

Energy Efficiency Experiments on Mali Powered Exynos 5 using OmpSs

Rune Holmgren

December 2014

PILOT PROJECT FOR MASTER THESIS

Department of Computer and Information Science

Norwegian University of Science and Technology

Supervisor 1: Professor Lasse Natvig

Co supervisor: Antonio Garcia Guirado

Problem statement

Here is the problem statement.

Acknowledgements

Here are the acknowledgements.

Abstract

This is the abstract.

Contents

	Prol	blem statement	j
	Ack	nowledgements	ii
	Abs	tract	iii
1	Intr	roduction	2
	1.1	Motivation	2
	1.2	Project Scope and Goal	2
	1.3	Problem Statement Interpretation and Approach	3
	1.4	Outline	3
2	Role	ated work	4
_	Reid	ated work	7
3	Bac	kground	5
	3.1	Energy measurement	5
	3.2	NEON	5
	3.3	The Performance API (PAPI)	6
	3.4	Task based programming	6
	3.5	OmpSs	6
	3.6	Heterognous multi-processor	6
	3.7	Experiment platforms	7
		3.7.1 Arendale Board	7
		3.7.2 ODROID-XU3	8
		3.7.3 ARM Cortex-A15	10
		2.7.4 ADM Conton A7	10

CONTENTS 1

		3.7.5 ARM Mali T604	11
		3.7.6 ARM Mali T628	12
	3.8	Algorithms	12
4	Setu	up and Methodology	13
	4.1	Test platfoms	13
		4.1.1 Arendale Duo	13
		4.1.2 ODROID-XU3	13
	4.2	Software	13
	4.3	Compilation and running of test benches	13
	4.4	Performance measurment	14
	4.5	Energy efficiency measurement	14
		4.5.1 ODROID-XU3	15
		4.5.2 Arendale Duo	16
5	Imp	plementation	17
6	Res	ult and Discussion	18
6 7		ult and Discussion aclution	18 19
	Con		
7	Con	nclution	19 20
7	Con Fut 8.1	ure Work	19 20 20
7	Con Fut 8.1	ure Work Experiment with heterogenousity	19 20 20 20
7	Fut: 8.1 8.2 8.3	ure Work Experiment with heterogenousity	19 20 20 20
7 8	Fut: 8.1 8.2 8.3 8.4	ure Work Experiment with heterogenousity	19 20 20 20 21
7 8	Fut: 8.1 8.2 8.3 8.4 Imp	ture Work Experiment with heterogenousity	19 20 20 20 21 21
7 8	Fut: 8.1 8.2 8.3 8.4 Imp	ture Work Experiment with heterogenousity	19 20 20 20 21 21
7 8 A	Fut: 8.1 8.2 8.3 8.4 Imp	nclution ure Work Experiment with heterogenousity	19 20 20 21 21 21 22

Introduction

1.1 Motivation

Increase of performance and power efficiency are the main goal of processor designers. Unfortunatly we are currently reaching the limits of the current strategies for further development. For some time, our processors have been strugeling to achive increased performance. Heat stops us from driving the clock frequency higher, while memory is lagging more and more behind. A solution to enable continued performance growth is multicore processors, and for the last decade this has been the focus. Unfortunatly adding cores will not be a sustainable solution forever. As the amount of cores grow, they are still competing for the same system resources and may have to wait for eachother to complete calculations on data with dependencies.

A promissing solution to this issue is heterogenous multi-processor systems. Heterognous multi-processor systems utilize multiple different processor cores in the same system. This allow different parts of a program to be executed on a suitable processor. By using a suitable core for each part of the program it is possible to achive better performance than homogenous multi-processor systems.

1.2 Project Scope and Goal

This pilot projects main goal is to do preliminary research and experiments on the energy efficiency of the Exynos 5 processor, with the intent to use the results next spring in my master

CHAPTER 1. INTRODUCTION

3

thesis. The goal of this research is to explore the potential of the task based programming model heterogenous multi-processor systems.

1.3 Problem Statement Interpretation and Approach

- Task 1: Implement or adapt suitable experiment applications for testing energy efficiency.
- Task 2: Implement some energy efficiency measurement application for both Arendale duo and Odroid-xu3.
- Task 3: Optimize experiment applications for both platforms.
- Task 4: Gather performance and energy efficiency results from the experiment applications on both platforms.
- Task 5: Analyze and evaluate experiment results.

TODO: Introduce how these tasks were solved.

1.4 Outline

TODO: This section need to be completed after the outline of the report is done.

Related work

Background

3.1 Energy measurement

3.2 NEON

NEON is a general-purpose single input multiple data (SIMD) technology implemented in the ARM Cortex A series of processors. It is able to run SIMD instructions on 128bit registers. By utilizing the NEON unit of the ARM processors, it is possible to achive paralellism in each seperate core. This will often open for great performance boost on problems like the ones explored in this paper. Each register may be filled with single precission floating point numbers ranging from 8 to 64 bit each. In future generations of the ARM ISA there will be support for other data types as well. Different implementations of NEON exist in the Cortex A cores, and while the even the simple implementations in smaller cores like the A7 can give great performance boost, the implementations present in the newest cores are performing even better. The A15 offer two NEON units, and the instruction pipeline to start the cores are shorter than in simpler implementations.

3.3 The Performance API (PAPI)

3.4 Task based programming

Task based programming allow a programmer to work with parallel programs, with an abstraction from the parallelization itself. When programming with this model, the program can be split into tasks which can run in parallel. When the program run, it will run a task manager as part of the program. This task manager can dynamically assign tasks to the processors, and the programmer does not have to handle all the time consuming tasks related to manual parallelisation. As long as the programmer correctly handle dependencies in the paralellized code, it will be possible to write this kind of code as if it was serial.

The task based programming model also allow simpler development of portable programs. When the program is running tasks on available CPUs, it is not a problem to allow it to run on larger or smaller numbers of processors, and even clusters can support the program. This model even allow the tasks to run on different types of processors in a hetrogenous environment.

3.5 OpenMP Super scalar

OpenMP Super scalar (OmpSs) is a extention of the OpenMP API to integrate features from the StarSs programming model. It is currently under development at the Barcelona Supercomputing Center. The goal of OmpSs is to extend the programming model to support a wide range og processors. The OmpSs programming model will run on a wide variety of different systems, such as traditional personal computers, clusters, shared memory systems and hetrogenous processors. While the software is not yet comlpeted or fully tested, there have been several reports exploring it's potentilal. The results have proven OmpSs as an efficient solution on both clusters and hetrogenous systems utilizing OpenCL and CUDA.

3.6 Heterognous multi-processor

Heterognous multi-processor systems have multiple different processors, opposed to traditional multi-processor systems. A typical modern processor have several processors, and a program

can run effectivly by having threads running parts of theis work on each of them. This work is often of such a nature that it can run better on a different processor. Sometimes it can run just as well on multiple simple processor, while using less die space and energy. In other instances, an advanced processor with some special capabilities, like vector instructions, can be more efficient.

This kind of processors have a potential to help us overcome the challenges that are emerging in processor development. Unfortunatly they also introduce several new challenges.

3.7 Experiment platforms

3.7.1 Arendale Board

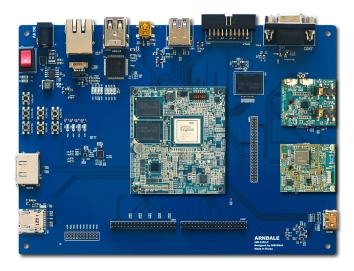


Figure 3.1: Arendale Duo

The Arendale Duo is a computing system mounted on a single board. It is fitted with an Exynos 5250 SoC, which contain a dualcore Arm Cortex-A15, as well as an ARM Mali T-604 GPU. This computer offer a range of supported linux distributions, as well as the OmpSs programming model. The computer was used in the 2014 master thesis "Acceleration with OmpSs and Neon/OpenCL on ARM Processor" by Trond Inge Lillesand. The thesis lay alot of the ground for this pilot project and planned master thesis.

3.7.2 ODROID-XU3



Figure 3.2: ODROID-XU3

The ODROID-XU3 is a new single-board computing system, offering interesting properties for these experiments. The system has an Exynos 5422 heterogenous Soc. Exynos 5422 has a quadcore ARM Cortex-A15 CPU and a ARM Mali T-628 GPU, but also a smaller quadcore ARM Cortex-A7 processor. These 3 different processing units can be used simultaniously to solve problems. In this paper, and the planned master thesis following it, the potency of this kind of heterogenous processor will be explored.

Power monitoring

The ODROID-XU3 commes with integrated power monitoring tools. Implemented in hardware, it have got 4 current sensors sitting on the power pins of the Exynos SoC. The energy monitors are indicated in figure 3.3. These monitor the current going through the large CPU cores, the small CPU cores, memory and GPU respectivly. In addition to the current sensors, the power management for the SoC is also available to the programmer, making supply voltage to the components known. By using the voltage and current, the power consumption is known. As a whole, the system offer frequency, voltage, current, temperature and power readings in real time. These fine energy and performance metrics make the system highly suitable for developers. They are able to run their programs, collect energy profiling data and optimize their software based on the result.

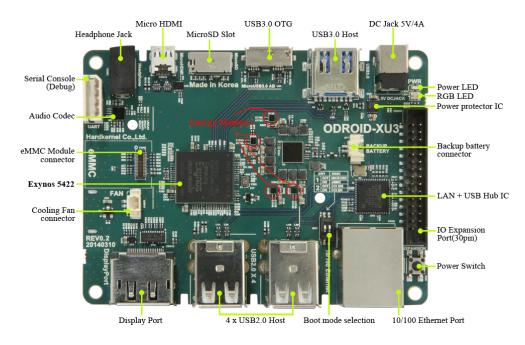


Figure 3.3: ODROID-XU3 anotated (hardkernel.com[1])

Performance

The ODROID-XU3 comes with the ARM Cortex-A15 limited to 2GHz and the ARM Cortex-A7 limited to 1.4G Hz. There is 2GB of memory available running at 800 MHz, and the ARM Mali T-628 GPU run at 600 MHz. This performance place it at the higher end of SoCs, but not quite in the top, as it is beaten by systems like TODO A reference system here). The ODROID-XU3 feature the new eMMC 5.0 standard for storage. The performance of this standard outform both older eMMC standards, as well as memory card readers, which other similar systems may contain. The ODROID-XU3 can achive a read/write performance of 198/74 MB/s[1], which alot better than older systems like the Arendale Duo.

In addition to the eight processor cores, the system also feature a ARM Mali T-628. The Mali T-628 function both as a regular GPU, as well as a GPGPU. It support both open GL and DirectX, and is able to produce graphics for all but the most demanding gaming and simulation purposes. In addition to this it is able to run computations with Open CL. This mean that problems with parallel parts, can be solved efficiently utilizing the GPU.

3.7.3 ARM Cortex-A15

Performance 1.0 GHz to 2.5GHz

L1 Cache 64KB

L2 Cache 4 MB

L3 Cache None in core, may be implemented shared in multicore system.

Architecture ARMv7-A

Architecture ARMv7-A

TrustZone® security technology

NEONTM Advanced SIMD

DSP & SIMD extensions

VFPv4 Floating point

Hardware virtualization support

Integer Divide

Fused MAC

Hypervisor debug instructions

Memory management 40-bit ARMv7 Memory Management Unit

3.7.4 ARM Cortex-A7

The ARM Cortex-A7 is designed to be a low power alternative to the ARM Cortex-A15 and ARM Cortex-A17, with the same supported ISA and features. This enable the ARM Cortex to be paired with it's largers relatives in a ARM big.LITTLE configuration.

Performance 1.2 GHz to 1.6GHz

L1 Cache 8-64KB

L2 Cache up to 1 MB

L3 Cache None in core, may be implemented shared in multicore system.

Architecture ARMv7-A

Supported features ARM Thumb-2

TrustZone® security technology

NEONTM Advanced SIMD

DSP & SIMD extensions

VFPv4 Floating point

Hardware virtualization support

Integer Divide

Fused MAC

Hypervisor debug instructions

Memory management 40-bit ARMv7 Memory Management Unit

3.7.5 ARM Mali T604

Performance 533 MHz

17 GFLOPS

Multicore support 1-4 cores

API Support OpenGL 1.1, 2.0, 3.0 and 3.1

OpenCL 1.1

DirectX 11

RenderScript

Anti-Aliasing 4xFSAA with minimal performance drop

16xFSAA

Cache 32-256KB L2 cache

3.7.6 ARM Mali T628

Performance 533/695 MHz

17/23.7 GFLOPS

Multicore support 1-8 cores

API Support OpenGL 1.1, 2.0, 3.0 and 3.1

OpenCL 1.1

DirectX 11

RenderScript

Anti-Aliasing 4xFSAA with minimal performance drop

16xFSAA

Cache 32-256KB L2 cache

3.8 Algorithms

Here I will write about the algorithms used in the experiments.

Setup and Methodology

4.1 Test platfoms

4.1.1 Arendale Duo

The Arendale Duo board was used for some perliminary research in this thesis. It was chosen because we had experience from earlier student projects using this board. It's feature set and properties are elaborated in section 3.7.1 Arendale Board.

4.1.2 ODROID-XU3

The ODROID-XU3 single board computing system was used for most of the experiments of this thesis. Fitted with an Samsung Exynos 5422 SoC it offer the heterogenous properties that will be explored in detail in the planned master thesis. In adition, the board offer multiple energy monitors, enabling precise data gathering. It's feature set and properties are elaborated in section 3.7.2 ODROID-XU3.

4.2 Software

4.3 Compilation and running of test benches

Mention frequency scaling here.

SoC Samsung Exynox 5250

CPU

Model ARM Cortex-A15

Manufacturing process 32nm Maxiumu clock frequency 1.7GHz

Number of cores 2 L2 Cache 1MB L1 Cache 32KB

GPU

Model ARM Mali-T604

Maxiumu clock frequency 600 MHz

Number of cores 4

Memory

Available memory 2 GB Maxiumu clock frequency 800MHz

Operating system

Distriubtion Linux Ubuntu

Version TODO

Table 4.1: Arendale Duo Spesifications

4.4 Performance measurment

In the experiments being run in this pilot project, execution time was used as the primary metric for performance. While running the experiments, the POSIX function gettimeofday() is used before and after running the main part of the application. The time difference is a used as a measurement of how good the performance is.

In the planned master thesis, PAPI (Performance API) may be used to measure more detailed aspects of the performance. PAPI is able to access alot of different performance measurements from systems. The available metrics include among others; Cache utilization and hit rate, memory utilization and bus utilization.

4.5 Energy efficiency measurement

Because of the different features of the two experiment platforms, two different energy efficiency measurement schemes was used.

SoC Samsung Exynox 5422

CPU 1

Model ARM Cortex-A15

Manufacturing process 32nm Maxiumu clock frequency 2.0GHz

Number of cores 4 L2 Cache 512KB

L1 Cache 32KB/32KB I/D

CPU 2

Model ARM Cortex-A7

Manufacturing process 32nm Maxiumu clock frequency 1.4GHz

Number of cores 4 L2 Cache 2MB

L1 Cache 32KB/32KB I/D

GPU

Model ARM Mali-T628 MP6

Maxiumu clock frequency 600 MHz

Number of cores 4

Memory

Available memory 2 GB Maxiumu clock frequency 933MHz

Operating system

Distriubtion Linux odroid Version 3.10.54+

Table 4.2: ODROID-XU3 Spesifications

4.5.1 **ODROID-XU3**

As elaborated in section 3.7.2 ODROID-XU3, the ODROID-XU3 feature 4 current sensors. These current sensors are used to monitor the current flow to the large CPU, small CPU, GPU and memory respectivly. In addition to the current sensors, it is possible to read the supply voltage to each of these components. Based on these two sensor readings, we can calculate the power consumption of each of the components in realtime while running our applications. Using this scheme for power measurement, we are able to get readings with high precission of each component. The results will show how the consumption vary with different application configurations both for the application as a whole and for different stages of execution.

 $Power\ consumption(Watt) = Current(Ampere) \times Voltage(Volt)$

4.5.2 Arendale Duo

Implementation

Result and Discussion

Conclution

Future Work

These are some suggestions for future work that may build uppon the work in this thesis.

8.1 Experiment with heterogenousity

In this thesis, there have been done experiments with the Exynos 5, which support ARM big.LITTLE. The heterogen properties of this processor was outside of the scope of this pilot project. The same applications can be adapted and optimized to explore the potential of this processor architecture. This is planned for the master thesis following this pilot project.

8.2 OmpSs with OpenCL kernels

A new feature of OmpSs is it's ability to manage OpenCL kernels as tasks. It is possible to issue OpenCL kernels as OmpSs tasks, and have the task manager assign them to GPUs and CPUs. This allow for portable code that can run effectively on a range of different system. It would be interesting to examine the potency of this way of utilizing the GPU, as it save the programmer from the job of manually tuning the loadbalance between GPU and CPU.

8.3 ARMv8-A 64-bit processors

ARM have created the next generation ARM processors. They run a new instruction set, with support for both 32- and 64-bit instructions. Running similar experiments on such a processor would be interesting.

8.4 Performance measurement

In this project, only simple forms of performance measurement was used. Running the same or similar experiments with access to other performance counters would be interesting. Cache hit rate and utilization, memory utilization, bus utilization and other counters could help understanding the strengths and weaknesses of the system.

Appendix A

Implementation

A.1 Introduction

A.1.1 Program 1

Bibliography

[1] Hardkernel.com. http://www.hardkernel.com/main/products/prdt_info.php?g_code=G140448267127&tab_idx=2. Accessed: 2014-10-25.

Curriculum Vitae

RUNE HOLMGREN

DATATEKNIKKSTUDENT OG **SYSTEMUTVIKLER**

Utdanning

08.2010 - d.d. Norges teknisk-naturvitenskaplige universitet (NTNU).

Sivilingeniør, Datateknikk. Ferdig juni 2015.

08.2007 - 06.2010 Heggen videregående skole.

Studiespessialisering med realfag.

Erfaring som svstemutvikler

06.2014 - 08.2014 Webutvikler for Visma Consulting

Sommerjobb på fulltid som utvikler og scrum master med angularJS og node.js med Leger uten grenser som kunde.

06.2013 - 09.2013 Systemutvikler hos Connome AS.

Sommerjobb på fulltid og deltid etter sommeren som androidutvikler mot hiemmeautomasion.

09.2012 - d.d. Webutvikler ved Studentersamfundet.

Webutvikling mot samfundet.no i Ruby on Rails. Ca. 12 timer i uken.

2011 - 2012 Androidutvikler i sammarbeid med Norsk Friidrett.

> Hobbyprosjekt som ble lansert i sammarbeid med Norsk Friidrett.

Øvrig arbeidserfaring

01.2013 - d.d. Undervisningsassistent ved NTNU.

Deltidsjobb som assistent i faget energieffektive datasvstemer.

08.2012 - 11.2012 Studentassistent ved NTNU.

Deltidsjobb som assistent i faget algoritmer og datastrukturer.

06.2012 - 07.2012

Drift og vedlikehold i Harstad Kommune.

Sommerjobb.

01.2012 - 03.2012 Studentassistent ved NTNU.

Deltidsjobb på fysikkløypa, et tiltak for å fremme realfag i barneskolen.

07.2011 - 08.2011

Driftslabratoriet hos Tine Meierier.

Sommerjobb.

06.2010 - 08.2010 Driftslabratoriet hos Tine Meierier.

Sommerjobb.

04.2010 - 08.2010 Plantasjen Harstad.

Deltidsjobb parallelt med videregående.

06.2009 - 08.2009 Privat ansatt i byggeprosjekt hos Byggmestrene Nilsen & Haukland.

Sommerjobb.

06.2008 - 08.2008

Byggmestrene Nilsen & Haukland.

Sommerjobb.

06.2007 - 08.2007 Byggmestrene Nilsen & Haukland.

Sommerjobb.

06.2006 - 07.2006 Aktiv byggpartner.

Sommeriobb.



Kontaktinformasjon

Telefon:

99 49 26 99

E-mail: runholm@stud.ntnu.no

Addresse:

Abels Gate 20A 7030 Trondheim

Kompetanseområder

Java Javascript Linux Python Ruby on Rails WebGL C Android **VHDL** HTML Photoshop Norsk CSS InDesign Engelsk SQL Vim Tysk

Idrettserfaring

Aktiv i trener og funksjonærroller.

Aktivitetslederkurs, friidrett. (en helg)

Trener 1 kurs rettet, friidrett. (to helger)

Kretsdommerkurs i friidrett. (en helg)

Kurslærerkur, friidrett.

Ung:leder utdanning, friidrett (fire langhelger, samt faglig oppfølgning av mentor)

Sertifikater

Førerkort KI. B

Referanser

Tilgjengelig ved forespørsel