HPC Cooling: A Flexible Modeling Tool for Effective Design and Management

HPC Cooling: A Flexible Modeling Tool for Effective Design and Management

Christian Conficoni, Andrea Bartolini, Andrea Tilli, Carlo Cavazzoni, and Luca Benini

Abstract—Complex computing platforms such as High Performance Computers and Data Centers are critical systems from the energy sustainability viewpoint, due to their high computational power and demanding thermal stability specifications. In this context, cooling is a crucial component to operate such systems efficiently. Adavanced solutions, based on liquid and hybrid topologies are available today, but they come with a twofold challenge. On one hand, as widely recognized in the literature, the cooling devices need to be operated in a coordinated and energy-efficient fashion. In addition, after design and deployment, the cooling system has to be dynamically managed to efficiently adapt to workload, and environmental conditions. On the other hand, at design time, the cooling hardware architecture has to be selected in order to fit in the best way the needs of the computing facility, also depending on the environmental conditions characterizing its location.
This work presents a flexible, low-complexity modeling tool to describe the overall thermal behavior of complex computational platforms, as well as the effect of the diverse cooling components, and the corresponding energy consumption. Analytical modeling equations, stemming from physical first principles, are used, thus providing a compact and computationally manageable tool. This can be then exploited to explore the design space, choosing the correct cooling configuration, and/or define energy-optimal holistic cooling strategies, for complex, multidimensional, and hard constrained systems such as today SuperComputers. The proposed method is presented in general terms, then validated on a case study of a real-life HPC system with a hybrid cooling architecture.

Index Terms—Thermal Modeling, Supercomputers, Cooling, Energy efficiency

1 Introduction

Large scale complex computational platforms, such as High Performance Computers (HPC) provide key services for nowadays society, being widely exploited in many crucial areas (industry, scientific research, finance, and so forth). In this context, the quest for higher computational power (and/or computing power density) platforms is on-going, and it is expected to continue in the future, fostered by IT technology advances (increased workloads, ultra-dense many-core chips) allowing to squeeze a great amount of computational power into single servers/blades [1].
Such trend comes with critical issues in terms of energy and power consumption. Indeed, increasing computational power and its density levels imply a greater cost, both from the energy and economic viewpoint, in removing the heat resulting from computation. In nowadays complex computational platforms, the cooling infrastructure is responsible for about $30-40\%$ of the total system's power consumption [2]. Such ratio, obtained with traditional cooling methods, is economically and environmentally unsustainable for next generation complex computing systems, due to the excessive amount of power it would require to operate them profitably [3]. Therefore, advanced cooling strategies are needed to curtail the cooling costs.
For what concerns technological enhancements, traditional air conditioning methods, based on *Computer Room Air Conditioners* (CRAC), or *Computer Room Air Handlers* (CRAH) (depending if the evaporator is built into the AC unit or an intermediate coolant is used [4]), have been endowed with *free cooling/economizer* mode, i.e. the capability to exploit the outside air, using only the AC blowers to circulate it in the room [5]. In addition, air-based systems are frequently combined with liquid cooling in the so-called *hybrid cooling solutions*. Liquid can be either conducted to heat exchangers directly connected to the most thermally critical IT equipment (usually the microprocessors and accelerators) [6], or made to flow through liquid-to-air *Rear Door Heat Exchangers* (RDHX), mounted at the rack level [7]. Also the liquid circuit can have free cooling capabilities, in this case the liquid coolant is refrigerated by ambient air. In any case, the key point is to exploit the superior heat removal capacity of liquid coolant (commonly water) with respect to air [8].
All these technology advances are not for free and require significant investment cost when moving from one technology to another. Considering the short lifetime (about 3 years on average) of a HPC infrastructure, the economical return of such investments must be carefully evaluated before making a decision. However, the final benefit depends on the machine usage profile as well as on the environmental conditions of the system location. Taking Europe as an example, according to June 2016 TOP500 list (which ranks the 500 most powerful supercomputers worldwide accordingly to their computational performance) Europe contains one fifth of worldwide supercomputers spread across the European territory: from NTNU: Norwegian University of Science and Technology in Trondheim (Norway) to Barcelona Supercomputing Center BSC (Spain). Clearly, due to diverse climate conditions, the optimal cooling strategy depends on the location of the HPC facility. In addition, to fully take

C. Conficoni and A. Tilli are with are with the Center of Complex Automated Systems (CASY) at Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Italy. e-mail: andrea.tilli, christian.conficoni3}@unibo.it.
A. Bartolini and L. Benini are with Integrated Systems Laboratory, Dept. of Information technology and ELectrical Engineering, ETH, Zurich, Switzerland. e-mail:barandre,lbenini}@iis.ee.eth.ch
C. Cavazzoni is with Supercomputing Innovation and Application Dept., Cineca, Bologna, Italy.e-mail: c.cavazzoni@cineca.it

advantage of technological advances, sophisticated cooling management algorithms need to be designed. The goal is to reduce the cooling infrastructure power consumption, improving the system overall efficiency, while also meeting all the critical thermal constraints (e.g. max die, package temperature limits) and, ideally, without impairing performance (or with minimal quality of service degradation).

Thermal management of complex computational platforms has been subject to a large research effort (see [9], [10] and references therein for comprehensive overviews) spanning diverse strategies and scale levels including workloads allocation and scheduling where the goal is to dispatch the jobs in time and space (among the computing units) efficiently [11], [12], possibly including communication constraints and cost [13], as well as workload power consumption profiles [14],[15] and ambient temperature [16]. Dynamic Voltage and Frequency Scaling (DVFS) is another popular strategy for energy optimization and thermal control [17].

However, the most common approach is to focus on cooling components' management. In this context, several solutions have been recently presented in the literature: in [18] and [19] multi-objective optimization is applied to select a air-cooled data-center cooling mode and liquid flow at the blade level (for 3D MPSOCs), respectively, with the aim of minimizing the system overall power, under quality of service and thermal requirements. Optimization is also exploited in [20] and [21] to select the internal rack fans speed, and in [22] where both the facility cooling devices and the local blades fans are controlled. In [23], [24], the nodes workload and the CRACs reference temperatures are used as control inputs to minimize the computational and thermal power of a data center. In [25], [26] the effort is put onto selecting the CRAHs blowers speed, and active sub-floor tiles opening, for minimizing the CRAHs consumption, while in [6] a power optimization tool, coordinating a hybrid, liquid plus air, cooling system, is presented. In [27], [28] direct liquid-based techniques have been investigated.

Beside such methods are technically sound, in the authors' opinion, a low complexity, but accurate modeling tool, capable of describing the overall facility thermal behavior could further improve efficiency, helping to formulate and tune energy optimization strategies for each specific HPC system, in a comprehensive and holistic fashion. Typically, Computational Fluid Dynamics (CFD) methods are used [29], [6], to model some parts of the system. However, such tools are computationally heavy and time consuming to be tuned [30]. Indeed, thermal analysis of system subparts with CFD tools takes several minutes [31], while approaching the overall system requires hours [32], unless some approximations are introduced [33]. In addition, CFD lack the flexibility to quickly represent and analyze different configurations.

For this reasons, most of works focus on a specific cooling level and the respective devices optimization ([20], [22] [26] for air-cooling, and [27] for liquid), or if the entire system is considered, heuristic [28], experimental-based efficiency characterization [34], [35] is carried out, or some options such as free-cooling, are not fully represented in model [6].

Bearing in mind these considerations, in this work we propose a tool to obtain a compact, analytical representation of the overall infrastructure thermal behavior. Such modeling tool is based on physical first-principles and lumped parameter representation, thus provides a low complexity, easy to tune, and flexible way to describe complex and possibly heterogeneous cooling topologies. At the same time, the system dominant thermal dynamics are accurately captured in a numerically tractable manner, highlighting the role of the cooling system devices. With this result at hand, efficiency analysis, evaluation of different cooling architectures impact can be quickly performed, with no need of burdensome retuning. Indeed, with the proposed tool, just a few lumped-parameters need to be known to characterize the overall system thermal behavior. Computing such coefficients by means of physical/geometrical considerations, or estimating them from experimental data, takes tens of milliseconds on a standard commercial laptop, which is much less than the aforementioned time required for accurate CFD analysis. Therefore, such representation can be flexibly adjusted to represent different cooling topologies, helping to define the best set-up, in terms of efficiency and investment cost trade-off. Furthermore, the analytical model can be exploited to design energy-optimal cooling management algorithms, which, beside thermal constraints and different environmental conditions, account for the combined effects of all the system components. Thanks to the compact modeling method, reasonable computational burden can be expected to run such algorithms, making them suitable for real-time implementation in modern HPC centers as well as for assessing, at design stage, the effects of a given cooling strategy, and the corresponding investment.

The paper is structured as follows. In Section 2 the macro-components (both computational hardware and cooling) constituting a HPC room are functionally described, detailing their main features and role played in the aforementioned holistic description. This constitutes the base framework to present the coarse-grain, analytical thermal modeling tool in Section 3, which allows to derive a powerful and flexible method for representing different cooling configurations, to be exploited, as, mentioned, for analysis, simulation, and optimization of the overall HPC operation. Such properties are shown and detailed in Section 4 where we consider Galileo, a Tier-1 HPC system, hosted at CINECA center in Italy, as an example to apply the presented approach. We extract the component parameters and create a reference model, representing the current system. We validate the modeling accuracy against real system data, and then proceed to build a generalized model, capable of highlighting the relative impact of different variants of the cooling structure, with respect to the current one. Several workload and ambient temperature scenarios are tested, as well as the aforementioned various system configurations, to assess the proposed solution capability in providing useful insights about best cooling architecture and options, depending on environmental conditions (i.e. system location). Section 5 ends the paper with some final remarks and considerations.

2 HPC Thermal System

In this Section, we introduce the components affecting the thermal behavior of complex computational platforms as HPCs. A qualitative description will be performed first, setting the basis for the mathematical elaboration carried out

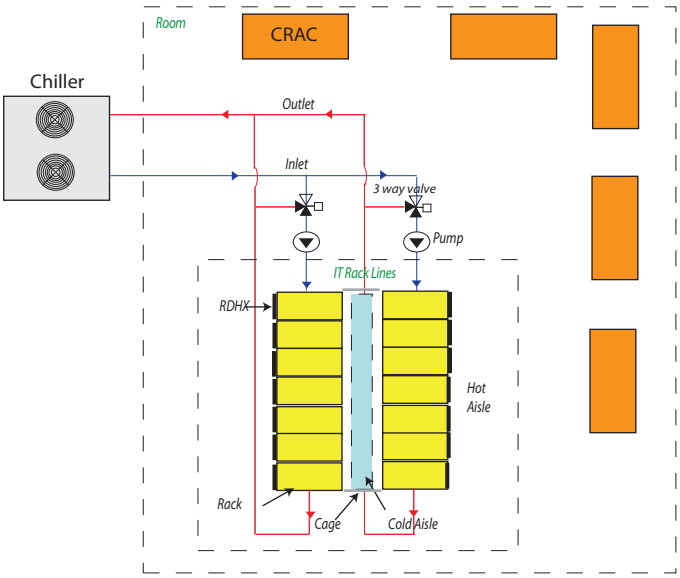

Fig. 1. Block scheme of an Hybrid Cooled Computational Platform.

in the next Section. Bearing in mind the goal of an holistic model, with limited complexity, a coarse grain resolution is adopted, looking at the cooling and computational devices as macro-blocks, and highlighting their role in the system thermal dynamics. A quite general floor-plan of an HPC room is portrayed in Fig. 1, hybrid cooling has been considered, for it will be the case for the benchmark studied in Section 4. If only direct liquid cooling is considered, the corresponding pipes will enter directly the racks, instead of the RDHXs, and the room air conditioner will not be present.

Starting from the computing nodes, these are typically deployed into blades or node servers, placed into chassis which are then hosted into racks. From the thermal viewpoint, these can be abstracted as heat generators, with heat energy as a function of the computational workload. Usually, local HW and SW power management, and thermal controllers are in charge of keeping the temperature distribution among the racks (or each rack line) as much uniform as possible. As mentioned in the Introduction, several methods, from internal fans control to DVFS and workload allocation, have been investigated for such purpose. Therefore, here we rely on such low-level strategies to describe the machine behavior at full-scale, assuming uniform local temperatures distribution, and focus on its interaction with the macro cooling components such as CRACs for air and chillers for liquid (usually water).

As concerns the room air-based conditioning, the racks are commonly deployed in two or multiple rows, in a standard cold-aisle hot-aisle arrangement; cold air delivered by CRACs' blowers is collected in a sub-floor plenum, then raises to the rack intake sides through perforated tiles, flowing into the machine (with the help of racks internal fans), then the hot air returns to the CRACs. To minimize cool and hot air mixing, a cage, providing a forced air path from the perforated tiles to the room roof and then back to the CRACs, can be mounted around the machine (see the gray bars in Fig. 1). CRACs performance and consumption is affected by the heat to be removed (roughly speaking the workload) and the environmental condition; for hot ambient temperature, the amount of work to keep the room cool would be clearly increased, while if free-cooling can be feasibly operated (i.e. complying with the system temperature constraints) the efficiency can be dramatically increased, switching off the thermal cycle and letting the blowers push the air in the subfloor.

Similar reasoning can be applied to chillers, required for liquid cooling. In this case, part of the return hot liquid coolant can be by-passed, and recirculated back into the inlet portion via three-way valves, thus letting the chiller to cool just part of the return flow. The chiller's pipe can also be split into multiple lines, each serving a portion of the IT equipment and with controlled flow rate through variable speed pumps (as depicted in Fig. 1), clearly this allows for more degrees of freedom, but implies larger cooling equipment costs. In case of direct free-cooling, such pipes will enter the system nodes, cooling them via dedicated cold plates. Clearly, a finer pipes splitting will take place inside the racks, however usually the flow rate among such micro-pipes cannot be actively controlled, thus, in terms of full scale modeling, they can be disregarded considering the flow rate of the pipes associated with active pumps as cooling control knobs.

If indirect liquid cooling is adopted, commonly associated with air in a hybrid topology, then the liquid will flow through rear-door heat exchangers, mounted on each rack (black bars in Fig. 1), with the purpose to help cooling the air heated up by the IT devices, before it flows back to the room conditioners.

In this work, the goal is to represent the behavior of all such systems and coupling them in a comprehensive fashion, to provide an effective, and relatively simple strategy to analyze thermal sustainability, feasibility, and power efficiency for modern HPC and Data Centers, with the goal to make suitable decisions at the design stage, thus achieving optimal energy performance at the operating stage. The proposed tool developed for such purposes is described in the next Section and it will be applied to the HPC selected as case study, in Section 4, both for optimal cooling management strategy, and, above all, for the overall efficiency analysis, according to different environmental situations (i.e. computational platform locations) and workload profiles.

3 Framework for Holistic HPC Thermal Modelling

Having defined the key macro-components for HPC/Data Center thermal regulation, in this Section we derive the corresponding models, which will allow to carry out cooling efficiency analysis, simulation, and optimization, considering the overall system in a holistic fashion. Given the system complexity, the modeling task is not trivial, and poses some challenges, in particular when the aim is to obtain accurate, but not too computationally heavy description. Indeed, CFD tools are typically adopted to accurately represent systems thermal behavior. However, as mentioned in the Introduction, despite the improvement in current CFD solvers performance, analyzing a HPC room at full-scale would be extremely time and computationally consuming, so typically just system sub-parts at small-scale level (e.g. server/blade) are described with these methods [6].

Even in this case, time-consuming set-up and tuning of

the sophisticated numerical algorithms involved in such approach can be required. In addition, CFD-based analysis/simulations are usually run for a limited set of nominal working points, which, in general, are not ensured to approximate all the possible system operating scenarios (e.g. different workload and environmental conditions) [30]. Finally, CFD models cannot be easily used to design optimal cooling strategies, due to the high number of variables involved, which would require heavy large-scale optimization techniques.

In order to avoid all these drawbacks, here we propose an analytical, lumped parameter thermal and energy characterization of the overall system, capturing the main thermal features of the elements described in the previous Section, as well as their power cost (for the active cooling components) and their respective influence, coupling and interaction. With such coarse grain modeling approach at hand, analysis, simulations, and integration with an optimization environment, can be performed with reasonable computational effort. Moreover, as it will be clarified in the following, the proposed modeling tool allows flexibility, that is, different cooling configurations can be easily represented, then tested and analyzed, by composing the macro-components equations, with just slight variations.

Before specifying the mathematical model for each component, it is worth to underscore the main idea behind the modeling approach, which stems from thermodynamics first principles [36] to characterize the components and their thermal interaction. In this respect, each HPC room component is described as a heat or cooling power source, and the thermal interaction among such elements is described by means of heat exchangers. This concept is sketched in the scheme shown in Fig. 2, which represents a hybrid cooled system, with rear-doors, similar to the case study which will be considered in Section 4. However, the same concepts and similar schemes can be drawn for direct liquid cooling, or standalone air cooling cases, as it will be shown in the remainder of the Section. Thermal capacitor and resistor symbols denote the heat exchange points, while heat and cooling power sources are denoted with the symbol P, entering with positive sign in case of heat sources, and minus for cooling devices (while the subscript defines the component, and thus its heating or cooling role, as it will be clarified later on).

The mathematical equations of such class of models will be derived assuming the ensuing hypothesis to hold true (the reader is referred to [37], [38] for detailed motivations and range of validity of the hypotheses)

- All coolant flows are turbulent;
- The air density and specific heat are constant in the range of the considered temperature values;
- No phase change takes place in liquid coolant during its cycle;
- The liquid coolant is incompressible, its density and specific heat are constant for the considered range of temperature values;
- The energy absorbed/rejected by heat exchange points depends on the inlet and outlet temperature average value.

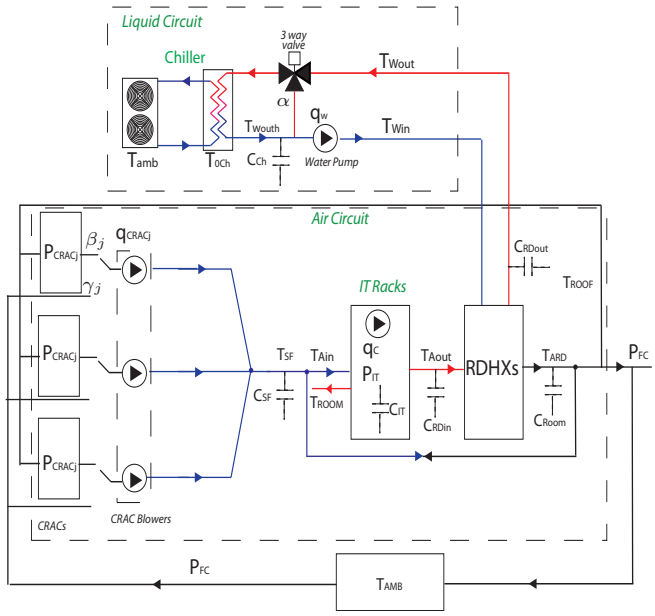

Fig. 2. Hybrid cooled HPC general scheme.

In addition, by virtue of what stated in Section 2 and similarly to what shown in related works (see for instance [34]), a uniform air temperature distribution is assumed for the racks, thus the computing system thermal behavior can be characterized by means of fewer state variables (temperatures), e.g. one for each line of racks.

With these clarifications at hand, we move to specify the mathematical description of each component described in Section 2.

3.1 IT Platform Thermal Model

From the thermal behavior viewpoint, the IT equipment of a complex computational platform can be seen as a heat source, or a set of heat sources, depending on the considered granularity, which is (are) directly related to the computational workload. Such heat is removed in part towards the room, via conduction, but mainly through forced convection of the coolant (air or liquid) flowing through the computing devices. In this context, the system thermal dynamics can be described by the following equations

$$\begin{aligned} C_{\mathrm{ITj}}\dot{T}_{\mathrm{ITj}} = P_{\mathrm{ITj}} - \left(\frac{T_{\mathrm{ITj}} - T_{\mathrm{ROOM}}}{R_{\mathrm{IT\text{-}Rj}}}\right) - \\ - \left(\frac{T_{\mathrm{ITj}}}{R_{\mathrm{IT\text{-}Cj}}} - \frac{T_{\mathrm{Coutj}} + T_{\mathrm{Cinj}}}{2R_{\mathrm{IT\text{-}Cj}}}\right) \\ \text{with } j = 1, \dots N,\ C = \{A, W\} \end{aligned} \tag{1}$$

where T_{ITj} are the aggregate temperatures representing the thermal status of the j^{th} computing system part, for instance one aggregate temperature for each the rack line which, in turn, can be obtained averaging the temperatures of all the nodes' cores belonging to the considered machine part, or, more conservatively, taking the maximum value. P_{ITj} denotes the corresponding computing part thermal power, produced by the computational workload, while capacitance C_{ITj} models the thermal inertia of the computing system parts, and R_{IT-Rj}, R_{IT-Cj} denote the thermal resistances modeling radiation towards the room, and forced convection area, respectively. Beside the specific system properties (flow rate, geometric topology), the latter

mainly depends on the coolant, which is commonly air but, in case of directly liquid cooled architectures, can also be water, thus the double subscript (A for air, W for water) in (1). Similarly, T_{Cinj}, T_{Coutj}, are the coolant inlet and outlet temperatures in the considered system subpart, while T_{ROOM} is the room temperature. It is further to remark that the resolution considered for partitioning the system can differ from the one of the cooling structure. For instance, in case the room air conditioning pushes the cool air in a common subfloor, or liquid pipes are shared among all the machine racks, the inlet temperatures T_{Cinj} will be the same for all j.

Thus, equation (1) can be easily and quickly adapted to different configurations, confirming the flexibility claimed introducing the proposed approach.

T_{Cinj} is affected by the CRACs or chiller operation, while temperatures T_{Coutj} will be affected by the flow rate of the coolant used to remove the j^{th} IT part q_{Cj}, according to

$$\begin{aligned} &C_{\text{outj}}\dot{T}_{\text{Coutj}} = \left(\tfrac{T_{\text{ITj}}}{R_{\text{IT-Cj}}} - \tfrac{T_{\text{Coutj}}+T_{\text{Cinj}}}{2R_{\text{IT-Cj}}}\right) - \\ &q_{Cj}c_{vC}\rho_C(T_{\text{Coutj}} - T_{\text{Cinj}}) - \underbrace{P_{\text{FC}} + \left(\tfrac{T_{\text{ITj}}-T_{\text{ROOM}}}{R_{\text{IT-Rj}}}\right)}_{optional} \\ &\text{with } P_{\text{FCj}} = q_{\text{FCj}}\rho_A c_{vA} T_{\text{ROOF}},\ j = 1,\dots N,\ C = \{A, W\} \end{aligned} \tag{2}$$

where ρ_C, c_{vC} are the coolant density and specific heat, respectively, C_{outj} are thermal capacitances modeling heat transfer from the IT components to the cooling mechanism whose value depend on the configuration. If standard air cooling is adopted then the room capacitance will be accounted for, in case of direct cooling it will be related to cold plates and liquid pipes, while if rear-doors are used it will correspond to the heat-exchangers thermal inertia.

The last two terms in the equation above can be needed, again depending on the cooling structure and options. P_{FCj} denotes the thermal powers removed by air free-cooling, accounting for the fact that part of the heat is let out of the room without having to be removed by the active cooling components. Such term is given by the flow rate portion of CRACs operating in free cooling q_{FCj}, which will exit from the room at temperature T_{ROOF}, i.e. that of the hot air raising towards the ceiling after reaching the hot aisle. Further details on such variables will be given later on. For now, we just remark that the air free cooling term has to be summed in the equation above only if air free cooling option is used, and no intermediate cooling stages (such as rear-doors) are present, thus T_{Coutj} are indeed affecting the air temperature to be cooled by CRACs (details will be provided in the following). Similarly, the power dissipated toward the room has to be considered only if no intermediate cooling components are available in between the computing system and the cooling devices (CRACs or chillers in case of liquid).

3.2 Rear Doors Thermal Model

As far as RDHXs are concerned, as the name says, we can regard them as liquid (usually water) to air heat exchange points. Applying similar reasoning to what in Subsection 3.1, the resolution at which the modeling is performed can be adjusted according to the cooling topology granularity. In general, we can express the RDHX blocks thermal behavior as

$$\begin{aligned} C_{\text{Roomj}}\dot{T}_{\text{ARDj}} &= q_{\text{Aj}}c_{vA}\rho_A(T_{\text{Aoutj}} - T_{\text{Ainj}}) - \\ &- \left(\tfrac{T_{\text{Aoutj}}+T_{\text{ARDj}}}{2R_{\text{RDHXj}}} - \tfrac{T_{\text{Winj}}+T_{\text{Woutj}}}{2R_{\text{RDHXj}}}\right) + \\ &+ P_{\text{FC}} + \left(\tfrac{T_{\text{ITj}}-T_{\text{ROOM}}}{R_{\text{IT-Rj}}}\right) \\ C_{\text{RDoutj}}\dot{T}_{\text{Woutj}} &= \left(\tfrac{T_{\text{Aout}}+T_{\text{ARDj}}}{2R_{\text{RDHXj}}} - \tfrac{T_{\text{Winj}}+T_{\text{Woutj}}}{2R_{\text{RDHX}}}\right) - \\ &- q_{Wj}c_{\text{vW}}\rho_W(T_{\text{Woutj}} - T_{\text{Winj}}),\ \ j = 1,\dots N. \end{aligned} \tag{3}$$

The first differential equation models the air side of the heat exchanger, with T_{ARDj} representing the aggregate temperature of the air exiting the racks, after the heat exchange with the j^{th} RDHX block (for instance the average temperature of a line of racks), C_{Roomj} the corresponding room thermal capacitance, and R_{RDHXj} the resistance of the heat exchangers. T_{Ainj}, T_{Aoutj} are the air temperatures at the input and output of the racks, c_{vA}, ρ_A the air specific heat and density, while q_{Aj} is the air flow rate of the racks internal fans. It is further to remark that, in case of hybrid cooling structure with rear-doors, the aforementioned variables will coincide with those denoted with the generic subscript C in eq. (2). Also, the possible free-cooling power and heat dissipated towards the room will be added in eq. (3) instead of eq. (2), as the air flows exiting towards the room will now be at temperatures T_{ARDj} instead of T_{Coutj}.

Finally T_{Winj}, T_{Woutj} are the RDHXs water side inlet and outlet temperatures, respectively, where inlet means the cold water coming from the chiller, and outlet indicates the warm one returning to the chiller after the heat exchange with the hot air. Such variables affect also the second differential equation, representing the liquid side of the heat exchangers, where C_{RDoutj} are the thermal capacitances at the water side of the RDHXs blocks, while q_{Wj} denote the water flow rate in the pipe lines feeding such blocks, and c_{vW}, ρ_W the water specific heat and density. Again, model resolution and cooling degrees of freedom can be set according to the available topology, e.g. in case of independently controlled liquid lines serving the rack lines (as depicted in Fig. 2), arbitrary (within the system limits) values for q_{Wj} could be set, otherwise the total liquid flow rate will be partitioned according to the pipes' hydraulic impedance.

3.3 Computer Room Air Conditioning Thermal Model

In case air cooling is adopted, clearly the CRACs can be seen as sources of cooling power (that is negative thermal power generators) affecting the temperature of the air entering the computing devices, and, in turn, affected by the return air temperature T_{ROOF}. Indeed, the cold air is typically pushed by the CRACs blowers in a sub-floor[1], and then raised through the cold aisle perforated tiles. Then, introducing a new temperature variable T_{SF} denoting the sub-floor air thermal status, we can express the CRACs effect as

$$\begin{aligned} C_{\text{SF}}\dot{T}_{\text{SF}} &= c_{vA}\rho_A q_{\text{CRAC}} T_{\text{ROOF}} - P_{\text{CRAC}} + q_{\text{FC}}c_{\text{vA}}\rho_A T_{\text{amb}} - \\ &- q_{\text{CRACtot}}c_{\text{vA}}\rho_A T_{\text{SF}},\ \text{with } q_{\text{CRACtot}} = q_{\text{CRAC}} + q_{\text{FC}} \end{aligned} \tag{4}$$

1. Actually, also the subfloor can be partitioned into multiple portions, serving different machines, or different parts of the same machines, since this is not the most common solution, here the case of a shared sub-floor is considered. However, multiple sub-floor divisions can be easily handled replicating eq. (4) for each part.

where C_{FC} represents the sub-floor capacitance, P_{CRAC} is the overall conditioners cooling power, and $q_{CRACtot}$ the total blowers flow rate, which is given by the sum of free cooling CRACs' blowers flow rate q_{FC}, and the flow of the CRACs operating in standard refrigerating mode q_{CRAC}. In turn, such variables are the result of the single CRACs contributions, which can be mathematically accounted as follows

$$\begin{aligned} P_{\text{CRAC}} &= \sum_{i=1}^{N_C} \beta_i P_{\text{CRACiM}},\ q_{\text{CRAC}} = \sum_{i=1}^{N_C} \beta_i q_{\text{CRACiM}} \\ q_{\text{CRACFC}} &= \sum_{k=1}^{N_{\text{FC}}} \gamma_k q_{\text{CRACkM}} \end{aligned} \tag{5}$$

with N_C the total number of CRACs, N_{FC} the number of CRACs endowed with free-cooling option, while P_{CRACiM}, q_{CRACiM}, q_{CRACjM} are the CRACs nominal cooling capacity and blower flow rates, respectively. β_i, γ_k denote the corresponding duty cycles modulating the CRAC cooling power (both ON/OFF hysteresis and PWM-like technique can be captured by such modeling). Finally, T_{amb} is the ambient temperature, and the corresponding term in eq. (4) accounts for the fact that, if air free cooling is available, then the same flow component that exited at T_{ROOF} is cooled by the ambient and return at T_{amb} to give a chilling contribution (see the arrow at the bottom of scheme in Fig. 2).

Beside the thermal effect on the system, another crucial information for what regards the air conditioners is the power consumption needed to produce such effect. CRACs are complicated thermal machines, and, in principle, sophisticated power consumption models should be used to characterize their power consumption. However, such models require accurate knowledge of several internal variables, which can be monitored internally to the CRAC if advanced local control strategies are implemented, but are usually not available at the considered scale level for the overall system analysis and control. Therefore, here the CRACs *Coefficient of Performance* (COP), defined as the ratio between the cooling load and its power consumption, is approximated by the refrigerating cycle *Carnot efficiency*, that is

$$COP_{\text{CRAC}} = \frac{T_{0\text{CRAC}}}{T_{\text{amb}} - T_{0\text{CRAC}}} \tag{6}$$

where T_{0CRAC} is the CRAC's evaporator side temperature. Even though Carnot coefficient does not account for non ideal thermodynamic cycles, at this modeling level, it can be used to represent the efficiency of cooling devices with reasonable accuracy [37]. In addition, also some power consumption has to be considered for spinning the CRACs blowers providing the required flow rate. In this respect, super-linear (usually quadratic) laws can be used to map the flow rate to the blowers consumption [39]. All that being given, the following power consumption model is introduced for what concern CRACs

$$\begin{aligned} P_{\text{consCRAC}} = &\sum_{i=1}^{N_C} COP_{\text{CRAC}}^{-1} \beta_i P_{\text{CRACiM}} + \sum_{i=1}^{N_C} k_{\text{BLi}} \beta_i q_{\text{CRACiM}}^2 + \\ &+ \sum_{k=1}^{N_{\text{FC}}} k_{\text{BLk}} \gamma_k q_{\text{CRACkM}}^2 \end{aligned} \tag{7}$$

where k_{BL} are components specific coefficients mapping the flow rate to the blower power.

3.4 Chiller Thermal Model

The chiller modeling follows similar steps to what presented about CRACs in the previous paragraph. In fact, the chiller will act on the liquid (water) inlet temperature, and will be affected by the heat produced by the machine through the return hot water temperature. Actually, the water temperature entering the system (RDHXs or cold plate depending on the kind of liquid cooling) could differ from the one exiting the chiller, due to the, quite standard, possibility to install by-pass valves (see the top of Fig. 2) recirculating part of the warm water back into the inlet pipes, without going through the chiller, as described in Section 2. For the sake of generality, in this work such option is considered. If it is not available, the model can be straightforwardly obtained by setting the chiller outlet temperature equal to the system inlet one.

Similarly to eq. (5), we define the temperature of water exiting the chiller T_{WoutCh}, then the corresponding dynamics reads as

$$\begin{aligned} C_{\text{Ch}} \dot{T}_{\text{Woutch}} = &\sum_{j=1}^{N} q_{\text{Wj}} c_{\text{vW}} \rho_W (T_{\text{Woutj}} - T_{\text{Winj}}) - \\ &- \left(\frac{\sum q_{\text{Wj}} T_{\text{Woutj}}}{\sum q_{\text{Wj}}} + T_{\text{Woutch}} \right) \\ &\qquad 2R_{Ch} - \frac{T_{0\text{Ch}}}{R_{\text{Ch}}} \end{aligned} \tag{8}$$

where C_{Ch} and R_{Ch} are the thermal capacitance and resistance of the chiller heat exchange point, respectively, while T_{0Ch} is the chiller evaporator side temperature. It is further to notice that effect of different inlet and outlet liquid temperatures in multiple pipe lines converging to a unique pipe (one for inlet and one for outlet) near the chiller is captured by considering the weighted sum, with weights given by the flow rates, of the different temperatures (see the second term in eq. (8)) similarly to what in eq. (4) for air flows at T_{amb} and T_{ROOF}. The thermal powers removed by the water lines are also summed up (first term on the right side of (8)) to compute the overall term to be rejected by the chiller. Clearly, such terms depends on the temperature gradient between the inlet and outlet liquid. The latter is provided by equation (3) if liquid is used fo feed rear-doors, or by expression (2) (with subscript W) if direct liquid cooling is used. The inlet temperature will be the same as the chiller output one if no by-pass valves are present, otherwise T_{Winj} can be expressed as

$$T_{\text{Winj}} = \alpha T_{\text{Woutch}} + (1 - \alpha_j) T_{\text{Woutj}},\ \alpha_j \in [0, 1] \tag{9}$$

where α_j denote the (normalized) position of the by-pass valve. The equation above stems from flow/energy balance considerations, making T_{Winj} the weighted sum of T_{Woutch} and T_{Woutj}, with weights being the flow rates of the corresponding liquid circuit portion.

Differently from the CRACs model, for chillers the variable T_{0Ch} can be assumed adjustable, assuming a reference value can be given to the chiller internal controller. Such value, will then determine the thermal power removed by the chiller from the returning water flow, and as in eq. (6)-(7), the chiller power consumption, approximated using the refrigerating cycle Carnot coefficient, that is

$$P_{\text{consCh}} = \sum_{j=1}^{N} COP_{Ch}^{-1} q_{Wj} c_{vW} \rho_W (T_{\text{Woutj}} - T_{\text{Winj}}) + \sum_{j=1}^{N} k_{Pj} q_{\text{Wj}}^2$$
$$COP_{ch} = \frac{T_{0Ch}}{T_{amb} - T_{0Ch}}. \tag{10}$$

Note that, since in this case T_{0Ch} can be considered as the control variable to be given to the chiller local controller as reference input, free cooling will be achieved setting $T_{0Ch} = T_{amb}$, thus making the chiller refrigerating work null. Also in this case, some cost has been associated to the pumps pushing the liquid through the pipes, using a quadratic law as for CRACs' blowers, just with a different flow-rate to power mapping coefficients k_{Pj}.

3.5 Coupling Variables and Constraints

Beside having defined the models of each crucial component on its own, trying to underline what are the interconnection with the other sub-systems, i.e. which variables enter as inputs from other parts, or affect other devices, still some "global" variables such as T_{ROOM}, T_{ROOF}, and possibly T_{Ainj} if air cooling is considered in eq. (1), need to be defined in terms of the others. In addition, in this paragraph, we summarize all the thermal and physical constraints the system is subject to, for these are crucial to implement cooling optimization strategies.

Starting with the variables coupling the subsystems described in the previous paragraphs, it is further to remark that the most involved situation concerns air cooling and indirect (via RDHXs) liquid cooling. Indeed, if direct liquid cooling is adopted, then conduction towards the room (second term in the right side of eq. (1)) can be neglected and variables T_{ROOF}, T_{ROOM} can be removed from the modeling, combining only equations (1), (2) (3), and (8) [40]. Instead, in case of air-cooling, possible recirculation between the hot and cool aisle can take place. This is mainly caused by the mismatch between the machine internal fans flow rates q_{Aj} (q_{Cj} in eq. (1) with subscript $C = A$), and those of the CRAC blowers, which are generally regulated independently one form the other; the former being typically set by the IT internal controllers and the latter by the room/building cooling management system. Let's start assuming, for simplicity, $j = 1$, i.e. the entire computing system is considered as a unique heat source with the cooling devices acting on it, applying the same considerations mentioned to obtain eq. (9), we can express the air flows inlet and outlet temperatures as follows

$$\begin{aligned} &T_{Ain} = T_{SF},\ T_{ROOF} = \tfrac{q_R T_{ARD}(\text{or} T_{Aout}) + (q_{CRACtot} - q_C) T_{Ain}}{q_{CRACtot}} \\ &\quad \text{if } q_{CRACtot} \geq q_A \\ &T_{Ain} = \frac{q_{CRACtot} T_{SF} + (q_R - q_{CRACtot}) T_{ROOF}}{q_R} \\ &T_{ROOF} = T_{ARD}(\text{or} T_{Aout}) \text{ if } q_{CRACtot} < q_A \end{aligned} \tag{11}$$

where the first line describes a surplus of airflow by the blowers w.r.t. the machine internal fans, and thus part of the cold air will go into the room, decreasing T_{ROOF}. On the other hand, if the blower flow rate is less than what required by the racks fans, then part of the room air will be absorbed by the fans and it will contribute to warm up the inlet air w.r.t. the sub-floor temperature, as modeled in the second line of eq. (11) (the blue and black arrows around the computing system block in Fig. (2) represent this two opposite recirculation conditions). As previously stated, sometimes cages are mounted to separate the hot and cold aisles and prevent such recirculation phenomena, in this case, clearly no mismatch in the flow rates can take place, therefore the simple condition $T_{\text{Ain}} = T_{\text{sf}}$, $T_{\text{ROOF}} = T_{\text{RD}}(T_{\text{Aout}})$ will hold. If multiple subparts are considered ($j > 1$) then binary condition (11) becomes more complicated, as different possibilities for each pair q_{Aj}, $q_{\text{CRACtotj}} = q_{\text{FCj}} + q_{\text{CRACj}}$[2]. Consider for instance the case with $j = 3$ and $q_{\text{CRACtot1}} > q_{\text{A1}}$, $q_{\text{CRACtot2}} > q_{\text{A2}}$, $q_{\text{CRACtot3}} < q_{\text{A3}}$, then, applying the same reasoning as before, the system coupling variables can be determined by solving the following equations

$$\begin{aligned} &\sum_{j=1}^{3} q_{\text{CRACtotj}} T_{\text{SF}} = \sum_{j=1}^{2} q_{\text{CRACtotj}} T_{\text{Ain}} + q_{\text{A3}} T_{\text{Ain}} - (q_{\text{A3}} - \\ &- q_{\text{CRACtot3}}) T_{\text{ROOF}} \\ &\sum_{j=1}^{3} q_{\text{CRACtotj}} T_{\text{RDj}} = \sum_{j=1}^{2} q_{\text{CRACtotj}} T_{\text{ROOM}} + q_{\text{A3}} T_{\text{ROOF}} - \\ &\sum_{j=1}^{2} (q_{\text{CRACtotj}} - q_{\text{Aj}}) T_{\text{Ain}}. \end{aligned} \tag{12}$$

Finally, for the room temperature, we assume it uniformly distributed with a value equal to the average $T_{\text{ROOM}} = \frac{T_{ROOF} + T_{Ain}}{2}$.

As far as thermal constraints are concerned, typically the following limitations need to be met, in order to ensure safe and effective operation of the computation platform

$$\begin{aligned} &T_{\text{ITj}} < T_{\text{ITmax}},\ \forall j,\ T_{ROOF} \leq T_{\text{ROOFmax}} \\ &T_{\text{Win}} > T_{\text{Winmin}},\ T_{\text{Wout}} < T_{\text{Woutmax}} \end{aligned} \tag{13}$$

where the first bound clearly enforces the thermal stability of the computing devices, while the limit on T_{ROOF} can be given to keep the room temperature within safe limits, and allow fast maintenance by human operators, or for some systems, due to CRAC return air temperature limits, as it will be detailed in the benchmark system of next Section. In case liquid cooling is used, then water inlet and outlet temperature need to be kept above and below, respectively, threshold values,in order to avoid dew point condensation, and damages to the pipes.

4 Case Study

With the modeling strategy introduced in the previous Section, a quite powerful tool can be built, integrating the different component models to flexibly describe various cooling architectures and strategies, evaluating their impact on the overall system efficiency to make decisions at the system design and deployment stage. Such features will be shown in this Section, taking a real hybrid-cooled Tier 1 HPC as benchmark.

4.1 Benchmark System Case Study

The system under study is a modern Tier 1 HPC, Galileo, hosted at CINECA, in Italy. Briefly speaking the computing system is based on an *IBM NeXtScale* cluster, while traditional air-cooling and rack level rear door heat exchangers

2. these variables denote the blowers flow rate entering the j^{th} portion of the system, i.e the partitioning of the overall variables described in eq. (5) according to the different hydraulic impedances.

$$
\begin{aligned}
C_{IT}\dot{T}_{IT} &= P_{IT} - \frac{T_{IT}-T_{ROOM}}{R_{IT-R}} - \left(\frac{T_{IT}}{R_{IT-A}} - \frac{T_{Aout}+T_{Ain}}{2R_{IT-A}}\right) \\
C_{RDin}\dot{T}_{Aout} &= \left(\frac{T_{IT}}{R_{IT-A}} - \frac{T_{Aout}+T_{Ain}}{2R_{IT-A}}\right) - q_A c_{vA}\rho_A (T_{Aout}-T_{Ain})\Big) \\
C_{Room}\dot{T}_{RD} &= q_A c_{vA}\rho_A (T_{Aout}-T_{Ain}) - \left(\frac{T_{Aout}+T_{ARD}}{2R_{RDHX}} - \frac{T_{Win}+T_{Wout}}{2R_{RDHX}}\right) + \\
&\quad + \frac{T_{IT}-T_{ROOM}}{R_{IT-R}} - q_{CRACFC}\rho_A c_{vA} T_{ROOF} \\
C_{SF}\dot{T}_{SF} &= c_{vA}\rho_A q_{CRAC} T_{ROOF} - P_{CRAC} + q_{CRACFC} c_{vA}\rho_A T_{amb} - q_{CRACtot} c_{vA}\rho_A T_{SF} \\
C_{RDout}\dot{T}_{Wout} &= \left(\frac{T_{Aout}+T_{ARD}}{2R_{RDHX}} - \frac{T_{Win}+T_{Wout}}{2R_{RDHX}}\right) - q_W c_{vW}\rho_W (T_{Wout}(t)-T_{Win}) \\
C_{Ch}\dot{T}_{Woutch} &= q_W c_{vW}\rho_W (T_{Wout}-T_{Win}) - \left(\frac{T_{Wout}+T_{Woutch}}{2R_{CH}} - \frac{T_{0Ch}}{R_{CH}}\right)
\end{aligned}
\tag{14}
$$

can be combined exploited. More specifically, the system consists of 14 racks, displaced in a single cool aisle hot aisle configuration, i.e. with two lines of racks facing each other, and absorbing cool air from perforated tiles. Each rack can host up to 6 stacked chassis, which in turn can contain 8 nodes cards each, for a total of 516 nodes. Each node is equipped with 2 Intel *Haswell E52630 v3* CPUs, with 8 cores of 1.8 GHz clock speed and 130W Thermal Design Power (TDP, [41]). In addition, 384 nodes mount two *Xeon Phi* 7120*p* accelerators, operating at 1.2GHz and 300W TDP. Taking into account also miscellaneous components, the overall system TDP is about 360 kW. As regards software infrastructure, SMP CentOS Linux distribution version 7.0 is executed on each node.

The HPC is put in a room portion containing 5 *Emerson 99UA* and 5 *Uniflair TDAV3342A* direct expansion CRAC units, for what concerns the air cooling part. A cage, forcing an air path from the perforated tiles to the room roof, then back to the CRACs, is mounted around the machine preventing recirculation and mixing of hot and cool air.

As regards the liquid circuit, RDHXs are mounted on each rack, and the water entering the heat exchangers is provided by a chiller, with adjustable cooling capabilities. A variable speed pump pushes the water into the pipes, while a three-way valve can be exploited to recirculate part of the warm return water, mixing it to the chiller outlet water, before re-entering the RDHXs.

Recalling the analysis and modeling effort carried out in Section 3, we can represent the HPC and cooling thermal behavior by equations in (14). In this case, the HPC has been considered as a unique heat source ($j = 1$), since balanced workload distribution for quite long jobs is typically ensured for this kind of machines [42], moreover, both air and liquid cooling granularity are not split for the two separate rack lines. Note that, according to what mentioned in paragraph 3.1, C_{RDin} denotes the thermal capacitance at the air side of the RDHXs, considered as a unique block. The equation parameters can be estimated by the system geometry and theoretical formulas[3], and/or identification procedures. The numerical values are reported in Tab. 4.1, along with the thermal bounds on the computing nodes and room temperature.

Table I Benchmark System's Parameters

Parameter	*Value*
$R_{IT-R}, R_{IT-A}, R_{RDHX}, R_{CH}$	86, 1.4, 1.1, 0.6 [mK/W]
C_{IT}, C_{RDin},C_{RDout}	2421, 786, 786, [J/K]
C_{Room}, C_{Ch}	12039, 549 [J/K]
$T_{0CRAC}, T_{Winmin}, T_{Woutmax}$	3, 18, 50 [°C]
T_{ITmax}, $T_{ROOFmax}$	85, 30 [°C]
P_{CRACmE},P_{CRACmU}	84.1,96.9 [kW]
c_{vA},c_{vW}	1005,4186 [J/(Kg× K)]
ρ_A,ρ_W	1.025, 1000 [Kg/m^3]

4.2 Model Validation on the Considered Benchmark

Before showing the potential usage of the proposed modeling tool, a necessary step is to validate its accuracy in predicting the real system behavior.

As the main focus is put onto energy efficiency, the capability of the model to capture such feature has been evaluated. A well established metric for measuring efficiency of complex computational platforms such as HPCs is the the Power Usage Effectiveness (PUE) [43], i.e. the ratio between the overall system power (cooling and computing power) and the workload. Therefore, the real 2015 (actually, from January 1^{st} to December 1^{st}) hourly PUE of Galileo have been compared to the results obtained using the mathematical representation in eq. (14). To this aim, the model has been parametrized according to the actual system cooling configuration which, for the considered period (the first year of life of the HPC) was equipped with just air cooling (therefore only the first, second and forth equation in (14) have been considered), and 2 Uniflair TDAV3342A CRACs could operate in free cooling mode.

For what concerns the cooling knobs in the considered period, the target of the actual cooling strategy was to keep the room temperature at 23°C in any condition. Therefore, such strategy has been replicated in our model. Results are shown in Fig. 3. For the sake of completeness, Figs. 3 (a) and (b) show the considered period ambient temperature and system workload profile, respectively. In Fig. 3 (c) the corresponding system hourly PUE (in red) and the values predicted by using our model are compared. Being the HPC workload rather constant, the changes in PUE are mainly dictated by the ambient temperature profile (see Figs. 3

3. For instance the computing part to room resistance can be computed as $R_{IT-R} = (hA_R)^{-1}$, where h is the heat transfer coefficient of the racks material and A_R is the overall racks side surface exposed to the room, while the thermal capacitance values have been determined as $C_{th} = \rho c_v V$, where V is the volume of the considered heat exchange point.

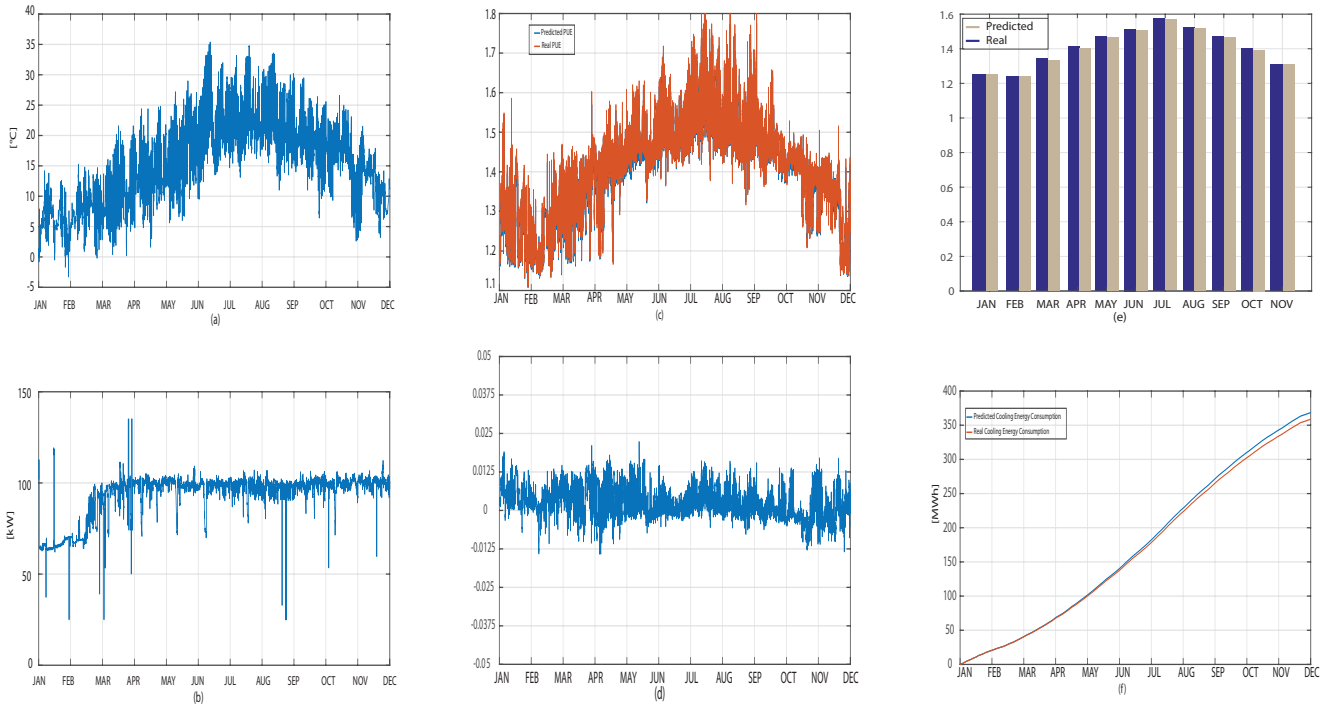

Fig. 3. Model accuracy validation for the case study. (a) 2015 Hourly annual temperature profile for Galileo location. (b) Galileo's 2015 hourly workload profile. (c) Predicted (blue) and real system PUE values. (d) PUE prediction error. (e) Monthly average PUE, real (blue) and predicted (gray).(f) Predicted (blue), and real (red) annual cooling energy consumption.

(a) and (c)). It can be seen how the values obtained using the proposed model (blue plot in Fig. 3 (c)) track the real data (in red in Fig. 3 (c)) pretty closely. This is confirmed and highlighted in Fig. 3 (d), showing the PUE prediction error (real minus predicted values). More specifically, in Fig. 3 (e) the monthly average PUE values are compared. It can be noted how, the maximum mismatch between the real average and the predicted one is about the 3%, in the worst cases. This is due to the approximations introduced by the coarse-grain modeling. However, the trend (PUE increase/decrease throughout the year) is correctly described, and the predicted average PUE pattern has the same shape of the real one. Such result shows that the proposed tool is capable of capturing the significant effects (environmental and computational) which impact a real HPC system efficiency. Therefore, it can be profitably exploited to make decisions about suitable cooling topologies depending on the workload and environmental conditions. Similarly, the model can be used to design energy-aware cooling strategies and thermal control policies, possibly with slight parameters fine tuning (based on identification with experimental data) when a very accurate behavior prediction has to be obtained. Finally, in Fig. 3 (f) the estimate of the annual cooling energy consumption, obtained integrating the power consumption throughout the year, is reported. Comparison against the real one show an an underestimation of the total cooling energy which is less then 2%, confirming again the reasonable accuracy of the proposed analytic description. In view of such results, the framework will be applied in the following to the considered case study, analyzing the effects of workload and environmental conditions on the system efficiency, and introducing suitable cooling design procedures.

4.3 Cooling Scenarios Analysis

Beside the capability of the proposed tool to predict the cooling consumption and system thermal behavior, thanks to its limited complexity, it can also be profitably exploited for designing model-based optimal cooling strategies.
In this respect, given (14), the power consumption (7)-(10), and constraints (13), an optimization problem has been formulated to define a energy-efficient cooling strategy, setting the corresponding control knobs $(\beta_j,\ \gamma_j,\ \alpha,\ q_w, T_{0Ch})$, with the goal to minimize the system PUE (i.e. minimizing the cooling power consumption). The problem is not convex (due to product between decision variables in both constraints and objective function), therefore a heuristic approach, based on multiple solutions obtained with a Sequential Quadratic Programming approach, starting from different initial points, has been exploited to approximate its solution[4].
Details of such procedure have been presented in [44] (we refer the interested reader to that work for a deeper elaboration on the optimization part), here such strategy is used as a common cooling management framework, to be applied to different topologies for the considered case study, using the proposed model, and the power consumption equations, to compare such various architectures, evaluating suitable (energy-efficient) cooling set-ups given different

4. With some abuse of notation we refer to the resulting cooling management strategy as "optimal". Although this term is not rigorously correct from the mathematical standpoint, for the solver algorithm is not ensured to reach the global optimum, it briefly capture the idea of a energy-aware cooling control, oriented to power consumption minimization.

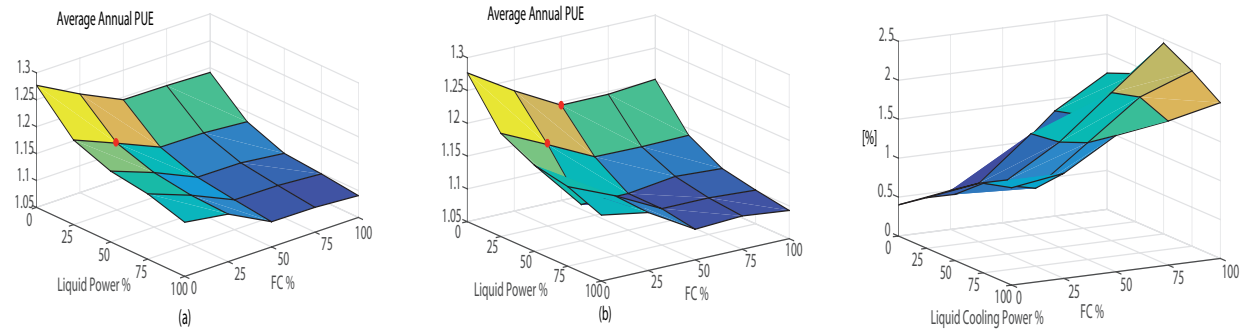

Fig. 4. Cooling efficiency comparison among different cooling endowments for the considered benchmark study, under a constant 80% TDP workload, and considering the machine location 2015 annual temperatures. Plot (a) annual average PUE for different cooling topologies. Plot (b) annual average PUE for different cooling topologies and speculative air free cooling strategy. Plot (c) annual PUE $\%$ improvement given by the speculative approach under the different cooling architectures.

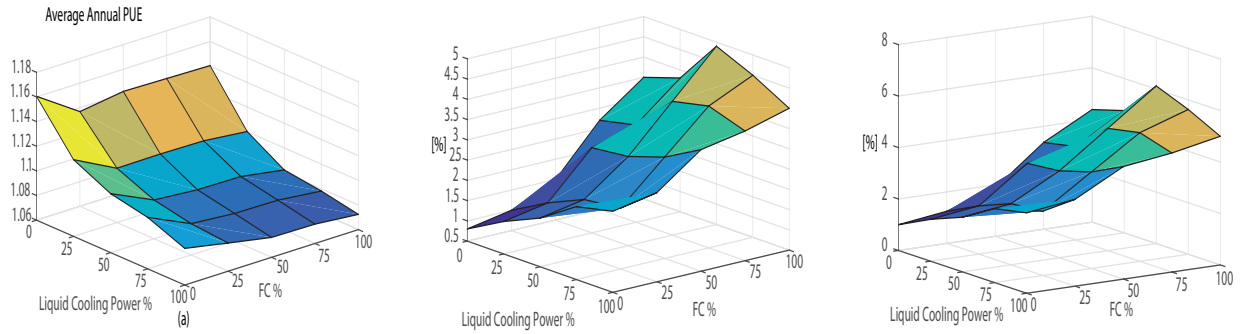

Fig. 5. Cooling efficiency comparison among different cooling endowments for the considered benchmark study, under an uneven night/day distribution of an average annual 80% TDP workload, unbalanced towards night with 40%TDP nightly workload and 60%TDP daily workload. Plot (a) annual average PUE for different cooling topologies with the consider unbalanced workload allocation. Plot (b) PUE $\%$ improvement given by the nightly workload increase w.r.t. the constant workload case. Plot (c) PUE $\%$ improvement achievable by combining nightly workload shifting and the speculative air free cooling, w.r.t. the constant workload case and conservative strategy.

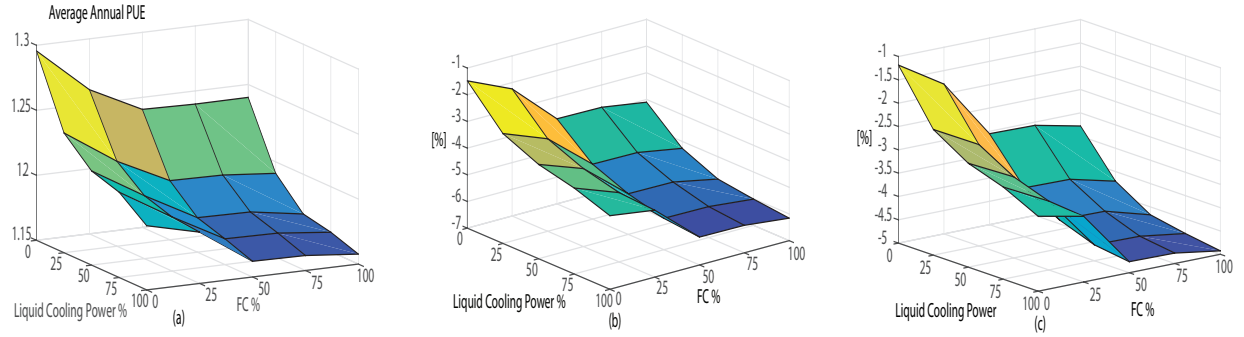

Fig. 6. Cooling efficiency comparison among different cooling endowments for the considered benchmark study, under an uneven night/day distribution of an average annual 80% TDP workload, unbalanced towards day, with 100%TDP daily workload and 60%TDP nightly workload. Plot (a) annual average PUE for different cooling topologies with the considered unbalanced workload allocation. Plot (b) PUE $\%$ worsening given by the daily workload increase w.r.t. the constant workload case. Plot (c) PUE $\%$ worsening by using speculative air free cooling under the daily unbalanced workload, w.r.t. the constant workload case and no speculative strategy.

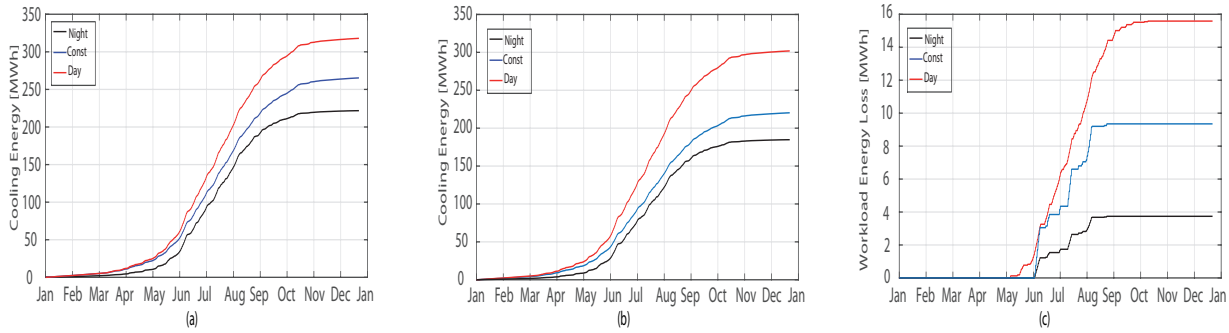

Fig. 7. Annual cooling power consumption for the considered benchmark HPC, with cooling configuration corresponding to 50% liquid cooling heat removal capability and 25% CRACs with free-cooling mode allowed. Comparison among workload uneven dispatch and air free-cooling speculative method. Plot (a) results with different workload allocation and no speculation on free-cooling. Plot (b) results with different workload allocation and air free cooling speculative management. Plot (c) Workload loss due to speculative strategy for the considered workload conditions.

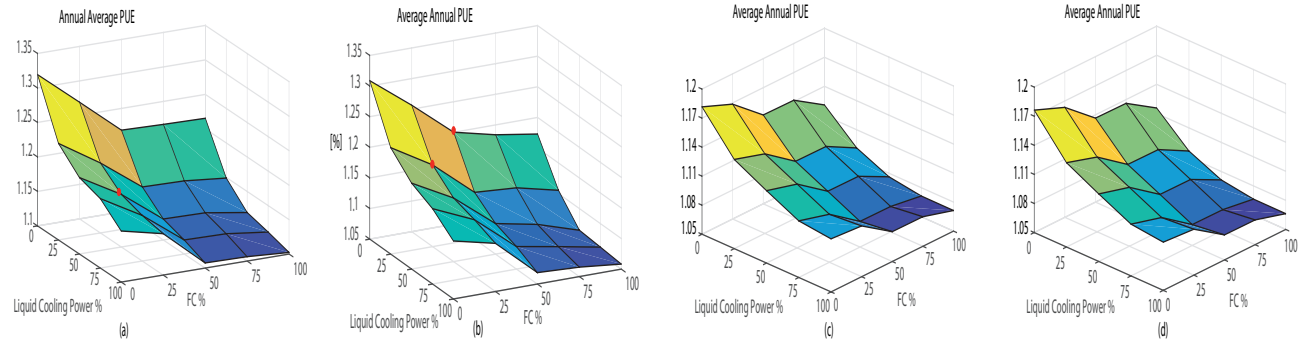

Fig. 8. Cooling efficiency comparison among different cooling endowments for the considered benchmark study under constant 80% TDP workload, and assuming a different geographic location. Plot (a), annual average PUE for different cooling topologies and Barcelona ambient profile. Plot (b), same as (a) but with speculative air-free cooling strategy. Plot (c) annual average PUE for different cooling topologies and Stockholm ambient profile. Plot (d), same as (c) but with speculative air-free cooling strategy.

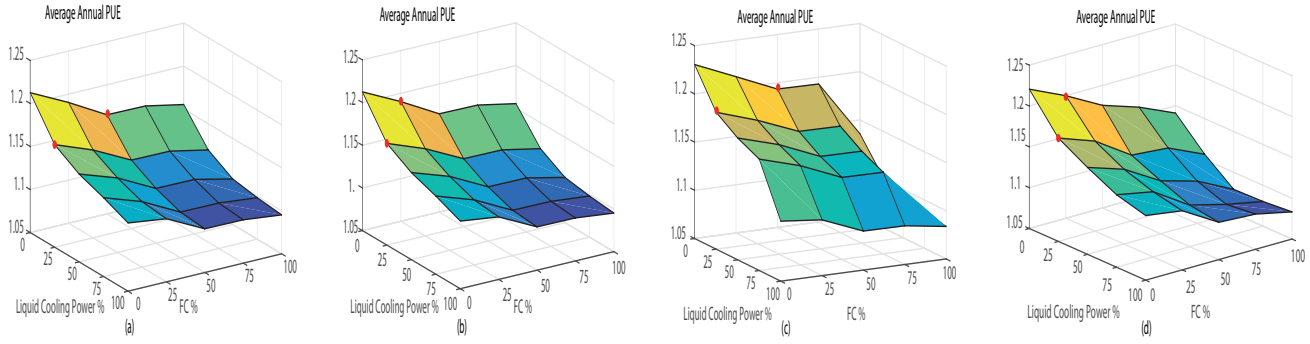

Fig. 9. Cooling efficiency comparison among different cooling endowments for the considered benchmark study under constant 80% TDP workload, and assuming a different geographic location. Plot (a), annual average PUE for different cooling topologies and Paris ambient profile. Plot (b), same as (a) but with speculative air-free cooling strategy. Plot (c) annual average PUE for different cooling topologies and Munich ambient profile. Plot (d), same as (c) but with speculative air-free cooling strategy.

expected workload profiles and environmental conditions. Actually, a subtle but crucial difference has been introduced with respect to the optimization carried out in [44]. Indeed here a "speculative" air free cooling strategy is allowed, running the optimized control with a less tight constraint on T_{ROOFMax} ($35°$ instead of $30°$) whenever a solution with only free-cooling CRACs are used is feasible. The rationale of this choice is to save cooling energy, exploiting the air free-cooling as much as possible, clearly keeping the computing device thermal stability. The price to pay for that is a possible workload penalization whenever a CRAC cooling cycle has to be switched on. In fact, we assume the T_{ROOFMax} original constraint is related to the CRACs requirement, not allowing to high return air temperature when operating in "electric cooling" mode, as is the case of the studied system devices, or to allow comfortable maintenance in the room by human operator if needed. In any case, it can happen that the workload P_{IT} has to be temporary cut, in order to bring back the room temperature (and then T_{ROOF}) within the original limits, either because otherwise the CRACs cannot be switched on and thermal stability would be lost (the temperature in the room is to high but the workload

and environmental conditions require active cooling), or to ensure a rapid operator intervention without waiting the CRACs to reduce the room temperature with the machine operating at high load (thus injecting considerable heat). The time of reduced workload is estimated using the proposed dynamic model, based on the actual system thermal state, the workload, the cooling knobs, and the desired final point (i.e. T_{ROOF}). The benefits (or drawbacks) of such strategy will be investigated in the following. At the same time, the modeling tool will be exploited to investigate the effect of heterogeneous cooling technologies, depending on workload allocation and varying environmental conditions, given by different geographical locations.

To this aim, a set of possible cooling architecture have been considered; for what concerns the liquid cooling part, the overall heat removal capability via the chiller and the rear-doors have been consider equal to $25, 50, 75, 100\%$ of the machine TDP, while for the air cooling circuit different free cooling options have been considered, again considering the $25, 50, 75, 100\%$ of CRACs endowed with this operation mode. Fig. 4 compares the efficiency of this set of configurations, computed as the average annual PUE, under a constant workload, corresponding to 80% of the considered system TDP, and hourly 2015 average temperatures in Bologna [45], Italy, where Galileo is hosted. Both traditional (Fig.4 (a)) and speculative algorithm (Fig.4 (b)) have been considered. It can be noted how, while obviously the option with liquid capable of removing all the heat, and 100% air free cooling is the best one, comparable efficiency can be achieved also limiting liquid and free cooling to 50%, thus reducing the investment in the infrastructure. The same reasoning holds for the speculative strategy, which clearly does not give much benefit if air free cooling is not allowed, still some saving can be obtained if the liquid is exploited, as confirmed by Figs. 4 (b), (c), the latter showing how speculative strategy can improve the average PUE by about 2% in the best case.

An interesting question which can be solved by the proposed modeling and analysis framework is what is the "minimum" cooling investment which is needed to ensure an annual PUE below a threshold, given an expected workload, and temperature profiles. In other words, how many CRACs need to be equipped with free-cooling, and/or how much cooling power the liquid part is able to remove.

In this respect, the red marks in Figs. 4 (a),(b), denote such minimal configurations, for a PUE threshold equal to 1.2. It can be noted how having the 50% of air conditioner with no liquid part is (with slight margin) enough to satisfy the PUE requirement, while having just the 25% of air free-cooling and adding an indirect liquid cooling with 25% heat removal capability allows to satisfy the condition with a larger margin, confirming the benefit of liquid cooling. Clearly, if no liquid pipes are installed in the HPC room, the cost of installation in this second configuration will be higher. Adopting speculative strategies does not help to move such threshold points (see Fig. 4 (b)), later on it will be shown how, under different ambient temperature profiles, speculative cooling can provide also a positive shift in such critical points.

To assess also the impact of smart workload allocation policies, the computing power profile has been unevenly dispatched toward day or night hours. Keeping the same mean value equal to 80% of TDP, two different distributions have been tested: 100% TDP workload during day and 60% denoted as "day", and the opposite condition (100% TDP at night and 60% TDP during day) denoted as "night". Results are shown in Figs. 5, 6, and compared to the constant workload scenario, both with speculative and traditional optimal strategies. Shifting the load toward cool nightly hours allow a considerable benefit, particularly for configurations allowing air free cooling (see Figs. 5 (a),(b)), furthermore, joining it to speculative cooling control, allows to increase the annual efficiency around 5% if all the CRACs are endowed with free cooling capabilities (see Fig. 5 (c)). On the contrary, loading the machine during the day worsen the efficiency performance no matter what the cooling configuration is, making the liquid and air free cooling option not be fully exploited, as confirmed by Fig. 6 (a), (b). However, adopting speculative control can mitigate such effects (see Fig. 6 (c)).

In Fig. 7 a more detailed analysis, corresponding to a fixed topology with 25% free-cooling and 50% liquid cooling power case is shown, regarding the annual cooling consumption under the aforementioned workload scenarios and speculative strategy. Comparing plots 7 (a) (b) it can be seen how speculation, combined with load shifting towards night, can reduce consumption by 25% w.r.t. the nominal case. However, for a fair investigation, the workload loss due speculation and possible reduction of P_{IT} should be evaluated as well. Plot 7 (c) shows such data; again, dispatching heavy load at night allows to get better result also in terms of performance loss, as power cut is almost never needed, while for the other two cases, especially in summer months, some workload loss can be noted. It is further to remark that the power cut starts earlier in the year for constant and day unbalanced workload, the latter increasing smoother due to the fact that the high day workload prevent free cooling activation and then also speculations and possible P_{IT} downgrade. However, in all cases, the overall loss is quite negligible (less then 1%) w.r.t. the overall annual workload achievable with no speculation.

Having established the effects of workload and cooling strategy, we move now to consider different environmental conditions. To this purpose, we assumed the benchmark HPC to be placed in other areas of Europe, corresponding to other computational centers locations such as: *Barcelona SuperComputing center* (Barcelona, Spain), *Leibniz Supercomputing Centre* (LRZ, Munich, Germany), *Trs Grand Centre de calcul du CEA* (TGCC, Paris, France), and *Center for High Performance Computing* at the KTH (Stockholm, Sweden). To emulate the 2015 ambient conditions we used the publicly available dataset ERA Interim forcasting data [46] from ECMWF which provides the temperature at 2m from the ground every 3 hours. We linearly interpolated these values to obtain the hourly temperature of each supercomputing site for the entire 2015. Supercomputer locations have been approximated to the nearest grid point. In order to better establish the effect of the geographic conditions on the cooling architecture, i.e. separating this contribution from possible advanced workload time distribution, the case with constant workload at 80% has been applied to all the aforementioned scenarios, while speculative air free cooling has been investigated as well. The results are summarized in

Figs. 8, 9, referred to the pairs Barcelona/Stockholm and Paris/Munich location, respectively.
Looking at Figs. 8 (a)-(b), and (c)-(d), it is easy to note that this corresponds to two quite different scenarios in terms of environmental conditions. The former, being associated with a Mediterranean place, shows worse efficiency, close to what in Fig. 4, corresponding to a similar geographic area. However, in this scenario, liquid cooling is needed if a PUE below 1.2 is required, and no speculative strategy is applied (see Fig. 8 (a)), with the best cooling configuration ensuring to not overcome the limit PUE requiring a 50% liquid cooling power w.r.t the machine TDP, and 25% of the CRACs with free cooling option. Indeed, if the previously described speculation with air free cooling is performed, then the same PUE bound can be ensured just with a 25% liquid cooling power capability and 25% of the CRACs able to operate in free cooling, or, more interestingly, with no liquid part at all, but having the 50% of CRACs equipped with free cooling (notice the two red marks in Fig. 8 (b)). This result is not surprising, indeed, being the considered region quite temperate during winter, and warm (but not too hot) in summer, there are many days of the year when speculative strategy allows to use free-cooling, while a more conservative solution would not.
For what concerns the scenario in Figs. 8, a complete different behavior can be noticed. In fact, being associated to a rather cold region, good efficiency can be ensured (provided optimal cooling control is implemented), even with standard air-based configurations. In addition, adding liquid cooling and air free-cooling does not improve, in relative terms, the efficiency as much as for the scenarios in Figs. 4, 8 (a)-(b). Indeed, comparing the best points (always 100% liquid cooling power, 100% air free cooling) in these scenarios, it can be seen how they all are pretty similar, tending to a PUE slightly lower then 1.1. This can be explained with the fact that, for cold temperature, the cooling system becomes oversized, and the cost of pumping water and blowing air becomes dominant, not allowing a further decrease in the PUE, while for warm temperatures, improving the cooling potential is actually contributing to save energy, avoiding inefficient refrigerating cycles in the CRACs and chiller.
Fig. 9 shows the results corresponding to the other considered possible locations of the Galileo HPC, since the environmental condition of the two geographical areas (Germany and north of France) are quite similar, so are the efficiency analysis of the possible cooling system arrangement (compare plots 9 (a),(c) and (b), (d)) respectively. As expected, the average annual PUE is improved w.r.t. southern Europe location such as those in Figs 4, 8 (a)-(b). However, differently from the case in 8 (c)-(d), some investment in air free cooling or hybrid (with also liquid part) can significantly improve the performance (note in particular the steep decrease for air free cooling higher then 75% of CRACs total in Fig. 9 (c)). Again, if the limit average annual PUE of 1.2 has to be satisfied, some advanced cooling infrastructure is needed, specifically, at least liquid circuit able to remove 25% of the heat generated by the computing system, or a standard air system with 50% free-cooling endowments would be needed for both the possible locations, as underscored by marks in Figs. 9 (a) and (c). When speculative free cooling is implemented, beside some improvements, no shift in the threshold point takes place for what regards LRZ location (Germany), as confirmed by plot 9 (d), while, if located in the TGCC location and operated with looser room temperature constraints, the system would operate under the given PUE also with just 25% of the conditioner able to exploit such strategy, and no liquid cooling would be required. it

5 Conclusions

An analytical, lumped parameters, and holistic thermal modeling approach for complex computational platforms has been presented, with the aim to provide an effective tool, with limited complexity, to be exploited for advanced computation centers cooling design and control. In this respect, the system thermal behavior has been considered at full scale, dealing with the complex energy interactions in a coarse grain but physically meaningful fashion. It has been remarked how different cooling architectures and their effects on the overall energy efficiency can be quickly and flexibly described adapting the model equations. The impact of external ambient conditions on the cooling efficiency has been explicitly taken into account, as well as the (obvious) effect of the computational workload.
The properties of the proposed approach have been investigated on a real case study, Galileo, a state-of-the-art hybrid-cooled Tier-1 Supercomputer. Specifically, different cooling configurations have been assumed, exploiting the proposed tool to model each of them. Then, the corresponding overall system efficiency has been evaluated, considering different possible workload profiles, and evaluating a less conservative free cooling mode decision strategy. In addition, the effect of different environmental conditions have been evaluated, assuming the studied system to be located in various geographical areas. It has been shown how the tools allow to make sensible decisions about the cooling set-up, which is crucial to ensure efficiency and sustainability of large scale computational centers. It is further to remark that all the cooling parts and the heat generating sources are taken into account, as well as the the aforementioned external factors. Another significant aspect is given by the light computational burden required to tune and run the model for analysis and evaluation. This not only is a clear advantage in terms of evaluating several possible cooling architectures in a short amount of time, but also can be exploited to implement energy-aware cooling management strategies, as the resulting optimization problem involves a reasonable number of decision variables and constraints.
In view of such features, the presented approach looks promising to provide valuable help in designing and control of today, and next generation HPC cooling systems. Economic criteria (e.g. cooling upgrade cost of investment) could be easily added to efficiency improvement in a sort of multi-objective decision process. Future developments will go in this direction. In addition, the characterization of speculative cooling control strategies could be extended, considering coupling and interaction with energy-aware schedulers and job dispatchers.

Acknowledgment

This work was supported by the EU FP7 ERC Advance Project MULTI-THERMAN (GA n. 291125).

REFERENCES

[1] K. Bergman, S. Borkar, D. Campbell, W. Carlson, W. Dally, M. Denneau, P. Franzon, W. Harrod, K. Hill, J. Hiller, S. Karp, S. Keckler, R. L. D. Klein, M. Richards, A. Scarpelli, S. Scott, A. Snavely, T. Sterling, and R. S. W. K. Yelick, "Exascale computing study: Technology challenges in achieving exascale systems," Tech. Rep., 2008, institution=DARPA IPTO, contract number FA8650-07-C-7724, Date-Added =2008,.

[2] G. I. Meijer, "Cooling energy-hungry data centers," *Science*, vol. 328, pp. 318–319, 2010.

[3] E. Or, V. Depoorter, A. Garcia, and J. Salom, "Energy efficiency and renewable energy integration in data centres. strategies and modelling review," *Renewable and Sustainable Energy Reviews*.

[4] T. Evans, "The different types of air conditioning equipment for it environments," *APC White Paper n. 59*, pp. 1–19, 2004.

[5] M. Pore, Z. Abbasi, S. K. S. Gupta, and G. Varsamopoulos, "Techniques to achieve energy proportionality in data centers: A survey," in *Handbook of Data Centers*, S. U. Khan and A. Y. Zomaya, Eds. Berlin: Springer, 2015.

[6] L. Li, X. W. W. Zheng, and X. Wang, "Coordinating liquid and free air cooling with workload allocation for data center power minimization," *Proc. of ICAC*, pp. 249–259, 2014.

[7] D. Watts, J. Caubet, D. Furniss, and D. Latino, *IBM NeXtScale System Planning and Implementation Guide*. IBM RedBooks, 2014.

[8] D. F. M. K. Patton, "The state of data center cooling: A review of current air and liquid cooling solutions." *Intel White Paper*, pp. 1–11, 2008.

[9] S. W. C. J. Kong and K. Skadron, "Recent thermal management techniques for microprocessors," *ACM Computing Surveys*, vol. 44(3), pp. 1065–1070, 2012.

[10] R. Rhomadon, M. Ali, A. M. Mahdzir, and Y. A. Abakr, "Energy efficiency and renewable energy integration in data centres. strategies and modelling review," *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 429–445, 2015.

[11] L. Wang and Y. Lu, "An efficient threshold-based power management mechanism for heterogeneous soft real-time clusters," *IEEE Trans. Ind. Informat.*, vol. 6(3), pp. 352–364, 2010.

[12] A. Sfrent and F. Pop , "Asymptotic scheduling for many task computing in Big Data platforms," *SInformation Sciences*, vol. 319, pp. 71–91, 2015.

[13] J. Menga, S. McCauleyb, F. Kaplana, V. J. Leungc, and A. K. Coskun, "Simulation and optimization of hpc job allocation for jointly reducing communication and cooling costs," *Sustainable Computing: Informatics and Systems*, vol. 6, pp. 48–57, 2015.

[14] A. Borghesi and A. Bartolini, "Predictive Modeling for Job Power Consumption in HPC Systems," *International Conference on High Performance Computing*, 2016.

[15] A. Borghesi and A. Bartolini , "Power capping in high performance computing systems," *International Conference on Principles and Practice of Constraint Programming*, 2015.

[16] A. Borghesi and C. Conficoni and M. Lombardi and A. Bartolini , "MS3: A Mediterranean-stile job scheduler for supercomputers - do less when it's too hot!" *2015 International Conference on High Performance Computing & Simulation (HPCS)*, pp. 88–95, 2015.

[17] V. Hanumaiah and S. Vrudhula, "Energy-efficient operation of multicore processors by dvfs, task migration, and active cooling," *IEEE Tran. on Computers*, vol. 63(2), pp. 349–360, 2014.

[18] J. Kim, M. Ruggiero, and D. Atienza, "Free cooling-aware dynamic power management for green datacenters," *Proc. of IEEE HPCS*, pp. 140–146, 2012.

[19] M. M. Sabry, A. K. Coskun, D. Atienza, T. S. Rosing, and T. Brunschwiler, "Modeling and dynamic management of 3d multicore systems with liquid cooling," *Proc. of VLSI-Soc*, pp. 1–10, 2009.

[20] R. Das, J. O. Kephart, J. Lenchner, and H. Hamann, "Utility-function-driven energy-efficient cooling in data centers," *Proc. of ICAC*, pp. 1526–1544, 2010.

[21] A. Banerjee, T. Mukherjee, G. Varsamopoulos, and S. K. Gupta, "Energy-optimal dynamic thermal management: Computation and cooling power co-optimization," *IEEE Trans. Ind. Informat.*, vol. 6(3), pp. 340–351, 2010.

[22] W. Huang, M. Allen-Ware, J. B. Carter, E. Elmootazbellah, H. Hamann, T. Caller, C. Lefurgy, J. Li, K. Rajamani, and J. Rubio, "Tapo: Thermal-aware power optimization techniques for servers and data centers," *Proc. of IEEE IGCC*, pp. 1–8, 2011.

[23] L. Parolini, E. Garone, B. Sinopoli, and B. H. Krogh, "A hierarchical approach to energy management in data centers," *Proc. of IEEE CDC*, pp. 1065–1070, 2010.

[24] L. Parolini, B. Sinopoli, B. H. Krogh, and Z. Wang, "A cyberphysical systems approach to data center modeling and control for energy efficiency," *Proc. of the IEEE*, vol. 100(1), pp. 255–268, 2012.

[25] Z. Wang, C. Bash, C. Hoover, , C. Felix, and R. Shih, "Integrated management of cooling resources in air-cooled data centers," *Proc. of IEEE Int. Conf. on Autom ad Eng.*, pp. 762–767, 2010.

[26] R. Zhou, Z. Wang, C. E. Bash, A. McReynolds, C. Hoover, R. Shih, N. Kumari, and R. K. Sharma, "A holistic and optimal approach for data center cooling management," *Proc. of IEEE Amer. Ctrl. Conf.*, pp. 1346–1351, 2011.

[27] M. M. Sabry and D. Atienza, "Temperature-aware design and management for 3d multi-core architectures," *Foundations and Trends in Electronic Design Automation*, vol. 8(2), pp. 117–197, 2014.

[28] P. R. Parida, T. J. Chainer, M. D. Schultz, and M. P. David, "Cooling energy reduction during dynamically controlled data center operation," *Proc. of ASME interPACK Conference*, pp. 457–471, 2013.

[29] D.C., Hwang, V. P. Manno, M. Hodes, and G. J. Chan, "Energy savings achievable through the liquid cooling: a rack level case study," *Proc. of IEEE ITherm*, pp. 1–9, 2010.

[30] M. Seymour, "The increasing challenge of data center design and management: Is cfd a must?" *Electronics Cooling*, pp. 28–33, 2011.

[31] K. C. Karki and A. Radmehr and S. V. Patankar, "Use of Computational Fluid Dynamics for Calculating Flow Rates Through Perforated Tiles in Raised-Floor Data Centers," *International Journal of Heating, Ventilation, Air-Conditioning, and Refrigeration Research*, vol. 9(2), pp. 153–166, 2003.

[32] G. Janiga, "A Few Illustrative Examples of CFD-based Optimization," in *Optimization and Computational Fluid Dynamics*, D. Thvenin and. G. Janiga, Ed. Berlin: Springer, 2008.

[33] A. Radmehr and B. Noll and J. Fitzpatrick and K. Karki , "CFD Modeling of an Existing Raised-Floor Data Center," *SEMI-THERM*, 2013.

[34] M. Iyengar, M. David, P. Parida, V. Kamath, B. Kochuparambil, D. Graybill, M. Schultz, M. Gaynes, R. Simons, R. Schmidt, and T. Chainer, "Extreme energy efficiency using water cooled servers inside a chiller-less data center," *Proc. of IEEE ITHERM*, pp. 137–150, 2012.

[35] M. P. David, M. Iyengar, P. Parida, R. Simons, M. Schultz, M. Gaynes, R. Schmidt, and T. Chainer, "Experimental characterization of an energy efficient chiller less data center test facility with warm water cooled servers," *Proc. of IEEE SEMI-THERM*, pp. 232–237, 2012.

[36] M. Massoud, *Engineering Thermofluids: Thermodynamics, Fluid Mechanics, and Heat Transfer*. Springer, 2005.

[37] T. J. Breen, E. J. Walsh, and J. Punch, "From chip to cooling tower data center modeling: Part i influence of server inlet temperature and temperature rise across cabinet," *Proc. of ITHERM*, pp. 1–10, 2010.

[38] A. Feschenko, Y. Kiselev, A. Kovalishin, L. Kravchuk, and A. Kvasha, "Spallation neutron source drift tube lineac resonance control cooling system modeling," *Proc. of IEEE Particle Accelerator Conference*, pp. 3754–3757, 2005.

[39] Y. Ma, J. Matusko, and F. Borrelli, "Stochastic model predictive control for building hvac systems: Complexity and conservatism," *IEEE Trans. Control Syst. Technol.*, pp. 1–16, 2015.

[40] C. Conficoni, A. Bartolini, A. Tilli, L. Benini, and G. Tecchiolli, "Energy-aware cooling for hot-water cooled supercomputers," *Proc. of IEEE DATE*, pp. 1353–1358, 2015.

[41] J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach*. Elsevier, 2012.

[42] A. Bartolini, A. Borghesi, T. Bridi, M. Lombardi, and M. Milano, "Proactive workload dispatching on the eurora supercomputer," in *Principles and practice of constrained programming*, B. O'Sullivan, Ed. Springer, 2014.

[43] V. Avelar, D. Azevedo, and A. French, "Pue: A compreehensive examination of the metric," The Green Grid, Tech. Rep., 2012.

[44] C. Conficoni, A. Bartolini, A. Tilli, C. Cavazzoni, and L. Benini, "Integrated energy-aware management of supercomputer hybrid cooling systems," *IEEE Tran. Ind. Inf*, vol. In Press, 2015.

[45] http://dexter-smr.arpa.emr.it.

[46] D. Dee, S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer *et al.*, "The era-interim reanalysis: Configuration and performance of the data assimilation system," *Quarterly Journal of the royal meteorological society*, vol. 137, no. 656, pp. 553–597, 2011.

Christian Conficoni received the Masters Degree in Electronic Engineering, from the University of Bologna, Italy in 2008. In 2013 he received the Ph.D. degree in automatic control, from the same institution. Currently he is a post doctoral researcher at the Department of Electrical, Electronic and Information Engineering (DEI), at University of Bologna. His research interests include applied nonlinear control solutions for power electronic and electromechanical systems oriented to power quality enhancement, adaptive observers for electric drives sensorless operation, modeling and energy-oriented optimal thermal management of advanced computing platforms.

Andrea Bartolini received a Ph.D. degree in Electrical Engineering from the University of Bologna, Italy, in 2011. He is currently a postdoctoral researcher in the Department of Electrical, Electronic and Information Engineering Guglielmo Marconi (DEI) at the University of Bologna. He also holds a postdoc position in the Integrated Systems Laboratory at ETH Zrich. His research interests concern dynamic resource management ranging from embedded to large scale HPC systems with special emphasis on software-level thermal and power-aware techniques. His research interest also includes ultra-low power design strategies for bio-sensors nodes operating in near-threshold.

Andrea Tilli is Associate Professor at the Department of Electrical, Electronic and Information Engineering Guglielmo Marconi (DEI) of the University of Bologna. In 2000, he received the Ph.D. degree in system science and engineering from the same university. He is member of the Center for Research on Complex Automated Systems Giuseppe Evangelisti (CASY), established within DEI. His current research interests include applied nonlinear control techniques, active power filters, wind turbines, electric drives for motion control and energy generation, thermal control of many-core systems-on-chip and supercomputers.

Carlo Cavazzoni has been graduated in physics at the University of Modena in the 1994, and subsequently he has attained the degree of Phd in Material Science at the International School for Advanced Studies (ISAS-SISSA) of Trieste in 1998 with a thesis about: Large Scale First-Principles Simulations of Water and Ammonia at High Pressure and Temperature. During his PhD, he has studied various problems concerning the implementation and the efficiency of parallel numerical algorithms being used in physical computer simulations. In Consorzio Interuniversitatio per la gestione del centro di Calcolo Elettronico dell'Itaila Nord-orientale (CINECA) he is presently Responsible of High Performance Computer (HPC) User support and R&D activities. He collaborate with different user communities to enable applications on massively parallel systems and innovative architecture solutions. In particular he is responsible for the parallel design of Quantum ESPRESSO suite of codes. He is a steering board member of the European Technology Platform for HPC (ETP4HPC). Co-Principal Investigator and Work Package leader in the Materials design at Exascale Center of Excellence (MaX CoE). Principal Investigator of the CINECA Intel Parallel Computing Centre. He his author and co-author of 50+ peer review articles,including Science, Physical Review Letters, Nature Materials, and many others. More than 5 PhD sponsored by Cineca, about 5 bachelor, more than 10 stages, more than 5 internship, and 1 master project supervised.

Luca Benini is full professor at the University of Bologna and he is the chair of Digital Circuits and Systems at ETHZ. He received the Ph.D degree in Electronic Engineering from Stanford University, USA, in 1997. He has served as chief architect for the Platform2012/STHORM project in STmicroelectronics, Grenoble in the period 2009-2013. He has held visiting and consulting researcher positions at École Polytechnique Fédérale de Lausanne (EPFL, Switzerland), Interuniversity MicroElectronics Center (IMEC, Belgium), Hewlett- Packard Laboratories (CA, USA), and Stanford University. Dr. Benini's research interests are in energy efficient system design and multi-core SoC design. He is also active in the area of energy efficient smart sensors and sensor networks for biomedical and ambient intelligence applications. He has published more than 700 papers in peer-reviewed international journals and conferences, four books and several book chapters. He is a fellow of the IEEE from 2007 and a member of the Academia Europaea.