AccelPower CAPE

A low-cost, open-hardware DC power measurement device for HPC computing nodes

Author Author ...

ArTeCS group
Dept. Arquitectura de Computadores y Automática
Universidad Complutense de Madrid
Madrid (Spain)

Rev. 1.1

April 6, 2016

Contents

List of Figures

List of Tables

Abstract

This report describes version 1.1 of AccelPower CAPE, a device (usually referred as CAPE) that extends the BeagleBone Black and provides DC power consumption monitoring capabilities. AccelPower CAPE provides accurate power, voltage and current measurements on each of the DC power rails suplying the main board or PCI-Express devices on modern compute nodes. It is designed and distributed under open-hardware license, featuring low acquisition cost and non-intrusive installation on existing compute nodes. This report also describes the necessary changes in `pmlib` to interface with AccelPower CAPE, and to easily instrument existing codes to gather energy consumption information from the device.

Chapter 1

Introduction

1.1 Motivation

Understanding the behavior of a computing architecture not only from the performance perspective, but also considering energy consumption, has become crucial in the road towards the exaflop barrier, even more with the advent of heterogeneous architectures, in which different computing units present dramatically different performance and power profiles. For that to happen, robust measurement environments that provide accurate and detailed power consumption profiles are mandatory.

In this report, we present and validate a complete low-cost hardware/software infrastructure based on open hardware/software components, for accurate power measurement of architectures based on PCIe accelerators. We describe the environment from two perspectives: (1) hardware elements and measurement instrumentation deployed in our system, and (2) software stack used to profile real applications using the aforementioned hardware elements. The validation is performed using a synthetic benchmark and on an Intel Xeon Phi 5110P PCIe accelerator card, comparing it with the power consumption information reported by the Intel SMC tool.

The environment is built around an ad-hoc extension shield (CAPE) for the BeagleBone Black, a low-cost open infrastructure supported by the community. The environment reports isolated information of all power sources for PCIe accelerators (and, optionally, other DC power sources in common computing nodes), including power lines of the PCIe bus (12V and 3.3V) and external 12V connectors.

From the software perspective, we complement the aforementioned environment by extending the `pmlib` infrastructure [1] to communicate with the developed device via I2C, with minimal effort for the developer interested in profiling the power consumption of a given application.

1.2 General environment overview

Figure 1.1 shows a schematic description of the developed hardware environment on a real server featuring a PCIe accelerator (in the following, we will refer to this server as the *Target Host*). Conceptually, following the guidelines presented in [4], the environment is built around a *Central Measurement Unit* (`CMU` in the following), that typically features a number of input channels, each one connected to a different *Measurement Source* (`MS`). Besides, there is a *Measurement Interface* (`MI`), that acts as an interface between the target architecture to be analyzed and the analysis environment.

At a glance, the illustrated setup uses different hardware components for each abstract element of the environment, namely:

MI: A BeagleBone Black gathers, timestamps, stores and serves power/current/voltage measurementes from the CMU via I2C (or natively from the ADCs), and serves them on

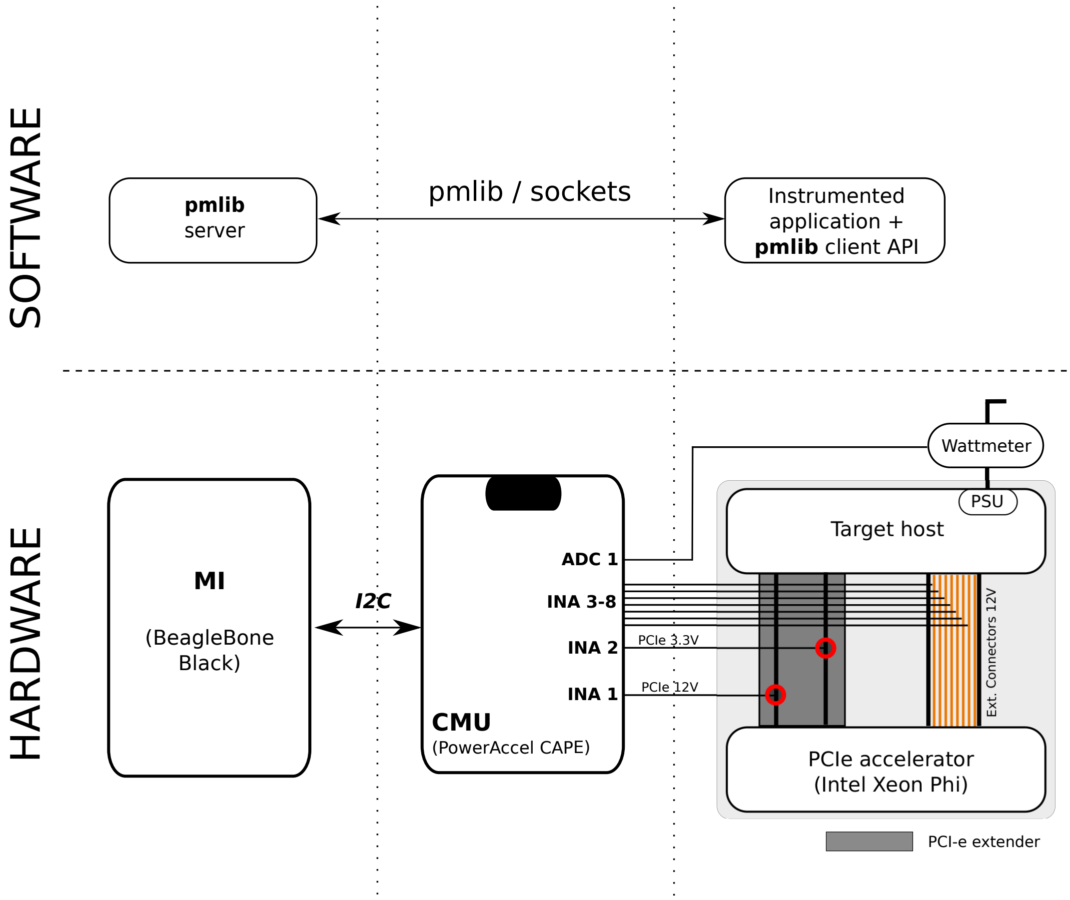

Figure 1.1: General environment overview.

demand to the instrumented application. For more information, see Section 2.1.

CMU: An ad-hoc AccelPower CAPE centralizes the reception of power/current/voltage data from the connected MIs. In our case, it features a number of on-board Texas Instruments INA219 sensors. For more information, see Section 2.2

MS1: One INA219 sensor, connected to the 12V power rail of the PCIe slot.

MS2: One INA219 sensor, connected to the 3.3V power rail of the PCIe slot.

MS3-8: One INA219 sensor per 12V external power lines. For a detailed description of the INA219 device, see Section 2.2.2.

MS9: Optionally, the ADC devices in the BeagleBone Black can be used to gather measurement from external devices providing analog outputs. Some examples include external Wattmeters (e.g. HIOKI XX), commercial PCIe extensors (e.g. FURAXA XX) or amperimetric clamps (e.g. Tectronix XX). For a detailed description of the ADC capabilities of our setup, see Section 2.2.3.

Chapter 2

Hardware design

2.1 BeagleBone Black

The BeagleBone Black is an open-hardware, low-cost and low-power SBC (*Single Board Computer*) easily expandable to fit the developer needs through CAPES (ad-hoc expansion shields), supported by the community. It features a low-power ARM Cortex-A8 processor and 512 MBytes of DDR3 RAM. The BeagleBone Black supports a plethora of Operating Systems, including different GNU/Linux distributions and Android. Table 2.1 summarizes the main features of the BeagleBone Black in its current release.

	Feature
Processor	Sitara AM3358BZCZ100, 1Ghz, 2000 MIPS
SDRAM Memory	512 MBytes DDR3L 800 Mhz
Onboard Flash	4 GBytes, 8bit embedded MMC
Debug Support	Optional Onboard 20-pin CTI JTAG
Power Source	miniUSB USB or DC jack (5V)
Indicators	1-Power, 2-Ethernet, 4-User Controllable LEDs
USB 2.0 Client Port	Access to USB0, Client mode via miniUSB
USB 2.0 Host Port	Access to USB1, Type A Socket, 500mA
Serial Port	UART0 access via 6 pin 3.3.V TTL Header
Ethernet	10/100, RJ45
SD/MMC Connector	microSD, 3.3V
Video Out	HDMI
Expansion Connectors	Power 5V, 3.3V, VDD_ADC (1.8V) 3.3V I/O on all signals McASP0, SPI1, I2C, GPIO, LCD, GPMC, MMC1, MMC2 AIN (1.8V MAX), 4 Timers, 4 Serial Ports, CAN0 EHRPWM, XDMA Interrupt, Power button Expansion Board ID (Up to 4 can be stacked)
Weight	39.68 grams
PCB dimensions	3.4” $\times$ 2.1”
Power	XX

Table 2.1: BeagleBone Black main features.

The BeagleBone Black can be run out-of-the-box by supplying power through a miniUSB port or, alternatively, through a 5V standard jack. In addition, the BeagleBone Black can transparently establish a network connection using Ethernet over USB using the miniUSB interface; on modern GNU/Linux distribution, this allows a quick setup and access to the board from a host computer in seconds.

One of the most appealing feature of the BeagleBone Black is its capability to be extended with plugin expansion boards (commonly, and in the following, referred as CAPES). Different types of CAPES have been designed by the community to support different types of functionality, including additional communication interfaces (e.g. CAN bus, RS232, XBee, GPS/GPRS), LCD screens, or industrial applications. Communication between the BeagleBone Black and the CAPE can be carried out using different communication interfaces (typically SPI or I2C). In the following, we describe AccelPower CAPE, an ad-hoc expansion CAPE that allows a precise measurement of DC voltage, power and current targeting modern HPC computing nodes.

2.2 AccelPower CAPE

2.2.1 Overview of the CAPE design

The AccelPower CAPE is an ad-hoc CAPE designed to extend the capabilities of the BeagleBone Black, adding power, voltage and current monitoring of up to eight DC power rails. The AccelPower CAPE features eight (8) input connectors at each side of the CAPE (and their corresponding outputs at the opposite side) that can be connected to internal DC power rails on commodity computing nodes. Internally, each connector is connected to a Texas Instruments INA219, that monitors the voltage drop through a precision *shunt resistor* and reports voltage, current and power through an I2C interface. Additionally, the CAPE provides six (6) external connectors that interface with the internal ADCs of the BeagleBone Black, and allow monitoring analog inputs.

2.2.2 Texas Instruments INA219

The core power monitoring system chosen for the AccelPower CAPE is based on the Texas Instruments INA219 component. The CAPE provides an independent INA219 sensor on each one of the 8 inputs of the board. Internally, the INA219 calculates the current flowing through a precision shunt resistor (in our case, $R_{shunt} = 0.05\Omega$) by first measuring the voltage across the resistor, and then applying Ohm's law.

The INA219 is a low cost and low-power device, featuring an easy callibration and usage. From the docs[1]:

> *"The INA219 is a high-side current shunt and power monitor with an I2C interface. The INA219 monitors both shunt drop and supply voltage, with programmable conversion times and filtering. A programmable calibration value, combined with an internal multiplier, enables direct readouts in amperes. An additional multiplying register calculates power in watts. The I2C interface features 16 programmable addresses"*

Even though each I2C bus in the BeagleBone Black supports up to 16 slave I2C devices, due to integration limitations (mainly related to the width of PCB traces to support typical currents in common DC power rails in current computing nodes), 8 INA219 have been integrated in the current version of AccelPower CAPE. Stackable CAPES are planned for the next version of the design, that will extend the number of devices (and hence the number of monitored power rails) visible to the BeagleBone Black, by additionally using the three I2C buses in the board. This design will allow a total number of 48 power rails monitored simultaneously.

INA219 callibration

(Add Technical Details of the INA219).

[1] `http://www.ti.com/lit/ds/symlink/ina219.pdf`

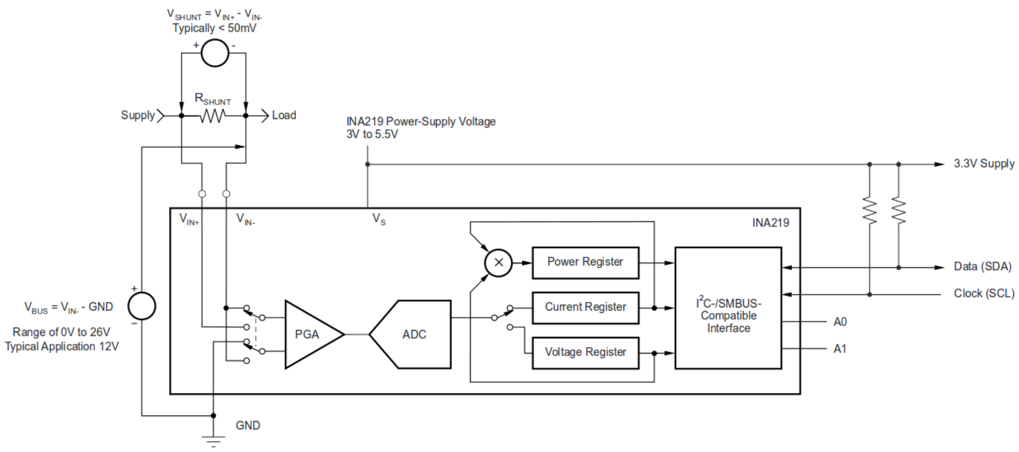

Figure 2.1: Texas Instruments INA219. Simplified schematic.

2.2.3 ADC

Details on ADC adaptation circuit (Luis).

2.2.4 Manufacturing details

In order to accommodate typical currents of up to XX Amps, PCB traces were designed and built to be XX inches wide, allowing currents of up to XX Amps with no significant heat dissipation. (More manufacturing details, GND plane, etc.)

Details on manufacturing (Juan Diego).

2.3 PCIe extensor

As said in Section 1, one of the main targets of AccelPower CAPE are host computing nodes equipped with PCIe accelerators. This type of devices are supplied not only by external 12V DC connectors directly connected to the host power supply, but also by 3.3V and 12V lines providing power through the PCIe slot. In its current version, we use

2.4 Comparison with similar efforts

Table 2.2. . .

2.5 Hardware Setup

Brief description on how to setup the AccelPower CAPE on a target host.

2.5.1 BeagleBone Black setup

The following guidelines consider an out-of-the-box BeagleBone Black with the default GNU/Linux instalation (Debian). Similar procedure should be followed with other OS distributions.

1. Connect the CAPE to the corresponding expansion header pins of the BeagleBone Black, following the indications in Figure **??**.

2. Connect the BeagleBone Black to the target host using the miniUSB cable provided.

	PowerMon 2	PowerInsight	ArduPower	AccelPower CAPE
Ref.	[2]	[5]	[3]	-
Host	ATmega168	X2	X3	ARM Cortex A8
Sensors	ADM1191	Y2	Y3	TI INA219
Technology	Shunt	Hall	Hall	Shunt
Lines	8	Y2	Y3	8
Voltage	Yes	Y2	Y3	Yes
Current	Yes	Y2	Y3	Yes
Power	No	Y2	Y3	Yes
Accuracy	12-bit	Y2	Y3	Y4
Frequency	Y1	Y2	Y3	Y4
SW support	Y1	Y2	Y3	Y4

Table 2.2: Comparative feature table with similar DC power measurement environments.

3. Automatically, the target host should detect the BeagleBone Black as an external disk, and provide a network address to the target host (typically 192.168.7.1) and to the BeagleBone Black (typically 192.168.7.2).
4. From the target host, access the BeagleBone Black establishing a remote connection through `ssh` with the following command:

```
ssh root@192.168.7.2
```

5. Detect the I2C devices in the CAPE (each INA219 is seen as an independent I2C device):

```
root@beaglebone:~\# i2cdetect -y -r 2
         0  1  2  3  4  5  6  7  8  9  a  b  c  d  e  f
    00:          -- -- -- -- -- -- -- -- -- -- -- -- --
    10: -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- --
    20: -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- --
    30: -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- --
    40: 40 41 42 43 44 45 46 47 -- -- -- -- -- -- -- --
    50: -- -- -- -- UU UU UU UU -- -- -- -- -- -- -- --
    60: -- -- -- -- -- -- -- -- -- -- -- -- -- -- -- --
    70: -- -- -- -- -- -- -- --
```

Note that each INA219 sensor is identified with a different address in the range 0x40 – 0x47.

6. If eight different INA219 have been detected, the hardware setup is ready.

2.5.2 External 12V connectors and PCIe extensor setup

Chapter 3

Software infrastructure

The goal of the proposed environment is not only to obtain accurate and detailed power consumption measurements of PCIe devices, but to be able to instrument existing codes on a target platform, obtaining power profiles of them as a whole, or focusing on specific code sections. This integration of power measurements needs to be essentially transparent for the programmer, including synchronization, data transfer and accuracy management. We describe next the main software components used to provide such functionality using the aforementioned hardware framework.

3.1 `pmlib`

`pmlib` [1] is a software infrastructure[1] that eases the process of measuring energy consumption and instrumenting applications running on different target architectures. In previous works, the developers of the library have mainly used `pmlib` combined with ad-hoc or simple commercial AC meters (e.g. Watts-up) in order to estimate the energy consumption of a number of target platforms. `pmlib` is a client-server framework in which the server module directly communicates with the power analysis instrumentation, and the client module (that can be embedded on the target code though a provided API) queries for gathered power measurements, synchronizing the obtained data with the execution of the target program.

`pmlib` is fully modular and extensible; our effort extends the framework to interact with the developed hardware environment, as described next. We have revamped the `pmlib` server running on the `MI` (BeagleBone Black) to directly interact with the `CMU` (AccelPower CAPE) via I2C. Alternatively, it also gathers measurements from the ADCs in the board. For that to happen, the `pmlib` server running on the `MI` establishes a connection with the remote `CMU` when it is started, and continuously gathers measurement samples from it via the corresponding mechanism (e.g. I2C).

3.2 Software setup and instrumentation examples

[1] `pmlib` was developed at University Jaume I, Castellón, Spain

Chapter 4

Experimental results

4.1 Synthetic validation

4.2 Comparison with software measurements

Chapter 5

License and availability

Appendix A

Schematics

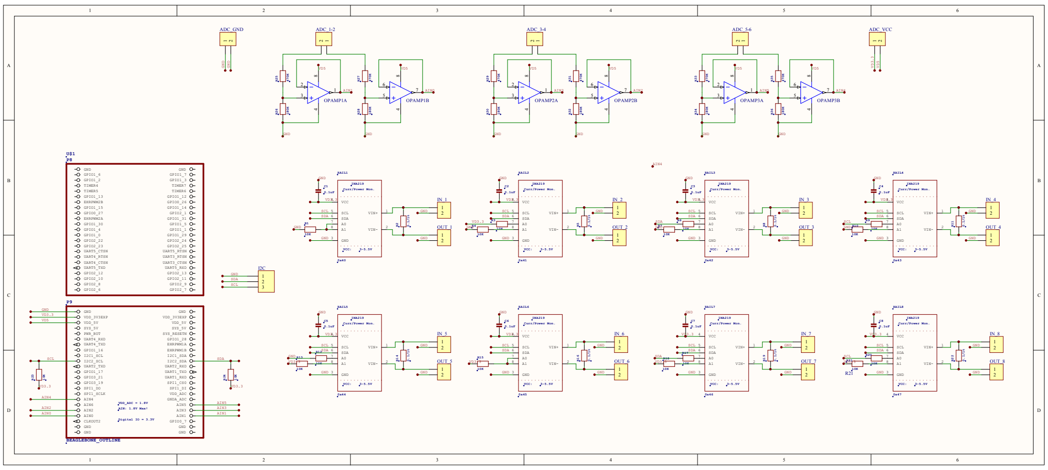

Figure A.1: AccelPower CAPE. Schematic.

Appendix B

Layout

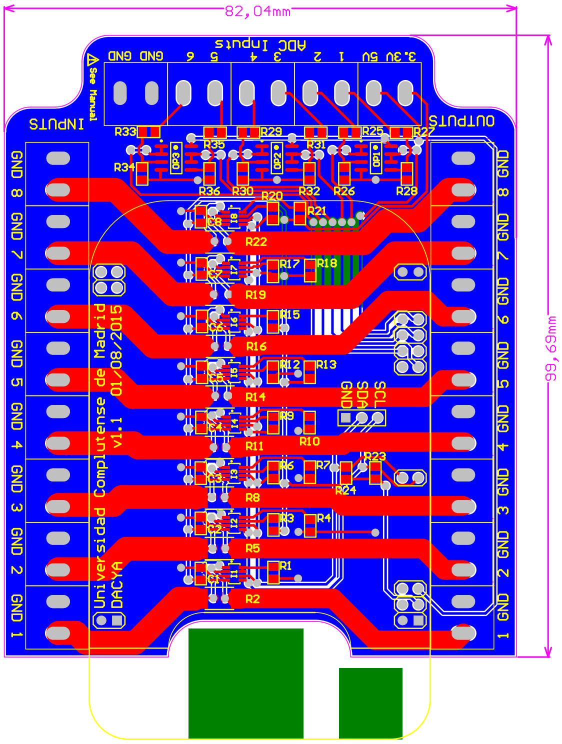

Figure B.1: AccelPower CAPE. Top layout.

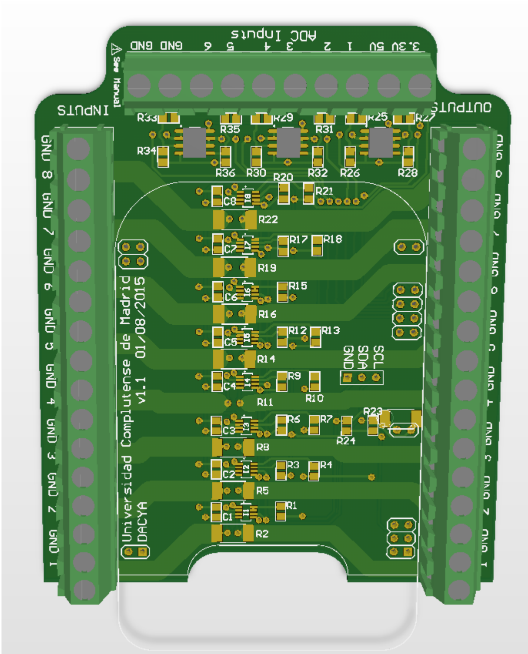

Figure B.2: AccelPower CAPE. 3D-view 1.

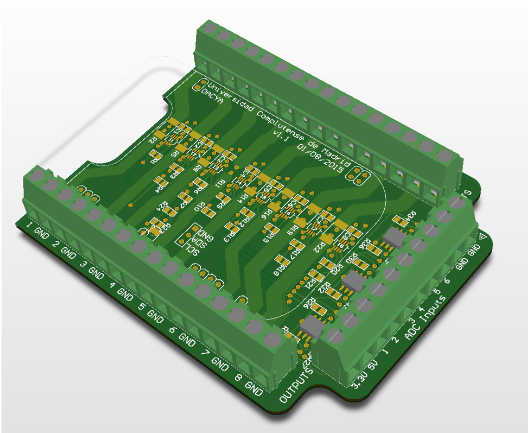

Figure B.3: AccelPower CAPE. 3D-view 2.

Appendix C

Bill of materials

Identifier	Part Number	Description	Quantity	Unit Cost	Total cost
XX	CRA2512-FZ-R050ELFCT-ND	Resistor, 0.05Ω	X	1.35	XX
XX	MC0805B104K160CT	Capacitor, $0.01\mu F$	X	0.029	XX
XX	INA219AIDCNR	INA219 Current/Power Monitor	8	1.53	XX
XX	MCWR08X1002FTL	Resistor, $10K\Omega$	X	0.0124	XX
XX	ERJ6RED1803V	Resistor, $180K\Omega$	X	0.163	XX
XX	ERJ6RED4703V	Resistor, $470K\Omega$	X	0.163	XX
XX	ERJ6RED8203V	Resistor, $820K\Omega$	X	0.163	XX
XX	BBONE-BLACK-4G	BeagleBone Black	1	45.74	45.74
XX	Delock PCI-e	PCEe riser	1	20.40	20.40

Bibliography

[1] S. Barrachina, M. Barreda, S. Catalán, M. F. Dolz, G. Fabregat, R. Mayo, and E. S. Quintana-Ortí. An integrated framework for power-performance analysis of parallel scientific workloads. *3rd Int. Conf. on Smart Grids, Green Communications and IT Energy-aware Technologies*, pages 114–119, 2013.

[2] Daniel Bedard, Allan Porterfield, Rob Fowler, and Min Yeol Lim. PowerMon 2: Fine-grained, integrated power measurement. Technical Report TR-09-04, RENCI, North Carolina, 2009.

[3] Manuel F. Dolz, Mohammad Reza Heidari, Michael Kuhn, Thomas Ludwig, and German Fabregat. Ardupower: A low-cost wattmeter to improve energy efficiency of hpc applications. In *Green Computing Conference and Sustainable Computing Conference (IGSC), 2015 Sixth International*, pages 1–8, Dec 2015.

[4] Francisco D. Igual, Luis M. Jara, José I. Gómez-Pérez, Luis Piñuel, and Manuel Prieto-Matías. A power measurement environment for pcie accelerators. *Computer Science - Research and Development*, 30(2):115–124, 2014.

[5] James H. Laros III, Phil Pokorny, and David Debonis. Powerinsight - A commodity power measurement capability. In *International Green Computing Conference, IGCC 2013, Arlington, VA, USA, June 27-29, 2013, Proceedings*, pages 1–6, 2013.