# A Survey of Homogeneous and Heterogeneous System Architectures in High Performance Computing

Yuxiang Gao
Cray Cluster Solution
Cray, Inc.
San Jose, CA 95112
sgao@cray.com

Peng Zhang
College of Engineering and Applied Sciences
Stony Brook University
Stony Brook, NY 11794
peng.zhang@stonybrook.edu

***Abstract*—The TOP500 and GREEN500 lists are two major resources to understand and forecast the future architecture design of high performance computing platform. Generally, supercomputer system design can be divided to two parts: single computing node and interconnection. Regardless interconnection, we categorize the systems into two types: homogeneous and heterogeneous; based on single node architecture. While general-purpose graphic processing unit (GPGPU) is introduced to parallel computing system, heterogeneous computing model is getting a more attractive solution to extreme scales system. Now about 25% of Top 100 supercomputers is designed by heterogeneous model. In this comparative study, we will focus on the Top 100 and Top 500 supercomputers and analyze the data to November 2015. Linpack and power efficiency will be the two main measurements to evaluate those systems. We also will highlight the latest progresses and point out the possible trends.**

***Keywords—high performance computing; system; power; homogeneous; heterogeneous;***

## I. INTRODUCTION

On July 29, 2015, Obama, the president of United State, issued "an Executive Order establishing the National Strategic Computing Initiative (NSCI) to ensure the United States continues leading in this field". At the same time, Department of Energy (DOE) announced three next generation supercomputers, two of them will base on CPU-GPU computing model; the other one will base on Intel next generation CPU, code name "Knights Landing". It appears that CPU and CPU-GPU computing models will be both critical for the future design of extreme scale systems. To face the coming innovation of high performance computing, we initially observe the existing supercomputers and technologies during the past years. That will help us to understand and forecast the trend and development in this area.

At first, we will generalize the concept of CPU and CPU-GPU computing as homogeneous and heterogeneous computing. CPU computing, or homogeneous computing is defined as interconnecting similar processing cores or units to build a high performance computer. It is a legacy principle of parallel system design. Comparing to homogeneous computing or homogeneous architecture, the systems that use more than one kind of processor or cores is referred to heterogeneous computing or heterogeneous architecture. These systems improve performance or energy efficiency not by adding the same type of processors, but by adding different processors or co-processors, which usually have specialized capabilities to speed up massive parallel tasks [1].

To overview the development of supercomputers, TOP500 and GREEN500 list will be essential resource for researchers. Since 1993, TOP500 [2] list and Linpack [3] have grown up to the most popular tool sets to evaluate supercomputing platform. Based on Linpack performance, TOP500 list, which includes the fastest 500 supercomputers, will be released biannually. Although driving performance improvement on supercomputer is still the top priority for the leading venders, this situation is slightly changing during the past 10 years.

As current top supercomputers have reached about 30 Petaflops in Linpack performance, energy consumption is becoming an extraordinary bottleneck of system development; thus, power efficiency has been dominant concern in the supercomputing platform design process. Inspired by the TOP500, the GREEN500 [4] has been released to promote the most power efficient supercomputers in TOP500 list and to encourage competition in power reduction. Overall, the TOP500 and GREEN500 provide a rich set of data for looking into the evolution of the supercomputing.

In this study, we will firstly categorize the systems to homogenous and heterogeneous in TOP500 and GREEN500 and compare the Linpack and performance efficiency. We will present the latest progress of those two variant design methods and show the advantages and shortages. Furthermore, regarding interconnections, we will subset the homogenous and heterogeneous systems by InfiniBand, GB Ethernet and custom interconnection and analyze the performance and power behaviors. We also present the distributions of supercomputers in the world by those two categories, homogeneous and heterogeneous. This paper, choosing the Top 100 and Top 500 supercomputers listed in the TOP500 and GREEN500, and

978-1-5090-5263-9/16 $31.00 © 2016 IEEE
DOI 10.1109/SmartCloud.2016.36

IEEE computer society

identifying development trends in supercomputer design, is organized as follows. In Section 2, we detail our methodology for analyzing the evolution of the top supercomputers. The power and performance representations are also presented; the latest technology and our analysis are discussed in Section 3. Finally, in Section 4, we discuss the characters of existing systems and possible future developments.

## II. EVALUATION OF SUPERCOMPUTER

Evaluating supercomputers is a challenging task, given the many aspects of the computers that are important to end users with different priorities. We apply the most popular evaluation tools such as the TOP500 [5-8] and GREEN500 [9-11] while developing our own plots.

### *A. TOP500 and GREEN500*

The TOP500 ranks the 500 supercomputers with the highest Linpack number - Rmax, a measure of maximum performance of a computer (in GFLOPS) when it runs the HPL benchmark [3, 12, 13]. In addition, for each submission the following responses are also measured: the theoretical peak performance - Rpeak, number of cores, co-processor cores, vendor, computer location, interconnection, CPU family. Another metric of interest, Linpack Efficiency, is determined by the ratio: Rmax/Rpeak. In summary, the web site has provided a rich set of data to further analyze and categorize the supercomputers.

The GREEN500 list organizers claim to rank "the most energy-efficient supercomputers in the world". Since introduced in 2007, the GREEN500 list was a reordering of the TOP500 list in order of highlighting power efficiency measured by Megaflops per Watt (MFLOPS/W), a metric used in the community. Due to the difficulty of measuring entire system, the GREEN500 allows to measure parts of the system and then multiply by a constant to get the power of entire system [9-11, 14]. For the purposes of this study, we are concerned with the power efficiency of the supercomputers appearing on the TOP500 list.

### *B. Development plots*

We develop the statistics of Top 100 and Top 500 supercomputers in TOP500 list from November 2011 to November 2015. In Figure 1, we plot the maximized and average power efficiency for Top 500 supercomputers. In Figure 2, we also plot Top 100 supercomputers in GREEN 500 list on November 2015. Using red points for homogenous system and blue points for heterogeneous system, we also normalized the Linpack performance by the minimized one 206,400 Gigaflops (GFLOPS) to show the trend that most of systems, which achieve the best power efficiency, deliver Linpack performance from 200,000 GFLOPS to 1,000,000 GFLOPS. In Figure 3, Figure 4, Figure 5 and Figure 6, we plot the average power and Linpack efficiency for homogeneous and heterogeneous systems, also all systems of Top 100 and Top 500 supercomputers from 2011 to 2015. In Figure 7 and Figure 8, we compared the Linpack and power efficiency for difference interconnection under the subsets of homogeneous and heterogeneous systems. In Table I, we list the number of systems, which are located in China, United States and other regions respectively.

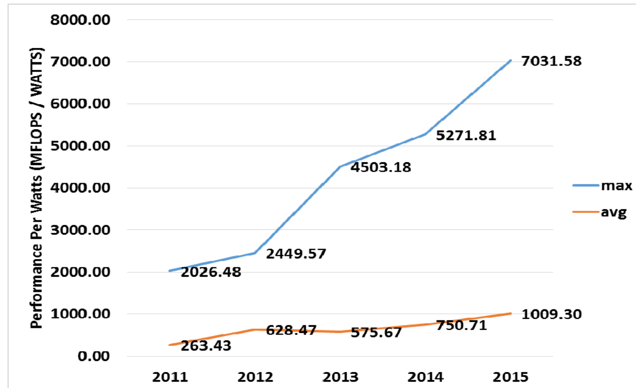


Figure 1: Top 500 supercomputers' performance Per watts

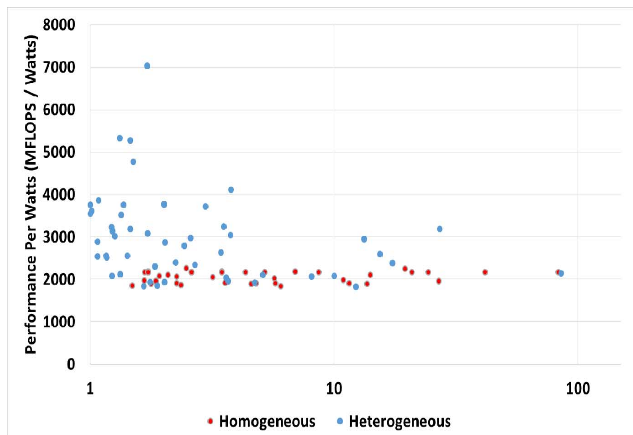


Figure 2. Performancce Per Watts by normalized Rmax in November GREEN500 list (Top 100)

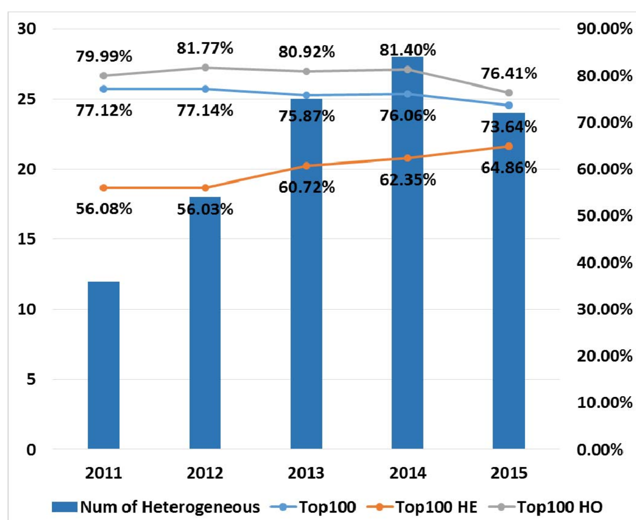


Figure 3. Average Linpack efficiency for homogeneous and heterogeneous systems (Top 100)

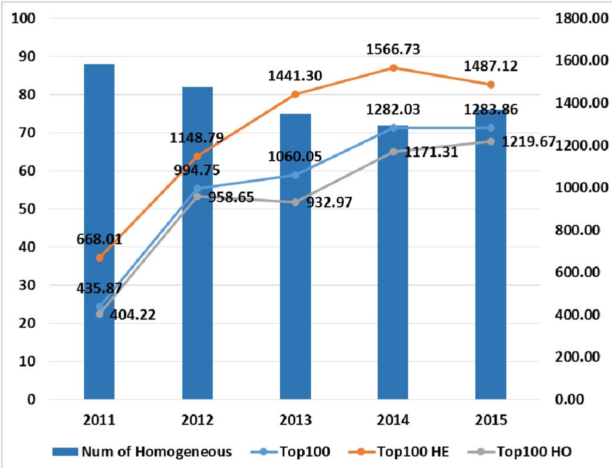


Figure 4. Average Performance Per Watts for homogeneous and heterogeneous systems (Top 100)

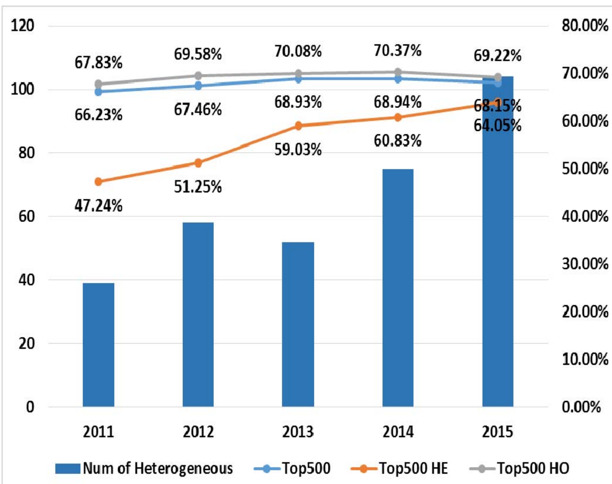


Figure 5 Average Linpack efficiency for homogeneous and heterogeneous systems (Top 500)

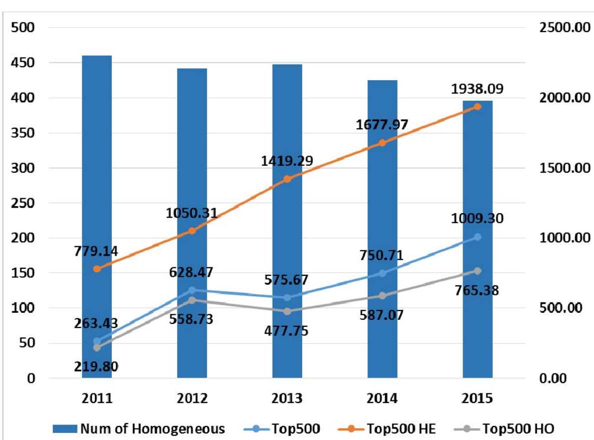


Figure 6 Average Performance Per Watts for homogeneous and heterogeneous systems (Top 500)

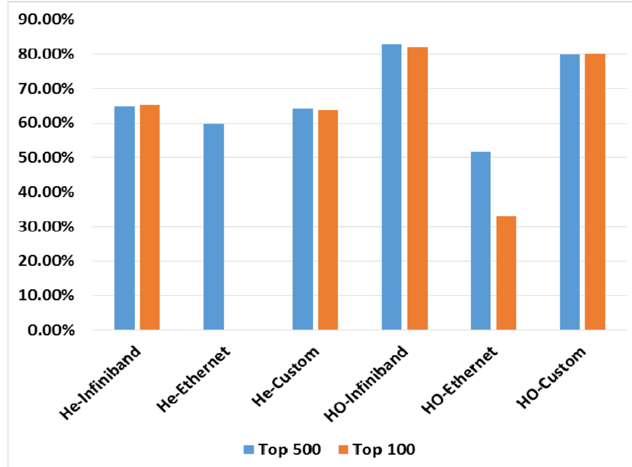


Figure 7 Average Linpack efficiency by interconnection in November 2015 Top 500 list

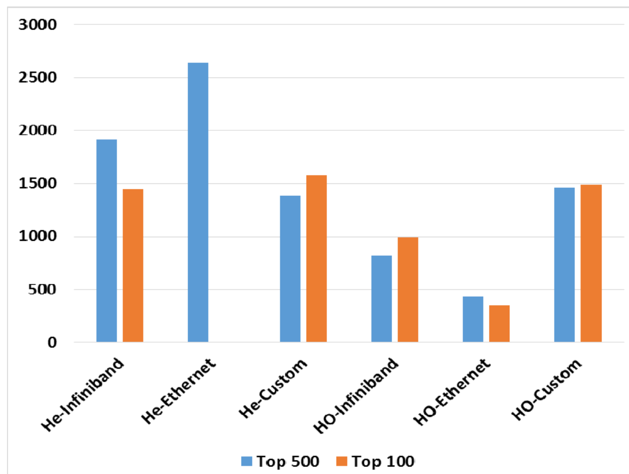


Figure 8 Average Performance Per Watts by interconnection in November 2015 Top 500 list

Table I

| Areas | Architecture | |
|---|---|---|
| | ***homogeneous*** | ***heterogeneous*** |
| US | 166 | 33 |
| China | 87 | 22 |
| Others | 143 | 49 |

## III. Analysis of Supercomputer Development

In this section, we present the Linpack and power efficiencies of the Top 100 and Top 500 supercomputers over the last five years in the TOP500 and GREEN500 lists [15]. The set of supercomputers from each list is partitioned into subsets – homogenous and heterogeneous. We compare Linpack and power efficiencies based on the following characteristics: locations and network interconnects [16]. In each case, we highlight the developmental trends of these categories [17].

In this section, we analyze the data presented in Section 2 further to draw conclusions in various other aspects.

## A. Linpack and power efficiencies

In this study, we focus on both Top 100 and Top 500 supercomputers listed in Top500 because Top 100 list is more sensitive to the advanced technology and will change more frequently; Top 500 list can present the main trend more stable.

Generally, the average Linpack efficiency of homogeneous systems is better; that of heterogeneous systems is being improved and is getting close to the average Linpack efficiency of Top 500 supercomputers. The average power efficiency of heterogeneous systems is more impressive, that of homogeneous system is getting improved, which are showed in Figure 3, Figure 4, Figure 5 and Figure 6.

In Figure 1, the average power efficiency of Top 500 supercomputers increased almost 5 times since 2011. The maximal power efficiency is more 5~10 times than the average of that.

In Figure 2 (Green500 list on November 2015), the heterogeneous systems less than 2,000,000 GFLOPS can achieve 7000 MFLOPS/W, which is 3 times of that of homogeneous systems with similar Linpack performance; both homogeneous and heterogeneous systems, which deliver more than 2,000,000 GFLOPS, can achieve 2000~3000 MFLOPS/W.

In Figure 3, the average Linpack efficiency in 2015 for homogeneous system drops due to a few large systems listed with low Linpack efficiency (~25%). That will be 81.39% if eliminating those machines.

In Figure 7 and Figure 8, we subset homogeneous and heterogeneous systems by three interconnection categories: InfiniBand, Ethernet and custom interconnection. The interesting highlights will be: the heterogeneous systems with Ethernet has good power efficiency in Top 500 supercomputers, but there is no same type of machines listed in Top 100 list; the heterogeneous systems built on InfiniBand has better average power efficiency compared with other architectures and interconnection types [17-21].

## B. Processor Family

The major processor families represented on the TOP500 list are AMD's x86-64, Intel's SandyBridge, IvyBridge and Haswell, Sparc and PowerPC.

In November 2015, only one systems used processors not from the four major processor families. The Sunway Blur Light Cluster, which built on InfiniBand and Shenwei processor in China, has 74% Linpack efficiency and 741 MFLOPS/W.

In total 104 heterogeneous systems listed in November 2015 list, there are 100 sites are built on Intel CPU and the rest of 4 sites built on AMD x86_64. And they achieved about 60% Linpack efficiency.

In total 394 homogeneous systems, there are 33 built on PowerPC or Sparc, which have made the Linpack efficiency about 85% ~90%. The rest of 363 sites built on Intel (346) or AMD x86_64 (17), can achieve about 65% Linpack efficiency.

Among the largest competitors, PowerPC, Intel's SandyBridge, IvyBridge and Haswell, AMD x86-64, Intel's CPUs is dominant with nearly 80% system share in latest release, with consistent increase in both power and Linpack efficiency. Moreover, the supercomputers like BlueGene/L, P and Q show strong scaling with near-constant efficiencies, regardless of system expansion.

In SuperComputing 2015, Intel presents new Xeon Phi processor, code name "Knights Landing", with the following features: binary compatible with Intel Xeon processor; more than 3,000 GFLOPS peak performance on double precision; 2D mesh architecture; out-of-order execute; up to 16 Gigabytes (GB) on package memory. One of the supercomputer promoted by Department of Energy will be built on the next generation of Intel Xeon Phi processor.

## C. GPGPU and Co-processor

As powerful parallel processing unit, since GPGPU and Co-processor are introduced to high performance computing, they made the big changes in every aspects of parallel computing architecture [22]. Three main vendors, Intel, NVIDIA and AMD, all provide efficient solutions and productions to improve the power efficiency of the supercomputers.

From programming perspective, Nvidia's CUDA [23] allows programmers to ignore the underlying graphical concepts in favor of more common high performance computing concepts. [24-27]. Meanwhile, Intel also provides flexible choice to end users to run program in offload, native and symmetric modes on Intel Xeon Phi Co-processors.

### 1) GPGPU

Nvidia and AMD are main vendors of GPGPU. In International SuperComputing 2015, AMD presents FirePro S9150/S9170 GPGPU, which is used to build No.3 supercomputer in GREEN500 in November 2015. And Nvidia's next generation GPGPU, Pascal GPU, will also be launched early 2016[23].

*a) NVIDIA:* Nvidia Tesla series GPGPU from K20, K40 and latest K80 provide strong computing power up to 2,910 GFLOPS. currently more than 90% heterogeneous systems in TOP500 list are built on Nvidia GPGPU. And NVIDIA has announcced next generation Pascal GPGPU, which assumes to deliver two times computing power than the previous generation. We highlight the features on current Tesla K80 GPGPU: 4992 Nvidia CUDA cores with dual-GPU design; up to 2,910 GFLOPS double-precision performance with Nvidia GPU boost; up to 8,730 GFLOPS singlr-precision performance with Nvidia GPU Boost; 24 GB of GDDR5 memory; 480 GB/s aggregate memory bandwidth [28, 29].

*b) AMD:* The AMD FirePro S9170 is equipped with ultra-fast 32GB of GDDR5 memory and is packed up to 2,620 GFLOPS of peak double precision and 9530 MFLOPS/W double precision performance. The supercomputers that are equipped with FirePro S9170 GPUs can achieve massive compute performance and optimized power efficiency.

### 2) Co-processor

In 2012, Intel announced three Xeon Phi coprocessor families: the Xeon Phi 3100, 5110P and 7120P. The Xeon Phi

7120P will be capable of 1,200 GFLOPS of double precision floating point instructions with 352 GB/s memory bandwidth at 300 W[30].

On November 2015, the No.1 supercomputer, built on Intel Xeon Phi Co-processors in China, has kept the position for 3 years. The next generation of Intel Xeon Phi Co-processor will launch late 2016 and deliver about 3,000 GFLOPS in double precision and 2-D mesh architecture[31].

### *D. Vendors*

#### *1) IBM*

In the November 2015 TOP500 list, IBM appears with 59 entries. Since June 2004, IBM is a main vendor of supercomputers. Except Blue gene series, IBM also build a CPU-GPU system in Italy, ranked 19 in TOP500 list. Highlights the features on latest IBM Blue Gen/Q system [32-34]: IBM PowerPC A2 1.6 GHz, 16 cores per node; 16 GB SDRAM-DDR3 per node; 5D Torus - 40 GB/s and 2.5 μsec latency [32].

IBM is going to build next generation supercomputers with Nvidia, which Department of Energy announced on November 2014.

#### *2) Cray*

In 2015, Cray launched high-density GPGPU or Co-processor server, CS-Storm. Two clusters built on CS-Storm are listed on Top 100 supercomputers and ranked in 13 and 14. Cray CS-Storm cluster, which is built on Intel E5-2660v2 10C 2.2 GHz with Nvidia K40, is capable of Linpack efficiency 0.583 and GREEN500 list capable of power efficiency 2386 MFLOPS/W. Also Cray XC-30 or 40 can be built with or without Nvidia GPU. In the meanwhile, Cray shows little change in system averages of Linpack efficiencies and power efficiency between November 2011 and November 2015.

#### *3) Intel*

Since Intel acquired QLogic and interconnection division of Cray, It has hold the whole hardware and software stack of high performance computing system. In SuperComputing 2015, Intel announced next generation of Xeon Phi processor, Knights Landing, which implements AVX-512, Multi-Channel DRAM (MCDRAM), and a new CPU core based on Intel's Silvermont architecture. Knights Landing is going to ship to Intel's first customers and developers as part of their early ship program, and pre-production systems for demonstrating supercomputer designs are up and running.

Intel's Omni-Path Architecture, a high bandwidth low-latency interconnect for HPC, is formally launched at SuperComputing 2015. Intel's Omni-Path implies on that Intel diverges from InfiniBand and build the market for interconnect fabrics.

#### *4) Others*

Others: vendors not part of the five are entitled "Others." On average, SGI is the leading vendor in Linpack efficiency and has remained competitive and improving power efficiency.

HP and Dell both present their systems by homogeneous and heterogeneous design. Stampede, located in Texas Supercomputer Center, is built on Intel Xeon Phi Co-processor and achieve 60.7% Linpack efficiency and 1145 MFLOPS/W.

Companies such as Fujitsu, Hitachi, NEC, etc., have shown consistent slight improvement in Linpack efficiency and have the greatest power efficiency change over the dominating vendors, especially Fuijitsu has become very competitive with their release of the K computer in 2012.

### *E. Interconnection*

In the Top 100 supercomputers, there are only seven machines built on 10G Ethernet and achieve about 25% to 30% Linpack efficiency. Most of those systems are used in industry or by service providers located in China. The rest of Top 100 systems use InfiniBand or custom interconnection. As of this date, no Ethernet supercomputers have achieved top 10% Linpack efficiency performance. In addition, although Gigabit Ethernet systems show consistent power efficiency as performance increases, only four supercomputers of this type exceeds 1,000,000 GFLOPS in performance[35].

In November 2015 TOP500 list, eight supercomputers was in the top 10% for both power efficiency and Linpack performance. The Top 40 systems in GREEN500 list are built on Intel processor with GPGPU or Co-processors.

## IV. CONCLUSION

This analysis has highlighted homogeneous and heterogeneous as main architectures of TOP500 system. All the analysis is based on this definition. It also points out other areas to explore, for example, as the TOP500 has begun to record number of cores per processor, would offer additional comparisons of the TOP500 systems by investigating to what extent of each design parameter affects Linpack performance and power efficiency. On Top500 list of June 2016, China announced a new supercomputer, Sunway TaihuLight, which delivered 93 Petaflops and captured the number one in the world. This supercomputer is composed of 40960 SW26010 processors, which has 260 cores and achieve over 3 Teraflops each. Especially the Sunway TaihuLight also stands in top three of the Green 500 list due to excellent power design. In the meanwhile, new memory technologies, 3D XPlot and software methods [36, 37], will affect the system architecture design in the future.

We have presented and analyzed the correlations of the Linpack and power efficiencies from 2011 to 2015 in the TOP500 list. We group such supercomputers according to their architectures and analyze the evolutions during the past few years. First, the heterogeneous systems are good in power efficiency, but are low in Linpack efficiency; and the homogeneous systems are low in power and good in Linpack on the contrary. Second, the heterogeneous systems with Ethernet interconnect are low in Linpack, but good in power, while systems using InfiniBand are good in both efficiencies. Third, Intel processors perform the best with high power efficiencies in GREEN 500 list. Finally, overall improvements to power and Linpack efficiency have been observed. The leading company will provide new and advanced technology to develop the next generation supercomputer systems.

# REFERENCES


[1] S. Mittal and J. S. Vetter, "A Survey of CPU-GPU Heterogeneous Computing Techniques," *ACM Comput. Surv.,* vol. 47, pp. 1-35, 2015.

[2] *Top 500 list*. Available: http://www.top500.org/

[3] A. Petitet, "HPL-a portable implementation of the high-performance Linpack benchmark for distributed-memory computers," *http://www.netlib.org/benchmark/hpl/,* 2004.

[4] *Green 500*. Available: http://www.green500.org/

[5] H. Meuer, E. Strohmaier, J. Dongarra, and H. D. Simon, "Top500 supercomputer sites," *Proceedings of SC,* pp. 10-16, 2001.

[6] E. Strohmaier, "TOP500 supercomputer," presented at the Proceedings of the 2006 ACM/IEEE conference on Supercomputing, Tampa, Florida, 2006.

[7] H. Meuer, E. Strohmaier, J. Dongarra, and H. Simon, "The Top500 project," *URL: http://www. top500. org/lists/2010/06,* 2010.

[8] J. Dongarra and P. Luszczek, "TOP500," in *Encyclopedia of Parallel Computing*, ed: Springer, 2011, pp. 2055-2057.

[9] W.-c. Feng and K. W. Cameron, "The green500 list: Encouraging sustainable supercomputing," *Computer,* vol. 40, pp. 50-55, 2007.

[10] B. Subramaniam and W.-c. Feng, "Understanding power measurement implications in the green500 list," in *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom)*, 2010, pp. 245-251.

[11] T. Scogland, B. Subramaniam, and W.-c. Feng, "The Green500 list: escapades to exascale," *Computer Science-Research and Development,* vol. 28, pp. 109-117, 2013.

[12] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK users' guide* vol. 8: Siam, 1979.

[13] J. Dongarra and P. Luszczek, "LINPACK Benchmark," in *Encyclopedia of Parallel Computing*, D. Padua, Ed., ed: Springer US, 2011, pp. 1033-1036.

[14] B. Subramaniam, W. Saunders, T. Scogland, and W.-c. Feng, "Trends in energy-efficient computing: A perspective from the Green500," in *Green Computing Conference (IGCC), 2013 International*, 2013, pp. 1-8.

[15] J. Dongarra and P. Luszczek, "HPC Challenge: Design, History, and Implementation Highlights," *Contemporary High Performance Computing: From Petascale Toward Exascale. CRC Computational Science Series. Taylor and Francis, Boca Raton,* 2013.

[16] P. M. Kogge and T. J. Dysart, "Using the TOP500 to trace and project technology and architecture trends," in *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, 2011, p. 28.

[17] Y. Deng, P. Zhang, C. Marques, R. Powell, and L. Zhang, "Analysis of Linpack and power efficiencies of the world's TOP500 supercomputers," *Parallel Computing,* vol. 39, pp. 271-279, 2013.

[18] S. Hemmert, "Green hpc: From nice to necessity," *Computing in Science & Engineering,* pp. 8-10, 2010.

[19] J. Dongarra and M. A. Heroux, "Toward a new metric for ranking high performance computing systems," *Sandia Report, SAND2013-4744,* vol. 312, 2013.

[20] Z. Peng, G. Yuxiang, and Q. Meikang, "A Data-Oriented Method for Scheduling Dependent Tasks on High-Density Multi-GPU Systems," in *High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on*, 2015, pp. 694-699.

[21] D. A. Reed and J. Dongarra, "Exascale computing and big data," *Commun. ACM,* vol. 58, pp. 56-68, 2015.

[22] S. W. Keckler, W. J. Dally, B. Khailany, M. Garland, and D. Glasco, "GPUs and the future of parallel computing," *IEEE Micro,* pp. 7-17, 2011.

[23] M. Fatica, "Accelerating linpack with CUDA on heterogenous clusters," presented at the Proceedings of 2nd Workshop on General Purpose Processing on Graphics Processing Units, Washington, D.C., USA, 2009.

[24] Y.-C. Fang, Y. Gao, and C. Stap, "Future enterprise computing looking into 2020," in *Frontier and Innovation in Future Computing and Communications*, ed: Springer, 2014, pp. 127-134.

[25] G. Yuxiang, S. Iqbal, Z. Peng, and Q. Meikang, "Performance and Power Analysis of High-Density Multi-GPGPU Architectures: A Preliminary Case Study," in *High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on*, 2015, pp. 66-71.

[26] P. Zhang and Y. Gao, "Matrix Multiplication on High-Density Multi-GPU Architectures: Theoretical and Experimental Investigations," in *High Performance Computing: 30th International Conference, ISC High Performance 2015, Frankfurt, Germany, July 12-16, 2015, Proceedings*, M. J. Kunkel and T. Ludwig, Eds., ed Cham: Springer International Publishing, 2015, pp. 17-30.

[27] P. Zhang, Y. Gao, J. Fierson, and Y. Deng, "Eigenanalysis-based task mapping on parallel computers with cellular networks," *Math. Comput,* vol. 83, 2014.

[28] J. Dongarra, M. Gates, A. Haidar, J. Kurzak, P. Luszczek, S. Tomov*, et al.*, "Accelerating numerical dense linear algebra calculations with GPUs," in *Numerical Computations with GPUs*, V. Kindratenko, Ed., ed: Springer International Publishing, 2014, pp. 3-28.

[29] F. Song and J. Dongarra, "A scalable approach to solving dense linear algebra problems on hybrid CPU - GPU systems," *Concurrency and Computation: Practice and Experience,* 2014.

[30] J. Dongarra, M. Gates, A. Haidar, Y. Jia, K. Kabir, P. Luszczek*, et al.*, "HPC programming on Intel Many-Integrated-Core hardware with MAGMA port to Xeon Phi," *Scientific Programming,* 2014.

[31] A. Gara, "Energy efficiency challenges for exascale computing," in *ACM/IEEE Conference on Supercomputing: Workshop on Power Efficiency and the Path to Exascale Computing*, 2008.

[32] "Overview of the IBM Blue Gene/P project," *IBM Journal of Research and Development,* vol. 52, pp. 199-220, 2008.

[33] D. Chen, N. Eisley, P. Heidelberger, S. Kumar, A. Mamidala, F. Petrini*, et al.*, "Looking under the hood of the IBM blue gene/Q network," presented at the Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, Salt Lake City, Utah, 2012.

[34] A. Wang, M. Gaudet, P. Wu, Jos, #233, N. Amaral*, et al.*, "Evaluation of Blue Gene/Q hardware support for transactional memories," presented at the Proceedings of the 21st international conference on Parallel architectures and compilation techniques, Minneapolis, Minnesota, USA, 2012.

[35] R. Feng, P. Zhang, and Y. Deng, "Network design considerations for exascale supercomputers," in *Proceedings of the 24th IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2012)*, 2012, pp. 86-93.

[36] K. Gai, M. Qiu, and H. Zhao, "Cost-Aware Multimedia Data Allocation for Heterogeneous Memory Using Genetic Algorithm in Cloud Computing," *IEEE Transactions on Cloud Computing,* vol. PP, pp. 1-1, 2016.

[37] M. Qiu, Z. Ming, J. Li, K. Gai, and Z. Zong, "Phase-Change Memory Optimization for Green Cloud with Genetic Algorithm," *IEEE Transactions on Computers,* vol. 64, pp. 3528-3540, 2015.