An Analytical Framework for Estimating TCO and Exploring Data Center Design Space

Damien Hardy
Marios Kleanthous
Isidoros Sideris
University of Cyprus

Ali G. Saidi
Emre Ozer
ARM

Yiannakis Sazeides
University of Cyprus

***Abstract*—In this paper, we present EETCO: an estimation and exploration tool that can be used to assess data center design decisions on Total-Cost-of-Ownership (TCO) and environmental impact. The tool can capture the implications of many parameters including server performance, power, cost, age, depreciation and Mean-Time-To-Failure (MTTF). The paper describes the tool model and its implementation, and presents experiments that explore tradeoffs offered by different server configurations, performance variability, MTTF, 2D vs 3D-stacked processors, and ambient temperature. These experiments reveal, for the data center configurations used in this study, several opportunities for profit and optimization in the datacenter ecosystem: (i) servers with different computing performance and power consumption merit exploration to minimize TCO and the environmental impact, (ii) performance variability is desirable if it comes with a drastic cost reduction, (iii) shorter processor MTTF is beneficial if it comes with a moderate processor cost reduction, (iv) increasing by few degrees the ambient datacenter temperature reduces the environmental impact with a minor increase in the TCO and (v) a higher cost for a 3D-stacked processor with shorter MTTF can be preferred, over a conventional 2D processor, if it offers a moderate performance increase.**

I. Introduction

During the last few years, datacenters have increased in numbers, size and uses [1]. In an effort to reduce costs and meet specific needs several configurations have come to market including micro-servers for I/O intensive workloads [2], [3] and blade-servers for space and power constrained environments. With these different systems comes a set of design decisions which effect the total cost of ownership. Consequently, to deliver a cost-efficient datacenter, designers should be aware of how different decisions affect the Total-Cost-of-Ownership (TCO) of a datacenter. Several TCO models have been proposed for guiding datacenters design [4], [5], [6], [7], [8] that mainly depend on the following five factors:

1) *Datacenter Infrastructure Cost*: the cost of acquisition of the datacenter building (real estate and development of building) and the power distribution and cooling equipment acquisition cost. The cost of the infrastructure is amortized over 10-20 years.

2) *Server Cost Expenses*: the cost of acquiring the servers, which depreciates within 3-4 years.

3) *Networking Equipment Cost Expenses*: the cost of acquiring the networking equipment, which depreciates within 4-5 years.

4) *Datacenter Operating Expenses*: the cost of electricity for servers, networking equipment and cooling.

5) *Maintenance and Staff Expenses*: the cost for repairs and the salaries of the personnel.

While the goal of datacenter designers is to minimize the TCO, another major concern is the energy consumption and the resulting environmental impact of such IT infrastructures. The CO_2 footprint is directly linked to the energy consumption, which corresponds to a substantial fraction of the TCO.

Research and commercial efforts are underway to reduce the energy consumption by choosing low-power based servers [9], [2], by reducing the server idle consumption [10] or by reducing the cooling power [11], which represents a significant part of the Power Usage Effectiveness (PUE) [12]. Also, an attempt is observed to reduce datacenters energy by optimizing their utilization with virtualization or more efficient co-location [13], [14], [15].

These trends render essential tools to assess the benefits and drawbacks of datacenter design choices on the TCO and the environmental impact. Only few tools, to the best of our knowledge, are publicly available to calculate TCO. APC [16] provides an online estimator tool while [7], [17] provide spreadsheets to estimate the TCO. Both tools do not allow easy exploration and fine grain design choices. Nevertheless, these tools outline the basic parameters and the framework that our tool is based on.

Other studies [18], [19], [20], [21], [22], [23] have developed their in-house model to assess the impact of their design solution on the TCO. Companies, like Facebook or Google, it is virtually certain they have their own models but they are unlikely to release their tools. Our publicly available tool[1] can offer a common framework for future research in this area and it can be combined with datacenter simulation tools [24] to enable more accurate exploration of datacenter design choices.

In this paper, we present EETCO: an estimation and exploration tool for assessing the implications of datacenter design decisions on TCO and the environment. This tool enables the exploration of the implications of several data center parameters including server performance, power, cost, age and mean-time-to-failure (MTTF).

The tool takes as inputs coarse and fine grain data center design parameters like PUE, racks organization, components cost, power consumption and MTTF, and produces outputs related to the organization and operation of a datacenter. The tool contains a kernel estimation component that is used by wrappers to explore design decision tradeoffs on TCO, which can reveal opportunities and challenges for the different parts of the datacenter ecosystem (hardware manufacturer, hardware vendor, datacenter designer).

[1] EETCO: http://www.cs.ucy.ac.cy/carch/xi/eetco.php

In the experimental section of the paper, wrappers are defined to explore high-performance vs. low-power based servers as well as the implications of performance variability, varying MTTF, changing ambient temperature and 2D vs 3D-stacked processors. These experiments reveal the conditions under which servers with different computing performance, power, cost and MTTF provide opportunity to reduce either or both the TCO and the CO_2 footprint.

The remainder of the paper is organized as follows. Section II overviews the proposed framework, while model details are given in Section III. The validation and experimental results are given in Section IV. Finally, Section V concludes the paper and gives directions for future work.

II. Framework overview

Our tool is built in two parts as illustrated in Figure 1. The first one is the kernel of the tool, which takes as inputs a datacenter configuration (land/building acquisition cost, cooling equipment cost per Watt) and configurations for different types of server modules (rack configurations, DRAM, processor and other components cost/power/MTTF).

The kernel produces the TCO and environmental impact estimation and other outputs related with the organization and operation of a datacenter (the five main factors of the TCO for the whole datacenter and per resource, the datacenter area, the number of racks, the total power consumption, for a rack the per resource total power consumption).

The second part, illustrated by the *exploration wrapper*, corresponds to a specific wrapper, which generates datacenter and server modules configurations for design space exploration, maintains the kernel's results for each configuration evaluated and returns the overall results. Different wrappers can be defined according to trade offs the user wants to explore. For instance, in the experimental results section, wrappers are defined to compare high-performance vs. low-power based servers and to investigate the effects of changing ambient temperature.

One wrapper, we would like to highlight, has the ability to produce what should be the value for a given input parameter, such as MTTF, while sweeping through a range of values for another input parameter, such as performance, to maintain constant a given output parameter, such as TCO. This wrapper helps produce a curve that divides a two dimensional exploration space into a region where the output parameter increases and another where it decreases, as compared to a reference design point.

An overview of the kernel framework is shown in Figure 2. For each different server configuration type (compute nodes, database nodes, storage nodes), the estimation starts with spares estimation that determines (i) the number of hot spares required to mitigate performance variability and ensure meeting performance requirement for the peak workload, and (ii) the number of cold spares needed due to server failures. The number of active servers, initial number of servers estimated assuming no variability plus the hot spares, will determine the costs for datacenter infrastructure, server acquisition, networking equipment, and power. The cold spares are used to determine the maintenance cost. These costs are then summed together to produce the contribution to the TCO of a given server type. The global TCO is the sum of the contribution from all server types.

III. TCO estimation

As shown in the previous section, the TCO estimation is the sum of the datacenter infrastructure cost ($C_{infrastructure}$), the server acquisition cost (C_{server}), the networking equipment cost ($C_{network}$), the power cost (C_{power}) and the maintenance cost ($C_{maintenance}$).

$$TCO = C_{infrastructure} + C_{server} + C_{network} + C_{power} + C_{maintenance} \quad (1)$$

In the above formula, the first line represents the capital expenses (CAPEX) and the second represents the operational expenses (OPEX).

In this section, we present the model used to determine these different factors. The list of input parameters and output results is shown in Tables I and II according to the following notation:

- N denotes **NUMBER** (e.g. number of required server modules, number of spares etc)
- C denotes **COST** (e.g. server module cost, electricity cost etc)
- A denotes **AREA** (e.g. datacenter area, cooling equipment area, etc)
- K denotes a **RATIO** (e.g. server modules per rack etc)
- P denotes **POWER** (e.g. total server power etc)
- D denotes **DEPRECIATION** (e.g. server, data center)

The resulting TCO with multiple server configurations can be easily determined under the assumption that, the contribution of each server configuration i is additive:

$$TCO = \sum_i [C_{infrastructure_i} + C_{server_i} + C_{network_i} + C_{power_i} + C_{maintenance_i}] \quad (2)$$

Without loss of generality and for ease of reading, a single server configuration is assumed in the following formulas.

In the next subsections, the different computation steps of the model estimation are described according to the flow in Figure 2, starting with spares estimation followed by the various cost estimations.

A. Hot and cold spares estimation

The distinction between hot and cold spares nodes is necessary, since the hot spares have to be accounted in the power consumption, the area, the cooling and power distribution requirements, whereas, the cold spares are only accounted in the maintenance cost.

1) Hot spares estimation: Various technological, operational and environmental conditions [25] can lead to processor performance variations. That is, in a population of processors, some of them are expected to be affected by a medium/high performance degradation while others will not be affected at all. This performance variation determines the need for hot spares to compensate the performance degradation. For instance, if the expected performance is at 90% of the maximum and the workload requirements are 10000x throughput (e.g. 10000 cores running separate threads), then we will need (10000/0.9 - 10000) 1111 extra cores to meet our requirements, which translates to extra server costs for acquisition, maintenance, power consumption and space.

To consider the performance variation, a *variability factor* (VF) is introduced. VF takes values from 0 to 1, with 0 meaning no degradation at all and 1 means no operation. The performance is thus given by $1 - VF$. With this factor, the

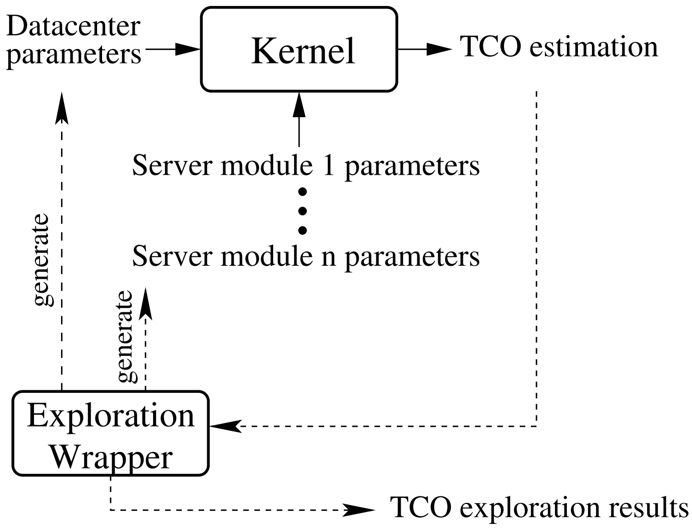

Fig. 1. EETCO tool structure

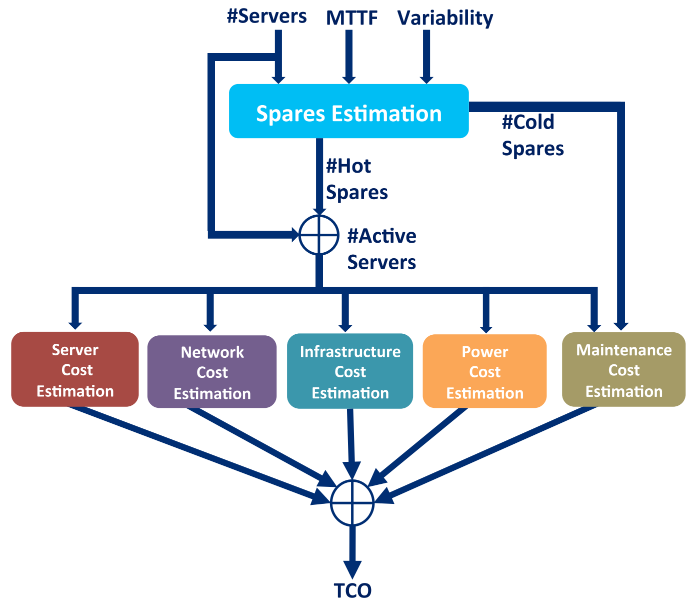

Fig. 2. Kernel framework overview

number of hot spares is determined as follows:

$$N_{hotspares} = \frac{N_{srvmodulesreq}}{1 - VF} - N_{srvmodulesreq}$$

where $N_{srvmodulesreq}$ is the number of server modules required for the peak workload when VF=0. The tool considers the performance variation of the hotspares and iteratively calculates their contribution to the extra servers required. For the rest of the paper, $N_{srvmodules}$ represents the number of active servers that is equal to the sum of $N_{hotspares}$ and $N_{srvmodulesreq}$ modules.

2) Cold spares estimation: Cold spares are server modules needed for replacement when active servers failed. The fault rate of a server can be determined by the MTTF of its components and the Mean Time To Repair (MTTR). By assuming a constant fault rate, an exponential distribution can be used to determine the number of cold spares required at a given time t as follows:

$$N_{coldspares} = N_{srvmodules} * \frac{t}{(MTTF_{server} + \frac{MTTR}{(24*365)})}$$

where $MTTF_{server}$ is the server's MTTF in years, and $MTTR$ is the mean time to repair a server module in hours.

If we have all the cold spares timely available, then the expected fraction of the servers (or Available Throughput) that are available at any given time will be:

$$AvailableThroughput = 1 - \frac{1}{1 + \frac{MTTF_{server}*24*365}{MTTR}}$$

B. Cost and environmental impact estimation

The different costs are simply derived from the number of server modules and number of cold spares as explained next.

1) Maintenance Cost: The maintenance cost per month is determined as follows:

$$C_{maintenance} = \frac{N_{coldspares} * C_{srvmodule}}{D_{srv} * 12} + N_{racks} * C_{salaryperrackpermonth} \quad (3)$$

where $C_{srvmodule}$ is the cost of one server module, $C_{salaryperrackpermonth}$ is the salary cost of datacenter staff per rack per month, $N_{coldspares}$ is calculated for t = D_{srv}, which is the server's depreciation in years, and N_{racks} is the number of racks determined as follows:

$$N_{racks} = \lceil \frac{N_{srvmodules}}{K_{modulesperrack}} \rceil$$

where $K_{modulesperrack}$ is the number of server modules per rack.

2) Networking Cost: The networking acquisition cost per month is determined as follows:

$$C_{network} = \frac{N_{racks} * C_{networkperrack}}{D_{network} * 12}$$

where $D_{network}$ is the networking equipment depreciation in years and $C_{networkperrack}$ is the networking equipment cost per rack. This cost account for the networking gear at the edge, aggregation, and core layers of the datacenter and assume that the cost scales linear with the number of racks.

3) Server Cost: The server acquisition cost per month is determined as follows:

$$C_{server} = \frac{(N_{srvmodules} + N_{agereplace}) * C_{srvmodule}}{D_{dc} * 12}$$

This equation gives the cost of buying the initial servers and the servers that get replaced, $N_{agereplace}$, during the data center's life because their age has exceeded the D_{srv}. More details for the server components, that compose the $C_{srvmodule}$, are given in Table I.

To calculate the $N_{agereplace}$ accurately we need to compute the age distribution at any given time in our datacenter and consider the servers with age greater than D_{srv} such they are replaced.

To compute the age distribution, and $N_{agereplace}$, the algorithm in Figure 3 is used that considers the total failures as a function of age up tp D_{dc} years.

This algorithm gives a breakdown of server's age at granularity of τ time length. The smaller the τ the more accurately the server replacements will be calculated. The tool also produces a breakdown of the age distributions of the servers based on τ and D_{srv} up to the point defined by the user. This tool feature can be used for calculating the age

TABLE I. INPUT PARAMETERS

Name	Description
$C_{buildingpersqm}$	cost of land acquisition/building deployment per square meter
$C_{cooling\&power\_eqperW}$	cost of cooling and power distribution infrastructure per Watt
$C_{elecperKWh}$	electricity cost per KWh
$K_{cooling\&powerarea}$	factor accounting for more space for the cooling and power distribution equipment
$C_{salaryperrackpermonth}$	the salary cost of datacenter staff per rack per month
D_{dc}	datacenter depreciation in years
PUE	power usage effectiveness of the datacenter
$A_{perrack}$	area of one rack in square meter
$K_{modulesperrack}$	number of server modules per rack
u_{srv}	average utilization of servers
$N_{srvmodulesreq}$	number of server modules required for the peak workload
$C_{srvmodule}$	cost of one server module determined with its components (processors, DRAM, disks, board, fans, power supply)
D_{srv}	server depreciation in years
$MTTF_{component_i}$	mean time to failure of a server module component in years
$MTTR$	mean time to repair a server module in hours
τ	time step per year for the age distribution computation
age_group	days to group by the age distribution output for better clarity
VF	variability factor
P_{srv_peak}	peak power consumed by a server determined with its components (processors, DRAM, disks, board)
P_{srv_idle}	power idle consumption of a server determined with its components (processors, DRAM, disks, board)
SPUE	Server Power Usage Effectiveness
$C_{networkperrack}$	cost of networking equipment per rack
$P_{networkperrack\_peak}$	peak power consumed by the networking equipment per rack
$P_{networkperrack\_idle}$	idle power consumed by the networking equipment per rack
$D_{network}$	networking equipment depreciation in years
$u_{network}$	average utilization of networking equipment
K_{loan}	interest rate of a loan

TABLE II. OUTPUT RESULTS

Name	Description
TCO	Total Cost of Ownership
Env_Impact	Environmental impact in kg CO_2 per year
$C_{infrastructure}$	datacenter infrastructure cost
$C_{building}$	land/building acquisition cost
$C_{cooling\&power\_equipment}$	cooling and power distribution equipment cost
N_{racks}	number of racks
C_{server}	server acquisition cost
$N_{srvmodules}$	number of active servers ($N_{srvmodulesreq} + N_{hotspares}$)
$N_{hotspares}$	number of hot spares
$C_{network}$	network cost
C_{power}	power cost
P_{total_peak}	total peak power consumption
$P_{total\_peak\_perrack}$	total peak power consumption per rack
P_{total_avg}	total average power consumption
$C_{maintenance}$	maintenance cost
$N_{coldspares}$	number of cold spares
$MTTF_{allunits}$	mean time to failure of a server module
$AvailableThroughput$	expected fraction of available servers in the datacenter

distribution of the servers at any time by varying τ and D_{dc}.

4) Power Cost: The power cost per month is determined as follows:

$$C_{power} = PUE * \frac{C_{elecperKWh} * 30 * 24}{1000} * (SPUE * P_{total\_srv} + P_{total\_network}) \quad (4)$$

where PUE is the power usage effectiveness of the datacenter (the ratio of total power of the datacenter to the IT power), $SPUE$ [12] is the Server Power Usage Effectiveness (The ratio of total power of a server to the power of pure electronic components) and $C_{elecperKWh}$ is the electricity cost per KWh. P_{total_srv} is the total power consumption of all the active servers considered in the power cost estimation. Depending on how the service provider is charged for the energy they consumed [26]: the peak power consumption or the actual consumption, the peak power ($P_{total\_srv\_peak}$) or the average power ($P_{total\_srv\_avg}$) has to be used.

$$P_{total\_srv\_peak} = N_{srvmodules} * P_{srv\_peak}$$

$$P_{total\_srv\_avg} = N_{srvmodules} * (u_{srv} * P_{srv\_peak} + (1 - u_{srv}) * P_{srv\_idle}) \quad (5)$$

where P_{srv_peak} is the peak power consumed by a server, P_{srv_idle} is the power idle consumption of a server and u_{srv} is the average server utilization. An interesting direction for future work is extending the tool to capture dynamic load behavior.

Finally, $P_{total\_network}$ is the total power consump-

```
# All servers are new at the beginning
ServersOfAge[0] = All servers;
TotalServerReplacements = 0;
failureRate = τ/serverMTTF;

for (timeStep = 0; timeStep <= Ddc; timeStep += τ)
{
  # Save new servers of previous step
  # in order to move them to age=1 later
  NewServers = ServersOfAge[0];

  # All servers with age=Dsrv are replaced due
  # to depreciation and they are now new (age=0)
  ServersOfAge[0] = ServersOfAge[Dsrv];
  ServersOfAge[Dsrv] = 0;

  # All depreciated servers are added. This
  # variable will define the extra cost from
  # server replacements due to depreciation
  TotalServerDepreciated += ServersOfAge[0];

  for (age = (Dsrv - 1);  age > 1; age--)
  {
    # Calculate failures of current server age
    failures = ServersOfAge[age] * failureRate;

    # Failed servers are replaced so are
    # becoming new and added to age = 0
    ServersOfAge[0] += failures;

    # Current servers are aged and moved to age+1
    # expect those that failed and were replaced
    ServersOfAge[age+1] = ServersOfAge[age]-failures;
  }

  # Servers of age=1 are the NewServers (age=0)
  # of the previous step (excluding failures)
  failures = NewServers * failureRate;
  ServersOfAge[0] += failures;
  ServersOfAge[1] = NewServers-failures;
}
```

Fig. 3. Age distribution and server replacements from depreciation

tion of the networking equipment and can be computed in a similar manner by replacing $N_{srvmodules}$ by N_{racks}, u_{srv} by $u_{network}$, P_{srv_peak} by $P_{networkperrack\_peak}$ and $P_{networkperrack\_idle}$ to obtain $P_{total\_network\_peak}$ and $P_{total\_network\_avg}$.

5) Infrastructure: The datacenter infrastructure cost per month is determined as follows:

$$C_{infrastructure} = \frac{C_{building} + C_{cooling\&power\_equipment}}{D_{dc} * 12}$$

where $C_{building}$ is the land/building acquisition cost, $C_{cooling\&power\_equipment}$ is the cooling and power distribution equipment cost and D_{dc} is the datacenter depreciation in years.

$$C_{building} = A_{perrack} * N_{racks} * K_{cooling\&powerarea} * C_{buildingpersqm} \quad (6)$$

where $A_{perrack}$ is the area of one rack, $K_{cooling\&powerarea}$ is a factor accounting for more space for the cooling and power distribution equipment and $C_{buildingpersqm}$ is the cost of land acquisition/building deployment per square meter.

$$C_{cooling\&power\_equipment} = C_{cooling\&power\_eqperW} * (P_{total\_srv\_peak} + P_{total\_network\_peak}) \quad (7)$$

where $C_{cooling\&power\_eqperW}$ is the cost of cooling and power distribution infrastructure per Watt.

TABLE III. EETCO MODEL VALIDATION

TCO Component	% of TCO in [17]	% EETCO model	Difference
$C_{infrastructure}$	22% ($763,672)	22% ($763,707)	+0.004%
C_{server}	57% ($1,998,097)	57% ($1,998,102)	+0.0003%
$C_{network}$	8% ($294,943)	8% ($295,081)	+0.04%
C_{power}	13% ($474,208)	13% ($473,784)	-0.09%
$C_{maintenance}$	-	-	-

TCO Component	% of TCO in [12]	% EETCO model	Difference
$C_{infrastructure}$	14%	12%	-2%
C_{server}	70%	72%	+2%
$C_{network}$	-	-	-
C_{power}	7%	7%	0%
$C_{maintenance}$	9%	9%	0%

6) Impact of Loan Interest: CAPEX are usually subject to loans based on an interest rate and a constant payment schedule. This cost is determined as follows:

$$\sum_i [\frac{C_i * \frac{K_{loan_i}}{12}}{1 - (1 + \frac{K_{loan_i}}{12})^{(-D_i * 12)}}]$$

where C represents each of the CAPEX (infrastructure, servers and networking equipment) cost over their depreciation period D and K_{loan} is the interest rate.

7) Environmental impact estimation: A conversion factor [27] can be used to translate the actual power consumption into the emission of CO_2 in kg. Thus, the environmental impact per year can be estimated as follows:

$$\frac{P_{total\_avg} * PUE * 24 * 365}{1000} * 0.54522$$

where

$$P_{total\_avg} = P_{total\_srv\_avg} * SPUE + P_{total\_network\_avg}$$

IV. VALIDATION AND CASE STUDIES

In this section we first validate the EETCO model (IV-A), and then we describe the experimental assumptions (IV-B), and use the model to present and analyze the experimental results (IV-C). The results include some case studies that reveal opportunities and challenges for different segments of the datacenter ecosystem.

A. Model validation

The model used in the proposed tool is validated by comparing its TCO breakdown against two previously published TCO breakdowns of large-scale data centers [17], [12]. The comparison is shown in Table III. For both comparisons, we use data center configurations as close as possible to the ones used in the previous studies. Our tool models the infrastucture, server, network, power and maintenance cost while Barosso et. al does not model the network cost and Hamilton does not model the maintenance cost. As such, when comparing our model to theirs we cannot compare with the missing data. The results of these comparisons show that our model produces similar breakdown and, therefore, increases our confidence about its accuracy.

The comparison against [17] using absolute values, shown in Table III, is also very accurate. In [12] the breakdown is only provided as percentage and, therefore, we could not assess the accuracy of the proposed model against absolute values.

B. Experimental setup

The experiments are conducted using two different server configurations named LPO and HPE. LPO represents a Low-Power High-Density server configuration, based on low-power

TABLE IV. HPE 1U BLADE SERVER CONFIGURATION.

Components	**Cost ($)**	**Power (W)**	**Power idle (W)**
2 Processors	2200	190	60
12 GB DRAM	300	6	1.5
2 Disks	360	20	10
Power supply, board and fans	900	43.2	14.3
Total	3760	259.2	85.8

TABLE V. LPO 2U BLADE SERVER CONFIGURATION.

Components	**Cost ($)**	**Power (W)**	**Power idle (W)**
48 Processors	4800	144	24
192 GB DRAM	4800	96	24
24 Disks	4320	240	120
Power supply, board and fans	1380	48	16.8
Total	15300	528	184.8

TABLE VI. COMMON SERVER CONFIGURATION

Parameter	**Value**
u_{srv}	0.2
D_{srv}	3 years
τ	1 day
VF	0

TABLE VII. DATA CENTER CONFIGURATION

Parameter	**Value**
$C_{buildingpersqm}$	3000$/$m^2$
$C_{cooling\&power\_eqperW}$	12.5$/W
$C_{elecperKWh}$	0.07$
$K_{coolingarea}$	1.2
$C_{salaryperrackpermonth}$	200$
D_{dc}	15 years
PUE	1.3 (HPE), 1.2 (LPO)

TABLE VIII. RACK AND NETWORK CONFIGURATIONS

parameter	**value**
Rack $A_{perrack}$ $K_{modulesperrack}$	42U 1.44m^2 (with: 0.6m ; depth 1.2m ; used distance 1.2m) LPO: 252 (21 blades per rack ; 12 servers per 2U blade) HPE: 42 (42 blades per rack ; 1 server per 1U blade)
$C_{networkperrack}$	10K$
$P_{networkperrack\_peak}$	360W
$u_{network}$	1
$D_{network}$	4 years

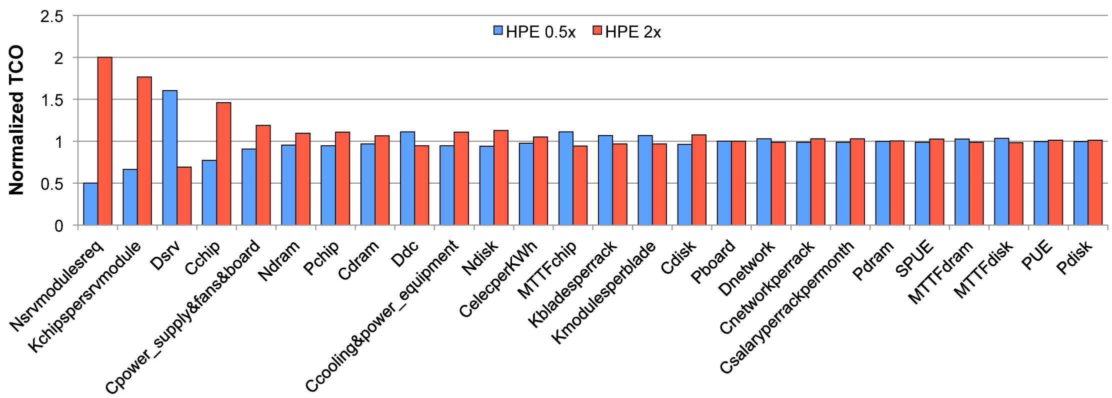

Fig. 4. TCO sensitivity analysis of HPE configuration

processors, like ARM and Atom [2], [3], while HPE represents a High-Performance server configuration based on high-performance Intel Xeon like processors. For the HPE server we consider 12GB DRAM and 2 disks on a dual socket motherboard in a 1U blade [28]. For the LPO server we consider 48 chips split on 12 motherboards (each motherboard with 16GB DRAM and 2 disks) in a 2U blade[29]. The decision of the DRAM capacity was based on the assumption that each core will be allocated 1GB of DRAM (LPO is based on a 4 core processor and HPE is based on a 6 core processor).

Tables IV, V and VI provide the breakdown of the cost and power consumption for both configurations and their common characteristics. For LPO, SPUE is assumed to be 1.1 to take into account the cooling cost reduction of the low-power configuration as compared to HPE (SPUE = 1.2). The power contribution of the power supply and fans for both configuration is directly determined in EETCO with SPUE and presented here for completeness.

For each experiment, unless noted otherwise, 50000 servers are assumed and the peak power consumption (noted peak) and the actual power consumption (noted average) is used to compute the power cost, when it makes a difference. Also we would like to note that the maintenance model assumes that the total blades MTTF is not affected on a failure and replacement of a server module. That means in the case of a server module failure, only that module will need to be replaced and not the whole blade.

Tables VII and VIII summarize the datacenter, the rack and network configurations. At the rack level, the LPO configuration contains 21 2U blade servers while the HPE configuration contains 42 1U blade servers.

We use publicly available data from published papers and industrial data to select representative values for the various parameters: [12], [7], [30], [17], [31], [5], [4] for the datacenter configuration, [32], [12], [17] for the common server configuration, and [29], [3], [33], [34], [10], [35], [18], [36] for the server configurations.

The $MTTF_{allunits}$ is computed assuming 100 years MTTF [35] per disk, 200 years MTTF [34] per 4GB DRAM DIMM. For processors the reported MTTF varies from 30 years [37] to 100 years [35]. We use 30 years for HPE processor and 100 years for LPO processor to account for

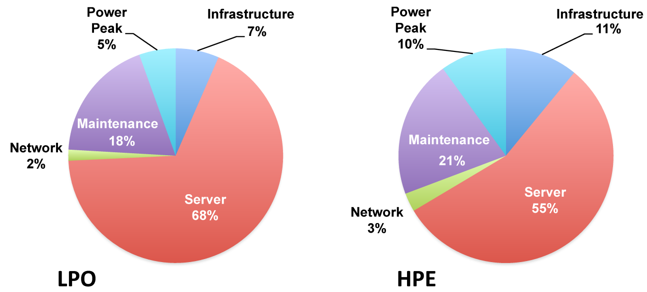

Fig. 5. Breakdown of LPO and HPE configurations

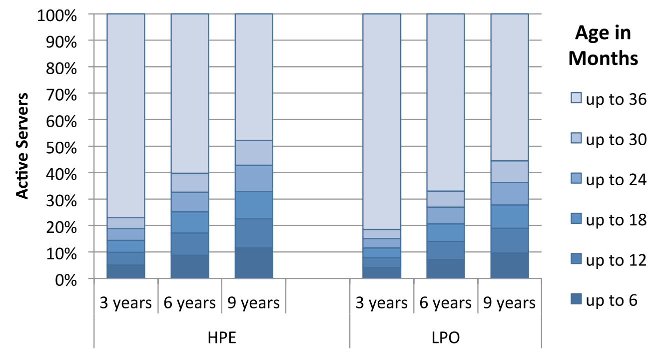

Fig. 6. Server's age distribution for periods of 6 months

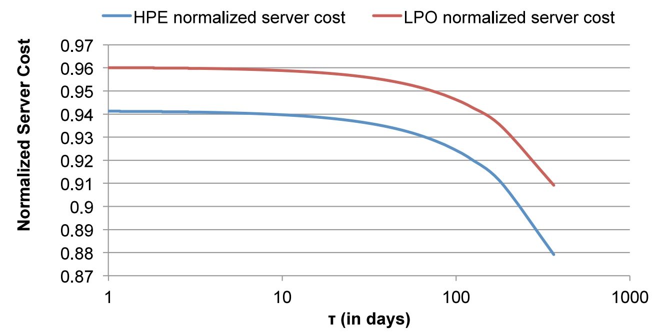

Fig. 7. Considering age for replacing servers after depreciation period

their difference in term of chip size and thermal behavior. The resulting $MTTF_{allunits}$ is 9.836 years and 12.5 years for HPE and LPO respectively.

A sensitivity analysis is performed on the baseline HPE configuration to show how changing the different parameters affects TCO. The baseline values are shown in Tables IV - VIII for the HPE server. In Figure 3, we show the sensitivity of the TCO value by halving (0.5x) and doubling (2x) the baseline value. Only parameters with an impact higher than 1% are shown and the parameters are sorted from high to low sensitivity. The Figure shows parameters that determine the server organization (processors, DRAM, disk) exhibit the largest sensitivity. This is explained by the large contribution of server cost to TCO (more than 50%). The results for the LPO configuration are almost identical (not shown for clarity).

In the next section we present analysis of various comparisons and case studies: the TCO breakdown of the HPE and LPO server configurations, the significance of more accurate age replacements, the impact of performance, power, cost and MTTF, the effect of performance variation, the implications of ambient temperature on the TCO and the environment and an initial analysis of the potential benefits of 3D integration.

C. Experimental results

TCO Breakdown for LPO and HPE: The TCO breakdown of a datacenter populated with LPO and HPE server configurations is shown in Figure 5. The average power cost is normalized with the peak power cost for each server configuration respectively.

As shown in the Figure, the server cost represents the most important part of the TCO, 68% and 55% for each configuration followed by the maintenance cost (18% LPO and 21% HPE). The power cost differs when the peak and the average is assumed. For the peak power consumption (shown in figure), the resulting cost is 5% for LPO and 10% for HPE. For the average power consumption (not shown here) the power cost is about 50% less than the peak power. This difference in power due to the power consumption at idle time, which is more significant for the HPE configuration.

Note that, the direct TCO comparison across the two server configurations is meaningless since the two configurations may have different performance. An exploration that considers the performance impact across configurations is performed subsequently.

Server's age distribution: Publicly available TCO models assume that at the end of the server's depreciation period all machines are replaced because they are considered aged. This is not true for all servers since some of them might have just recently been replaced due to failures. The tool estimates replacements based on an age distribution of the servers at time steps of τ per year. Figure 6 shows a plot of this distribution for τ = 4 days after the age of 3, 6 and 9 years of the datacenter's life. The results suggest that the age of the server's is important to be considered during the replacements at the end of the depreciation period because a significant number of servers have already been replaced due to failures. As the datacenter age progresses the age distribution of the servers spreads more and thus the TCO error that assumes replacements of all servers at the end of the depreciation period gets larger.

EETCO uses this additional information to more accurately calculate the number of servers that will be replaced due to depreciation. Figure 7 shows the Server Cost estimation considering the age of the servers for different time steps τ and normalized to the case where all server's are replaced every 3 years. Results indicate that for large steps, ($\tau > 10$), we might underestimated the cost of the servers because we only sample the age few times per year. As the time step τ gets smaller the server's cost converge's to about 96% for the HPE configuration and 94% for the LPO configuration. These results suggest that estimating the server's cost considering the age distribution will result to 4 - 6% less in the server's cost estimation normalized to the case where all servers are replaced after a constant depreciation period.

Considering a time step smaller than 4 days make a small difference in the results and thus we consider τ = 4 for the rest of the experiments.

Impact of the processor's MTTF: Attempting to improve a processor's MTTF may increase its cost due to the use of more expensive and reliable components. In this experiment, the trade-off between the processor's MTTF and the processor's cost is explored. The selected range for processor's MTTF is 20 to 150 years which examine the trends near the range reported in previous work [37], [35].

Figure 8 shows what should be the processor's cost to keep

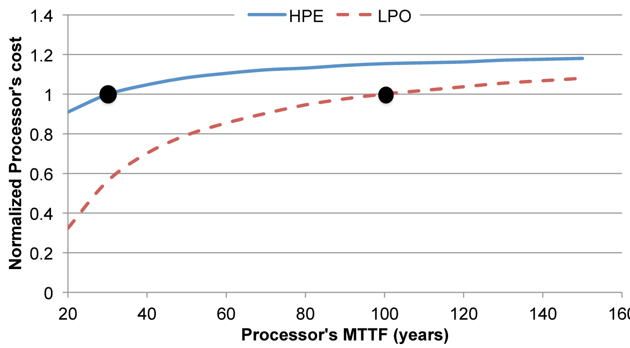

Fig. 8. Impact of processor's MTTF. Results normalized with those obtained with the reference value

the TCO constant when the MTTF varies relative to a reference value (30 and 100 years for HPE and LPO respectively, shown with the black dots in the figure).

As shown in the figure, for the HPE configuration, an increase (up to 2x) in terms of MTTF budget may be interesting. For 2x MTTF, a price increase near 20% is affordable, while above this region the price cost stays nearly constant. The LPO can benefit by decreasing the MTTF, but the processor cost reduction has to be significant when the MTTF is below $2/3$ (66 years) of the reference value. Smaller changes of MTTF, in both directions, require moderate changes in the cost.

These observations can be useful for: (i) processor manufacturers to assess how the MTTF of processor affects the TCO and to estimate the potential profit for a given design and MTTF budget; (ii) hardware vendors to increase their margin by selecting the appropriate processor; (iii) datacenter designers to reduce the TCO when they have the choice between processors with equivalent performance but different prices and MTTF, and to define their maintenance model.

Different computing performance between servers: The TCO breakdown is not sufficient to compare the two server configurations since they may not have the same computing performance. Let us assume that LPO configuration will require more servers to reach the same computing performance as HPE. We use an equivalent performance coefficient (epc), defined to be how many LPO server modules are required to reach the computing performance of one HPE server module. We vary epc from 1 to 6, which is a representative range across servers with different processors for cloud applications derived from [38], to observe the trends. Results are presented in Figure 9 and the values are normalized with the TCO and the environmental impact obtained with HPE.

As shown in Figure 9, when epc is relatively small, the TCO obtained with the low-power configuration (LPO) is better. At a given point ($epc \sim 3.5$ in our case for 1GB per core of DRAM) the TCO of both configurations is equal. Nevertheless, in that case the resulting environmental impact is lower with LPO. LPO is, thus, preferable for the environment for equivalent TCO. After that point, HPE is a better choice for both the TCO and the environmental impact. The results also indicate that when the total datacenter's DRAM is kept equal for both configuration the benefits of the LPO server is higher. Awareness of such trends can be useful: (i) for processor manufacturers to design processors that can trade-off between performance and cost and (ii) for datacenter designers to optimize for both the TCO and CO_2.

Note that considering QoS issues is beyond the scope of this analysis.

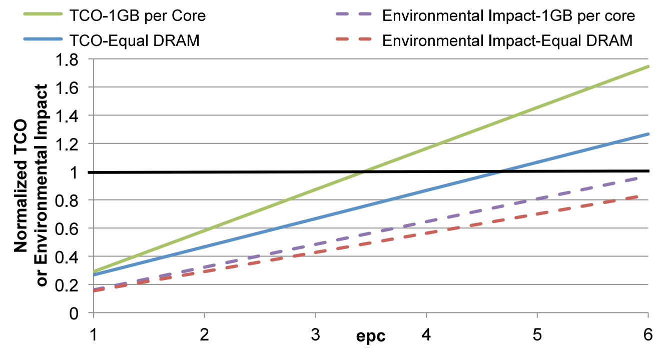

Fig. 9. Different computing performance between LPO and HPE. LPO results normalized with HPE results. $N_{srvmodulesreq}(LPO) = epc * N_{srvmodulesreq}(HPE)$

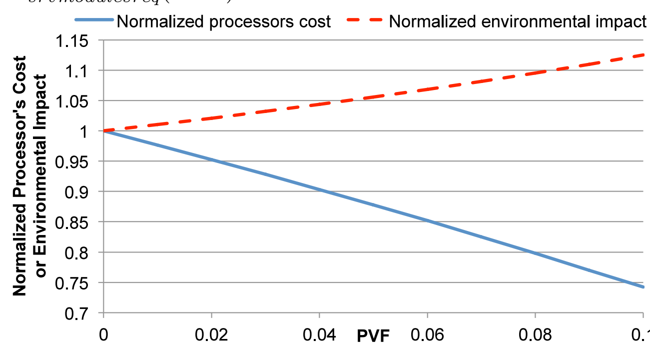

Fig. 10. Impact of performance variation for the HPE configuration. Results are normalized with those obtained when VF=0

Impact of performance variation: As mentioned in the previous section, there are various sources of processor performance variability. This variation may affect the processor's cost in addition to performance [39] (i.e. the higher the variation, the lower the processor's cost). In this experiment, a variability factor (VF) is assumed to range from 0 to 0.1 while power remains unchanged. The results, illustrated in Figure 10, show what should be the processor's cost to keep the TCO constant. The figure also show the environmental impact of performance variability.

As shown in the figure, if the processor's cost reduction is higher than the reduction needed to keep the TCO constant (i.e. below the iso curve), there is an opportunity to reduce the TCO. This positive impact of performance variability comes at the price of an environmental impact increase. In fact, the higher the performance variability, the higher the number of active servers needed, which results inevitably in a higher energy consumption and thus higher CO_2 emissions.

This data presents:

(i) For processor manufacturers: an opportunity to sell (or even design) processor with performance variability instead of throwing away such processors. A key challenge is the design of processors with performance guarantees and less power consumption

(ii) For hardware vendors: a challenge to define business models to deal with performance variability

(iii) For datacenter designers: an opportunity to reduce the TCO.

Impact of ambient temperature: This experiment addresses the effect of ambient temperature (assumed 20°C) on the TCO and the CO_2 emissions. An increase in the ambient temperature from 20°C to 30°C has a positive impact on

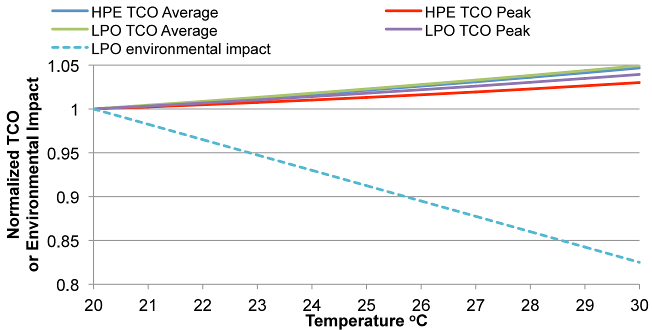

Fig. 11. Impact of temperature. Results are normalized with those obtained when T=20°C

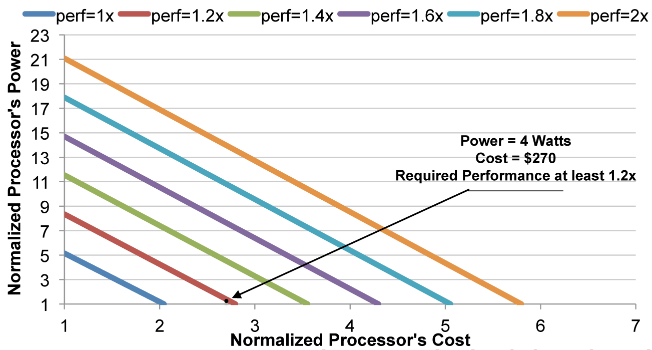

Fig. 12. 2D vs. 3D processor. Results are normalized with those obtained with LPO configuration

the cooling power consumption (in a previous study the PUE scales from 2 to 1.65 [40]) while the MTTF is reduced [40], [41]. In this experiment, we assess this positive and negative impact by assuming a linear reduction of the PUE per degree and a constant server's power consumption. We also use the same values for PUE and ambient temperature as in [40].

The effect of the ambient temperature on the MTTF is modeled using the Arrhenius equation that predicts the acceleration factor (AF) due to the temperature:

$$AF = e^{\frac{Ea}{k}(\frac{1}{T_r} - \frac{1}{T_a})}$$

where Ea is the activation energy in electron-volts (0.3 in our case), k is the Boltzmanns constant (8.617E-05), T_r and T_a are the reference temperature (20°C + 273) and the actual temperature, in degrees Kelvin.

With AF, the $MTTF$ resulting from the actual operating temperature can be determined as follows:

$$MTTF = \frac{MTTF_{ref}}{AF}$$

where $MTTF_{ref}$ is the $MTTF$ at the reference temperature.

As shown in Figure 11, the CO_2 emissions is significantly reduced while we can observe a small TCO increase (HPE and LPO TCO average lines overlap). Consequently, increasing by few degrees the ambient datacenter temperature appears to be a good trade-off to reduce the environmental impact without increasing significantly the TCO.

Comparison between 2D and 3D processors: To overcome the memory wall, 3D-stacking architectures have received significant attention by the architecture community [42], [43], [9], [44]. One proposition is to improve performance by stacking multiple DRAM layers on top of a logic layer. This approach provides higher performance as compared to 2D processors but with the trade-off of (i) higher processor cost and processor power consumption, (ii) chip temperature increase and (iii) probably a lower MTTF due to the stacking of multiple layers.

In this experiment, we try to assess the overall benefits of 3D-stacked chips as compared to 2D processors. To the best of our knowledge, this is the first time such a comparison is performed with the datacenter TCO perspective in mind. The basic idea behind a 3D chip design is that the increased performance and reduced overall server power due to the 3D-integrated DRAM will cover the extra cost of stacking 3D chips and possible reduction in the MTTF.

For the 3D server configuration we use the LPO configuration as baseline with the difference that the 4GB off-chip DRAM per chip is now integrated with the 3D chip. In Figure 12 we attempt to project what should be the performance increase for tolerating the cost and power increase to keep the TCO constant equal to the LPO datacenter configuration.

The 3D chip cost increase is due to the 3D stacking process and the 3D-integrated DRAM and the power increase is due to the additional power of the 3D-integrated DRAM, assuming that the off-chip DRAM interface is still maintained. For example, assuming that a 3D chip cost will be at least the cost of the LPO chip ($100) + the cost of the DRAM ($100) + a cost for 3D stacking, testing and packing, extra provisions for MTTF and possible additional cooling solutions (35% increase) that equals to a minimum price of $270.

On the other hand, the overall server power decreases because, for the same capacity of DRAM per server, the on-chip DRAM has lower power as compared to the off-chip DRAM for the same capacity. We assume that the 3D-integrated DRAM has 1 Watt power as compared to the off-chip DRAM which has 2 Watts power. That makes the total power of the 3D chip equal to the chip power (3 Watts) + the 3D-integrated DRAM (1 Watt) = 4 Watts.

As shown in Figure 12, the performance increase should be at least 1.2X to have enough room (below the curve) to support the cost and power consumption increases due to 3D stacking and to improve the Performance/TCO. Also, Figure 12 reveals that the cost can be increased up to 200% when the power stays the same and the power increase up to 500% when the server cost stays constant and with the same performance.

This initial comparison of 2D and 3D processors, from a datacenter TCO perspective, shows interesting trends that motivates examining the trade-offs between performance, cost, power and MTTF for profitable 3D processor deployment in servers for datacenters. This experiment merits to be explored in more detail with more precise models for MTTF, thermal, power consumption and 3D processors cost and performance which is part of our ongoing work.

V. Conclusion

In this paper, we have presented EETCO: an estimation and exploration tool that can be used to assess datacenter design decisions on TCO and the environment. The tool considers many of the key datacenter parameters and is shown to be quite accurate when compared with previous published TCO breakdowns. Different case studies have been performed to assess tradeoffs between server configurations, age, performance variability, datacenter ambient temperature, and 3D processor integration.

This paper reveals opportunities and challenges for how to tune and optimize the datacenter design.

The plans for future extensions to the TCO tool are:

- a model to take into account the contribution of the networking equipment to the TCO

- a model for the interest rate a business must pay on loans
- heterogeneous processor types
- different hardware maintenance models
- a model for the virtual machine, software and the software maintenance contributions to the TCO
- a model at the service level based on different kind of server configurations and utilization
- validation of our model with data coming from available information on datacenters
- federated data centers, consider TCO trade-offs of using different number of facilities and locations
- combine EETCO tool with a datacenter load simulation tool.

Acknowledgment

The research leading to this paper is supported by the European Commission FP7 project ”Energy-conscious 3D Server-on-Chip for Green Cloud Services (Project No:247779 ”EuroCloud”)”.

References

[1] Cisco, “Cisco global cloud index: Forecast and methodology, 20112016,” 2012.

[2] Calxeda, “http://www.calxeda.com.”

[3] seamicro, “http://www.seamicro.com/.”

[4] C. Patel and A. Shah, “Cost model for planning, development and operation of a data center,” HP Laboratories Palo Alto, Tech. Rep., 2005.

[5] J. Karidis, J. E. Moreira, and J. Moreno, “True value: assessing and optimizing the cost of computing at the data center level,” in *Proceedings of the 6th ACM conference on Computing frontiers*, ser. CF ’09. New York, NY, USA: ACM, 2009, pp. 185–192.

[6] J. Moore, J. Chase, P. Ranganathan, and R. Sharma, “Making scheduling ”cool”: temperature-aware workload placement in data centers,” in *Proceedings of the annual conference on USENIX Annual Technical Conference*, ser. ATEC ’05. Berkeley, CA, USA: USENIX Association, 2005, pp. 5–5.

[7] J. Koomey, K. Brill, P. Turner, J. Stanley, and B. Taylor, “A simple model for determining true total cost of ownership for data centers,” white paper, Uptime Institute, 2007.

[8] K. V. Vishwanath, A. Greenberg, and D. A. Reed, “Modular data centers: how to design them?” in *Proceedings of the 1st ACM workshop on Large-Scale system and application performance*, ser. LSAP ’09. New York, NY, USA: ACM, 2009, pp. 3–10.

[9] E. Ozer and et al., “Eurocloud: Energy-conscious 3d server-on-chip for green cloud services,” in *2nd Workshop on Architectural Concerns in Large Datacenters*, 2010.

[10] D. Meisner, B. T. Gold, and T. F. Wenisch, “Powernap: eliminating server idle power,” in *Proceeding of the 14th international conference on Architectural support for programming languages and operating systems*, ser. ASPLOS ’09. New York, NY, USA: ACM, 2009, pp. 205–216.

[11] N. El-Sayed, I. Stefanovici, G. Amvrosiadis, and A. A. Hwang, “Temperature management in data centers: Why some (might) like it hot,” in *Proceedings of SIGMETRICS 2012*, 2012.

[12] L. A. Barroso and U. Holzle, “The datacenter as a computer: An introduction to the design of warehouse-scale machines, morgan and claypool publishers, 2009.”

[13] S. Govindan, J. Liu, A. Kansal, and A. Sivasubramaniam, “Cuanta: Quantifying effects of shared on-chip resource interference for consolidated virtual machines,” in *Proceedings of 2011 ACM Symposium on Cloud Computing*, 2011.

[14] J. Mars, L. Tang, R. Hundt, K. Skadron, and M. L. Soffa, “Bubble-up: Increasing utilization in modern warehouse scale computers via sensible co-locations,” in *Proceedings of the 44th annual IEEE/ACM International Symposium on Microarchitecture*, 2011.

[15] L. Tang, J. Mars, N. Vachharajani, R. Hundt, and M. L. Soffa, “The impact of memory subsystem resource sharing on datacenter applications,” in *Proceedings of the 38th International Symposium on Computer Architecture*, 2011.

[16] APC, “http://www.apc.com/tools/isx/tco/.”

[17] J. Hamilton, “Overall data center costs http://perspectives.mvdirona.com/2010/09/18/ overalldatacenter-costs.aspx.”

[18] K. Lim, P. Ranganathan, J. Chang, C. Patel, T. Mudge, and S. Reinhardt, “Understanding and designing new server architectures for emerging warehouse-computing environments,” in *Proceedings of the 35th Annual International Symposium on Computer Architecture*, ser. ISCA ’08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 315–326.

[19] S. Polfliet, F. Ryckbosch, and L. Eeckhout, “Optimizing the datacenter for data-centric workloads,” in *Proceedings of the international conference on Supercomputing*, ser. ICS ’11. New York, NY, USA: ACM, 2011, pp. 182–191.

[20] S. Li and al, “System-level integrated server architectures for scale-out datacenters.in 44th annual intl. symposium on microarchitecture,” in *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, 2011, Porto Alegre, Brazil.

[21] V. J. Reddi, B. Lee, T. Chilimbi, and K. Vaid, “Web search using mobile cores: Quantifying and mitigating the price of efficiency,” in *Proceedings of the 37th International Symposium on Computer Architecture*, 2010.

[22] Y. Chen and R. Sion, “To cloud or not to cloud? musings on costs and viability,” in *Proceedings of 2011 ACM Symposium on Cloud Computing*, 2011.

[23] K. T. Malladi, F. A. Nothaft, and K. Periyathambi, “Towards energy-proportional datacenter memory with mobile dram,” in *Proceedings of the 39th International Symposium on Computer Architecture*, 2012.

[24] D. Meisner, J. Wu, and T. F. Wenisch, “Bighouse: A simulation infrastructure for data center systems,” in *Proceedings of the 2012 IEEE International Symposium on Performance Analysis of Systems and Software*, 2012.

[25] O. S. Unsal, J. Tschanz, K. A. Bowman, V. De, X. Vera, A. González, and O. Ergin, “Impact of parameter variations on circuits and microarchitecture,” *IEEE Micro*, vol. 26, no. 6, pp. 30–39, 2006.

[26] K. Le, J. Zhang, J. Meng, R. Bianchini, T. D. Nguyen, and Y. Jaluria, “Reducing electricity cost through virtual machine placement in high performance computing clouds,” in *Proceedings of Super Computing (SC11)*, 2011.

[27] Defra, “http://archive.defra.gov.uk/environment/business/ reporting/pdf/101006-guidelines-ghg-conversion-factors.pdf.”

[28] Hewlett-Packard, “Hp proliant bl280c generation 6 (g6) server blade.”

[29] B. Limited, “Boston viridis - arm microservers.”

[30] P. Turner and K. Brill, “Cost model: Dollars per kw plus dollars per square foot of computer floor,” white paper, Uptime Institute.

[31] J. Hamilton, “Internet-scale service efficiency,” in *Large-Scale Distributed Systems and Middleware (LADIS) Workshop*, 2008.

[32] “Barcelona supercomputing center http://www.bsc.es/plantillaa.php?cat_id=202.”

[33] Dell, “http://www.dell.com/.”

[34] Micron, “http://download.micron.com/pdf/technotes/tn0018.pdf.”

[35] P. Kogge, K. Bergman, S. Borkar, D. Campbell, W. Carson, W. Dally, M. Denneau, P. Franzon, W. Harrod, K. Hill *et al.*, “Exascale computing study: Technology challenges in achieving exascale systems,” 2008.

[36] Micron, “http://www.micron.com/products/support/power-calc.”

[37] J. Srinivasan, S. V. Adve, P. Bose, S. V. A. P. Bose, and J. A. Rivers, “The case for lifetime reliability-aware microprocessors,” in *Proceedings of the 31st International Symposium on Computer Architecture*, 2004, pp. 276–287.

[38] P. Lotfi-Kamran, B. Grot, M. Ferdman, S. Volos, O. Kocberber, J. Picorel, A. Adileh, D. Jevdjic, S. Idgunji, E. Ozer, and B. Falsafi, “Scale-out processors,” in *Proceedings of the 39th International Symposium on Computer Architecture*, 2012.

[39] A. Das, B. Ozisikyilmaz, S. Ozdemir, G. Memik, J. Zambreno, and A. Choudhary, “Evaluating the effects of cache redundancy on profit,” in *Proceedings of the 41st annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO 41. Washington, DC, USA: IEEE Computer Society, 2008, pp. 388–398.

[40] M. K. Patterson, “The effect of data center temperature on energy efficiency,” in *11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, 2008, pp. pages 1167–1174.

[41] “Western digital wdc.ph/wdproducts/library/other/2579-001134.pdf.”

[42] G. H. Loh, “3d-stacked memory architectures for multi-core proces-

sors," *SIGARCH Comput. Archit. News*, vol. 36, pp. 453–464, June 2008.

[43] N. Hardavellas, M. Ferdman, B. Falsafi, and A. Ailamaki, "Toward dark silicon in servers," *Micro, IEEE*, vol. 31, no. 4, pp. 6 –15, july-aug. 2011.

[44] D. Milojevic, S. Idgunji, D. Jevdjic, E. Ozer, P. Lotfi-Kamran, A. Panteli, A. Prodromou, C. Nicopoulos, D. Hardy, B. Falsafi, and Y. Sazeides, "Thermal characterization of cloud workloads on a power-efficient server-on-chip," in *Proceedings of 30th International Conference on Computer Design (ICCD-2012)*, 2012.