Energy-Efficient, Thermal-Aware Modeling and Simulation of Datacenters: The CoolEmAll Approach and Evaluation Results

Leandro Fontoura Cupertino, Georges Da Costa, Ariel Oleksiak, Wojciech Pia Tek, Jean-Marc Pierson, Jaume Salom, Laura Siso, Patricia Stolf, Hongyang Sun, Thomas Zilio

To link to this article : DOI :10.1016/j.adhoc.2014.11.002
URL : http://dx.doi.org/10.1016/j.adhoc.2014.11.002

Energy-efficient, thermal-aware modeling and simulation of data centers: The CoolEmAll approach and evaluation results

Leandro Cupertino [a], Georges Da Costa [a], Ariel Oleksiak [b], Wojciech Piątek [b], Jean-Marc Pierson [a], Jaume Salom [c], Laura Sisó [c], Patricia Stolf [a,*], Hongyang Sun [a], Thomas Zilio [a]

[a] *Toulouse Institute of Computer Science Research (IRIT), University of Toulouse, France*
[b] *Poznan Supercomputing and Networking Center (PSNC), Poznan, Poland*
[c] *Catalonia Institute for Energy Research (IREC), Barcelona, Spain*

ABSTRACT

This paper describes the CoolEmAll project and its approach for modeling and simulating energy-efficient and thermal-aware data centers. The aim of the project was to address energy-thermal efficiency of data centers by combining the optimization of IT, cooling and workload management. This paper provides a complete data center model considering the workload profiles, the applications profiling, the power model and a cooling model. Different energy efficiency metrics are proposed and various resource management and scheduling policies are presented. The proposed strategies are validated through simulation at different levels of a data center.

1. Introduction

IT energy impact is now a major concern from the economical point of view but also from the sustainability one. IT was responsible for around 2% of the global energy consumption making it equal to the demand of aviation industry in 2008 [1]. Focusing on data centers, late 2012 numbers from the European Commission [2] shows that European data centers consumed 60 TW h during 2012. The same study expects this number to double before 2020.

While this aggregated consumption is high, still nearly a third of organizations (29%) owning data centers did not measure their efficiency in 2012 [3]. Out of this study, for the data centers that measure their Power Usage Effectiveness (PUE) [4], more than a third (34%) have a PUE over or equal to 2, meaning they consume more power on cooling, air movement and infrastructure than on computing itself. The average PUE over all data centers is between 1.8 and 1.89.

Large energy needs and significant CO_2 emissions cause the issues related to cooling, heat transfer, and IT infrastructure location more and more carefully studied during planning and operation of data centers. Even if we take ecological and footprint issues aside, the amount of consumed energy can impose strict limits on data centers. First of all, energy bills may reach millions of euros making computations expensive. Furthermore, available power supply is usually limited so it also may reduce data center development capabilities, especially looking at challenges

* Corresponding author.
E-mail addresses: fontoura@irit.fr (L. Cupertino), dacosta@irit.fr (G. Da Costa), ariel@man.poznan.pl (A. Oleksiak), piatek@man.poznan.pl (W. Piątek), pierson@irit.fr (J.-M. Pierson), jsalom@irec.cat (J. Salom), lsiso@irec.cat (L. Sisó), stolf@irit.fr (P. Stolf), sun@irit.fr (H. Sun), zilio@irit.fr (T. Zilio).

related to exascale computing breakthrough foreseen within this decade. For these reasons many efforts were undertaken to measure and improve energy efficiency of data centers. Some of those projects focused on data center monitoring and management [5–7] whereas others on prototypes of low power computing infrastructures [8,9]. Additionally, vendors offer a wide spectrum of energy efficient solutions for computing and cooling [10,11].

A variety of possibilities exist at the design level, which have to be simulated in order to be compared and to select the best one. During the lifetime of a data center, smart management can lead to better visibility of the platform behavior and to reduce energy consumption.

In order to optimize the design or configuration of a data center we need a thorough study using appropriate metrics and tools evaluating how much computation or data processing can be done within a given power and energy budget and how it affects temperatures, heat transfers, and airflows within the data center. Therefore, there is a need for simulation tools and models that approach the problem from a perspective of end users and take into account all the factors that are critical to understanding and improving the energy efficiency of data centers, in particular, hardware characteristics, applications, management policies, and cooling. To address these issues the CoolEmAll project [12] aimed at decreasing energy consumption of data centers by allowing data center designers, planners, and administrators to model and analyze energy efficiency of various configurations and solutions. To this end, the project provides models of data center building blocks and tools that apply these models to simulate, visualize and analyze data center energy efficiency.

The structure of the paper is as follows. Section 2 presents relevant related works. Section 3 contains a brief description of the CoolEmAll project. In Section 4 we present the models that are used in the design and management tools. In Section 5 the metrics used to assess the quality of design and management are presented. Section 6 describes smart data center management techniques. In Section 7 we show the results of the simulation experiments and the impact of the proposed models and tools. Section 8 concludes the paper.

2. Related work

Issues related to cooling, heat transfer, IT infrastructure configuration, IT-management, arrangement of IT-infrastructure as well as workload management are gaining more and more interest and importance, as reflected in many ongoing works both in industry and research. There are already software tools available on the market capable to simulate and analyze thermal processes in data centers. Examples of such software include simulation codes along with more than 600 models of servers from Future Facilities [13] with its DC6sigma products, CA tools [14], or the TileFlow [15] application. In most cases these simulation tools are complex and expensive solutions that allow modeling and simulation of heat transfer processes in data centers. To simplify the analysis process Romonet [16] introduced a simulator, which concentrates only on costs analysis using simplified computational and cost models, disclaiming analysis of heat transfer processes using Computational Fluid Dynamics (CFD) simulations. Common problem in case of commercial data center modeling tools is that they use closed limited databases of data center hardware. Although some of providers as Future Facilities have impressive databases, extensions of these databases and use of models across various tools is limited. To cope with this issue Schneider have introduced the GENOME Project that aims at collecting "genes" which are used to build data centers. They contain details of data center components and are publicly available on the Schneider website [17]. Nevertheless, the components are described by static parameters such as "nameplate" power values rather than details that enable simulating and assessing their energy efficiency in various conditions. Another initiative aiming at collection of designs of data centers is the Open Compute Project [18]. Started by Facebook which published its data center design details, it consists of multiple members describing data centers' designs. However, Open Compute Project blueprints are designed for description of good practices rather than to be applied to simulations.

In addition to industrial solutions significant research effort was performed in the area of energy efficiency modeling and optimization. For example, models of servers' power usage were presented in [19] whereas application of these models to energy-aware scheduling in [20]. Additionally, authors in [21,22] proposed methodologies of modeling and estimation of power by specific application classes. There were also attempts to use thermodynamic information in scheduling [23]. Nevertheless, the above works are focused on research aspects and optimization rather than providing models to simulate real data centers. In [24], the authors propose a power management solution that coordinates different individual approaches. The solution is validated using simulations based on 180 server traces from nine different real-world enterprises. Second, using a unified architecture as the base, they perform a quantitative sensitivity analysis on the impact of different architectures, implementations, workloads, and system design choices. Shah and Krishnan [25] explores the possibility of globally staggering compute workloads to take advantage of local climatic conditions as a means to reduce cooling energy costs, by performing an in-depth analysis of the environmental and economic burden of managing the thermal infrastructure of a globally connected data center network. SimWare [26] is a data warehouse simulator which compute its energy efficiency by: (a) decoupling the fan power from the computer power by using a fan power model; (b) taking into account the air travel time from the CRAC to the nodes; and (c) considering the relationship between nodes by the use of a heat distribution matrix.

3. The CoolEmAll project

CoolEmAll was an European Commission funded project which addresses the complex problem of how to make data centers more energy and resource efficient. CoolEmAll developed a range of tools to enable data center designers, operators, suppliers and researchers to plan and operate facilities more efficiently. The participants in the project included a range of scientific and commercial organizations with expertise in data centers, high performance computing, energy efficient server design, and energy efficient metrics.

The defining characteristic of the CoolEmAll project is that it bridges this traditional gap between IT and facilities approaches to efficiency. The main outcomes of CoolEmAll are based on a holistic rethinking of data center efficiency that is crucially based on the interaction of all the factors involved rather just one set of technologies. The expected results of the project included a data center monitoring, simulation and visualization software, namely SVD toolkit, designs of energy efficient IT hardware, contribution to existing (and help define new) energy efficiency metrics.

Some commercial suppliers (most notably Data center Infrastructure Management suppliers) and consultants have recently begun to take a more all-encompassing approach to the problem by straddling both IT and facilities equipment. However, few suppliers or researchers up to now have attempted to include the crucial role of workloads and applications. That is beginning to change, and it is likely that projects such as CoolEmAll can advance the state of the art in this area.

As noted in [27], the objective of the CoolEmAll project was to enable designers and operators of a data center to reduce its energy impact by combining the optimization of IT, cooling and workload management. For this purpose CoolEmAll investigated in a holistic approach on how cooling, heat transfer, IT infrastructure, and application-workloads influence overall cooling- and energy-efficiency of data centers, taking into account various aspects that traditionally have been considered separately.

In order to achieve this objective CoolEmAll provided two main outcomes: (i) design of diverse types of computing building blocks well defined by hardware specifications, physical dimensions, and energy efficiency metrics, and (ii) development of simulation, visualization and decision support toolkit (SVD Toolkit) that enables analysis and optimization of IT infrastructures built of these building blocks. Both building blocks and the toolkit take into account four aspects that have a major impact on the actual energy consumption: characteristics of building blocks under variable loads, cooling models, properties of applications, applications workloads, and workload and resource management policies. To simplify selection of right building blocks used to design data centers adjusted to particular needs, data center efficiency building blocks are precisely defined by a set of metrics expressing relations between the energy efficiency and essential factors listed above. In addition to common static approaches, the CoolEmAll approach also enables studies and assessment of dynamic states of data centers based on changing workloads, management policies, cooling methods, environmental conditions and ambient temperature. This enables assessment and optimization of data center energy/cooling efficiency also for low and variable loads rather than just for peak loads as it is usually done today. The main concept of the project is presented in Fig. 1.

4. Data center modeling

4.1. Data Center Efficiency Building Block (DEBB)

One of the main results of the CoolEmAll project is the design of diverse types of data center building blocks on different granularity levels, following a blueprint-specification format called Data center Efficiency Building Block (DEBB).

A DEBB is an abstract description of a piece of hardware and other components reflecting a data center building block on different granularity levels. To illustrate the concept, the DEBB in CoolEmAll was constructed around the RECS (Resource Efficient Computing & Storage) unit [28], a multi-node computer system developed by Christmann [29] with high energy-efficiency and density. The following describes the different granularity levels defined in the DEBB:

1. *Node unit* is the finest granularity of building blocks to be modeled within CoolEmAll. This smallest unit reflects a single CPU module in a RECS.
2. *Node group* is an ensemble of building blocks of level 1, e.g. a complete RECS unit (currently consisting of 18 computing nodes in RECS2.0).
3. *Rack (ComputeBox1)* is a typical rack within an IT service center, including building blocks of level 2 (Node Groups), power supply units and integrated cooling devices.
4. *Room (ComputeBox2)* is an ensemble of building blocks of level 3, placed in a container or compute rooms, with the corresponding CRAC/CRAH (Compute Room Air Conditioner or Air-Handling Unit), chiller, power distribution units, lighting and other auxiliary facilities.

A DEBB on each granularity level is described in the following. More details the on definitions of these components can be found in [30].

- Specification of components and sub-building blocks,
- Outer physical dimensions (black-box description), and optionally arrangements of components and sub-building blocks within particular DEBB (white-box description),
- Power consumption for different load-levels concerning mainly CPU and memory, and optionally IO and storage,
- Thermal profile describing air-flow (including direction and intensity) and temperature on inlets and outlets for different load-levels,
- Metrics describing energy efficiency of a DEBB.

A computing node will be the smallest unit of the modeling process in DEBB. The models established at a lower level, e.g., a Node unit or Node group should provide building blocks to the modeling of larger modules, e.g. full racks or server rooms, for simulations. In this way, DEBBs can improve and facilitate the process of modeling,

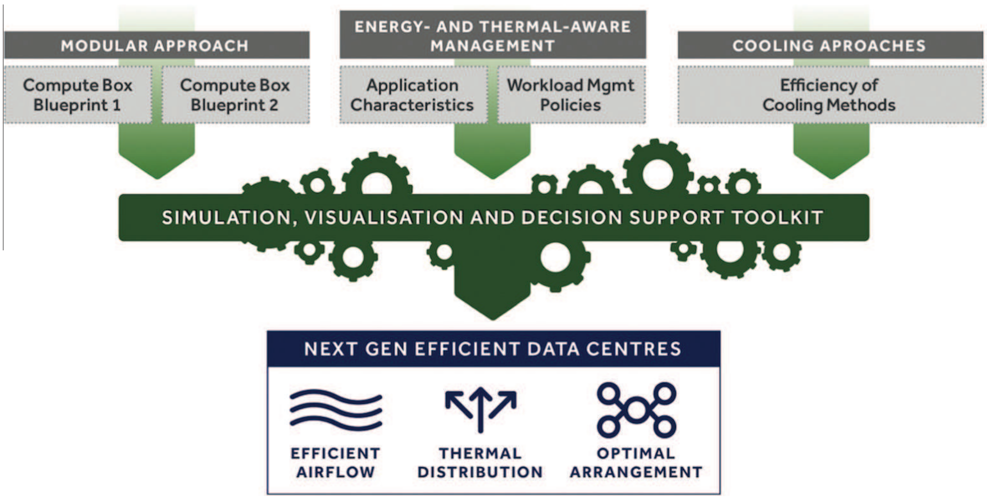


Fig. 1. The CoolEmAll concept.

simulation, and visualization of data centers by delivering predefined models with comprehensive information concerning performance, power consumption, thermal behavior, and shape of data center components.

4.2. Workload characterization

4.2.1. Workload specification

In terms of workload management, workload items are defined as jobs that are submitted by users [31]. In general, workloads may have various shapes and levels of complexity ranging from multiple independent jobs, through large-scale parallel tasks, up to single applications that require single resources. That allows distinguishing several levels of information about incoming jobs. These levels are presented in Fig. 2. It is assumed that there is a queue of jobs submitted to the resource manager, and each job consists of one or more tasks. If preceding constraints between tasks are defined, a job may constitute a whole workflow.

The aim of workload profile is to provide information about structure, resource requirements, relationships and time intervals of jobs and tasks that will be scheduled during the workload simulation phase. Having these dependencies established, it is possible to express the impact of each workload item on the system. To this end, each job specified within the workload has to be extended with the particular application characteristic describing its behavior on the hardware. Thus, workload profile contains the references to the corresponding application profiles that are linked during the simulation. To model the application profile in more detail, CoolEmAll follows the DNA approach proposed in [32]. Accordingly, each task can be defined as a sequence of phases that show the impact of the application on the resources that run it. Phases are then periods of time within which the system is stable (cpu load, network, memory, etc.) given a certain threshold. More details concerning application profiles are provided in the next section. This form of description enables definition of a wide range of workloads: HPC (long jobs, computational-intensive, hard to migrate) or web service (short requests) that are typical for virtualized data center environments.

For the purposes of the workload description within the CoolEmAll project we adopted Standard Workload Format (SWF) [33] that is used for the traces stored in the Parallel Workloads Archive. For now it is one of the main and commonly used formats providing unitary description of both workloads models as well as logs obtained from real systems. In addition to the predefined labels in the header comments, described by Feitelson in [33], we introduce support of a new header label that is used to provide information about types of applications. An example of a workload expressed in SWF is presented in Fig. 3.

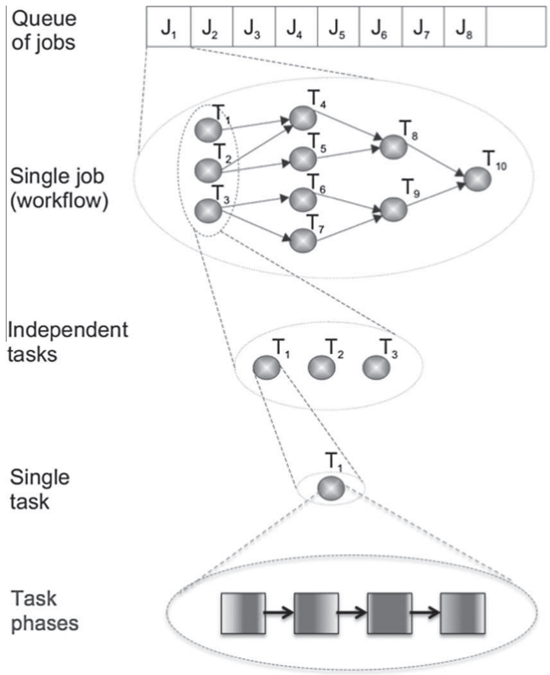


Fig. 2. Workload model.

```
;StartTime: Thu Jan 17 10:00:00 CET 2013
;Application: 1 Linpack
;Application: 2 Abinit
;Application: 3 Mencoder
```

1	514	-1	8720	1	-1	-1	1	-1	-1	-1	-1	-1	3	-1	-1	-1	-1
2	1057	-1	4513	8	-1	-1	8	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
3	1589	-1	7266	2	-1	-1	2	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
4	2161	-1	6636	4	-1	-1	4	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
5	2742	-1	12845	2	-1	-1	2	-1	-1	-1	-1	-1	3	-1	-1	-1	-1
6	3284	-1	7867	1	-1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
7	3831	-1	8361	4	-1	-1	4	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
8	4409	-1	5644	8	-1	-1	8	-1	-1	-1	-1	-1	3	-1	-1	-1	-1
9	4968	-1	8493	8	-1	-1	8	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
10	5524	-1	3780	4	-1	-1	4	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
11	6103	-1	7568	1	-1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
12	6634	-1	6818	16	-1	-1	16	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
13	7225	-1	5663	8	-1	-1	8	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
14	7810	-1	11140	2	-1	-1	2	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
15	8391	-1	8333	2	-1	-1	2	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
16	8941	-1	4164	8	-1	-1	8	-1	-1	-1	-1	-1	3	-1	-1	-1	-1
17	9527	-1	7568	8	-1	-1	8	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
18	10103	-1	10431	1	-1	-1	1	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
19	10674	-1	6209	4	-1	-1	4	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
20	11279	-1	5760	4	-1	-1	4	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
21	11825	-1	7871	8	-1	-1	8	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
22	12369	-1	5709	8	-1	-1	8	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
23	12912	-1	8616	16	-1	-1	16	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
24	13452	-1	1958	8	-1	-1	8	-1	-1	-1	-1	-1	3	-1	-1	-1	-1
25	14038	-1	7770	4	-1	-1	4	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
26	14573	-1	2370	8	-1	-1	8	-1	-1	-1	-1	-1	3	-1	-1	-1	-1
27	15147	-1	11933	8	-1	-1	8	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
28	15708	-1	7730	2	-1	-1	2	-1	-1	-1	-1	-1	2	-1	-1	-1	-1
29	16238	-1	3576	8	-1	-1	8	-1	-1	-1	-1	-1	3	-1	-1	-1	-1
30	16815	-1	8602	4	-1	-1	4	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
31	17352	-1	8785	4	-1	-1	4	-1	-1	-1	-1	-1	1	-1	-1	-1	-1
32	17936	-1	10676	1	-1	-1	1	-1	-1	-1	-1	-1	2	-1	-1	-1	-1

Fig. 3. SWF workload definition.

In general, workload profiles may be taken from real systems or generated synthetically. The main goal of synthetic workloads is to capture the behavior of real observed workloads and to characterize them at the desired level of detail. On the other hand, they are also commonly adopted to evaluate the system performance for the modified or completely theoretical workload models. Usage of synthetic workloads and their comparison to the real ones have been the subject of research for many years. In [34], the authors analyzed both types of workloads in terms of their accuracy and applicability. Today, several synthetic workload models have been proposed [35,36], which are based on workload logs collected from large scale parallel systems in production use. In a set of experiments depicted in Section 7, we define workloads using arrival rate based on the Poisson process as it has been typically adopted to reflect the task arrivals in supercomputing clusters [35,36] as well as in web servers [37].

4.2.2. Application specification

For the purpose of CoolEmAll, applications behavior can be assimilated to its resource consumption. Indeed, CoolEmAll project aims at evaluating the impact of applications from a thermal and energy point of view. Using resources consumption allows evaluating this impact. As applications are usually complex, their resource consumption cannot be assimilated as constant during their lifetime. Applications will be considered as a sequence of phases. One phase will be considered as a duration during which resource consumption can be considered as constant. As the same application will consume different amount of resources depending on the hardware on which it runs, application profiles will encompass the hardware on which it ran. Using a translator it will be possible to take a profile obtained on a particular hardware and to translate it to the probable resource consumption on different hardware. Exact values will be in percentage of maximum available resource. For instance, for CPU this will be the load, and for the network this will be the ratio (in percentage) between the actual bandwidth and the maximum on the platform.

One phase will be characterized by its duration, by the mean resource consumption during this duration and by the reference hardware used to obtain those values. As an example, a simplified XML description of an application could include the section shown in Fig. 4 where there are two phases, one of 4 s, one of 40 s. The first one uses mainly the CPU while not using the memory infrastructure and thus can be labeled as CPU-intensive. The second one loads at the maximum of CPU and memory. Such phase is usually labeled as memory-intensive. The current available resources are shown in Table 1. Once a profile is acquired, it can be displayed as shown in Fig. 5.

4.2.3. Power model

Power measurement is a key feature for the development and maintenance of energy efficient data centers. The use of power models enables the estimation of application's power dissipation, offering a higher granularity for the measurement and leveraging the application scheduling with energy consumption. Besides, even power meters can present some inaccuracy and the use of such models can enhance the measurements.

In CoolEmAll, we target at the RECS compute box, a high density computing system with embedded power meters for each of its computing nodes. In this study, we used a RECS version 2.0 with two types of modules: Intel Core i7-3615QE processor with 16 GB of RAM and Intel Atom N2600 processor with 2 GB of RAM. The compute box is populated with 18 nodes in total – 6 i7 and 12 atom nodes.

The embedded RECS's power meters have a precision of 1 W, which for some usages may not be enough. Even

```xml
<resourceConsumptionProfile>
    <resourceConsumption>
        <referenceHardware>Intel_i7</referenceHardware>
        <duration>PT4S</duration>
        <behaviour name="cpu">
            <value>99</value>
        </behaviour>
        <behaviour name="memory">
            <value>29</value>
        </behaviour>
    </resourceConsumption>
    <resourceConsumption>
        <referenceHardware>Intel_i7</referenceHardware>
        <duration>PT40S</duration>
        <behaviour name="cpu">
            <value>98</value>
        </behaviour>
        <behaviour name="memory">
            <value>92</value>
        </behaviour>
    </resourceConsumption>
</resourceConsumptionProfile>
```

Fig. 4. Example of profile of an application composed of two phases.

Table 1
Resources monitored for profiling applications.

Resource name	Content
CPU	Percentage of processor load
Network	Percentage of network bandwidth
IO	Percentage of disk bandwidth
Memory	Percentage of memory bandwidth

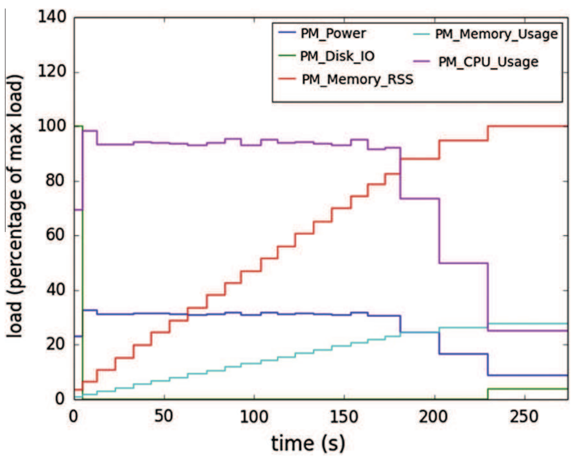


Fig. 5. Graphical representation of application profile for a ray tracing benchmark (c-ray).

more, when running some experiments, we noticed that the power meter measurement inaccuracy can reach up to 24 W. These experiments were executed by stressing one node, while the others remains turned off; in this configuration, the power of the turned off nodes reported by the power meter varies according to the stressed node and reaches from 0 to 9 W maximum. So the three most erroneous nodes were stressed and the power of the turned off nodes summed up 24 W. This enforces the need for estimating the power even if we dispose of a physical meter. For modeling the power dissipation of a node, we chose one of the nodes which presented less noise in other nodes and included an external power meter to provide higher precision measurements.

A usual way of modeling the power consumption of a node is to create a CPU proportional estimator. As the processor is claimed to be the most power hungry device on a computing system [38], capacitive models are greatly used as follows:

$$P = (cv^2 f)u, \tag{1}$$

where the power (P) is estimated based on the CPU's voltage (v), frequency (f), capacitance (c) and usage (u). This model does not take into account the type of operation that is executed by the CPU. For instance, the same CPU load can provide different power consumption according to the device it uses [39]. Previous work verified that CPU temperature has a high correlation with the dissipated power [40], even though one cannot decouple the temperature between applications.

In CoolEmAll we use an application level estimator based on performance counters, model specific registers (MSR) and system information. It has been shown that calibrated linear models can provide estimation to generic applications with an accuracy error smaller than 10% [41,40]. Performance counters (PC) are CPU counters that quantify the number of events done by the processor per core, e.g. cache misses. These counters can be fetched at process level, making the transition to an application level modeling straight forward. MSR provides precise information regarding the processor's operating frequency and C-states. Although MSRs cannot provide process level information, we can join its information to other process dependent variables such as PC. System information is fetched from the operating system and provides data for networking and memory usage. A complete list of evaluated variables can be founded in Table 2.

A set of synthetic benchmarks was created to simulate a generic application running on this platform. This benchmark set consists of four phases; first we progressively stress the CPU by increasing its usage in 20% steps. This procedure is repeated for three frequencies (1.2, 2.3 GHz and Boost) and three CPU intensive benchmarks which stress the control unity, floating point unity and the random number generator. The second phase stressed the memory access, by forcing read/write access to all caches (L1d, L2 and L3) and the RAM. Finally the network is stressed by running Linux's `iperf` tool and limiting the download/upload to 200, 400 and 1000 Mbit/s. A detailed analysis of these benchmarks can be found in [39]. These data were then used to calibrate the capacitive model and to create a linear model using the above mentioned variables. The power profile of this synthetic workload is shown in Fig. 6.

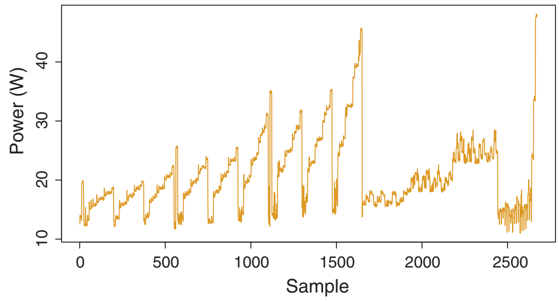


Fig. 6. Synthetic workload's power profile.

The results of the calibration of the two models can be seen in Fig. 7. One can see that the capacitive model fails when different programs present the same CPU load and lacks the power dissipation due to RAM memory access, presenting an mean absolute error (MAE) of 2.38 W and a correlation factor of 0.6961. The use of performance counters, even as a black box provides a better estimation with a MAE of 0.51 W and a correlation of 0.9831. The results of the black box present a better precision than the embedded power meter, which has 1 W precision.

4.3. Cooling model

The cooling model defined in CoolEmAll has the objective of calculating the power of the cooling equipment and other electric devices in a data center as a function of IT workload, ambient temperature and room set-up operation temperature. The model is based on a simple data center with a computer room air handler (CRAH), e.g., fan and air–water coil, power distribution unit (PDU) and lighting. All these elements generate thermal load and provide the cooling and power requirements for operating the IT components. Outside the data center, a chiller provides cooling water to the CRAH and dissipates the exhausted heat from the data center to the atmosphere by a dry-cooler (Fig. 8 shows details). Other electric components such as uninterruptible power supply (UPS), back-up generator and transformer are excluded from the present model.

The following describes the cooling model at a single time stamp. In this model, Q is referred to as the heat dissipation and P the power consumption. The variables that are varying with time are indicated by the time index t, otherwise they have constant values in the model. As an overview, the model has been constructed based on basic thermodynamic equations of conservation of mass and energy. The total power consumption in the data center (P_{DC}) is calculated from the knowledge of the IT load in the data center (P_{load_DC}), some boundary conditions such as the inlet air room operation temperature (T_{R_in}), the ambient temperature (T_{amb}), the relation between other loads and IT load (α), and the performance parameters of chiller, fan, cooling coil of CRAH and coil of dry-cooler.

Table 2
Power modeling variables.

Name	Name	Name
perf.cycles	perf.instructions	perf.cache_references
perf.cache_misses	perf.branch_instructions	perf.branch_misses
perf.bus_cycles	perf.idle_cycles_frontend	perf.idle_cycles_frontend
perf.cpu_clock	perf.task_clock	perf.page_faults
perf.context_switches	perf.cpu_migrations	perf.minor_faults
perf.major_faults	perf.alignment_faults	perf.emulation_faults
perf.L1d_loads	perf.L1d_load_misses	perf.L1d_stores
perf.L1d_store_misses	perf.L1d_prefetch_misses	perf.L1i_load_misses
perf.LLC_loads	perf.LLC_load_misses	perf.LLC_stores
perf.LLC_stores_misses	perf.L1d_prefetches	perf.LLC_prefetch_misses
perf.dTLB_loads	perf.dTLB_load_misses	perf.dTLB_stores
perf.dTLB_store_misses	perf.iTLB_loads	perf.iTLB_load_misses
perf.branch_loads	perf.branch_load_misses	perf.node_loads
perf.node_load_misses	perf.node_stores	perf.node_store_misses
perf.node_prefetches	perf.node_prefetch_misses	msr.cpu_freq
msr.cpu_time_in_c0	sys.cpu_use	sys.RRS
sys.received_bytes	sys.sent_bytes	

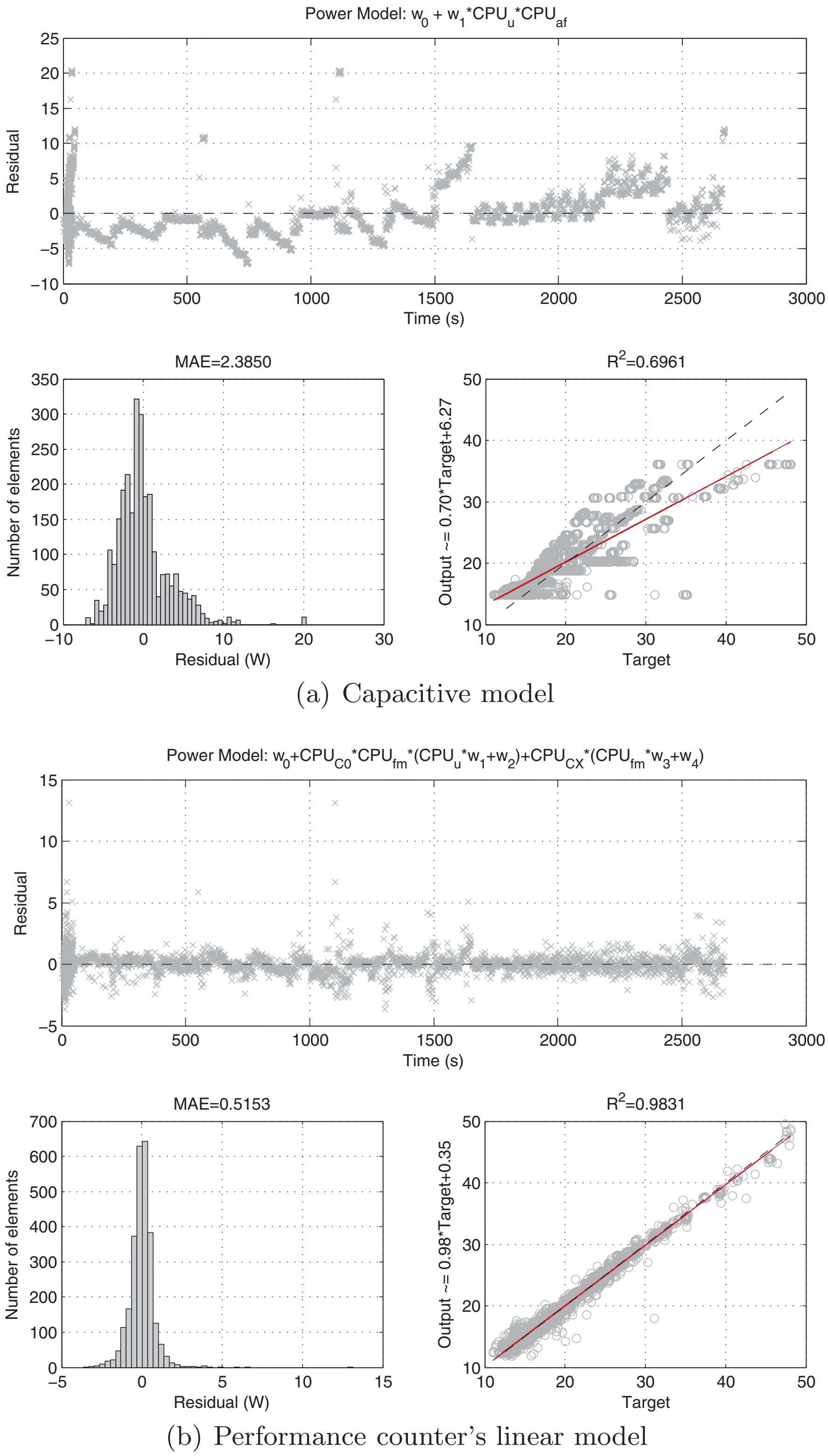


Fig. 7. Model comparison of the power models. The subfigures plot the residual of each data, the histogram of these residuals and the correlation between target data and model's output.

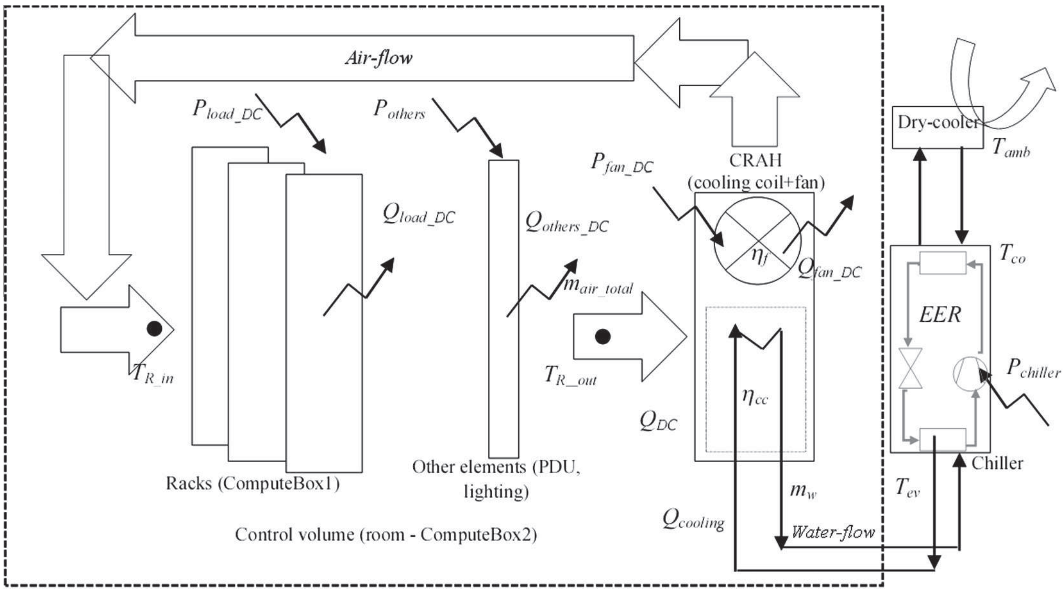


Fig. 8. Cooling model of a ComputeBox-based data center.

On the one hand, different thermal loads are defined in the data center, and the total thermal load Q_{DC} is the sum of them — the heat associated to IT load Q_{load_DC}, the heat from other loads such as PDU and lighting Q_{others_DC} and the heat from fans distributing air inside the data center room Q_{fan_DC} — that are defined as follows:

$$Q_{DC}(t) = Q_{load\_DC}(t) + Q_{others\_DC}(t) + Q_{fan\_DC}(t) \tag{2}$$

On the other hand, the total power consumption of the data center will be sum of powers from these components and the power consumed by the chiller:

$$P_{DC}(t) = P_{load\_DC}(t) + P_{chiller}(t) + P_{fans\_DC}(t) + P_{others}(t) \tag{3}$$

The following shows how to calculate each of these power consumptions.

The heat associated to the IT load Q_{load_DC} is assumed to be equal to the power of IT load P_{load_DC} [42], which is calculated as the sum of the IT load of each rack P_{load_rack}. As it is stated in [42], this assumption is possible since the power transmitted by the information technology equipment through the data lines can be neglected.

$$P_{load\_DC}(t) = Q_{load\_DC}(t) \tag{4}$$

$$P_{load\_DC}(t) = \sum_{j=1}^{N_r} P_{load\_rack}(t) \tag{5}$$

For the loads on lighting and PDU, a factor α is used to relate their power consumption with IT load. According to [43], α is estimated to be around 20%, for a typical data center with $2N$ power and $N+1$ cooling equipment, operating at approximately 30% of rated capacity. Therefore, this value for α can be considered as an example of current energy use in data centers. It is assumed that this power is also transformed into heat inside the data center:

$$P_{others\_DC}(t) = \alpha \cdot P_{load\_DC}(t) \tag{6}$$

$$Q_{others\_DC}(t) = P_{others\_DC}(t) \tag{7}$$

The power consumed by the fan P_{fans_DC} is related to the pressure rise over the fan Δp that is equal to the pressure drop provided by CFD calculations, the air density ρ, and the fan efficiency η_f as follows:

$$P_{fans\_DC}(t) = \frac{\Delta p(t) \cdot m_{air\_total}(t)}{\eta_f \cdot \rho} \tag{8}$$

The heat dissipated is the power consumed that is not transformed in pressure energy as stated below:

$$Q_{fan\_DC} = (1 - \eta_f) \cdot P_{fans\_DC}(t) \tag{9}$$

The total load of a data center $Q_{DC}(t)$ is determined by the air flow rate m_{air_total}, the specific heat C_p, and the difference between the inlet and outlet air temperatures, i.e., T_{R_in} and T_{R_out}, as shown in the following equation. Note that the air flow also affects the power consumed by the fans P_{fan_DC}, and consequently the heat generated by them inside the room Q_{fan_DC}.

$$Q_{DC}(t) = m_{air\_total}(t) \cdot C_p \cdot (T_{R\_out}(t) - T_{R\_in}) \tag{10}$$

The cooling demand faced by the chiller $Q_{cooling}$ includes the thermal load in the data center and the inefficiency η_{cc} in the coil of the CRAH:

$$Q_{cooling}(t) = \frac{Q_{DC}(t)}{\eta_{cc}} \tag{11}$$

To get the power consumption of the chiller, we have to consider a generic performance profile that is function of the condenser temperature (T_{co}), the evaporator temperature (T_{ev}) and the partial load (PLR). This performance is usually based in certified catalog data from manufacturer. The following shows directly the relation between the cooling load and the power consumed in the chiller $P_{chiller}$ by means of energy efficiency ratio (EER):

$$P_{chiller}(t) = \frac{Q_{cooling}(t)}{\mathrm{EER}(t)} \tag{12}$$

The partial load ratio (PLR) specifies the relation between the cooling demand in a certain condition and the cooling load in nominal conditions ($Q_{cooling\_nom}$), which corresponds to the operation of the chiller at the chilled water temperature T_{ev} and condenser water temperature T_{co}. In addition, PLR is also related to the cooling capacity rated $Q_{cooling\_rated}$, which corresponds to load of the chiller in standard condition (full load; temperature of chilled water leaving the chiller at 7 °C and temperature of condenser water entering the chiller at 30 °C), in which EER is named EER_{rated}. The following shows the relations:

$$\text{PLR}(t) = \frac{Q_{cooling}(t)}{Q_{cooling\_nom}} \tag{13}$$

$$Q_{cooling\_nom} = Q_{cooling\_rated} \cdot COOL(T_{ev}, T_{co}) \tag{14}$$

$$CoolPR(t) = CoolPR(T_{ev}, T_{co}, \text{PLR}(t), \text{EER}_{rated}) = \frac{1}{\text{EER}(t)} \tag{15}$$

To calculate the chilled water temperature T_{ev}, it is necessary to know the room inlet air temperature T_{R_in} and the minimum temperature difference ΔT_{h-ex} on the coil of CRAH between the output of air and the inlet of water:

$$T_{ev} = T_{R\_in} - \Delta T_{h-ex} \tag{16}$$

Common values of ΔT_{h-ex} on the commercial coils are between 5 °C and 15 °C. Since the chiller performance is also affected by T_{R_in}, higher operation temperature in the data center room will need less power consumption from the chiller, and hence increase the cooling efficiency.

Figs. 9 and 10 show the performance of cooling models. It is presented the relation of power consumption in cooling devices (chiller, fans) with ambient temperature, inlet air room operation temperature and partial load. Also PUE3 (see Section 5) is shown.

5. Energy efficiency metrics

CoolEmAll uses a set of metrics at different levels of analysis defined in Section 4.1. Different metrics have been selected depending on the level at which the experiments are conducted and the purpose of assessment. The following classify the metrics considered:

- *Resource usage metrics* refer to the utilization of a certain resource (CPU, memory, bandwidth, storage capacity, etc.), concerning a component (node) or a set of components (node-group, rack).
- *Heat-aware metrics* take temperature as the main indicator for the behavior of a data center.
- *Energy-based metrics* are defined as the consumption of power along a period of time.
- *Impact metrics* that are used to assess the performance of data center in environmental and economic terms.

The complete set of metrics defined in CoolEmAll was described in the public report of the project [44] as well as in some articles [45,46]. To assess the impact of different strategies used in the simulations conducted in this paper, the following ones are selected: Total energy consumed, Power Usage Effectiveness (PUE), productivity, energy wasted ratio, carbon emissions, electricity costs. The following defines these metrics.

Total energy consumption (in W h): This corresponds to the total power consumed by the data center over a certain period of time.

$$E_{DC} = \int_{t_1}^{t_2} P_{DC}(t)dt \tag{17}$$

Productivity: This metric indicates the relation between the useful work (W_{DC}) in the data center and the energy required to obtain this useful work during a certain period of time. Useful work [47] identifies the measurable work done by a data center while providing a given service. Useful work is defined on the application level and depending on the application purpose it might be expressed by the number of floating-point operations, number of service invocations, number of transactions, etc.

$$Productivity = W_{DC}/E_{DC} \tag{18}$$

Power Usage Effectiveness (PUE): As defined by The Green Grid [48], this metric (defined as PUE_3) is the ratio of the total power consumption in the data center and

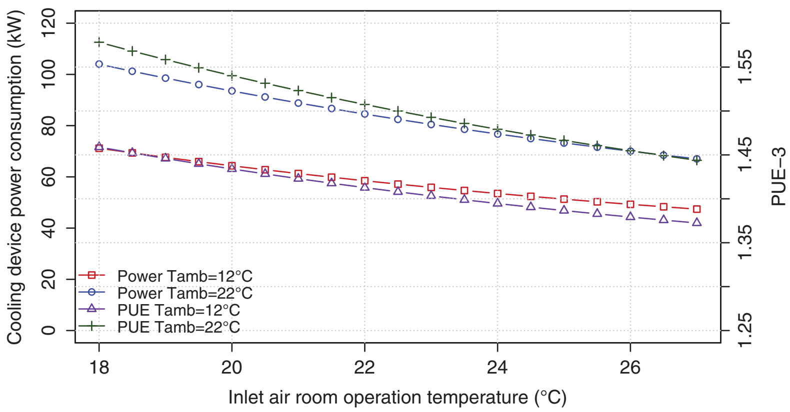


Fig. 9. Cooling devices (chiller and fans) power consumption and PUE3 dependence of ambient temperature and inlet air room operation temperature (maximum IT load 274 kW; rated chiller cooling capacity (30 °C outside air) 250 kW).

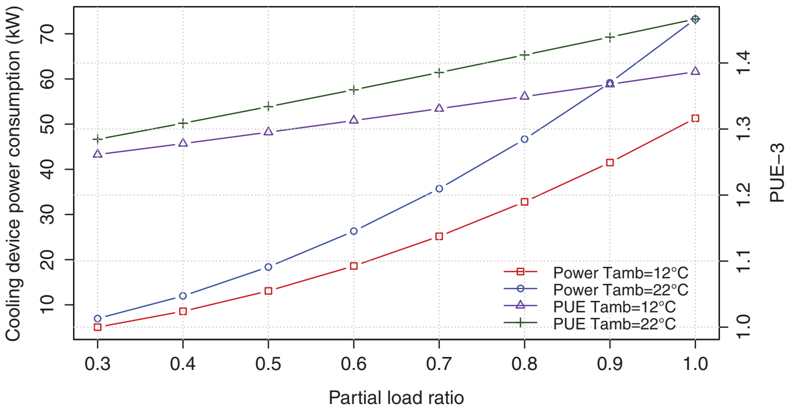


Fig. 10. Cooling devices (chiller and fans) power consumption and PUE3 dependence of ambient temperature and partial load (maximum IT load 274 kW; rated chiller cooling capacity (30 °C outside air) 250 kW).

the power used by the IT equipment. It can be defined at an instantaneous point in time or at the aggregated level over a period of time (in terms of energy).

$$PUE_3 = E_{DC}/E_{IT} \tag{19}$$

In the framework of CoolEmAll and to assess the impact of load management with fans that will stop when they are not used, another level of PUE (referred to as PUE_4) is defined, where the consumption of fans in racks is excluded from the IT load. For practical monitoring of this metric, it should be necessary to have separated power meters for fans or a signal to detect its operation mode with an assumption of the fan power consumption. The formula to calculate this metric is expressed as follows:

$$PUE_4 = E_{DC}/(E_{IT} - E_{fans} - E_{PSU}) \tag{20}$$

Energy Wasted Ratio (EWR): This metric assesses how much energy is wasted and is not used for producing useful work.

$$EWR = E_{DC\_not\_useful\_work}/E_{DC} \tag{21}$$

Carbon emission (in kg CO_2): This metric converts the total power consumed to CO_2 emissions using carbon emissions factor (CEF), which depends on the country since it is a function of the participation of the different energy sources and technologies (carbon, nuclear, natural gas, wind, hydro, solar, biomass, etc.) in the total electricity generation and the efficiency of conversion.

$$\text{Carbon emission} = E_{DC} \cdot \text{CEF} \tag{22}$$

Electricity cost (in €): This metric is calculated by multiplying the total energy consumed by the price of electricity.

$$\text{Electricity cost} = E_{DC} \cdot \text{Electricity price} \tag{23}$$

6. Resource management and scheduling policies

In the scope of resource management and scheduling policies, we can usually distinguish three basic components they consist of. These components include scheduling, resource allocation, and resource management. Scheduling is responsible for defining the order of execution for the ready tasks. Resource allocation selects the specific resource(s) for each job to be executed. Finally, resource management means the configuration of the resource states, usually related to their energy efficiency. Quite commonly, these components form the separated phases of various policies as presented in Fig. 11. In the following subsections, we will present strategies classified with respect to the convention described above.

6.1. Scheduling algorithms

A scheduling algorithm specifies the order in which tasks are served during the scheduling process (alternatively – it defines the order in which tasks are placed in the queues). The following shows some widely used algorithms, which can be applied to the scheduling of tasks in data centers.

- First Come First Served (FCFS) – a basic scheduling policy in which tasks are served in the order of their arrival in the system. This strategy reduces the waiting time of tasks.
- Last Come First Served (LCFS) – a policy contrary to FCFS, in which the tasks that arrive at the system later are scheduled first.
- Largest Job First (LJF) – tasks are scheduled in order of decreasing size, wherein the size specifies the number of requested processors. The main aim of this strategy is to optimize the utilization of the system.
- Smallest Job First (SJF) – tasks are ordered according to the number of requested processors. This strategy increases the throughput of the system.

The aforementioned scheduling algorithms can be extended with one of the backfilling approaches [49,50], which exchange the positions of the jobs in the queue based on the availability of the resources and the priorities of the tasks.

- Conservative Backfill – allows a lower priority task to run only if it does not delay any of the higher priority waiting tasks.

Fig. 11. Components of workload and resource management policies.

- Aggressive Backfill – allows a lower priority task to run if it does not delay the highest priority task.
- Relaxed Backfill – allows a lower priority task to run if it does not delay the highest priority task in the manner that does not exceed a predefined factor.

6.2. Resource allocation strategies

Resource allocation strategies define the manner in which tasks are assigned to resources. Since tasks are submitted by different users over time, the decision of where to execute each arriving task are usually made in an online manner without knowledge of the future task arrivals. First, we describe three basic allocation strategies that are commonly used to balance the loads of different computing nodes in the system.

- Random – each task is assigned to a randomly chosen node.
- Round-Robin – the tasks are assigned to the nodes in a round robin manner.
- Load balancing – each task is assigned to a node in order to balance the overall load of the system.

While the previous strategies do not explicitly consider any objective related to the tasks, the following describes three greedy allocation strategies that are performance-, energy- and thermal-aware, respectively.

- Execution Time Optimization (ExecTimeOpt) – each task is assigned to a node that minimizes its execution/response time.
- Energy Usage Optimization (EnergyOpt) – each task is assigned to a node that minimizes the energy consumed by the task.
- Maximum Temperature Optimization (MaxTemp-Opt) – each task is assigned to a node that leads to the lowest maximum outlet temperature.

The aims of the above three strategies are to minimize the average task response time, the overall energy consumption, and the maximum outlet temperature. For the thermal-aware strategy, the maximum outlet temperature is used as an objective because it has been shown to directly impact the cooling cost of data centers in both homogeneous and heterogeneous environments [51,52].

Finally, tasks can also be assigned to resources in order to consolidate the workload in a predefined allocation manner. The following describes some consolidation strategies.

- High performance – tasks are assigned to nodes starting from high performance ones.
- Low power – tasks are assigned to nodes starting from low power ones.
- Location-aware – tasks are assigned to nodes with respect to their physical locations.

Depending on the implemented scheduling model (single or multi-level), the presented resource allocation strategies might have different impact on the final allocation of the resources. In case of scheduling at the RECS level, the above strategies are responsible for assigning tasks directly to the nodes with respect to their resource requirements. For scheduling at the room level, a scheduler has to first choose the rack where the task will be assigned, and then the RECS or nodes within which further allocation will be performed. In this case it is possible to mix two or more strategies by applying, for example, the location-aware strategy in order to select a rack, followed by the load balancing strategy to balance the load within the chosen rack, and finally the thermal-aware strategy to minimize the outlet temperature of the chosen RECS.

6.3. Resource management policies

Resource management policies specify a set of operations performed on the resources during the scheduling process. They usually require supports from the underlying hardware layer and their effectiveness is closely related to the managed IT equipment. The following describes two most popular policies.

- Switching nodes ON/OFF – a node is switched on or off, depending on if it is used or not.
- Dynamic Voltage and Frequency Scaling (DVFS) – the frequency of a processor is scaled up or down, depending on if the processor is used or not.

7. Simulations

This section presents simulations and the results obtained by using the Data Center Workload and Resource Management Simulator (DCworms) [53]. Different models of the data center components presented in Section 4 and various resource scheduling policies presented in Section 6 are evaluated.

7.1. Simulation setup

Resource description. In our experiments, five different types of processors are used and their technical specifications are presented in Table 3. All five types of processors were previously profiled in order to obtain their detailed power and performance characteristics. More information can be found in [54]. Moreover, to perform comprehensive

Table 3
Technical specifications of the simulated processors.

Processor	Max. frequency (GHz)	RAM memory (GB)	Number of cores
Intel Core i7-3615QE	2.3	16	4 (8 logical)
Intel Core i7-2715QE	2.1	16	4 (8 logical)
Intel Atom D510	1.66	4	2 (4 logical)
Intel Atom N2600	1.6	2	2 (4 logical)
AMD G-T40N	1	4	2 (2 logical)

studies, different processor configurations are simulated at three different levels, namely the RECS level, the rack level, and the room level. The cooling model adopts that described in Section 4.3.

Benchmarks and workloads. Several types of benchmarks can be used to demonstrate the gains of the proposed system. The three most classical kinds of benchmarks are:

- Micro benchmarks, testing only one particular sub-system like memory accesses.
- Single-host benchmarks, usually used to test a particular host.
- Classical distributed benchmarks from the HPC community like NPB (Nas Parallel Benchmarks).

Benchmarks of the first category were used during development and tests of the monitoring infrastructure and of the application profiling tools. Benchmarks presented in this article include: `fft`, `abinit`, `c-ray`, `lin_lgb`, `lin_3gb`, `lin_tiny`, `tar`, `openssl`. Specifically, `fft` is a tool to compute Fast Fourier Transforms. `abinit` is a scientific tool for electronic simulation at the atomic level. `c-ray` is a raytracing tool. `lin_lgb`, `lin_3gb` and `lin_tiny` are different instances of Linpack (classical high performance computing benchmark). `tar` is an archive manipulation tool. Finally, `openssl` is an open-source implementation of cryptographic protocols.

Power consumption model. To estimate power consumption of the given processor we followed the model proposed in Section 4.2.3 supported with the gathered application profiles. We replayed the tasks execution, adjusting the frequency level and assumed linear dependency between power processor power drawn and its utilization. Our previous studies [53] show that such an approach presents reliable accuracy, with respect to the data gathered on real hardware, and might be boldly used as an power consumption estimator.

7.2. Simulation results

7.2.1. Results for the RECS level

This subsection shows the results of the simulations performed at the RECS level. Specifically, experiments were conducted to evaluate a system with one single

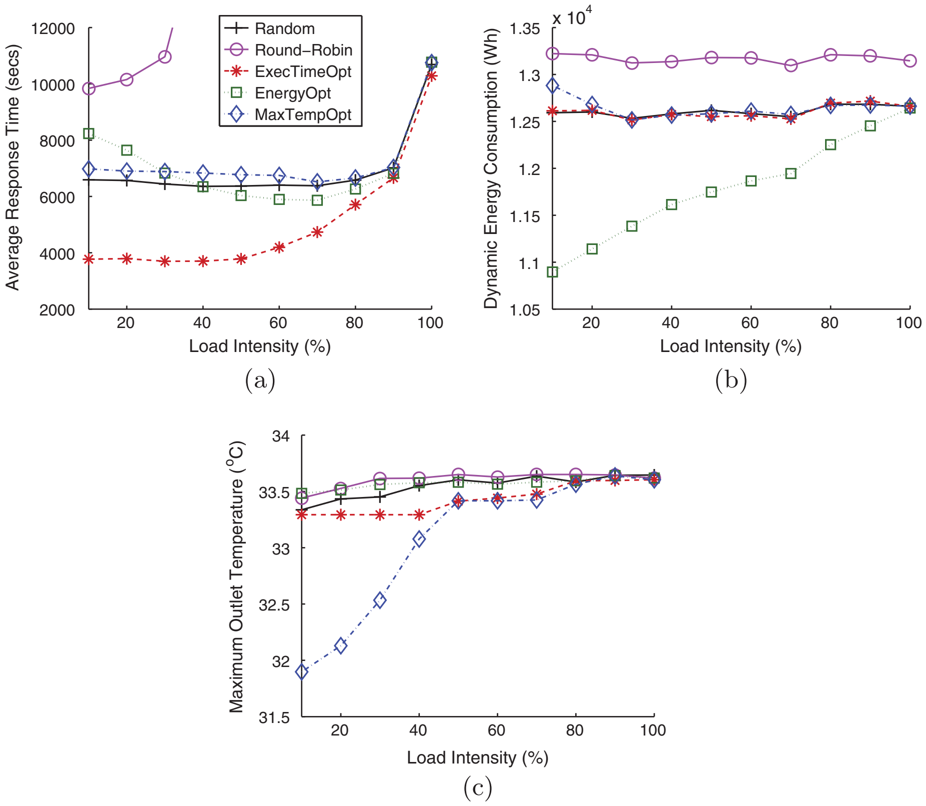


Fig. 12. Performance of five resource allocation strategies at the RECS level.

RECS2.0 unit consisting of 18 processors/nodes. The following describes the processor configuration used in the experiment:

- 8 Intel Core i7-2715QE nodes.
- 4 Intel Atom D510 nodes.
- 6 AMD Fusion G-T40N nodes.

The workload consists of 1000 tasks randomly drawn from the benchmarks described in Section 7.1. Tasks arrive according to the Poisson process. The load intensity used in the simulation is proportional to the average arrival rate λ (#jobs/h), and it is defined as $\lambda/10$. Five resource allocation strategies – Random, Round-Robin, ExecTimeOpt, EnergyOpt and MaxTempOpt – were evaluated with FCFS scheduling (which is also used in all subsequent experiments). Besides energy consumption, average response time of the jobs are used as a performance metric, and maximum outlet temperature is used as an indicator for the cooling cost. No resource management technique was applied, so all processors are switched on at all times.

Fig. 12 shows the simulation results. Note that only the dynamic energy consumption is shown in the figure, since the nodes are not switched off even when they are idle, so the static part will be identical for all strategies. The simulation results confirm our intuition that ExecTimeOpt provides better average response time, EnergyOpt provides less dynamic energy consumption, and MaxTempOpt provides the lowest maximum outlet temperature. The other two strategies (especially Round-Robin) perform badly for all three metrics, since they are oblivious to the platform and workload characteristics. Moreover, a tradeoff can be observed among the conflicting objectives of performance, energy and temperature (more details concerning such tradeoff can be found in [55]). In particular, the MaxTempOpt strategy reduces the maximum outlet temperature by about 1–1.5 °C under light system load. Although the difference in the outlet temperature is small, it can have a strong impact on the cost of cooling, especially when more RECS units are present in the system. The next two subsections study this more general case by applying the ON/OFF resource management policy to save more energy.

7.2.2. Results for the rack level

This subsection shows the results of the simulations performed at the rack level. Experiments were conducted to evaluate a rack consisting of three RECS2.0 units. The following shows the processor configurations used in the experiments:

- 18 Intel Core i7 nodes: 14 Intel Core i7-3615QE nodes and 4 Intel Core i7-2715QE nodes.
- 18 Intel Atom nodes: 14 Intel Atom N2600 nodes and 4 Intel Atom D510 nodes.
- 18 AMD Fusion G-T40N nodes.

The workload contains 600 openssl tasks with a fixed load intensity. Tasks arrive according to the Poisson process with a submission time range (difference between submission of last and first task) of 2760 s. Two types of consolidated resource allocation strategies – high performance and low power – are evaluated with the ON/OFF resource management policy. The load balancing strategy is used as a reference for comparison.

Table 4 present the results according to various energy-efficiency criteria, and Table 5 compare the impact of studied policies on the evaluation criteria. One can see the significant improvement in terms of useful work and productivity for the consolidation on high performance nodes approach. It is obscured, however, by the increase in the scope of energy usage. An improvement on this criterion can be achieved by benefiting from the possibility of switching off unused nodes. Consolidation on high performance resources with additional power management seems to be a good trade-off between energy usage and

Table 4
Energy-efficiency metrics for consolidation policies at a rack level.

Metrics	Policy consolidation				
	Load balancing	High performance		Low power	
			On/off		On/off
Total processors energy consumption (W h)	687	853	628	553	380
Total IT energy consumption (W h)	1671	1838	1613	1538	1365
Total node group fans energy consumption (W h)	412	412	145	411	246
Total rack energy consumption (W h)	2394	2586	2021	2240	1852
Total data center fans energy consumption (W h)	113	113	113	113	113
Total cooling device energy consumption (W h)	423	423	423	423	423
Total other devices energy consumption (W h)	48	52	40	45	37
Total energy consumption (W h)	2978	3174	2597	2821	2425
Mean rack power (W)	1884	2035	1591	1764	1458
Mean power (W)	2344	2498	2044	2221	1909
Max rack power (W)	1976	2115	1748	1866	1651
Max power (W)	2438	2579	2205	2325	2106
PUE	1.244	1.227	1.285	1.259	1.309
PUE Level 4	1.782	1.727	1.61	1.835	1.777
Energy waste rate (%)	19.03	13.31	1.22	11.78	0.59
Useful work (10^{-6} UW units)	556.70	1057.94	1057.94	244.37	244.37
Productivity (UW units/W h)	232,570	409,154	523,517	109,075	131,971

Table 5
Comparison of the obtained results with reference to the load balancing policy.

Metrics	Policy consolidation				
	Load balancing	High performance		Low power	
			On/off		On/off
Total processors energy consumption (%)	0	+24.28	−8.48	−19.38	−44.59
Total IT energy consumption (%)	0	+9.99	−3.47	−7.98	−18.33
Total node group fans energy consumption (%)	0	+0.02	−64.73	−0.02	−40.13
Total rack energy consumption (%)	0	+8.02	−15.58	−6.40	−22.64
Total data center fans energy consumption (%)	0	+0.02	+0.02	−0.02	−0.02
Total cooling device energy consumption (%)	0	+0.02	+0.02	−0.02	−0.02
Total other devices energy consumption (%)	0	+8.02	−15.58	−6.40	−22.64
Total energy consumption (%)	0	+6.58	−12.77	−5.25	−18.57
Mean rack power (%)	0	+8.00	−15.60	−6.38	−22.62
Mean power (%)	0	+6.56	−12.79	−5.23	−18.55
Max rack power (%)	0	+6.99	−11.57	−5.59	−16.48
Max power (%)	0	+5.78	−9.56	−4.62	−13.63
PUE (%)	0	−1.37	+3.30	+1.21	+5.23
PUE Level 4 (%)	0	−3.09	−9.65	+2.97	−0.28
Energy waste rate (%)	0	−30.04	−93.58	−38.09	−96.91
Useful work (%)	0	+90.04	+90.04	−56.10	−56.10
Productivity (%)	0	+75.93	+125.10	−53.10	−43.26

productivity. On the other hand, Consolidation on low power CPUs can be a good approach to decrease total power usage or to increase capacity. However, it should be noted that this leads to noticeable deterioration of the performance factors.

7.2.3. Results for the room level

This subsection shows the results of the simulations performed at the room level. Experiments were conducted to evaluate a server room populated with 10 racks. The following shows the configurations of the racks:

- 5 racks equipped with 10 4-unit chassis, each chassis provides a node group containing 4 Intel Core i7 nodes (Intel Core i7-3615QE).
- 5 racks equipped with 40 1-unit chassis, each provides a node group containing 1 AMD Fusion G-T40N node.

The following are the parameters used for the cooling devices:

- Computer Room Air-Handling Unit (CRAH): fan efficiency = 0.6, cooling coil efficiency = 0.95, deltaThEx = 10.
- Chiller: max cooling capacity = 10,000, cooling capacity rate = 40,000.
- Dry cooler: deltaThDryCooler = 10. Dry cooler efficiency = 0.02. Details related to the cooling parameters can be found in [54].

A workload containing 6000 openssl tasks is used to drive the simulation. Tasks arrive at the system according to the Poisson process with an average inter-arrival time of 1 s. The same set of resource allocation strategies as in the rack-level case are evaluated, again, according to the following criteria: PUE, PUE-Level 4, Productivity, Energy waste rate, max IT Power and Total energy used. Table 6 summarizes the results.

According to Table 6, the consolidation policy that favors high performance nodes (Intel i7 in this case) with additional node power management outperforms other strategies with respect to the evaluation criteria. Accumulating load on the most efficient (in terms of performance) nodes allows to improve both PUE-related metrics as well as the productivity factor. However, one should note the increase in maximum power consumption and total energy usage for the high performance consolidation, which should be carefully watched in terms of cooling devices capacity. Therefore, the high performance strategies are

Table 6
Assessment of resource allocation policies at room level.

Policy	Metrics					
	PUE	PUE Level-4	Productivity (rsa1024sign/W h)	Energy waste rate (%)	Max. IT Power (W)	Total energy (W h)
Load balancing	1.478	1.983	406,269	42.78	22,966	30,275
HighPerfConsolidation	1.478	1.968	449,726	25.9	23,424	31,649
HighPerfConsolidation + NodePowMan	1.383	1.786	534,816	5.639	22,027	24,909
LowPowerConsolidation	1.479	1.993	391,227	29.17	22,318	29,508
LowPowerConsolidation + NodePowMan	1.365	1.798	435,131	2.77	21,885	24,495

Table 7
Impact assessment of room-level resource management policies.

Policy	Total energy (W h)	Carbon emissions (kg CO_2)	Electricity cost (€)	Savings (%)
Load balancing	30,275	10.29	4.54	–
HighPerfConsolidation	31,649	10.76	4.75	−5
HighPerfConsolidation + NodePowMan	24,909	8.47	3.74	18
LowPowerConsolidation	29,508	10.03	4.43	3
LowPowerConsolidation + NodePowMan	24,495	8.33	3.67	19

good for minimization of the total energy consumption and maximization of productivity (useful work per Joule). On the other hand, low power consolidation strategies are better in cases when power usage should be constrained. In this case power capping can be applied to save additional power. As the cooling capacity in a data center is based on the maximum power consumption of the IT infrastructure, power capping can leverage facility's total cooling capacity (more details can be found in [56]). Besides the evaluation of the specific policies, the presented experiments demonstrate the usefulness as well as the drawbacks of the presented energy-efficiency metrics. For example, values of PUE are reasonably good, but the IT part of PUE includes constant speed fans. These fans are source of inefficiency, that is, their work when nodes are idle is a waste of energy. Thus, PUE Level-4 expresses the actual efficiency in a more adequate way. Applying node-level power management policies affects a ratio between components effectively taking part in computing to other overheads (such as useless fans work, power supply lost) and thereby improves PUE Level-4 to a greater extent than PUE. In this experiment, PUE decreases when node power management techniques are used, because the applied model assumes significant correlation between heat load and power usage of the cooling devices. In other settings, with bigger server room and cooling systems, PUE could even raise after improving efficiency of the IT part. Additional insight is provided by the proposed Energy Waste Rate (EWR), which estimates factor of energy that is wasted in the studied period. It can be easily seen that introducing power management techniques improves its value significantly.

7.3. Impact assessment

The experiments presented in the previous subsection point out the potential of obtaining relevant energy savings when resource management and scheduling policies are applied, especially at the room level. These energy savings can be converted to carbon emissions and electricity cost, as it is shown in Table 7. The carbon emission factor (CEF) used to calculate the carbon emissions is 0.34 kg CO_2/kW h and the electricity price used to calculate the operation costs is 0.15 €/kW h according to [57].

As we can see, the consolidation policy that favors high performance nodes increases energy consumption by 5%. However, the amount of savings reaches 18% when node power management is applied. The strategy of low performance nodes consolidation provides savings of 3% that increases until 19% when node power management is included. Furthermore, when extending these strategies to large-scale data centers with size bigger than this model and where the operation runs for 8760 h per year, the amount of total carbon emissions and operation cost reduced would be substantially worthy.

8. Conclusion

In this paper we have presented the approaches and evaluation results of the CoolEmAll project with the aim of making data centers more energy and resource efficient. We have presented workload profiles, application-workloads, power and cooling models used in the approach. In addition, different energy efficiency metrics at different levels of analysis were proposed. Various resource management and scheduling policies, including performance, energy, thermal-aware policies and consolidation policies were presented. Simulations were conducted by using the Data Center Workload and Resource Management Simulator (DCworms) for different levels in a data center, i.e., RECS level, rack level and room level. The experiments validate the specific resource management policies proposed and the energy-efficiency metrics. In future works, CFD simulations will be conducted to validate the simulation results. We will also study resource management allocation considering heat recirculation in the data center and other resource management policies such as DVFS.

Acknowledgement

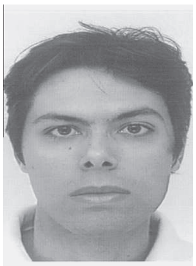
This research is funded by the European Commission under contract 288701 through the project CoolEmAll.

References

[1] J.G. Koomey, Worldwide electricity used in data centers, Environ. Res. Lett. 3 (3) (2008) 034008, http://dx.doi.org/10.1088/1748-9326/3/3/034008.
[2] EU Joint Research Centre, Harmonised Global Metrics to Measure Data Centre Energy Efficiency, November 2012. <https://ec.europa.eu/jrc/en/news/harmonised-global-metrics-measure-data-centre-energy-efficiency-7014>.
[3] M. Stansberry, J. Kudritzki, Uptime Institute 2012 Data Center Industry Survey, Tech. Rep., 2012.
[4] C. Belady, A. Rawson, J. Pfleuger, T. Cader, Green Grid Data Center Power Efficiency Metrics: PUE and DCiE, Technical Report, The Green Grid, 2008.
[5] D. Borgetto, H. Casanova, G. Da Costa, J.-M. Pierson, Energy-Efficient Job Placement on Clusters, Grids, and Clouds, John Wiley & Sons Inc., 2012.
[6] A. Kipp, L. Schubert, J. Liu, T. Jiang, W. Christmann, M. vor dem Berge, Energy consumption optimisation in HPC service centres, in: Proc. of the Second International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering, Civil-Comp Press, Stirlingshire, UK, 2011, pp. 1–16, http://dx.doi.org/10.4203/ccp.95.33.

[7] A. Berl, E. Gelenbe, M. Di Girolamo, G. Giuliani, H. De Meer, M.Q. Dang, K. Pentikousis, Energy-efficient cloud computing, Comput. J. 53 (7) (2010) 1045–1051, http://dx.doi.org/10.1093/comjnl/bxp080.
[8] D. Careglio, G. Da Costa, R. Kat, A. Mendelson, J.-M. Pierson, Y. Sazeides, Hardware Leverages for Energy Reduction in Large Scale Distributed Systems, Technical Report IRIT/RT-2010-2-FR, IRIT, University Paul Sabatier, Toulouse, France, May 2010.
[9] A. Ramirez, European scalable and power efficient HPC platform based on low-power embedded technology, in: International Supercomputing Conference, Leipzig, Germany, 2013.
[10] EcoCooling Inc. <http://www.ecocooling.org>.
[11] Colt Technology Services Group Ltd., Modular Data Centre. <http://www.colt.net/uk/en/products-services/data-centre-services/modular-data-centre-en.htm>.
[12] CoolEmAll Project Web-site. <http://www.coolemall.eu>.
[13] Future Facilities Ltd. <http://www.futurefacilities.com>.
[14] CA Technologies. <http://www.ca.com>.
[15] Innovative Research Inc. <TileFlow software, <http://inres.com/products/tileflow>.
[16] Romonet. <http://www.romonet.com/overview>.
[17] Schneider Electric, Data Center Genome Project. <http://datacentergenome.com>.
[18] Facebook, The Open Compute Project Foundation. <http://www.opencompute.org>.
[19] R. Basmadjian, N. Ali, F. Niedermeier, H. de Meer, G. Giuliani, A methodology to predict the power consumption of servers in data centres, in: Proceedings of the 2Nd International Conference on Energy-Efficient Computing and Networking, e-Energy '11, ACM, New York, NY, USA, 2011, pp. 1–10, http://dx.doi.org/10.1145/2318716.2318718.
[20] O. Mammela, M. Majanen, R. Basmadjian, H.D. Meer, A. Giesler, W. Homberg, Energy-aware job scheduler for high-performance computing, Comput. Sci. – Res. Dev. 27 (4) (2012) 265–275, http://dx.doi.org/10.1007/s00450-011-0189-6.
[21] G. Da Costa, H. Hlavacs, K. Hummel, J.-M. Pierson, Modeling the energy consumption of distributed applications, in: I. Ahmad, S. Ranka (Eds.), Handbook of Energy-Aware and Green Computing, Chapman & Hall, CRC Press, 2012. <http://www.crcpress.com/product/isbn/9781466501164>, <http://eprints.cs.univie.ac.at/3541/> (Chapter 29).
[22] M. Witkowski, A. Oleksiak, T. Piontek, J. Weglarz, Practical power consumption estimation for real life HPC applications, Future Gener. Comput. Syst. 29 (1) (2013) 208–217, http://dx.doi.org/10.1016/j.future.2012.06.003 (including Special section: AIRCC-NetCoM 2009 and Special section: Clouds and Service-Oriented Architectures).
[23] T. Mukherjee, A. Banerjee, G. Varsamopoulos, S.K.S. Gupta, Model-driven coordinated management of data centers, Comput. Networks 54 (16) (2010) 2869–2886, http://dx.doi.org/10.1016/j.comnet.2010.08.011.
[24] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, X. Zhu, No "power" struggles: coordinated multi-level power management for the data center, SIGARCH Comput. Archit. News 36 (1) (2008) 48–59, http://dx.doi.org/10.1145/1353534.1346289.
[25] A. Shah, N. Krishnan, Optimization of global data center thermal management workload for minimal environmental and economic burden, IEEE Trans. Compon. Packag. Technol. 31 (1) (2008) 39–45, http://dx.doi.org/10.1109/TCAPT.2007.906721.
[26] S. Yeo, H.-H. Lee, SimWare: a holistic warehouse-scale computer simulator, Computer 45 (9) (2012) 48–55, http://dx.doi.org/10.1109/MC.2012.251.
[27] E. Volk, D. Rathgeb, A. Oleksiak, CoolEmAll – optimising cooling efficiency in data centres, Comput. Sci. – Res. Dev. 29 (3–4) (2014) 253–261, http://dx.doi.org/10.1007/s00450-013-0246-4.
[28] Christmann, Description for Resource Efficient Computing System (RECS), 2009. <http://shared.christmann.info/download/project-recs.pdf>.
[29] Christmann. <http://www.christmann.info>.
[30] M. vor dem Berge, G. Da Costa, M. Jarus, A. Oleksiak, W. Piatek, E. Volk, Modeling Data Center Building Blocks for Energy-efficiency and Thermal Solutions, Springer, 2013.
[31] D.G. Feitelson, Workload Modeling for Computer Systems Performance Evaluation, version 1.0.2, September 2014. <http://www.cs.huji.ac.il/feit/wlmod/>.
[32] G.L.T. Chetsa, L. Lefevre, J.-M. Pierson, P. Stolf, G. Da Costa, DNA-inspired scheme for building the energy profile of HPC systems, in: Proceedings of the First International Conference on Energy Efficient Data Centers, E2DC'12, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 141–152, http://dx.doi.org/10.1007/978-3-642-33645-4_13.
[33] S. Chapin, W. Cirne, D. Feitelson, J. Jones, S. Leutenegger, U. Schwiegelshohn, W. Smith, D. Talby, Benchmarks and standards for the evaluation of parallel job schedulers, in: D. Feitelson, L. Rudolph (Eds.), Job Scheduling Strategies for Parallel Processing, Lecture Notes in Computer Science, vol. 1659, Springer, Berlin Heidelberg, 1999, pp. 67–90, http://dx.doi.org/10.1007/3-540-47954-6_4 (up-to-dated version of the standard workload format (SWF) definition available in http://www.cs.huji.ac.il/labs/parallel/workload/swf.html).
[34] V.M. Lo, J. Mache, K.J. Windisch, A comparative study of real workload traces and synthetic workload models for parallel job scheduling, in: Proceedings of the Workshop on Job Scheduling Strategies for Parallel Processing, IPPS/SPDP '98, Springer-Verlag, London, UK, UK, 1998, pp. 25–46.
[35] D. Feitelson, Packing schemes for gang scheduling, in: D. Feitelson, L. Rudolph (Eds.), Job Scheduling Strategies for Parallel Processing, Lecture Notes in Computer Science, vol. 1162, Springer, Berlin Heidelberg, 1996, pp. 89–110, http://dx.doi.org/10.1007/BFb0022289.
[36] A.B. Downey, A parallel workload model and its implications for processor allocation, Cluster Comput. 1 (1) (1998) 133–145, http://dx.doi.org/10.1023/A:1019077214124.
[37] B.P.K. Psounis, P. Molinero-Fernandez, F. Papadopoulos, Systems with multiple servers under heavy-tailed workloads, Perform. Eval. 62 (2005) 456–474.
[38] X. Fan, W.-D. Weber, L.A. Barroso, Power provisioning for a warehouse-sized computer, SIGARCH Comput. Archit. News 35 (2) (2007) 13–23.
[39] L.F. Cupertino, G. Da Costa, J.-M. Pierson, Towards a generic power estimator, Comput. Sci. – Res. Dev. (2014) 1–9, http://dx.doi.org/10.1007/s00450-014-0264-x.
[40] M. Jarus, A. Oleksiak, T. Piontek, J. Węglarz, Runtime power usage estimation of HPC servers for various classes of real-life applications, Future Gener. Comput. Syst. 36 (2014) 299–310, http://dx.doi.org/10.1016/j.future.2013.07.012.
[41] G. Da Costa, H. Hlavacs, Methodology of measurement for energy consumption of applications, in: Grid Computing (GRID), 2010 11th IEEE/ACM International Conference on, 2010, pp. 290–297, http://dx.doi.org/10.1109/GRID.2010.5697987.
[42] N. Rasmussen, Calculating Total Cooling Requirements for Data Centers, White Paper 25 rev 3, Schneider Electric's Data Center Science Center. <http://www.apcmedia.com/salestools/NRAN-5TE6HE/NRAN-5TE6HE_R3_EN.pdf>.
[43] Guidelines for Energy-efficient Datacenters, White Paper, The Green Grid, 2007.
[44] L. Sisó, R. Fornós, A. Napolitano, J. Salom, D5.1. White Paper on Energy and Heat-aware Metrics for Computing Modules, Tech. Rep., CoolEmAll, v1.4 7th FP GA no 288701, 2012.
[45] L. Sisó, J. Salom, M. Jarus, A. Oleksiak, T. Zilio, Energy and heat-aware metrics for data centers: metrics analysis in the framework of coolemall project, in: Cloud and Green Computing (CGC), 2013 Third International Conference on, 2013, pp. 428–434, http://dx.doi.org/10.1109/CGC.2013.74.
[46] E. Volk, A. Tenschert, M. Gienger, A. Oleksiak, L. Siso, J. Salom, Improving energy efficiency in data centers and federated cloud environments: comparison of CoolEmAll and Eco2Clouds approaches and metrics, in: Cloud and Green Computing (CGC), 2013 Third International Conference on, 2013, pp. 443–450, http://dx.doi.org/10.1109/CGC.2013.76.
[47] Harmonizing Global Metrics for Data Center Energy Efficiency, Tech. Rep., Global Taskforce Reaches Agreement Regarding Data Center Productivity, March 2014.
[48] V. Avelar, D. Azevedo, A. French, PUE™: A Comprehensive Examination of the Metric, White Paper 49, The Green Grid, 2012.
[49] S. Srinivasan, R. Kettimuthu, V. Subramani, P. Sadayappan, Characterization of backfilling strategies for parallel job scheduling, in: Parallel Processing Workshops, 2002. Proceedings. International Conference on, 2002, pp. 514–519, http://dx.doi.org/10.1109/ICPPW.2002.1039773.
[50] J. Ward, A. William, C. Mahood, J. West, Scheduling jobs on parallel systems using a relaxed backfill strategy, in: D. Feitelson, L. Rudolph, U. Schwiegelshohn (Eds.), Job Scheduling Strategies for Parallel Processing, Lecture Notes in Computer Science, vol. 2537, Springer, Berlin, Heidelberg, 2002, pp. 88–102, http://dx.doi.org/10.1007/3-540-36180-4_6.
[51] Q. Tang, S.K.S. Gupta, G. Varsamopoulos, Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: a cyber-physical approach, IEEE Trans.

Parallel Distrib. Syst. 19 (11) (2008) 1458–1472, http://dx.doi.org/10.1109/TPDS.2008.111.
[52] H. Sun, P. Stolf, J.-M. Pierson, G. Da Costa, Energy-efficient and thermal-aware resource management for heterogeneous datacenters, Sustain. Comput.: Inform. Syst. (2014), http://dx.doi.org/10.1016/j.suscom.2014.08.005.
[53] K. Kurowski, A. Oleksiak, W. Piatek, T. Piontek, A. Przybyszewski, J. Weglarz, DCworms – a tool for simulation of energy efficiency in distributed computing infrastructures, Simul. Model. Pract. Theory 39 (2013) 135–151, http://dx.doi.org/10.1016/j.simpat.2013.08.007.
[54] E. Volk, W. Piatek, M. Jarus, G. Da Costa, L. Siso, M. vor dem Berge, Update on Definition of the Hardware and Software Models, Deliverable D2.3.1, CoolEmAll, 2013.
[55] H. Sun, P. Stolf, J.-M. Pierson, G. Da Costa, Multi-objective scheduling for heterogeneous server systems with machine placement, in: Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on, 2014, pp. 334–343, http://dx.doi.org/10.1109/CCGrid.2014.53.
[56] A. Oleksiak, W. Piatek, T. Piontek, H. Sun, E.P.P. Montanera, Second Set of Resource Management and Scheduling Policies, Deliverable D4.6, CoolEmAll, 2013.
[57] R. Kemma, D. Park, Methodology Study Eco-design of Energy-using Products Meeup, Tech. Rep., Final Report. VHK. Delft, The Netherlands, 2005. <http://ec.europa.eu/enterprise/policies/sustainable-business/ecodesign/methodology/index_en.htm>.



Leandro Cupertino received the BSc degree in Computer Engineering in 2007 from the Rio de Janeiro State University (UERJ), Brazil. In 2009 he received MSc degree in Electrical Engineering from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). He joined the research group of Toulouse Institute of Computer Science Research (IRIT) in 2012, where he is working on the energy modeling of applications. There, he is pursuing a PhD degree in Computer Science at the University of Toulouse III, France. His research interests include energy aware computing, cloud and high performance computing, optimization techniques, computational intelligence and artificial neural networks.

Georges Da Costa is a permanent Assistant Professor in Computer Science at the University of Toulouse. He received his PhD from the LIG HPC research laboratory (Grenoble, France) in 2005. He is member of the IRIT Laboratory. His main interests are related to large-scale distributed systems, algorithmic, performance evaluation and energy-aware systems. He is Work Package leader of the European project CoolEmAll which aims at providing advanced simulation, visualization and decision support tools along with blueprints of computing building blocks for modular data center environment. He is working group chair of the European COST0804 Action on 'Energy efficiency in large scale distributed systems'. His research currently focus on energy aware distributed systems. He serves on several PCs in the Energy aware systems, Cluster, Grid, Cloud and Peer to Peer fields. His research highlights are grid cluster & cloud computing, hybrid computing (CPU/GPU), large scale energy aware distributed systems, performance evaluation and ambient systems.

Ariel Oleksiak works in the Applications Department at Poznan Supercomputing and Networking Center since 2002. He received his MSc degree in Computer Science, Intelligent Decision Support Systems, from the Pozna University of Technology, Poland in 2001, and PhD in Computer Science from the same institution in 2009. His research interests include mainly resource management, scheduling, simulations, and performance analysis in distributed computing systems including energy efficiency issues. He has actively contributed to many research projects such as GridLab, HPC-Europa, QosCosGrid, and BREIN. He was a member of the management committee of the European COST Action IC0804: Energy Efficiency in Large Scale Distributed Systems and is a member in COST Action IC1305 on Sustainable Ultrascale Computing where he is a chair of the energy-efficiency working group. He coordinates the EU-funded CoolEmAll project dealing with energy-efficiency of data centers.

Wojciech Piatek received his MSc degree in Computer Science from the Poznan University of Technology in 2010. His masters dissertation concerned the management of computational and network resources on a grid. Since 2008, he has been working in the Applications Department in the Poznan Supercomputing and Networking Center (PSNC) as a system analyst and developer. He participated in the PL-Grid project, and he is currently involved in the CoolEmAll project and a grant with Poznan University of Technology: Energy-efficient Models and Algorithms for Future Generation Computing and Networking Systems. He is also one of the developers of the Grid Scheduling Simulator (GSSIM) and Data Center Workload and Resource Management Simulator (DCworms). During the summer of 2010, he was a member of the HPC research group in the Center for Research Computing at the University of Notre Dame. His research activities concern energy efficient resource management in grids, clouds and HPC systems. Additionally, his research interests are focused on mathematics and algorithms.

Jean-Marc Pierson serves as a Full Professor in Computer Science at the University of Toulouse (France) since 2006. He received his PhD from the ENS-Lyon, France in 1996. He was an Associate Professor at the University Littoral Cote-d'Opale (1997–2001) in Calais, then at INSA-Lyon (2001–2006). Since 2013, he is Vice-Dean of the University in charge of IT. He is a member of the IRIT Laboratory and Chair of the SEPIA Team on distributed systems. His main interests are related to large-scale distributed systems. He serves on several PCs and editorial boards in the Cloud, Grid, Pervasive, and Energy-aware computing area. Since the last years, his researches focus on energy aware distributed systems (in particular for Cloud and HPC), monitoring, job placement and scheduling, virtual machines techniques, green networking, autonomic computing, mathematical modeling. He was chairing the EU funded COST IC804 Action on Energy Efficiency in Large Scale Distributed Systems (2009–2013) and currently participates in national and european projects on clouds or energy efficiency (ANR SOP, FP7 CoolEmAll).

Jaume Salom joined the Catalonia Institute for Energy Research (IREC www.irec.cat) in January 2010, where he currently leads the research group in Thermal Energy and Building Performance. Main activities and responsibilities of Dr. Jaume Salom is to lead the research activities and projects in the following fields: Energy efficiency and integration of renewable energy sources in buildings & communities, energy efficiency in thermal industrial processes and data centers and optimization and advanced simulation of energy systems and building physics. Several international and collaborative projects are in process.
Previously, in 1999, he has co-founded AIGUASOL, a cooperative firm, which has became an international reference in the fields of thermal energy efficiency, renewable energies, building physics and software development. Since 2006 until the end of 2009, he has been the General Manager of AIGUASOL and he has leaded several departments in the company over the last decade: the Building Physics group (1999–2005), the Software group (1999–2008) and Financial & Economical Department (since 2003). He holds a doctorate degree in Thermal Engineering from the Polytechnical University of Catalonia and he has research and professional experience in the fields of heat and mass transfer, fluid mechanics, numerical and dynamic simulation, energy efficiency, building physics, thermal comfort, software development and daylighting. Additionally, he is founder and minor partner of a new company which is devoted to develop a solar thermal concentrating collector with stationary reflector.
He collaborates in several national and international research and consultancy projects. He has leaded the promotion and consolidation of the use of advanced simulation tools in Spain and Portugal like TRNSYS (more than 100 users), EES or Meteonorm. He is one of the main developers of the TRANSOL tool (more than 400 users over the world) for the simulation and design of solar thermal systems. He teaches in master and postgraduate courses and also has interest on project management, leading & organizational aspects of work groups and business intelligence. He was born in 1969 in Mallorca.

Laura Siso is MSc in Energy Engineering. She has been working on issues related to energy efficiency and renewable energy for more than 18 years, through the realization of studies, projects and systems monitoring, in several reputed companies of Spain. She has technically delved into the technologies of sustainable building, air conditioning, solar thermal, dynamic simulation of thermal performance of systems, co-generation and district heating and cooling networks. Focus of the application of these technologies has been building, industry and data center sectors. Her experience has been developed on both engineering and research fields, as well as, a global vision about energy applications, economics and savings has been achieved during all of this time. Along her professional career, she developed skills of leading technical teams, managing long-term projects and participating in international research projects (European Commission) and networks (International Energy Agency). Nowadays she leads her own consultancy company MIRA ENERGIA (www.miraenergia.com) which offers consultancy services of energy efficiency in building sector and implementation of renewable energy technologies. Particularly, the global sustainability concepts achieve the practical application through the implementation of LEED Building Certification. In parallel, she is acting as senior researcher in the field of data centers, as in-house consultant at IREC (Catalonia Institute for Energy Research), to find out best-practices to reduce the carbon emissions and operation costs associated to the high energy consumption of those facilities.

Patricia Stolf is an assistant professor since 2005. She teaches at the Toulouse University (France). She obtained a PhD in 2004 in the LAAS-CNRS laboratory (Toulouse-France) on Tasks scheduling on clusters for remote services with quality of service. She now works in the IRIT laboratory in the SEPIA team and is currently working in the field of distributed algorithms and autonomic computing in large scale distributed systems like grid and clouds. She studies resources management, load-balancing, energy aware autonomic systems and energy and thermal-aware task scheduling. She is involved in different research projects: in the ACTION COST IC0804 Energy Efficiency in Large Scale Distributed Systems, in theEuropean CoolEmAll project and in the national ANR SOP project.

Hongyang Sun received his PhD in Computer Science from Nanyang Technological University in 2011. Prior to that, he completed his BEng in Computer Engineering from Nanyang Technological University, Singapore in 2005, and his MSc in Computer Science from National University of Singapore under Singapore-MIT Alliance in 2006. He is currently a postdoc researcher at the IRIT laboratory, France. His research interests include scheduling and resource management, parallel and high-performance computing, design and analysis of algorithms.

Thomas Zilio is a computer science engineer. He received his Master's degree in computer science at the University of Toulouse in 2012. He is a member of the IRIT Laboratory and the SEPIA Team and has been working on the CoolEmAll project which aims at providing advanced simulation, visualization and decision support tools for modular data centre environment since late 2012.