REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* AUG 2011	**2. REPORT TYPE** CONFERENCE PAPER (Post Print)	**3. DATES COVERED** *(From - To)* JAN 2011 – JUN 2011

4. TITLE AND SUBTITLE ENERGY EFFICIENCY EVALUATION AND BENCHMARKING OF AFRL'S CONDOR HIGH PERFORMANCE COMPUTER	**5a. CONTRACT NUMBER** IN-HOUSE
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) RYAN LULEY, COURTNEY USMAIL, MARK BARNELL	**5d. PROJECT NUMBER** HPCC
	5e. TASK NUMBER IN
	5f. WORK UNIT NUMBER HO
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFRL/RITB 525 BROOKS RD ROME, NY 13441-4505	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-RI-RS-TP-2011-9

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA #: 88ABW-2011-2798, 20110519
CLEARED ON: 19 May 2011

13. SUPPLEMENTARY NOTES
Publication at DoD High Performance Modernization Program's 2011 User Group Conference, Location: Portland, OR, Submission deadline: 20110520, Conference Dates: 20110620-20110623 . This is a work of the United States Government and is not subject to copyright protection in the United States.

14. ABSTRACT
Emerging supercomputers strive to achieve an ever increasing performance metric at the cost of excessive power consumption and heat production. This expensive trend has prompted an increased interest in green computing. Green computing emphasizes the importance of energy conservation, minimizing the negative impact on the environment while achieving maximum performance and minimizing operating costs. The Condor Cluster, a heterogeneous supercomputer composed of Intel Xeon X5650 processors, Cell Broadband Engine processors, and NVIDIA general purpose graphical processing units was engineered by the Air Force Research Laboratory's Information Directorate and funded with a DoD Dedicated High Performance Computer Project Investment (DHPI). The 500 TeraFLOPS Condor was designed to be comparable to the top performing supercomputers using only a fraction of the power. The objective of this project was to determine the energy efficiency as a function of performance per Watt of Condor.

15. SUBJECT TERMS
SUPERCOMPUTING, CLUSTERS, ENERGY EFFECIENCY, BENCHMARKING

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** COURTNEY USMAIL
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	U	11	**19b. TELEPONE NUMBER (*Include area code)*** 315-330-3133

Energy Efficiency Evaluation and Benchmarking of AFRL's Condor High Performance Computer

Ryan Luley, Courtney Usmail, and Mark Barnell
US Air Force Research Laboratory, Information Directorate, Computing Technology Applications Branch (AFRL/RITB), Rome, NY
{ryan.luley, courtney.usmail, mark.barnell}@rl.af.mil

Abstract

Emerging supercomputers strive to achieve an ever increasing performance metric at the cost of excessive power consumption and heat production. This expensive trend has prompted an increased interest in green computing. Green computing emphasizes the importance of energy conservation, minimizing the negative impact on the environment while achieving maximum performance and minimizing operating costs.

The Condor Cluster, a heterogeneous supercomputer composed of Intel Xeon X5650 processors, Cell Broadband Engine processors, and NVIDIA general purpose graphical processing units was engineered by the Air Force Research Laboratory's Information Directorate and funded with a DoD Dedicated High Performance Computer Project Investment (DHPI). The 500 TeraFLOPS Condor was designed to be comparable to the top performing supercomputers using only a fraction of the power. The objective of this project was to determine the energy efficiency as a function of performance per Watt of Condor.

The energy efficiency of Condor was determined using the Green500 test methodology, in particular measuring power consumption during maximum performance on the High Performance LINPACK (HPL) Benchmark. The HPL Benchmark measures computing performance in floating point operations per second while solving random dense linear equations. A power meter was used to measure the average energy consumption of a single node of the system over the duration of the execution time of the benchmark. Using the energy consumption from a single node and assuming each node to draw equal amounts of energy, the efficiency performance of the entire system was calculated. We demonstrate that Condor achieves an energy efficiency performance comparable to the top supercomputers on the Green500 List.

1. Introduction

The past 20 years have seen a seemingly unstoppable increase in computer performance; we have witnessed a remarkable 10,000 fold improvement in the peak performance of a high end supercomputer (Feng and Cameron, 2007). The drive in computer advancements has been strictly performance-based, doing anything necessary to achieve a maximum number of floating point operations per second (FLOPS). What has not been heavily considered throughout these advancements however, is the energy efficiency of the supercomputer. Green computing takes a new view of high performance computing by considering the energy consumption required to achieve maximum performance goals (Feng et al., 2008).

The drive for energy efficient computing has been increasingly present in the past several years for a number of reasons. We continue to observe an immense increase in the peak performance of computers on the TOP500 list over the years. While this speedup is a great feat, the cost to run these powerful computers is an unavoidable roadblock for sustainability in terms of total cost of ownership. Consider that the price of electrical energy per megawatt is estimated to be approximately $1 million per year (Feng et al., 2008). According to the TOP500 list of November 2010, the top performing supercomputer in the world used 4.04 MW of power; and this system was not the most power hungry on the list (http://www.top500.org/list/2010/22/100).

Since 2006, there has been an evident drive for energy efficient computing by the US Government. In December of 2006 the U.S. Congress passed Public Law 109-431 “to study and promote the use of energy efficient computer servers in the United States” (http://energystar.gov). The law emphasizes the need for energy efficient improvements for government and commercial servers and data centers and required a study be done by the Environmental Protection Agency (EPA) Energy Star Program to analyze the areas of potential impacts in energy efficiency improvements, as well as recommendations for incentive programs to advance the transition to energy efficient computing. The EPA Energy Star program submitted the “Report to Congress on Server and Data Center Energy Efficiency” in 2007 where energy use and cost for data centers in the U.S. was extensively examined and prospective areas of improvement were addressed (http://www.energystar.gov). In addition, AMD, Dell, IBM, Sun Microsystems, and VMware formed the Green Grid consortium in 2007. The mission of the Green Grid is to improve the energy efficiency of data servers and computer ecosystems (Kurp, 2008).

The Green500 List was started in April 2005 to encourage energy efficiency as a first-class design consideration in emerging supercomputer construction and to provide a ranking of the top performing supercomputers with respect to an energy efficiency metric (www.green500.org). Similar to the well-known TOP500 List that ranks high performance computers based on peak performance, the Green500 list measures the peak performance of a system running the High Performance LINPACK (HPL) benchmark while also measuring the energy consumed to achieve such performance. Supercomputers are ranked by MegaFLOPS (MFLOPS) per Watt, with the minimum criteria to be accepted on the Green500 List being that the supercomputer must achieve HPL performance great enough to appear on the most recent TOP500 list.

With energy efficiency in mind, the Air Force Research Laboratory’s Information Directorate engineered the *Condor* Cluster, a heterogeneous supercomputer composed of Intel Xeon X5650 processors, Cell Broadband Engine (Cell BE) processors, and NVIDIA general purpose graphical processing units. This project was funded with a DoD Dedicated High Performance Computer Project Investment (DHPI) and has a theoretical single precision peak performance of 500 TeraFLOPS (TFLOPS). This paper examines the energy efficiency of *Condor* using the Run Rules for the Green500 List, to demonstrate the total cost of ownership efficiency of this unique system design.

2. The *Condor* Cluster

The *Condor* Cluster is a heterogeneous supercomputer composed of 94 NVIDIA Tesla C2050’s, 62 NVIDIA Tesla C1060’s, 78 Intel Xeon X5650 dual socket processors, and 1716 Sony PlayStation 3s (PS3s), adding up to a total of 69,940 cores and a theoretical peak performance of 500 TFLOPS. There are 84 subcluster head nodes, of which six are gateway nodes that do not perform computations, while the other 78 compute head nodes are capable of 230 TFLOPS of theoretical peak processing performance. Each of the 78 compute head nodes are composed of two NVIDIA general purpose graphical processing units (GPGPUs) and one Intel Xeon X5650 dual socket hexa-core processor (i.e. 12 cores per Xeon). Of the 78 compute head nodes, 47 contain dual NVIDIA Tesla C2050 GPGPUs while 31 contain dual NVIDIA Tesla C1060 GPGPUs. The head nodes are connected to each other via 40 Gbps InfiniBand and 10Gb Ethernet. Additionally, each compute node is connected to a 10GbE/1GbE aggregator that provides communication to a subcluster of 22 PS3s. In total, the PS3s can achieve a theoretical peak performance of 270 TFLOPS.

The NVIDIA GPGPUs in *Condor*, the Tesla C1060 and the newer model Tesla C2050, share similar architectures but vary in performance. Both the C1060 and C2050 have the same Tesla architecture based on a scalable processor array (Lindholm, 2008). The architecture can be broken down into independent processing units called texture/processor clusters (TPCs). The TPCs are made up of streaming multiprocessors which perform the calculations for the GPGPU. The streaming multiprocessors can be broken down further into streaming processors or cores; these are the main units of the architecture (Maciol, 2008). The GPGPU communicates with the CPU via the host interface (Lindholm, 2008). However, the C1060 model has 240 cores while the C2050 has 448 cores (http://www.nvidia.com).

The Intel Xeon processors on each head node are built on the energy efficient Intel Nehalem microarchitecture. This architecture was made with several Intel technologies that adjust performance and power usage based on application needs. When not in use, the processor is capable of drawing a minimal amount of power and also capable of operating above the rated frequency when necessary. The *Condor* Cluster is equipped with the six-core Intel Xeon dual socket X5650, giving a total of 12 cores per processor (http://www.intel.com).

The Sony Toshiba IBM (STI) Cell BE is a nine core heterogeneous processor that consists of one PowerPC Processing Element (PPE) and eight Synergistic Processing Elements (SPEs). The PPE is based on the open source IBM Power Architecture processor and is responsible for controlling and coordinating the SPE tasks and runs the operating systems on the processor (Buttari et al., 2007). The eight SPEs are responsible for the majority of the compute power on the processor (Gschwind et al., 2007). All code executed by the SPE is done in the 256 KB software controlled local store (Buttari et al., 2007). The SPEs consist of a Synergistic Processing Unit (SPU) and Memory Flow Controller (MFC). The MFC transfers data between the SPE cores as well as between the local store and the system memory (Gschwind et al., 2007). Connection from PPE to SPEs is made via the Element Interconnect Bus (EIB) which has a peak bandwidth of 204.8 GB/s (Buttari et al., 2007).

Condor utilizes the PS3 as a computing platform for access to the Cell BE. The PS3 is equipped with the Cell BE with minor alterations. Only six of the eight SPEs available for use in the PS3; one SPE is disabled for yield reasons at the hardware level and one SPE is reserved solely for the GameOS (Buttari et al., 2007). For use in the *Condor* Cluster, CentOS Linux was installed on the PS3s. Additionally, of the total 256 MB of available memory for the Cell Broadband Engine only 200 MB is accessible to Linux (Buttari et al., 2007).

The *Condor* Cluster was engineered to increase the combat effectiveness of the Department of Defense through technological advances supported by high performance computing. Next generation synthetic aperture radar (SAR) sensors strive to provide surveillance of larger areas (30 km diameter) with smaller targets at resolutions close to one foot. Applications such as this demand real-time processing of over 200 sustained TFLOPS. This surveillance capability can be achieved using the SAR backprojection algorithm, a computationally intensive algorithm that enables every pixel to focus on a different elevation to match the contour of the scene. The SAR backprojection algorithm has been optimized by the AFRL Information Directorate to eliminate nearly all double precision operations, favoring application on the Cell Broadband Engine. NVIDIA Tesla GPGPU cards also have a preference for single precision operations which critically enhances the algorithm and consequently the number of pixels generated for a 30 km surveillance circle.

3. High Performance LINPACK

The LINPACK benchmark has become the de facto standard for measuring real peak computational performance of high-performance computers for nearly twenty years. HPL introduced the ability to address scalability in the LINPACK testing environment, in order to accurately measure the performance of larger, parallel distributed memory systems. Since 1993, HPL has been used to formulate the TOP500 list of the most powerful supercomputers in the world (Dongarra et al., 2001).

HPL provides an implementation of the LU decomposition for solving a system of equations. The benchmark includes the ability to measure the accuracy of the solution, as well as the time required to compute it. In addition, HPL requires the use of the Message Passing Interface (MPI) for providing inter-process communication, and an implementation of the Basic Linear Algebra Subprograms (BLAS) for the linear algebra operations library.

Because of the general acceptance of HPL as the standard measure of computational performance, Feng et. al. chose to adopt the benchmark to provide the FLOPS metric for scalable system performance as it relates to energy efficiency (Feng and Cameron, 2007).

3.1. HPL CUDA

Fatica (2009) describes an implementation of HPL for NVIDIA Tesla series graphics processing units (GPUs). The approach described utilizes the CUBLAS library for the BLAS implementation and requires only minor modifications to the HPL code. In particular, the implementation utilizes the GPU as a co-processor to the CPU, executing the benchmark simultaneously on both architectures. Thus, a critical component to achieving maximum performance is to find the optimum division of processing load between the CPU and GPU.

The only modification to the HPL source code required to enable execution on the Tesla series GPUs was changing memory allocation calls to *cudaMallocHost* calls. Subsequent acceleration of the benchmark is achieved by intercepting calls to DGEMM and DTRSM to utilize the CUBLAS library routines. Fatica's implementation exploits the independence of DGEMM operations, by overlapping them on the CPU and GPU.

We used CUDA 3.2 and Open MPI 1.4.3 to execute the implementation of HPL on *Condor's* GPGPU compute nodes.

3.2. HPL Cell Broadband Engine Architecture

To execute HPL on the Cell BE of the PS3 we used a modified implementation of the one described by Kistler, et. al. (2009). The approach described was targeted for the IBM BladeCenter QS22, with two IBM PowerXCell 8i processors. The PowerXCell 8i is a component of several of the top 10 computers on the Green500 List (http://www.green500.org). Our implementation has been modified to run on the Cell BE available in the PS3, a variant similar to the IBM BladeCenter QS21. As previously mentioned, the PS3 Cell only has 6 synergistic processing elements (SPEs) available for computation, as opposed to the eight SPEs available on the PowerXCell 8i. In addition, the PowerXCell 8i has an enhanced double precision unit which the PS3 Cell does not have (Kistler et al, 2009).

Contrary to the approach used to implement HPL for the Tesla series GPUs, Kistler et. al. implemented the benchmark through multiple kernel modifications. In particular, the most compute-intensive kernels were modified to exploit the key architectural characteristics of the PowerXCell 8i. The result was the creation of an HPL acceleration library (Kistler et al, 2009).

We used the IBM Cell SDK 3.1 and Open MPI 1.4.3 to execute the implementation of HPL on *Condor's* PS3 nodes.

4. Test Methodology

To measure the energy efficiency of the *Condor* Cluster, we followed the Run Rules for submission to the Green500 List. This consists of two basic steps: (1) executing the HPL benchmark capable of achieving peak performance on the supercomputer and (2) measuring the energy consumption of the supercomputer while running the benchmark. It is understood that in many cases measuring the total system energy consumption is not feasible. Therefore, the Run Rules allow for measuring power at a subcomponent (e.g. 1U node, rack, etc.) and then extrapolating this measurement across the entire system (Run Rules, http://www.green500.org).

Given the uniqueness of the system and its heterogeneous nature, the HPL benchmark could not be run across the entire system at one time. Additionally, we were not able to measure the power for the entire system at a central location. Furthermore, there is a significant difference in the power draw between the PS3's and head compute nodes as well as the computational performance, particularly as a result of the memory limitations of the PS3 architecture. Therefore, in order to measure the total power consumed by the system, the supercomputer had to be broken down into three subcomponents: two PS3's, one NVIDIA C1060 compute node with two NVIDIA C1060s and one Intel Xeon processor, and one NVIDIA C2050 compute node with two NVIDIA C2050s and one Intel Xeon processor. The benchmark was executed on each of the three subcomponents and the power for each unit was measured in isolation. The total power for *Condor* was then determined using the following equation where P is power, R_{max} is the maximum performance achieved by HPL, and N is the number of units:

$$P_{total}(R_{max}) = N_{PS3}\cdot P_{PS3}(R_{maxPS3}) + N_{C1060}\cdot P_{C1060}(R_{maxC1060}) + N_{C2050}\cdot P_{C2050}(R_{maxC2050}) \quad (1)$$

We use a similar equation to (1) to estimate the peak performance R_{max} of *Condor*.

Prior to obtaining the results reported below, the HPL benchmark was optimized for each of the three subcomponents. Tuning HPL to achieve the maximum performance on each subcomponent consisted of varying a selection of parameters and running several cases to observe the peak FLOPS that could be attained. Documentation on performance tuning and setting up the input data file for HPL was referenced to assist in this process. One of the most critical parameters is determining the matrix size, N, to run. This decision is largely determined by the size of RAM for the processor being tested. A listing of the parameters used for our study is seen in Table 1. In addition, we show the memory available to each subcomponent and the percentage of memory which the matrix requires when running HPL.

Table 1 – Parameters used for HPL execution

Subcomponent	**Problem Size**	**Block Size**	**NBMIN**	**NDIV**	**Panel factorization**	**Recursive factorization**	**Broadcast**	**RAM**	**%RAM for *NxN***
SONY PS3	5440	128	4	2	R	L	Bandwidth Reducing	256MB	88%
NVIDIA C2050 compute node	51080	512	8	2	L	R	Increasing Ring	24GB	81%
NVIDIA C1060 compute node	51080	256	2	2	R	L	Increasing Ring	24GB	81%

To measure the power consumption of the subcomponents we used the “Watts Up? Pro ES” and followed the Power Measurement Tutorial by Ge et. al. (2006). The Watts Up? Pro ES is a digital power meter with a PC interface. The meter collects data in one second intervals and stores the results in internal memory until connected to a PC. Upon completion of a set of tests the data was downloaded to the PC via USB for recording and processing power data; Watts Up? Download Software was used to collect the data from the device.

The same method was used for capturing the power consumption of each subcomponent. Prior to powering on and executing HPL, the subcomponent power cord was connected to the power meter, which was subsequently connected to the on-rack power strip. The only difference was for the PS3s, in which we connect two PS3s to a power strip and then connected the power strip to the meter. Two PS3s were monitored because the HPL implementation used was written for a QS22 containing two Cell BE processors. Each subcomponent was then powered on and allowed to run for approximately 15 minutes. This allowed the computers to stabilize and to get accurate readings of the average idle power consumption of each subcomponent. After the stabilization period, we executed the HPL code for each particular subcomponent using the parameters determined above for achieving maximum performance. Though the Green500 Run Rules state that it is sufficient to measure power consumption for a minimum of 20% of the HPL runtime, we measured consumption over the entire run. In addition, the Run Rules state that only two runs are necessary – given a tolerance of less than 1% in power variation between the two – yet we chose to run these tests 10 and 20 times for the Tesla GPUs and PS3s, respectively.

5. Results

The results presented below show the energy efficiency performance of *Condor* at the subcomponent level. We present the energy consumption of the subcomponent running HPL versus the average idle consumption, and calculate the energy efficiency in GFLOPS/W using the peak performance achieved on HPL.

Over the course of 20 runs on two PS3 nodes the average power consumption showed little variation while executing the HPL benchmark. The average power draw for two PS3s while running the benchmark at peak performance was observed to be 199.95 W. As compared to the power draw while idle, the increase in the amount of power required to execute the peak performance of the HPL benchmark is very low, as shown in Figure 1. This demonstrates the efficiency of the PS3 while running computationally intensive problems.

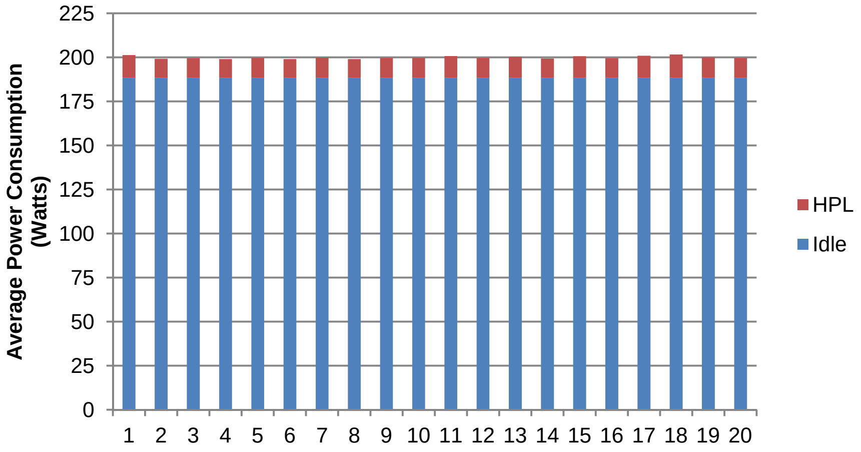

Figure 1 - Power consumption of two PlayStation 3 nodes executing the HPL benchmark. When idle, the two PS3s consume 188.49 W on average. At peak HPL performance, the nodes draw an average of 199.95 W, an additional load of approximately 5.73 W per node.

Figure 2 shows the results of each run on the PS3 in terms of GigaFLOPS (GFLOPS) achieved and the average power consumption over the entire run. There is an apparent relationship between the peak performance that is achieved and the power consumed by the nodes. In most cases, slightly higher power consumption was witnessed when the performance was greater. A similar relationship was observed on each of the subcomponents tested.

The experimental average peak performance of the PS3s was determined to be 10.46 GFLOPS. Thus, at an average rate of 199.95 W consumed, the energy efficiency for the PS3s can be calculated as .052 GFLOPS/W (52 MFLOPS/W). Such a rating would be sufficient to place the PS3 nodes in the 20th percentile of the November 2010 Green 500 List.

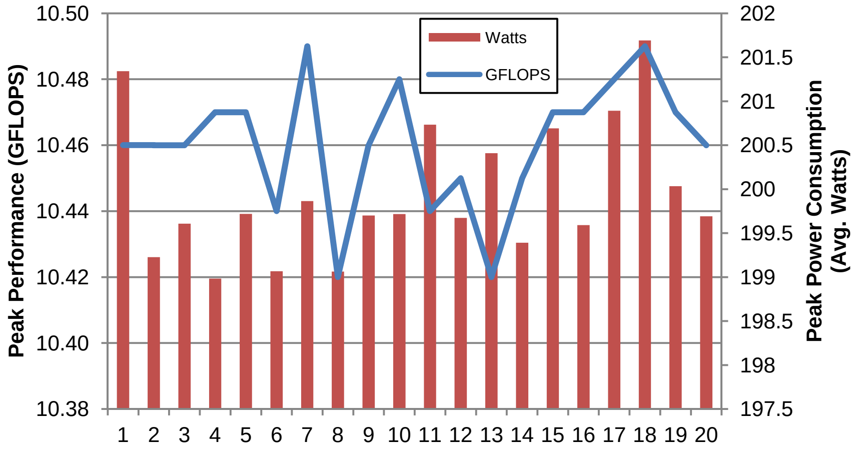

Figure 2 - Performance of the HPL benchmark on two PlayStation 3 nodes. Peak performance measured as output from HPL, while power consumption is measured as the average over the duration of the HPL execution.

For comparison, the theoretical peak performance of a single PS3 node is 10.97 GFLOPS. Thus, the peak performance for two PS3s is 21.9 GFLOPS. Experimentally, we achieved 48% of peak performance for the HPL benchmark. However, we expected this poor performance because the PS3 Cell BE is not optimized for double precision computation. On the other hand, a single PS3 node could achieve 153 GFLOPS in single precision.

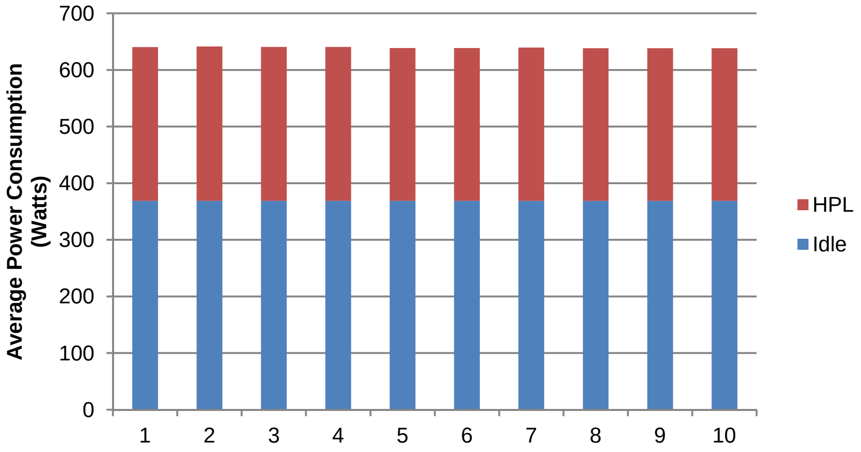

Figure 3 - Power consumption of a compute node with dual NVIDIA C2050 GPUs executing the HPL benchmark. When idle, the node consumes 368.991 W on average. At peak HPL performance, the node draws an average of 639.59 W, an additional load of approximately 270.6 W.

The NVIDIA C2050 compute nodes demonstrated higher power consumption, particularly when compared to consumption over idle use, but also showed significant improvements in HPL performance. Figure 3 shows that the average idle power consumption of the NVIDIA C2050 compute nodes is 368 W. When operating at peak performance, we observed that the nodes consumed 639 W on average. This represents a 73% increase in consumption.

Figure 4 shows the results of each run on the C2050 compute. The experimental average peak performance for the C2050 compute node was observed to be 619.5 GFLOPS, which equates to 54% of the theoretical 1.158 TFLOPS for these nodes (i.e. 128 GFLOPS for the Intel processor and 515 GFLOPS per NVIDIA C2050). The energy efficiency for the C2050 compute nodes can be calculated as .966 GFLOPS/W (966 MFLOPS/W). This efficiency would place the C2050 compute nodes in the 99th percentile of the November 2010 Green500 List.

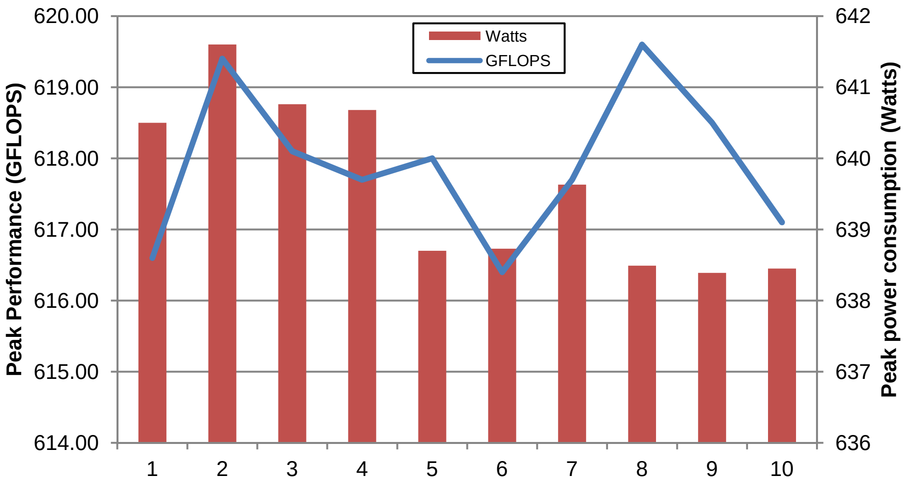

Figure 4 - Performance of the HPL benchmark on a compute node with dual NVIDIA C2050 GPUs. Peak performance measured as output from HPL, while power consumption is measured as the average over the duration of the HPL execution.

The NVIDIA C1060 compute nodes demonstrated lesser power consumption to the C2050 compute nodes. Figure 5 shows the average consumption of the C1060 compute nodes when idle as compared to the average consumption for each run of HPL. The average idle power consumption of the NVIDIA C1060 compute nodes is 337 W. When operating at peak performance, we observed that the nodes consumed 506 W on average. This represents an approximate 50% increase over idle performance.

However, unlike the C2050 compute nodes, the C1060 nodes are not fully optimized for double precision computations. In particular it is the C1060 which does not perform optimally, as the Intel processors are the same as those on the C2050 nodes. The theoretical peak performance of a C1060 for single precision is 933 GFLOPS. However, the theoretical peak performance for double precision is 78 GFLOPS (http://www.nvidia.com). Conversely, the C2050 performs at 1.3 TFLOPS in single precision and 515 GFLOPS for double precision (http://www.nvidia.com). As a result, we observed much lower performance on the C1060 compute nodes at an average of 118 GFLOPS, or 42% of the peak.

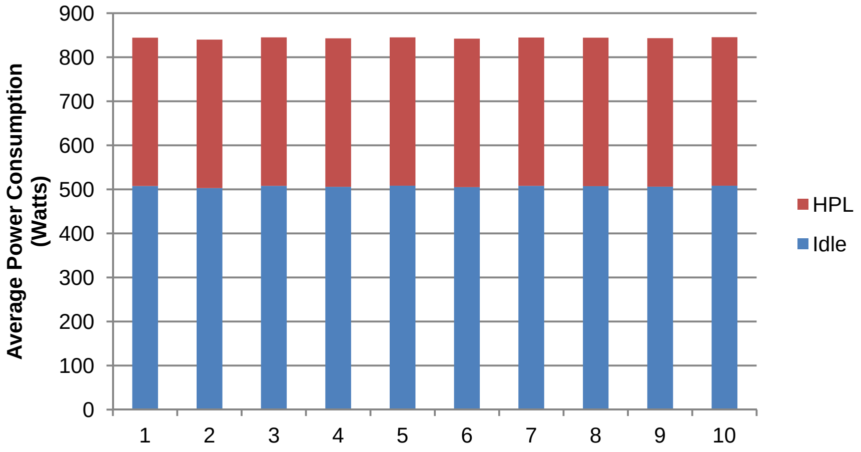

Figure 5 - Power consumption of a compute node with dual NVIDIA C1060 GPUs executing the HPL benchmark. When idle, the node consumes 336.94 W on average. At peak HPL performance, the node draws an average of 506.85 W, an additional load of approximately 169.85 W.

Figure 6 shows the results of each run on the C1060 compute. With an average performance of 118 GFLOPS and average power consumption of 506 W, the energy efficiency for the C1060 compute nodes can be calculated as .223 GFLOPS/W (223 MFLOPS/W). This efficiency would place the C1060 compute nodes in the 75th percentile of the November 2010 Green500 List.

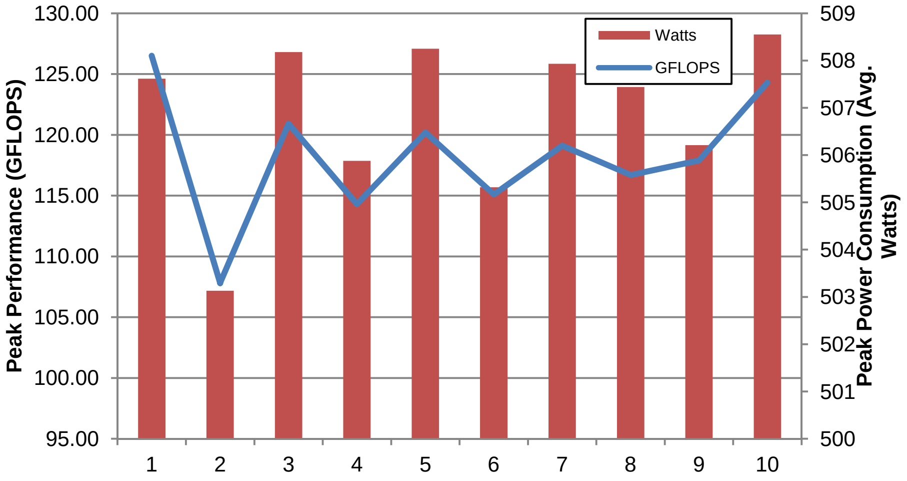

Figure 6 - Performance of the HPL benchmark on a compute node with dual NVIDIA C1060 GPUs. Peak performance measured as output from HPL, while power consumption is measured as the average over the duration of the HPL execution.

Our results across all three node classes are shown in Table 2. Using Equation (1) from above, we can calculate the overall energy efficiency of *Condor* to be approximately .192 GFLOPS/W (192 MFLOPS/W). This rating reflects 41.7 TFLOPS of double precision performance and 217.3 KW of consumed power.

Table 2 - Observed Energy Efficiency of *Condor* by Subcomponent

Subcomponent	# of Nodes	Avg Watts Per Node	GFLOPS Per Node	GFLOPS/W
SONY Playstation 3	1716	99.98	5.23	.052
NVIDIA C2050 compute node	47	639.59	619.5	.966
NVIDIA C1060 compute node	31	506.85	118.3	.233

While our method for measuring the average power consumption of the nodes is consistent with the methodology prescribed by the Green500, we realize that isolation of a single node for running HPL and then extrapolating the results across the entire supercomputer is not consistent with the TOP500 run rules. Parallelization of the benchmark across the entire supercomputer would introduce degradations on the overall performance, e.g. due to communication and coordination between the nodes. What we present here can thus be described as an experimentally-rooted theoretical maximum for the energy efficiency performance of *Condor*. In practice, we would expect the overall peak performance of HPL to drop slightly when utilizing the full cluster.

6. Conclusions and Future Work

In a time where the drive for advancing computer systems has been dominated by peak performance at any cost, the Green500 List challenges emerging developers to examine another key aspect to advanced computing, namely, energy efficiency. Not only has the cost to operate top-of-the line supercomputers soared beyond a million dollars per year, but the excessive power consumption of these emerging supercomputer is negatively impacting the environment, making energy efficiency a necessity in system design. The Green500 List provides a ranking system where the performance per Watt metric has not only taken precedence over other metrics, but has been encouraged as a primary consideration in new designs.

We demonstrated here that the *Condor* Cluster is capable of achieving energy efficiency performance that would place in the top 35% of the most recent Green500 list (http://www.green500.org). However, the computational performance is limited with respect to HPL because the Cell BE and NVIDIA C1060 are not optimized for double precision floating point operations.

However, for the majority of the applications run on the *Condor* Cluster single precision operation is sufficient; as such the design model for the supercomputer was not intended to achieve extraordinary double precision performance. We consider exploration of mixed-precision approaches to HPL (Kurzak & Dongarra, 2006) or other single precision benchmarks as an area of future research to demonstrate the efficiency of *Condor* in its targeted niche of computation.

A key design concept of *Condor* was to bring the three critical drivers in supercomputer design – peak performance, price/performance, and performance/Watt – together into a unique and highly sustainable system capable of solving some of the military’s most critical information processing problems. High performance computing systems are designed to achieve a peak performance based on their desired applications. *Condor* is capable of sustaining a peak performance of 200-300 TFLOPS required to perform several important military applications. While the cost of engineering a high performing supercomputer can be very expensive, *Condor* was built using commodity game consoles and graphics processors that achieve performance comparable to specialized architectures at a fraction of the cost. With a total cost of $2.5M, the price/performance ratio far exceeds that of comparable systems. Finally, we have demonstrated the energy efficiency of *Condor* to be 0.192 GFLOPS/W. The energy needs of *Condor* can be translated into sustainability costs on the order of $0.5M per year. Thus, the *Condor* Cluster is a powerful, yet highly sustainable asset for the Air Force Research Laboratory and Department of Defense.

Acknowledgements

The authors would like to acknowledge Qing Wu and Gaurav Khanna for their assistance in getting the HPL implementations running on *Condor*, and to NVIDIA for providing the Tesla implementation.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of AFRL.

References

"Board Specification Tesla C1060 Computing Processor Board." *NVIDIA Corporation,* January 2009. http://www.nvidia.com/docs/IO/56483/Tesla_C1060_boardSpec_v03.pdf.

"Board Specfications Tesla C2050 and Tesla C2070 Computing Processor Board." *NIVIDIA Corporation*, July 2010. http://www.nvidia.com/docs/IO/43395/BD-04983-001_v03.pdf.

Buttari, A., J. Dongarra, and J. Kurzak, "Limitations of the PlayStation 3 for High Performance Cluster Computing" LAPACK Working Note 185, 2007.

Dongarra, J., P. Luszczek, and A. Petit, "The LINPACK Benchmark: Past, Present, and Future." *Concurrency and Computation: Practice and Experience*, 15, 9, pp. 803-820, 2003.

Fatica, M., "Accelerating Linpack with CUDA on heterogeneous clusters." GPGPU '09, Washington D.C. 2009.

Feng, W., and K. Cameron, "The Green500 List: Encouraging Sustainable Supercomputing." *Computer,* 40, 12, pp. 50-55, 2007.

Feng, W., X. Feng, and R. Ge, "Green Supercomputing Comes of Age.' *IT Professional*, 10, 1, pp. 17-23, 2008.

Ge, R., H. Pyla, K. Cameron, and W. Feng, "Power Measurement Tutorial for the Green500 List." http://www.green500.org/docs/pubs/tutorial.pdf. 2007.

Gschwind, M., D. Erb, S. Manning, and M. Nutter, "An Open Source Environment for Cell Broadband Engine System Software." *Computer*, 40, 6, pp. 37-37, 2007.

"Intel Xeon Processor 5600 Series: The Next Generation of Intelligent Server Processors." *Intel Corporation*, 2010. http://www.intel.com/assets/PDF/prodbrief/323501.pdf

Kistler, M., J. Gunnels, D. Brokenshire, and B. Brenton, "Programing the Linpack benchmark for the IBM PowerCXell 8i processor." *Scientific Programming*, 17, pp. 43-47, 2009.

Kurp, P., "Green Computing: Are you ready for a personal energy meter?" *Communications of the ACM*, 51, 10, pp. 11-13, 2008.

Kurzak, J., and J. Dongarra, "Implementation of the Mixed-Precision High Performance LINPACK on the CELL Processor." LAPACK Working Note 177, 2006.

Lindholm, E., J. Nickolls, S. Obermand, J. Montrym, "NVIDIA Tesla: A Unified Graphics and Comptuing Architecture." *IEEE Micro*, 28, 2, pp. 39-55, 2008.

Maciol, P., K. Banaś, "Testing Tesla Architectuer for Scientific Computing: the Performance of Matrix-Vector Product." *Proceedings of the International Multiconference on Computer Science and Information Technology*, pp. 285-291, 2008.

"Public Law 109-431." (120 Stat. 2920, Date: 20 December 2006) 109th Congress, http://www.energystar.gov/ia/products/downloads/Public_Law109-431.pdf.

"Report to Congress on Server and Data Centery Energy Efficiency Public Law 109-431." *U.S. Environmental Protection Agency ENERGY STAR Program*, 2007. http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf.

"Run Rules for the Green500: Power Measurement of Supercomputers" Version 0.9, http://www.green500.org/docs/pubs/RunRules_Ver0.9.pdf.

The Green500 List. http://www.green500.org.

Top500 Supercomputing Sites. http://www.top500.org.