SAND2008-4754C

Investigating Real Power Usage on High Performance Computing Platforms

James H. Laros III
Kevin Pedretti, Sue Kelly
John Vandyke, Courtenay Vaughn

Sandia National Labs[1]
P.O. Box 5800, MS-1319
Albuquerque, NM 87185

Mark Swan

Cray Inc.
1340 Mendota Heights Road
Mendota Heights, MN 55120

Abstract—Power is a major consideration in the design of current High Performance (HPC) Computing platforms and is likely to be a limiting factor in the architecture of next generation HPC platforms. Current estimates for a one to two petaflop system in the year 2010 range from five to fifteen megawatts. With power costs of approximately $1,000,000 per megawatt annually, it is important to understand and ultimately affect both the amount and how power is used on HPC platforms. To understand real power usage we have instrumented an HPC platform to collect power data at a per-socket, per-second granularity. We show how we have positively affected power consumption on multi-core processors using our Light Weight Kernel. Additionally, we will show analysis of real application power usage and present future research topics instigated by our findings.

Index Terms—Power, High Performance Computing (HPC)

1 Introduction

The proposed poster will illustrate our results to date and future research directions on the topic of real power use as it relates to High Performance Computing (HPC) platforms and scientific applications. Power is proportional to the product of the Capacitance, Frequency and Voltage squared. Individual or combinations of these factors can be manipulated to effect system power usage. Hardware vendors have been addressing the issue of power for quite some time. In recent years, vendors have decided to leverage Moores Law by increasing the number of processor cores on a chip rather than increasing the frequency of a single processor core. Additionally, vendors have made available an expanding number of features that allow software to exploit numerous power saving states. Our approach has been to augment and or leverage, whenever possible, the work done by the commodity vendor. We have executed a systems level approach described in the following sections.

1. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000. Contact: jhlaros@sandia.gov

2 Instrumentation

Our first effort, necessarily, involved implementing the instrumentation necessary to measure the real power consumption of an HPC platform at the finest granularity and highest frequency possible. The proposed poster will outline how we have accomplished this on a Cray XT4/5, enabling granular, scalable, high frequency collection of current and voltage measurements.

3 Light Weight Kernel Efficiency

With the ability to observe power consumption, we first focused on maximizing the power efficiency of our Light Weight Kernel on multi-core sockets. Our goal was to ensure that we use as little power as possible during idle states. As the number of cores increase, the likely hood of one or more cores of a multi-core socket remaining unused increases (in particular, bandwidth bound highly scalable scientific

computing applications). The proposed poster will illustrate how only cores in use will be employed during the execution of an application(reference Fig 1).

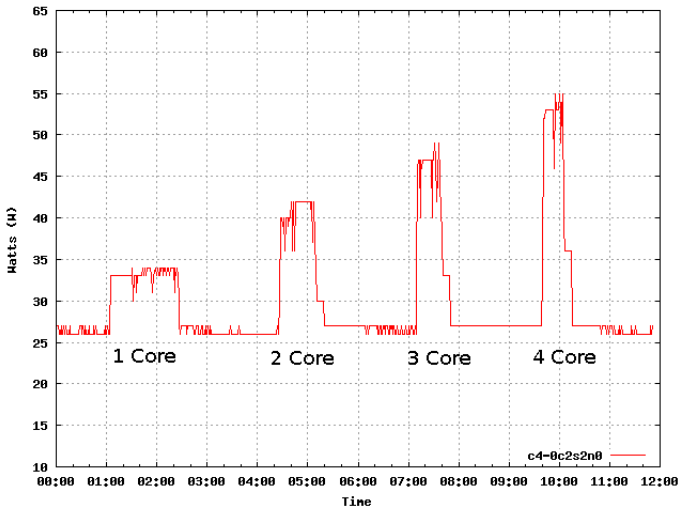

Fig. 1. Multi-Core Light Weight Kernel Power Use

4 Application Power Signatures

Our next effort involved quantifying real application power use. Our instrumentation allowed us to collect detailed information on the power use of individual applications. We have collected *Application Power Signatures* for a number of real scientific applications and common HPC benchmarks. The following graphs depict the real power used by two different applications (SAGE and PARTISN). In addition, we calculate the total energy used throughout the duration of the application by approximatting the definite integral below the curve. (reference Figs 2 & 3)

5 Planned Research Areas

We are currently in the process of investigating a number of areas that build on the work presented. Listed are a few of the areas of research currently in progress.

5.1 Adjusting Frequency While Maintaining Performance

It is quite common for scientific applications to be either memory or communication bound.

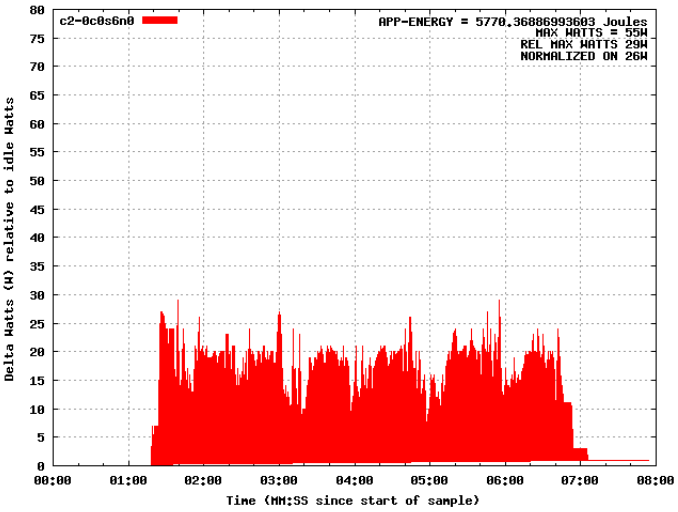

Fig. 2. SAGE Application Power Signature

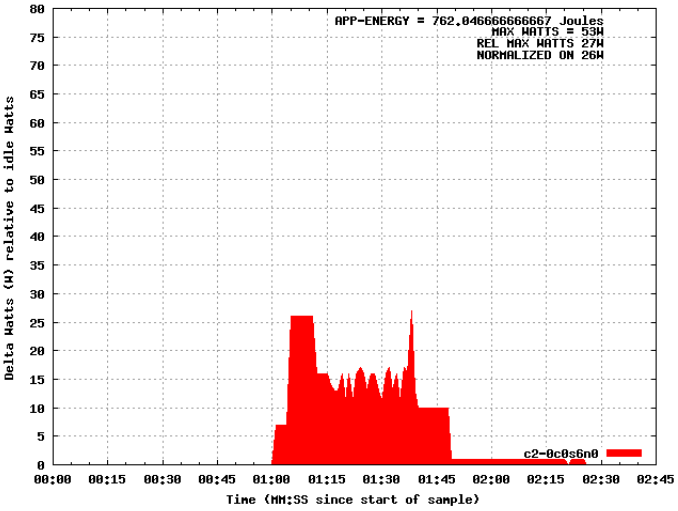

Fig. 3. PARTISN Application Power Signature

The ability to measure real energy use of an application will allow us to investigate whether we can dynamically reduce cpu frequency (resulting in lower input voltage) with minimal to no impact on application performance.

5.2 Message Passing Interface Instrumentation (MPI)

We are investigating the possibility of instrumenting MPI to enter architecture supported power savings states in cases, for example, when nodes are blocked awaiting communication. Cumulatively these small periods of power saving could potentially be very significant in total application energy usage.