Balancing Power and Performance: A Multi-Generational Analysis of Enterprise Server BIOS Profiles

Nour Rteil[†*], Kat Burdett[‡*], Stephen Clement[‡*], Astrid Wynne[†‡], and Rich Kenny[†‡]
[*]Interact,
Harrogate, U.K
Email:nour@interactdc.com
[†]Techbuyer,
Harrogate, U.K
Email:{k.burdett,s.clement,a.wynne,r.kenny}@techbuyer.com

***Abstract*—There is growing concern over data center energy consumption and the largest energy consumer in a data center is invariably the IT equipment. Power management systems in a server allow for a balanced or energy efficient modes but often there is concern that efficiency mode will reduce performance. This paper profiles multiple generations of server from the last decade, comparing balanced settings against performance settings. We analyse the potential savings available through using a more efficient BIOS setting and given typical server utilisation conclude that there is very little reason to use the maximum performance settings outside exceptional circumstances.**

***Index Terms*—server benchmarking, data centers, energy efficiency, sustainable computing, power consumption**

1. Introduction

A data center (DC) is an energy intensive facility designed to support large amounts of information technology (IT) hardware. The DC sector consumes 1-2.4% of global electrical power [1] and is on par with the aviation sector in terms of carbon emissions [2]. The unique requirements for a DC - water, high-voltage power, skilled personnel, and in particular low latency internet connections - lead these facilities to cluster around certain locations. For example in Europe, DCs are concentrated around 5 cities: Frankfurt, London, Amsterdam, Paris and more recently Dublin [3]. This concentrates pressure on local utilities and infrastructure to the extent that some have enforced moratoriums on new data centers [4] or require independent power generation [5].

IT equipment is the largest consumer of power within a DC, typically accounting for 60-90% of facilities' electricity consumption. Since most other systems in the DC scale based on IT load, reducing the IT load is vital in curtailing data center energy consumption. Measures like the EU Ecodesign requirements [6] and Energy Star [7] have been successful in improving the energy efficiency of the IT. DC IT tends to be under-utilised and idle power is a major factor in the overall IT load, prompting local governments in Amsterdam to recently mandate servers be put in *balanced power management* modes [8]. In many DCs the IT is not accessible by the operators and customers are reluctant to utilise the more efficient power management modes due to fears of lost performance and extra latency.

In this paper we will analyze the effect of using different BIOS profiles in a variety of servers. Section 2 discusses the available profile settings and how these alter the operation of the server. Section 3 discusses the results of profiling the difference between the two primary bios profiles for the server: performance and efficiency.

2. BIOS Modes and Mechanisms

In this section, we explain the different profiles present in Dell PowerEdge and HPE ProLiant servers. We then explore some of the key system settings and present related findings on how these settings impact power and performance.

2.1. System Power and Performance Profiles

Each server manufacturer has a set of predefined system profiles that the user can utilise on their server. A system profile is a collection of BIOS system settings that are grouped to optimize energy efficiency, performance, or reliability depending on application needs.

Dell PowerEdge Gen 12 and Gen 13 servers have 4 predefined system profiles: [9] [10].

Performance Per Watt Optimized (DAPC): The BIOS controls the processor power states to achieve the maximum Performance/Watt at all utilization levels and workload types while still meeting performance requirements. BIOS also manages system Power Capping in this mode.

Performance Per Watt Optimized (OS): The (OS) controls the processors power management. The main controls are the processor frequency or performance states (P-states), and the processor clock throttling (T-states). The OS modifies the power states to achieve the best operating performance, based on the Node Manager inputs and the processor utilization.

Performance: The BIOS prioritises processor performance, disabling any scaling in clock-speed or c-states.

[‡] *Authors contributed equally to this work*
Corresponding author: Nour Rteil

TABLE 1. COMPARISON OF SYSTEM PERFORMANCE PROFILE SETTINGS FOR DELL POWEREDGE SERVERS

Profile / Settings	Generation	Performance Per Watt Optimized (DAPC)	Performance	Performance Per Watt Optimized (OS)
CPU Power Management	All	System DBPM (DAPC)	Maximum Performance	OS DBPM
Memory Frequency	All	Maximum Performance	Maximum Performance	Maximum Performance
Turbo Boost	All	Enabled	Enabled	Enabled
C1E	All	Enabled	Disabled	Enabled
C States	All	Enabled	Disabled	Enabled
Memory Patrol Scrub	All	Standard	Standard	Standard
Memory Refresh Rate	All	1x	1x	1x
Monitor/ Mwait	All	Enabled	Enabled	Enabled
Thermal Algorithm	All	Minimum Power	Maximum Performance	Minimum Power
Memory Operating Voltage	Gen 12	Auto	Auto	Auto
Collaborative CPU Performance Control	Gen 12 & 13	Disabled	Disabled	Disabled
Energy Efficiency Turbo	Gen 13	Enabled	Disabled	Enabled
Uncore Frequency	Gen 13 & 14	Dynamic	Maximum	Dynamic
Energy Efficiency Policy	Gen 13 & 14	Balanced Performance	Performance	Balanced Performance
No. of Turbo Boost Enabled Cores	Gen 13 & 14	All	All	All
Write Data CRC	Gen 14	Disabled	Disabled	Disabled
CPU Interconnect Bus Link Power Management	Gen 14	Enabled	Disabled	Enabled
PCI ASPM L1 Link Power Management	Gen 14	Enabled	Disabled	Enabled
Processor EIST	Gen 14	Enabled	Enabled	Enabled

TABLE 2. DELL DPAC VS PERFORMANCE ANALYSIS OVERVIEW

Model	12^{th} Gen PowerEdge [9]	EMC 14^{th} Gen PowerEdge [13]	R740 14^{th} Gen PowerEdge [14]
Benchmark	SPEC power_ssj2008	STREAM, WRF, Fluent, HPL (HPC Workloads)	SPEC CPU2017
Idle Power Change	66% more in Perf.	DAPC: -23%	Perf. 3x
Energy Efficiency	DAPC 44% better @ 50% load	N/A	DAPC 8% more
Performance	DAPC 2% less	HPL: Perf +4%, Others: No Change	No Change

Dense Configuration Optimized: Settings are optimized to achieve maximum reliability. This mode is typically selected for systems with high DIMM count configurations where reliability is favoured over power or performance considerations.

Dell introduced the Workstation Performance profile and removed the Dense Configuration Optimized profile in 14th generation PowerEdge servers. [11] [12].

Table 1 summarizes the different settings applied for DAPC, Performance and OS System Profiles for Gen 12, Gen 13 and Gen 14 servers. PowerEdge servers are set to DAPC by default.

HPE ProLiant servers have a different set of predefined profiles [15]:

- Balanced Power and Performance : provides optimum settings to maximize power savings with minimal impact to performance for most operating systems and applications.
- Minimum Power Usage: enables power reduction mechanisms that can negatively affect performance. This mode guarantees a lower maximum power usage by the system.
- Maximum Performance : disables all power management options that can negatively affect performance.

The default profile in HPE ProLiant servers is set to Balanced Power and Performance.

Server profiles differ from one manufacturer to the other. Performance mode in Dell PowerEdge does not encompass the same settings and configurations as the Maximum Performance mode in HPE ProLiant servers, though both modes are designed by the manufacturers to serve the same purpose: maximize performance. Users also have the option to configure each of the system settings separately as a custom profile. This is why it is important to look into some of the key settings separately and understand how their mechanisms impact power and performance.

2.2. System Power and Performance Technologies and Settings

Multiple profile options impact the performance and power draw of the CPU. Some common technologies and settings can be seen below.

Turbo Boost Technology allows the processor to engage to a higher frequency than the processors rated frequency if the processor has available power and is within temperature specifications, which results in a higher system performance.

Turbo Boost is engaged on a per-socket basis. If some of the cores of a socket are idle, then other cores of the same socket can go to a higher processor performance state. Disabling Turbo Boost reduces power usage and the maximum achievable performance under some workloads.

Intel suggests enabling Turbo mode at high server workloads to extract higher performance and lower power consumption, and disabling it at lower workloads to save power. [16]

Non-uniform Memory Access (NUMA) Technology/ Node Interleaving: NUMA links a series of nodes together with a high-speed interconnect. When node interleaving is enabled, memory addresses are interleaved across the memory installed for each processor and some workloads might experience improved performance. Hypervisors usually perform better with memory interleaving disabled as it allows them to allocate all of a VMs memory within a single NUMA node where possible.

Idle sleep states (C-states): a power-saving feature found on some Intel Xeon servers. It works by cutting the clock signal and power from idle units inside the CPU. C-states start at C0, which is the normal CPU operating mode. C1 is often referred to as a Halt state, because the processor stops executing instructions but it can still return to its executing state (C0) almost instantaneously. The higher the C number, the lower the power usage of that idle state and deeper the sleep mode the CPU goes into. This causes the CPU to take more time to fully ”wake up” from deeper states. In some cases C-states can also increase performance in workloads by allowing other cores to turbo boost to higher frequencies.

Dynamic Voltage and Frequency Scaling (DVFS) and P-states: DVFS adjusts the clock frequency of a processor by adjusting the supply voltage depending on the workload. Unlike Turbo Boost, the dependency between the idle state and clock frequency is not accounted for in DVFS.

P-states represent the voltage-frequency control states, which change depending on the workload.The operating system (OS) requests specific P-states based on the current workload and the processor may accept or reject the request and set the P-state based on its own state. The higher the P-state number, the lower the power consumption but slower the processor speed.

If the OS policy is aimed at high performance, it selects higher frequencies and uses more energy. If the policy is aimed at saving energy, it selects lower frequencies and thus lower performance. Lowering the frequency when resources are not needed can lead to significant reduction in the energy required for a computation, particularly for memory-bound workloads.

For this study, we considered the system profiles as a whole instead of their underlying and interconnected attributes.

2.3. Profile Tuning

The primary consideration when tuning the power and performance profile of a server is the type of workload it's running. Latency-critical workloads require faster computations which pose a unique set of challenges towards increasing energy efficiency. These servers waste huge amounts of energy during idle and are energy inefficient when running at low loads [17].

Dell compared various system profiles in terms of performance and power efficiency for PowerEdge 12th generation and 14th generation servers using different benchmarks. The difference between DAPC and Performance profiles are summarized in Table 2. The results show little difference in performance between the two modes across the 3 models while using significantly more power, especially at idle for the Performance profile. For this reason, they suggested that DAPC is the most suitable system profile for HPC workloads even though Performance profile provides the highest performance scores.

3. Impact of BIOS Profile

We used industry standard benchmark SPEC SERT [18] to gauge changes in performance based on system profile. This is comprised of 10 worklets, as seen in Table 3.

Figure 2 shows efficiency mode consistently uses less power to do the same, or nearly the same work, as expressed in operations across all CPU worklets and utilisation levels less than 100%. The amount to which this can be seen varies by server, with the R640 Silver 4116 showing greatest similarity between bios profiles across CPU worklets, closely followed by R640 Gold 6114 for the majority of the worklets.

TABLE 3. SERT WORKLETS AND EXAMPLE APPLICATIONS

Workload	Worklet	Applications
CPU	Compress	Used for serving GIF image formats from a website.
	CryptoAES	Used by governmental bodies to protect classified/sensitive data. It is also adopted by many mobile applications, including WhatsApp, VPN implementations and OS file systems.
	LU	Used for solving mathematical equations that are usually used in data analysis.
	SOR	Used to solve differential equations to model the behaviour of complex systems. These computations are usually used in some machine learning algorithms.
	Sort	Sorting is used in searching applications.
	SHA256	SHA-256 hash: blockchain, cryptocurrency and SSL certificates.
	SSJ	Used in most enterprise applications and eCommerce websites.
Storage	Sequential	Used in applications that store and retreive data (files, pictures, videos) such as Dropbox, Google Drive, etc.
	Random	Used in applications that store and retreive data (files, pictures, videos) such as Dropbox, Google Drive, etc.
Memory	Flood3	In-memory cache, message brokers
	Capacity3	

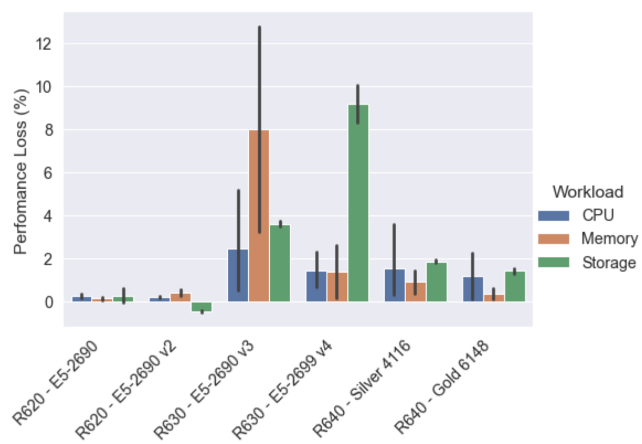

Figure 1. Average percentage loss of operations for each workload and scenario from changing bios profile from performance to efficiency.

There is a stark deviation in behavior between the scalable and pre-scalable processors across all CPU worklets - this can be best seen in the curved trendline for efficiency bios across these tests at partial utilisations. The R620 V1 and V2 seem to have very similar responses and variations between server profiles across all 6 CPU worklets - the performance mode showing a linear relationship between power consumption and performance, while the curved response of efficiency mode sees significant gains in performance at lower increases in power consumption for low utilisations tending toward more significant power costs as utilisation increases further.

There is a variation in performance at 100% on all servers and for all worklets as bios server profile changes

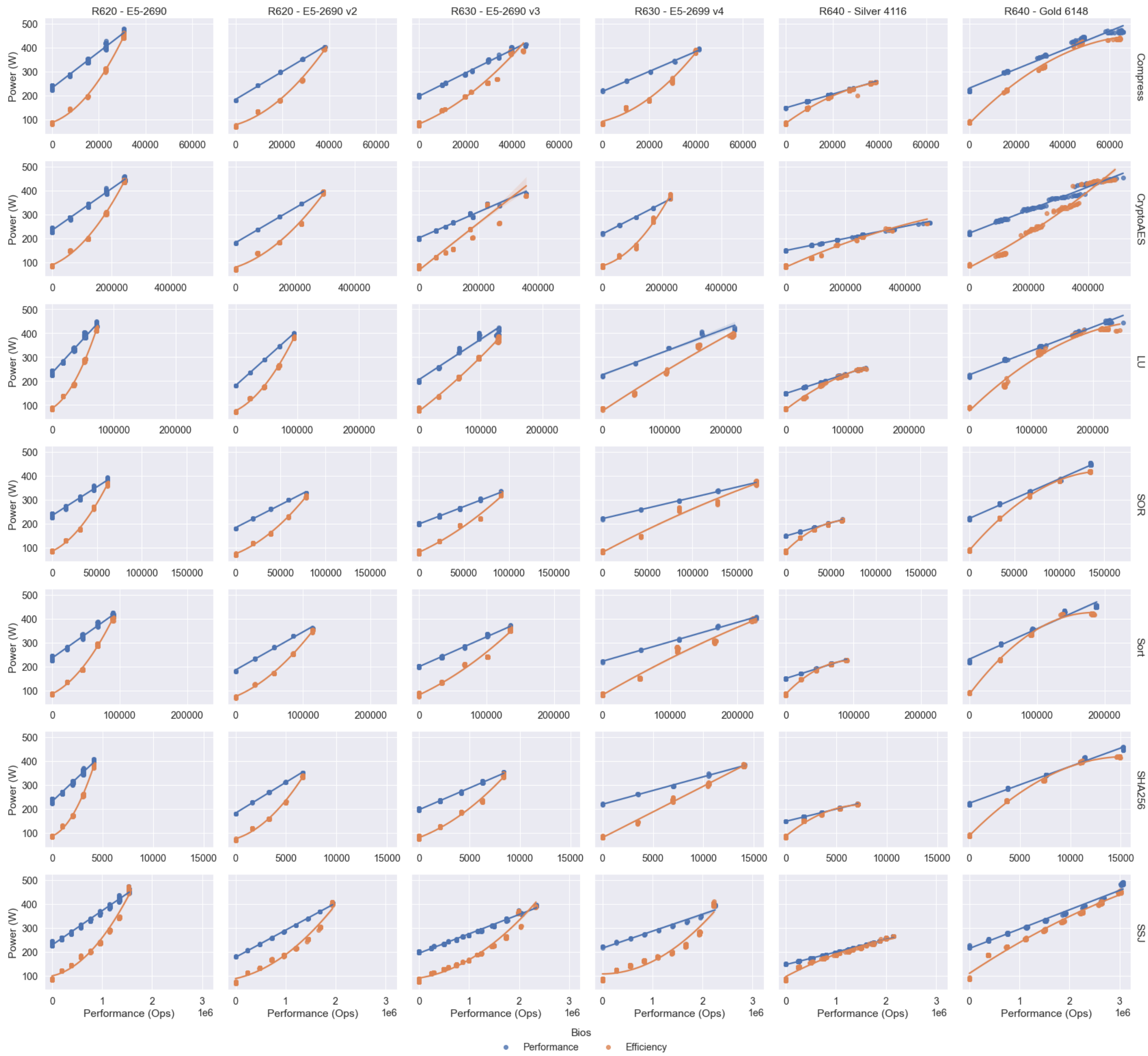

Figure 2. Benchmark performance against power consumption for every CPU worklet (rows) and server (columns) for both performance and efficiency bios'

from performance to mode, as seen in Figure 1. The average drop in total operations performed at 100% is 1.56% when considered all scenerios and servers together. When considering CPU, memory, and storage worklets separate the average drop in performance from changing bios is 1.16%, 1.85%, and 2.63% respectively - although there is considerable variation between worklets. The greatest drops in performance seem to occur on the R630 v3 for CryptoAES and Flood3, followed by both storage worklets on the R630 v4. There is an increase in performance on the R640 Gold for CryptoAES and both storage worklets for the R620 V2. While changing from efficiency to performance bios often results in a small loss in performance, that is not always the case and when it will happen and by much is very much a case by case basis.

4. Conclusion

In conclusion, for CPU bound workloads the balanced power setting represents a significant energy saving with negligible impact on peak performance, typically 1-2%, though this can vary significantly depending on workload and server. Considering the typical enterprise server has low utilisation, any impact on CPU workload performance is negated by the large savings available in energy at lower utilisation levels. The difference in power consumption between the two BIOS settings is greatest when the server is idling, up to 250W in some cases. This difference reduces as utilisation increases, with both BIOS profiles typically drawing similar power at maximum load. If a server is being consistently and highly utilised (averaging 75+%) then the case for a balanced power mode is much weaker as the power savings are less and the performance impact can start to manifest, this would most likely be the case for HPC clusters. There are also generational differences in the changes of BIOS profile for CPU workloads, the newer R640 servers exhibited a smaller increase on power consumption from performance mode and an impact on performance from the balanced power mode, so the impact of changing BIOS profiles for newer servers is less beneficial. If the difference between the BIOS profiles remains small for newer servers then legislation around server BIOS might not be as beneficial as expected.

We note that the generational differences are exaggerated in memory and storage workloads. Here we see a much larger impact on performance when switching to the balanced power mode BIOS for the R630 v3 and v4 servers. Since CPU utilisation in these workloads is low, the difference in power consumption between the two BIOS profiles tested is approximately the same as when they are idle. It depends on the use case whether an increase in memory and storage latency is offset by the energy saving, especially with R630 servers.

Though manufacturers may differ slightly in how BIOS profiles are defined, the underlying hardware is largely the same so we'd expect similar results across all servers. The significance of the savings available in IT through simple changes like BIOS profiles should not be viewed in isolation. Data centers are typically designed for peak IT loads, and partial loads exaggerate inefficiencies in other systems like cooling and power delivery, the power savings available through BIOS changes may impact other systems adversely. Therefore, data center energy saving measures should never happen in isolation or without validation.

Acknowledgments

This work was supported by a UKRI Future Leaders Fellowship [grant number MR/T04389X/1].

References

[1] IEA. (2021) Data centres and data transmission networks. [Online]. Available: https://www.iea.org/reports/data-centres-and-data-transmission-networks

[2] M. Avgerinou, P. Bertoldi, and L. Castellazzi, "Trends in data centre energy consumption under the european code of conduct for data centre energy efficiency," *Energies*, vol. 10, p. 1470, 9 2017.

[3] CBRE. (2022) Europe Data Centres Q1 2022. [Online]. Available: https://www.cbre.com/insights/viewpoints/europe-data-centres-q1-2022

[4] [Online]. Available: https://www.dutchdatacenters.nl/en/nieuws/amsterdam_moratorium_jul20/

[5] "CRU Direction to the System Operators related to Data Centre grid connection processing," Commission for Regulation of Utilities (CRU), Tech. Rep. CRU/21/124, nov 2021, available at: https://www.cru.ie/document_group/data-centre-grid-connection/.

[6] "Commission regulation (eu) 2019/424," *OJ*, vol. L 74, pp. 46–66, 2019-03-18.

[7] [Online]. Available: https://www.energystar.gov/products/enterprise_servers

[8] "Power management guide for data centers," Netherlands Enterprise Agency, Tech. Rep., 2022. [Online]. Available: https://www.rvo.nl/sites/default/files/2022-04/Guide%20Power%20Management%20for%20Data%20Centers%20-%20English.pdf

[9] J. Beckett, "Bios performance and power tuning guidelines for dell poweredge 12th generation servers," Dell, Tech. Rep., Dec 2012. [Online]. Available: https://downloads.dell.com/solutions/general-solution-resources/White%20Papers/12g_bios_tuning_for_performance_power.pdf

[10] W. Liu, "Bios setup user guide for 13th generation dell poweredge servers," Dell, Tech. Rep., Sep 2014. [Online]. Available: https://downloads.dell.com/solutions/general-solution-resources/White%20Papers/BIOS%20Setup%20User%20Guide.pdf

[11] W. Liu, M. Shutt, and P. Rubin, "Setting up bios on 14th generation (14g) dell emc poweredge servers," Dell, Tech. Rep., Apr 2018. [Online]. Available: https://docplayer.net/85002265-Setting-up-bios-on-14th-generation-14g-dell-emc-poweredge-servers.html

[12] W. Liu, M. Shutt, H. Wei, P. Rubin, and H. Liu, "Setting up bios on 15th generation (15g) dell emc poweredge servers," Dell, Tech. Rep., Aug 2021. [Online]. Available: https://dl.dell.com/manuals/common/dell-emc-poweredge-15g-set-up-bios.pdf

[13] A. K. Singh, "Bios characterization for hpc with intel skylake processor," Dell EMC HPC Innovation Lab, Tech. Rep., Aug 2017. [Online]. Available: https://dl.dell.com/manuals/all-products/esuprt_software/esuprt_it_ops_datcentr_mgmt/high-computing-solution-resources_white-papers5_en-us.pdf

[14] B. Wagner, "Dell emc poweredge server system profile performance comparison," Dell EMC, Tech. Rep., May 2018. [Online]. Available: https://downloads.dell.com/solutions/general-solution-resources/White%20Papers/SystemProfileEnergyPerformanceComparison_v1p0.pdf

[15] "Configuring and tuningh pe proliant servers for low-latency applications," HPE, Tech. Rep., Oct 2017. [Online]. Available: https://support.hpe.com/hpesc/public/docDisplay?cc=us&docId=emr_na-c01804533

[16] "Power management in intel architecture servers," Intel, Tech. Rep., Apr 2009. [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/motherboards/server/sb/power_management_of_intel_architecture_servers.pdf

[17] D. Lo, L. Cheng, R. Govindaraju, L. A. Barroso, and C. Kozyrakis, "Towards energy proportionality for large-scale latency-critical workloads," *2014 ACM/IEEE 41st International Symposium on Computer Architecture (ISCA)*, 2014.

[18] K.-D. Lange, M. G. Tricker, J. A. Arnold, H. Block, and C. Koopmann, "The implementation of the server efficiency rating tool." ACM Press, 2012, p. 133. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2188286.2188307