Draft Whitepaper: Energy Efficient HPC Working Group, Energy and Power Aware Job Scheduling and Resource Management. November 9, 2017

The purpose of this draft whitepaper is to solicit broader community feedback on the trends and implications that can be drawn from understanding the results of a global survey on large-scale HPC energy and power aware job scheduling and resource management. It is being published in conjuction with a Birds of Feather at SC17.

TITLE: A Survey of Energy and Power Aware Job Scheduling and Resource Management Techniques in Supercomputing Centers

1.0 Introduction

One of the major challenges that supercomputing centers face in building systems for high performance computing involves issues of energy consumption. Contemporary petascale systems can have peak power demands that exceed 20 megawatts and instantaneous power fluctuations of 8 megawatts. Despite ongoing improvements in microarchitectures and the use of high degrees of parallelization found in accelerator-based systems, the expectation is that the energy draw of large-scale systems will continue to increase as the community moves toward exascale systems.

The Energy Efficient High Performance Computing Working Group (EE HPC WG) has previously published work that surveys how supercomputing centers in the United States [1] and Europe [2] have been approaching problems related to energy consumption. Generally, the approaches can be broken into two broad categories: (1) approaches that involve a supercomputing center's physical plant, and (2) approaches that involve controlling characteristics of the supercomputers themselves. The first category considers practices such as lighting control (e.g., shutting off datacenter lights) or thermal management (e.g., widening the datacenter temperature setpoint levels and humidity ranges for short periods of time). The second category considers practices such as fine-grained power management (e.g., setting CPU voltage and frequency scaling settings on specific CPUs within the supercomputer), coarse-grained power management (e.g., power capping), load shifting (e.g., moving part of a workload to another facility that has more power available), or job scheduling techniques (e.g., understanding the power profiles of applications and queueing them based on those profiles to achieve some overall power or energy objective, such as recognizing that increasing the number of computational nodes dedicated to a job might increase the job's power consumption

but might decrease the job's overall energy consumption by allowing the job to be completed in less time). While techniques that fall into the first category are important, their effectiveness is limited. That is to say, once all of the extraneous equipment within the datacenter has been shut down and the temperature and humidity setpoints have been adjusted to their maximum bounds, nothing more can be achieved in this category of approaches. To that end, approaches that fall within the second category seem more likely to be able to have far-reaching impacts on energy and power consumption within supercomputing centers.

In mid-2016, the EE HPC WG formed a team focused on energy and power management through the use of job scheduling, resource management, and associated tools. The Energy and Power Aware Job Scheduling and Resource Management (EPA JSRM) team is comprised of approximately 70 members from supercomputing centers, various academic and laboratory research centers, and the vendor community, particularly focusing on job scheduling and resource management software vendors and system integrators. Most of the members are from North America and Europe, however there are members from Asia as well.

During its work, the team identified a number of supercomputing centers that have developed, or are currently developing, technologies that use EPA JSRM techniques on one or more large-scale systems. Overall, eleven sites were identified and nine sites agreed to participate in a survey that asked questions about each site's supercomputer installation, typical utilization metrics and the types of jobs the site typically runs, and details of the use of EPA JSRM techniques employed by the site. After examining responses to the survey questionnaire from each site, a three-person sub-team interviewed personnel from the site to clarify details in the responses or to ask for further technical details of responses that seemed especially noteworthy. We present a high-level evaluation of these survey responses, including unique characteristics of individual sites as well as common characteristics across sites. Based on this evaluation, we present recommendations for system software researchers and scheduler vendors who are working in this area in an effort to help guide these endeavors. To the best of our knowledge, the sites that were studied comprise the set of centers that are actively involved in developing EPA JSRM techniques for large-scale deployment, however the EE HPC WG team is open to expanding this set based on feedback to this whitepaper.

Throughout this paper we refer to two types of system software that we specifically define here. *Job schedulers* allow high-performance computing users to efficiently share the computing resources that comprise an HPC system. Users submit batch jobs into one or more batch queues that are defined within the job scheduler. The job scheduler examines the overall set of pending work waiting to run on the computer and makes decisions about which jobs to place next onto computational nodes within the computer. Generally speaking, the job scheduler attempts to optimize some characteristic such as overall system utilization or fast access to resources for some subset of batch jobs within the computing center's overall workload. The various queues that are defined within the job scheduler may be designated as having higher or lower priorities and may be restricted to some subset of the center's users, thus allowing the job scheduler to understand distinctions of importance of certain jobs within the overall workflow.

To carry out its work, a job scheduler typically interacts with one or more *resource managers*, which are pieces of system software that have privileged ability to control various resources within a datacenter. These resources can include things such as the physical nodes that make up a high-performance computer's computational resources; disks, disk channels, or burst buffer hardware that comprise I/O resources; or network interfaces, network channels, or switches that comprise interconnect resources. For example, a job scheduler might use resource management software to configure the processing cores, memory, disk, and networking resources within one or more computational nodes in accordance with the requested resources for a specific batch job prior to launching that job onto the allocated computational nodes. Finally, in some cases, resource management software might have the ability to actuate pieces of the physical plant that are responsible for delivering electricity to the datacenter or cooling the datacenter.

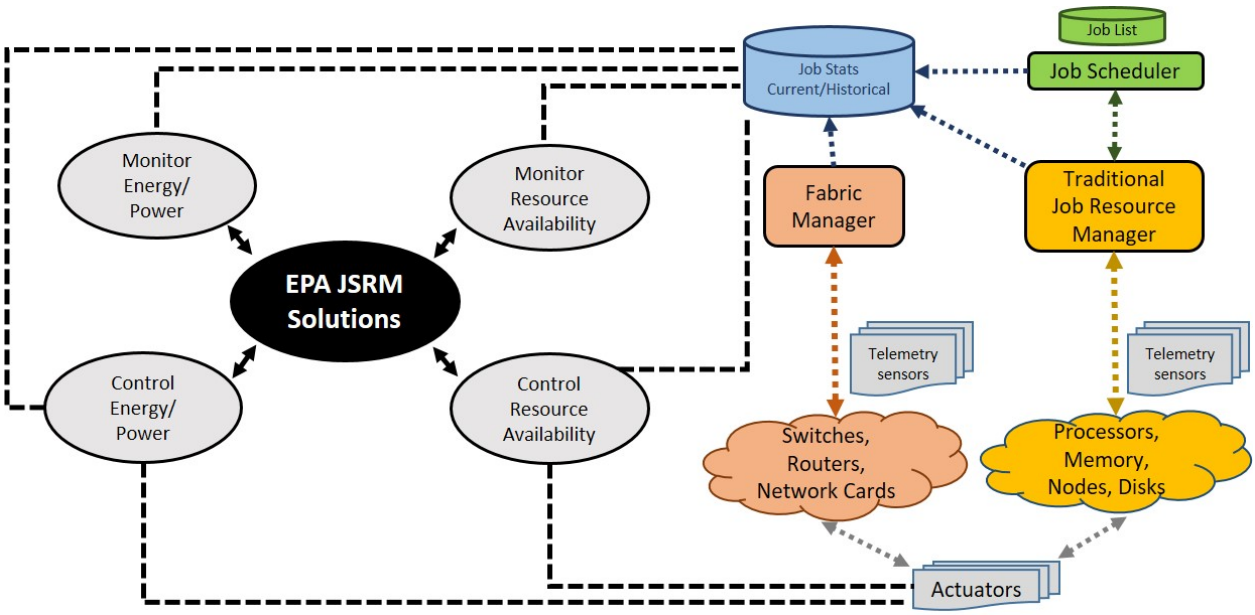

Figure 1.1: Interactions among multiple components that make up a typical EPA_JSRM (Energy and Power Aware Job Scheduling and Resource Management) solution

This paper considers the synthesized use of job schedulers and resource managers to provide *energy and power aware job scheduling and resource management* capabilities within a high-performance computing datacenter. Figure 1.1 presents an overview of the different components that may participate in such a solution. As shown, depending on the complexity of the implementation, the tasks of an EPA-JSRM solution can be divided into four functional categories - the monitoring and control of energy/power consumed by the resources, and their availability. Energy/Power 'monitoring' techniques complement traditional resource management of processors, memory, nodes, disks, and networks. The 'control' of

energy/power is heavily dependent on telemetry sensors that are responsible for constantly monitoring the activity of the system resources. Examples of such control techniques could range from simple human-controlled actuation of processor dynamic voltage and frequency settings to reduce power to much more complex scenarios where the job scheduler has detailed historical knowledge of job characteristics and schedules multiple jobs simultaneously in a way that optimizes for certain energy- or power-specific objectives. Because system-wide software agents like the job scheduler have access to details of a supercomputing center's entire workload, and can potentially apply advanced data analytics to the problem, they have the potential for improving the energy and power consumption of supercomputers in ways that are unlikely to be possible for human-controlled processes. Accordingly, we expect that a trend in coming years will be to have system-wide techniques play an increasing role in these endeavors.

The remainder of this draft white-paper describes details of the survey that was conducted as the major contribution of this paper. A future version of the paper will (1) present a brief background of energy and power aware job scheduling and resource management including related work, (2) analyze the survey results to identify characteristics common among multiple sites as well as characteristics unique to individual sites, (3) discuss opportunities and recommendations based on the analysis of survey results and, finally (4) draw conclusions and remarks about potential future work.

2.0 Survey

This section of the paper describes details of responses to the survey that was conducted as the major contribution of this paper. Criteria for inclusion in the survey included that a site (1) should be actively pursuing an energy and power aware job scheduling and resource management solution, (2) on a large-scale high-performance computing system, and (3) actively pursuing technology development with the intention of using the EPA JSRM solution in the site's production computing environment. Specifically considered as out-of-bounds for purposes of the survey were sites that were only investing EPA JSRM solutions in a research capacity or on small clusters. All of the sites identified for participation in the survey have at least one high-performance computing system within the top 100 ranks of the June 2017 Top500 list. In total, eleven sites were invited to participate in the survey with nine sites agreeing to respond. Genci/CINES in France and an anonymous secure site in the United States were identified but chose not to participate. Table 3.1 lists the sites that completed the survey. The survey process spanned eleven months from initial request to final reviewed response (September 2016 through August 2017).

Site Name	**Geographic Location**
RIKEN	Japan
Tokyo Institute of Technology	Japan
CEA	France
KAUST	Saudi Arabia
LRZ	Germany
STFC	United Kingdom
Trinity (LANL + Sandia)	United States
CINECA	Italy
JCAHPC	Japan

Table 3.1: Sites that participated in the energy and power aware job scheduling and resource management survey

Table 3.2 presents a high-level summary of the site responses to the survey, categorized into capabilities that each site is considering in the context of research, technology development with the intent to eventually deploy into production, and those that are actively deployed into the

site's production computing environment. Due to the selection process described above, some sites may not have research or technology development efforts, however all sites have some type of production deployment of energy and power aware job scheduling and resource management in place.

	Research Activities	Technology Development with Intent to Deploy	Production Deployment
RIKEN	* Integrating job scheduler info with decision to use grid vs. gas turbine energy	* Power-aware job scheduling for Post-K, with Fujitsu.	* 3 days for large jobs each month * Automated emergency job killing if power limit exceeded * Pre-run estimate of power usage of each job, based on temp
Tokyo Institute of Technology	--	* Inter-system power capping. TSUBAME2 and TSUBAME3 will need to share the facility power budget.	* Resource manager dynamically boots or shut downs nodes to stay under power cap (summer only, enforced over ~30 min window). Interacts with job scheduler to avoid killing running jobs. NEC implemented, works cooperatively with PBS Pro. * Resource manager shuts down nodes that have been idle for a long time. * Uses virtual machines to split compute nodes, complicates physical node shut down
CEA	* Analyze collected power and energy info archived long term and use for EPA scheduling. * Investigating mpi_yield_when_idle * Investigating power capping and DVFS with BULL	* Give users mark on how well they used power and energy * Developing power adaptive scheduling in SLURM, together with BULL * Developing 'layout logic' in SLURM, be able to tell what PDUs/Chillers a node or rack depends on and avoid scheduling jobs on them when maintenance is being done.	* Energy use provided to users at end of every job. * Manually shutting down nodes to shift power budget between systems.
KAUST	Monitoring and managing power usage under data center power and cooling limits	Analyzing and detecting most power hungry applications in production. Deploying the optimal power limit constraint strategy for users on Shaheen Cray XC40, while maintaining	* Static power capping via Cray CAPMC. 30% of nodes run uncapped, 70% run with 270 W power cap * SLURM Dynamic Power Management (SDPM),

		several HPC systems in production (BG/P and clusters)	interfaces with Cray CAPMC (KAUST worked with SchedMD to develop SDPM)
LRZ	* Investigating merging SLURM and GEOPM for system energy & power control. * Investigating scheduling for power, rather than energy. * Linking job scheduler with IT infrastructure + cooling; scheduler may delay jobs when IT infrastructure is particularly inefficient	* Adding energy-aware scheduling capabilities to SLURM, similar to what they have with LoadLeveler today.	* First time new app runs, characterized for frequency, runtime, and energy. * Administrator selects job scheduling goal, energy to solution or best performance. * LRZ worked with IBM on energy-aware scheduling support in LoadLeveler, now ported to LSF.
STFC	* IBM/LSF energy-aware scheduling being experimented with on small-scale (360 node) system. * Programmable interface (based on PowerAPI) for application power measurements of code segements (with interface to JSRM) * Investigation of power aware policies using higher level abstract e.g. GEOPM and Job Scheduler.	* Deployment of reporting tool for user power consumption at the job level. (Fine as well as coarse granularity)	* Continuously collecting power and energy system monitoring info, data center, machine, and job levels
Trinity (LANL + Sandia)	* Analyzing power system monitoring info to assess potential of EPA scheduling, gather traces for evaluating EPA approaches.	* Developed EPA job scheduling support with Adaptive Inc. for MOAB/Torque, interfaces with Cray CAPMC and Power API. Trinity now using SLURM, but MOAB work remains available for future use. * Developed Power API implementation with Cray, utilized by MOAB/Torque for EPA job scheduling.	* Cray CAPMC power capping infrastructure, out-of-band control, administrator ability to set system-wide and node-level power caps (available on all Cray XC systems).
CINECA	* Scalable power monitoring, used to predict per-job power	* Developing EPA job scheduling support in SLURM, with E4. Also tracking EPA	* EPA job scheduling on Eurora system (now decommissioned) using

	use and used to generate predictive models for node power and temperature evolution (with University of Bologna)	SLURM work being done by BULL and SchedMD.	PBSPro, collaboration with Altair
JCAHPC (University of Tsukuba and the University of Tokyo)	--	--	* Ability to set power caps for groups of nodes via the resource manager (Fujitsu proprietary product) * Manual emergency response, admin sets power cap. * Delivering post-job energy use reports to users

Table 3.2: Summary of site responses to the survey

In the following subsections, each of the eight questions in the survey are considered individually along with responses from specific sites.

Question 1

Question 1: *What motivated your site's development and implementation of energy or power aware job scheduling or resource management capabilities?*

Asking the sites to describe their motivation for developing and implementing EPA JSRM technology was an attempt to understand what they were hoping to accomplish, what were their goals and what was the reasoning behind the choices they made. The following paragraphs examine the specific centers' responses regarding motivation

Riken, Tokyo Institute of Technology and JCAHPC (University of Tsukuba and the University of Tokyo) all have externally-driven motivation to reduce power costs and limit power consumption, in accordance with their power supplier, or during high use times (e.g., summer) and during emergencies. In these cases, limiting power consumption was a response to shortages of electricity supply after the nuclear accident following the March 11, 2011 tsunami.

RIKEN also has explicit incentives to avoid exceeding contractual upper limits of electrical power consumption. Additionally the center uses power from a gas turbine generator as a co-generation system, which should be used optimally. Both of these are incentives based on operation costs.

CEA as well as the Trinity (LANL and Sandia) sites are not currently restricted, but anticipate the need to operate within power consumption constraints in the future. These efforts also include predictability and stability of power consumption, for example forecasting future power usage to utility providers and controlling power consumption ramp rates and band management.

KAUST is periodically limited by their site's power and cooling capacity relative to the wall-socket demands of their HPC systems. This is based on staging of multiple systems during installations as well as high power demands for acceptance testing (e.g., when running compute intensive high performance Linpack), as well as in production with full scale codes reaching power peaks higher than HPL.

LRZ has the incentive to save electricity (kilowatt per hour) cost. Further, the center has high motivation to follow the German government's push to be "green".

STFC was motivated by a mandate from the British government to improve energy efficiency across the entire computing spectrum. On the STFC center specific level this translates to optimizing compute resources with low costs for infrastructure and operations. Their immediate priority has been similar to LRZ, that is to save electricity (kilowatt hour) cost.

Cineca also has the goal of saving costs, with the strong and explicit goal of saving costs for operating cooling. Like KAUST, Cineca has been periodically limited by their site's power and cooling capacity.

Question 2

> Question 2: *Please describe your data center and major high-performance computing system or systems where energy or power aware job scheduling and resource management capabilities have been deployed in a way that covers some or all of the following points of interest. (a) Total site power budget or capacity in watts. (b) Total site cooling capacity. (c) Major high-performance computing system or systems in terms related to: number of cabinets, nodes, and cores; peak performance; node architecture, high-speed network type, memory; peak, average, and idle power draw. Other information to help describe site/system level drivers for energy or power aware job scheduling and resource management.*

The sites questioned are multi-megawatt sites with the supercomputer driving major power consumption. In some cases, while the major HPC system is the largest power consumer, it should be noted that there are several smaller systems located either on the same site or within the same building. These smaller systems share some of the power resources. Some of the tested efforts are evaluated on the smaller systems but considerations are in general to be implemented for the major system due to its power footprint. Table 3.3 summarizes the site responses. The data provided in this table was provided in response to the general questions

above, without detailed methodology about measurement requirements. As such, it should only be used as a coarse-grained indicator. It should not be used for fine-grained comparison.

Organization	(a) Site Power Budget	(b) Site Cooling capacity	(c) Major HPC System	System draw
RIKEN	12-13MW +2x5MW (Gas Turbine Co-Generation)	36MW (10500 RT)	K computer 82,944 nodes	max:15MW Avg: 12MW Idle: 10MW
Tokyo Tech	2MW	2MW	TSUBAME2.5 1400 nodes	Max:1.4 Avg: 0.8MW Idle 0.55MW
CEA	10MW	7.5MW	Anticipated 25PF System in 2017	Avg: 5 MW
KAUST	3.6MW	2.9MW	Shaheen 2 6174 nodes	max:3MW Avg: 2MW Idle: 0.55MW
LRZ	10MW	10MW +	SuperMUC Phase 1 / 2	Max 2.9MW /1.5 MW Avg 2.2MW / 1.2MW Idle: 0.7MW/0.4MW
STFC	4.5MW	2MW	846 x dual Skylake (128GB), 840 x KNL 64 core (96 GB) 24x dual Skylake (1TB)	Up to 1 MW.
LANL + SNL (Trinity)	19.2MW	15MW warm water + 12MW air	Trinity (9436 HSW nodes + 9984 KNL nodes)	Possible Peak: 8.9MW Observed Peak: 8.4MW Idle: 2.4MW
Cineca	6.5MW	Up to 4 MW	Marconi	*Only

			(7500 nodes) (1500BDW + 3600KNL + 2500SKL nodes)	subsystem evaluated at time of writing.
JCAHPC	8MW (for OFP: 4.2MW)	4.2MW for OFP	Oakforest-PACS (8208 nodes)	HPL 2.7 MW max 3.2 MW avg. 2.3-2.4MW

Table 3.3: Summary of supercomputing site power budgets, major HPC systems, and power draw (updates as of October 2017)

Question 3

Question 3: *Describe the general workload on your high-performance computing system or systems. Specifically, any or all of the following would be useful: (a) What is running right now, or what does a typical snapshot look like? How many jobs are running? What sizes are these jobs? Generally how long do jobs run? (b) What does the backlog of queued jobs look like? How many jobs are currently waiting? What are the sizes of waiting jobs? How long will they run? (c) What is the throughput of your system? Approximately how many jobs per month? (d) In simple terms, describe your main scheduling goal. Possible examples of scheduling goals might include priority, turn-around time, fairness, efficiency, or system utilization. What percentage of your system's use would you consider to be "capability" (using the maximum computing power to solve a single large problem in the shortest amount of time) or "capacity" (using efficient cost-effective computing power to solve a few somewhat large problems or many small problems)? (e) If you have statistical information available, what is the minimum, median, maximum, and 10th, 25th, 75th, and 90th percentile job size and wallclock time?*

The primary goals for all centers are optimization of utilization, serving the job mix that is a priority at the specific center as well as providing fair scheduling to the user. Table 3.4 provides an overview of general operation metrics of the various sites' systems. Information in this table is admittedly somewhat sparse due to the fact that some sites either do not specifically track these metrics or are unable to release specific details of their site workloads.

Institution	**Average running jobs**	**Average job size (nodes)**	**Average wall clock**	**Average queue size**	**System throughput**

RIKEN	110	467	5396s	300	32,000 jobs/month
Tokyo Tech	420	1.92	2.08h	650	146,727 jobs/month
CEA					
KAUST		128	13h		150,000 jobs/month
LRZ					
Trinity[1]					
STFC	50		1h	400	30,000 jobs/month
Cineca		64-128			
JCAHPC					75,000 jobs/month

Table 3.4: Job-related statistics for each center's major HPC system or systems

In terms of utilization, some centers, such as LRZ, have no specific application focus and supply the compute power to each project supported by the center. Other sites have specific production codes that are critical to the center's mission and may, accordingly, apply measures to prioritize these mission critical codes.

The following paragraphs explore details of operating goals, special operations, job mix, and utilization at specific centers.

RIKEN operates with the goal of maximizing utilization under a power constraint. Their center operates 27 days in normal operation mode where jobs can occupy up to half of the system. During the remaining 3 days of the month full system runs are possible to be scheduled.

Tokyo Institute of Technology also has the goal of maximizing utilization and throughput under a power constraint. The major difference is that 92% of their jobs are single node jobs. However, it is to be noted that 32% of their resource is consumed by jobs larger than 16 nodes. Typical jobs are Gaussian and Gromacs.

CEA also aims to maximize system utilization and achieves 85% utilization. A speciality for them is the usage of a Meta-scheduler developed in-house. This abstracts away the underlying scheduler and gives priorities according to projects and groups within the organisation. Also checkpoint-restart and mechanisms to split long running jobs are implemented using this. Functionality of the meta-scheduler can and will be reused to integrate energy considerations.

[1] Values given in the interview were derived from a different system at the cite, at the time of writing, these are reevaluated and did not reflect the state of the system, thus removed.

KAUST uses vendor provided tools and relies on SLURM and tools provided by Cray, for system measurements. The workloads are diverse as is the job mix with several small and large jobs queued 24h to fill the gaps and not drain too many nodes for a run.

LRZ splits its queues into job categories according to size. This consists of a test queue (jobs requesting fewer than 32 nodes, 30 minute limit), a micro queue (jobs requesting fewer than 32 nodes, 48 hour limit), a general queue (jobs requesting more than 32 and fewer than 512 nodes, 48 hour limit), a large queue (jobs requesting more than 512 and fewer than 2048 nodes, 48 hour limit). System nodes are shared between queues with an average node usage of 86.5% which is equivalent to 315.5 days during the year (2014). The application mix is very broad with up to 240 different applications out of all fields with a majority in fluid dynamics and astrophysics (30% each).

Trinity uses fair share scheduling according to projects, similar to CEA, with a focus on large jobs. This scheduling policy is due to the fact that capacity jobs are run on secondary clusters. A management council sets priorities according to project or deadline needs.

The system at STFC has been set up for capacity where the computing power is used to run many small problems. There is an equal split between academic and industry users and the fair share scheduling policy is used. The job can run at most 24 hours after which it is killed. Longer jobs have to be sliced in several 24h jobs for which checkpoint restart mechanisms are used. The consecutive jobs are launched from the restart files. The application mix can be very broad, but majority of applications that run is in-house developed scientific software such as Code_Saturne (35%), DL_MESO(33%), and DL_POLY(13%).

Cineca has a mixture of academic (90%) and industry (10%) users. During normal operation half of the jobs are large jobs (utilizing more than 2,000 Broadwell cores, while no statistics on the 11PF KNL section of the Marconi cluster are available, yet). The maximum job size is 1/5th of the system during normal operation. Large jobs are typically between 256 and 512 nodes (16k-32k cores). For the industry partners there are service level agreements (SLA) in place. Some academic projects also have strict SLA's such as the projects with the community of fusion energy research. The major concern for the center is fairness and system utilization.

Question 4

> Question 4: *Describe the energy and power aware job scheduling and resource management capabilities of your large-scale high-performance computing system or systems.*

Site responses to Question 4 highlight the fact that the different centers have various incentives for investigating and deploying energy and power aware job scheduling and resource management, thus the capabilities also vary.

In order not to exceed the contracted power limit, RIKEN uses some power management strategies that do not rely on job scheduling. If a job would exceed the limits, the critical job is canceled. For large jobs a preliminary evaluation is executed to estimate the consumption behavior.

Tokyo Institute of Technology implements real-time power monitoring. In addition, capabilities exist to shut down complete nodes if necessary. With these power restrictions the resource manager decides at 90% saturation not to start new jobs and shuts down jobs at 95% saturation and higher. The node level measurements are provided by Hewlett Packard, and node level power caps are set at 950W.

CEA provides power and energy information to the user through SLURM and has tools for generating reports to educate users.

KAUST is the only site using SLURM dynamic power management for large scale operation. They are also using Cray's static capping.

LRZ uses IBM LoadLeveler capabilities to operate with focus on energy to solution. They use node level measurements and energy tags in combination with a static predictive model (database). According to these measurements and tags the appropriate maximum frequencies are selected.

The energy tags technique used at STFC was developed and evaluated with IBM LoadLeveler and was finally implemented via IBM Spectrum Platform LSF which is used in production. STFC uses a smaller evaluation cluster and also looks into techniques such as DVFS, lowering power of idle nodes and operating CPUs in S3 state for underutilized nodes. The most recently installed system will also have the capability to power down idle nodes.

Cineca is testing power aware job schedulers on a smaller system with the goal of saving energy and optimizing queuing time. The effort also includes a scalable monitoring infrastructure to generate job power prediction for node power and temperature evolution. These efforts are supported from the vendor side.

JCAHPC has system power caps enforced via the resource manager. The resource management software is implemented by Fujitsu and utilizes underlying hardware capability. The current use cases are data collection and user information, with no automatic response system in place as of yet. This capability is planned.

Question 5

Question 5: *List and briefly describe all of the elements that comprise your energy and power aware job scheduling and resource management capabilities. (a) Include an implementation time component to your answer (this is, when was*

it implemented?). (b) Are these elements commercially available supported products? (c) Has there been much non-portable/non-product work done to implement your capabilities?

RIKEN designed and developed a prototype system that estimates the power of jobs and, in the future with the post K-computer, they will tackle power aware job scheduling. Riken and Fujitsu are jointly developing this capability.

Tokyo Institute of Technology's power management system has been jointly designed by the site and NEC, with implementation by NEC. It works cooperatively with the existing Job scheduler, PBS professional. "System power capping" was first implemented during April to June in 2011, and upgraded in 2013. An "energy saving" capability was implemented in 2016. These capabilties are implemented as open source software, except the existing PBS professional.

CEA has been engaged with projects related to energy aware scheduling for more than 5 years. They provide the energy usage for each job (only for those having used dedicated resources) at the end of the job. Three years ago, they started working on power capping in order to limit the datacenter global energy usage through job scheduling techniques. All of the main developments concerning energy-aware scheduling are located in the SLURM resource manager and can be used on all Linux clusters.

KAUST has the static capping capabilities of the SLURM-based Dynamic Power Capping (SDPC) installed with help of Cray, and SchedMD.

In the case of LRZ, IBM's implementation for SuperMUC Phase 1 has been available since the first day of the production system, which is also commercially available in LSF. LRZ is planning on moving towards open source with SLURM and GEOPM for these capabilities. SLURM is already in use on all other LRZ systems.

STFC's capabilities were also developed with IBM, in this case extending LSF. The power capping is available in the scheduler, with future work going toward predictive models and reporting tools for user awareness.

For Trinity a system-wide power monitoring and control infrastructure is in place. The system's power management capabilities are developed in collaboration with Cray and are currently under testing.

Cineca's first implementation was with their Eurora system two years ago, which is now being moved to the Galileo system. In both cases the queue manager is PBS Professinoal from Altair, with their own scheduler logic. PBS Professinonal is proprietary software, but its source is available upon request. A linear optimization solver for the scheduler was based on a proprietary library. This power aware scheduling system is now based on SLURM and the

prototype system entered normal operation phase named DAVIDE, featuring OpenPower processors and Nvidia GPUs, integrated by E4.

The JCAHPC capability was implemented by Fujitsu as part of the system as requested in their RFP. Fujitsu would not have done it without the request. The system estimates power consumption and, if the estimated power consumption exceeds the predetermined threshold or predetermined power limit, then the system gives warning to the administrator. Fujitsu did all the development and testing.

Question 6

> Question 6: *Do you have application/task level joint optimization, such as topology-aware task allocation, as a way of directly improving energy consumption or indirectly improving energy consumption (for example, by improving application performance, resulting in reduced wallclock time)? Did you engage software development communities to improve your energy and power aware job scheduling and resource management solution for this capability?*

Currently these efforts are not in place, but research is going on in this direction. The noteworthy exception is IBM's development using static frequencies on a per job basis. These are used in both LRZ and STFC with either LoadLeveler and LSF. A major concern is stability and testing for these systems and can be seen within all centers.

Riken has no application-awareness in their system's JSRM. The current power-estimation model and the paths envisioned build on assumptions and observations that similar jobs and same users applications have strongly similar power consumption.

Tokyo Institute of Technology uses no application aware optimization either and argue that by using full-bisection fat-tree network topologies, these considerations can be avoided and time focussed on other issues.

CEA has no power-aware scheduling in place, thus the question can be answered with their research in SLURM extensions. Stability is the main issue for bringing this into a production environment.

KAUST follows their observation that optimizing for time (performance) is analogous to optimizing for energy and focuses on the traditional goals which will bring the inherited benefit of optimized energy consumption.

LRZ is similar to KAUST; optimizing for performance is optimizing for energy. If you have optimized for performance, than the only thing left is adjusting frequencies. That is what the scheduler does. It picks the best frequency out of those available for your job.

For STFC, energy-aware scheduling research with LSF is an ongoing project carried out in conjunction with IBM. The current systems in place already show worthwhile improvements and collecting energy profiles and building models to drive control policies is part of the roadmap.

Trinity tries to focus on workload managers with minimal user involvement. This should be at most hints or region markings, but no rework of the applications.

CINECA has an active user group that is incentivised to work on this. These are the developers of the QuantumESPRESSO code. Profiling and effective energy management is evaluated and appropriate APIs are evaluated. From a center perspective, the CINECA team also is bringing the annual centers budget into the equation. This is to evaluate if the center's budget and capacity can be better planned for and resources spent more efficiently without impacting users and time to solution.

University of Tsukuba and the University of Tokyo (JCAHPC) has no such capabilities in place and still is working on an easy way to incentivise users.

Question 7

> Question 7: *How well does your solution work? What are the advantages and disadvantages of your implementation? Describe any results, benefits, or unintended consequences of your implementation.*

Many sites reported that their solutions work "well". These sites include Riken, Tokyo Institute of Technology, and LRZ. KAUST additionally reported "well with qualifications" that were related to high (up to 2x) variability in job performance from run to run. The main identified problem is increased performance variability, however queue wait times are generally reduced. LANL + SNL (Trinity) reported a similar observation.

The fact that for the most part sites developed and deployed their solutions in-house with some kind of NRE-type funding suggests that reporting that their solutions work "well" is somewhat expected. That is to say, they probably would have either continued development until their solutions work "well" or would have abandoned the effort altogether.

Other sites were really too early along their development cycle to be able to give broad feedback. These sites include CEA and LANL + SNL (Trinity).

LRZ's response highlights that challenges exist in exposing job profiles to the JSRM solution so it can make decisions about job "types" and, accordingly, decisions about how to schedule the job with energy and power in mind. LRZ's solution involves users tagging their jobs at submission time with metadata that the scheduler can use. The challenge is that this can sometimes be difficult for users to get right if scale or input decks change.

When STFC implemented their initial energy and power aware job scheduling and resource management solution, using energy tags and DVFS, they were surprised to see very small improvements to energy consumption. The main reason was found to be a low impact of this solution on short running jobs. The solution worked well for long running jobs. This was a motivation to move towards analyzing job profiles and implementing deep sleep states for idle nodes. With this combination, good results can be observed at STFC.

Cineca's response highlighted the notion that the competing objectives of the datacenter vs. users means that users are probably not ready to allow performance loss unless they get some kind of direct benefit (e.g., by making CPU hours cost less if energy-aware scheduling policies are invoked for a given job).

Question 8

> Question 8: *What are the next steps for the energy or power aware job scheduling and resource management capability you have developed? (a) Do you intend to continue site development and/or product deployment? (b) Will your planned next steps drive new requirements in procurement documents, NRE funding, etc.?*

Almost all the questioned sites have a roadmap for their power aware developments.

RIKEN's long term goal is to implement their power estimator for jobs and moving it from the initial prototype stage to production. The second goal is to implement a power-aware job scheduler that uses these estimates for the scheduling decisions.

A short term goal involves the implementation of canceling jobs that exceed the power limit set according to the contractual power limit of the power company. That is an alternative way of selecting the right application to cancel during periods of over-consumption. The upcoming procurement documents will not be affected directly until the success of these programs can be evaluated.

Tokyo Institute of Technology has the long term goal of extending the single system power capping mechanisms among multiple systems at their site. They have a shared short term goal with RIKEN, which is the selection process for job cancellation if the power limit is saturated. These considerations are already in the procurement documents for the upcoming Tsubame-3 system.

CEA's work on power aware scheduling is an ongoing effort as mentioned above and a major task is to integrate the energy-awareness into the software stack and tools, with the goal of reducing unnecessary data-movement and replication.

Kaust uses their capabilities only under special conditions since they are not necessary during normal operation currently. These special situations include transitions to new systems and during battery upgrades to avoid unnecessary downtimes. Upcoming procurements do have energy-aware JSRM included.

LRZ's goal is towards moving to an open solution, not to be tied to a proprietary solution. For upcoming systems they would require the vendors to make these developments open and not vendor IP or closed source. Going forward the integration of facility power and cooling into JSRM scheduling is a major milestone. These considerations are part of upcoming procurements.

STFCs focus is on integrating three components to enable a detailed view of system usage at different levels of granularity: (1) DCIM, a site-wide infrastructure management tool, (2) EAS, the energy aware scheduler and (3) Platform Analytics: IBM's application monitoring tool that helps maintain historic data of application performance including power measurements. These efforts are part of upcoming procurements.

LANL and SNL with their Trinity system see themselves in the learning mode and try to identify which approach to JSRM makes sense and is beneficial at which level. They have identified that this can not be offloaded to the users without having a minimally invasive solution. Thus both work towards pushing for open APIs to measure the control power in large-scale HPC systems. This means an ongoing effort is planned and will be included in upcoming systems.

CINECA is convinced that only by reducing operational expenses can future systems be operated with high efficiency. Otherwise the capital expenses cannot be justified for building larger and larger machines if they cannot be fully utilized. These systems are publicly funded.

JCAHPC (University of Tsukuba and the University of Tokyo) has just deployed their system and will be using these new capabilities as needed. They currently don't have any immediate further development plans.