section covers the results and issues which arose when using RAPL to measure the energy consumption in a NUMA environment.

5.1 First Set of Experiments

ARM server vs ATOM and QUAD, using clap and software-based experiments

In figures 4.1, 4.2 and 4.3, it is plotted the physical measurements from the beginning of the workload until the end.

Stages

> All the experiment sets show 3 stages. The stages can be better identified when plotting the memory workload against cpu usage, rather than the energy consumption measurements (see Figures in GDrive-Add?). The three stages consist in different phase of the experiment. The first stage consists on the initialization process. During this stage mostly memory is being used, rather than cpu workload. The second stage is the connection phase. It has the goal of fetching the meta data fetching from the CERN servers needed to perform the reconstruction of the events. Anew, during this stage, the cpu load is low when compared to the memory workload. Lastly, the third stage corresponds to the event processing phase. Therefore, the last stage is cpu intensive and the

one that is performing the useful computation for the reconstruction of events.

Stages comparison

Regardless the number of processes running, the time for the three stages is constant in all the experiment sets, if the cpu is not overcommitted. When the number of processes exceed the number of available cores, the time to process the events increases since there are no available cores to process the events concurrently. In the overcommitted situation, the time increase follows a ratio *nr_of_processes/nr_of_cores_available*. For example, if the number of processes running is 6 and the number of cores available is 4, the time needed to process the events increases roughly 2/3 compared to when the cpu is not overcommitted.

Importance of the stages

Unarguably, the most important stage when studying the energy efficiency of workload in CERN is the third stage. There are two main reasons for that: first, the CMSSW configuration at either CERN, 2nd and 3rd tiers has proxies and caches that speedup the second stage [refs]. Lastly, given the amount of data to be processed in the last phase and thus the energy consumed by the event processing stage, the energy consumed by the former stages becomes irrelevant. Therefore, in the remainder of the chapter we focus our analysis on the event processing stage only. The energy measurements of the third stage are shown in the figures 4.4, 4.5 and 4.6.

5.1.1 Comparison ARM and Intel architecures

5.1.2 Tools and techniques

5.2 Second Set of Experiments

ARM board and Intel Xeon, using on chip and external measurements

5.2.1 Comparison ARM and Intel architecures

5.2.2 Tools and techniques

5.3 Third Set of Experiments

Intel Xeon, using RAPL to measure energy consumed by the different nodes, with different types of binding

Chapter 6

Future Work

Chapter 7

Conclusions

Bibliography

[1] Stack overflow. `http://stackoverflow.com/`. Accessed: 2014-10-27.

[2] Abdurachmanov, D., Elmer, P., Eulisse, G., and Muzaffar, S. Initial explorations of arm processors for scientific computing. *Journal of Physics: Conference Series 523*, 1 (2014), 012009.

[3] Pinto, G., Castor, F., and Liu, Y. D. Mining questions about software energy consumption. In *Proceedings of the 11th Working Conference on Mining Software Repositories* (New York, NY, USA, 2014), MSR 2014, ACM, pp. 22–31.

Appendix A

First appendix

This is the first appendix. You could put some test images or verbose data in an appendix, if there is too much data to fit in the actual text nicely.