EUROSERVER: Energy Efficient Node for European Micro-servers

Yves Durand,[1] Paul M. Carpenter,[2] Stefano Adami,[3] Angelos Bilas,[4]
Denis Dutoit,[1] Alexis Farcy,[5] Georgi Gaydadjiev,[6] John Goodacre,[7] Manolis Katevenis,[4]
Manolis Marazakis,[4] Emil Matus,[8] Iakovos Mavroidis,[4] John Thomson[9]

[1]Univ. Grenoble Alpes, CEA–LETI, MINATEC Campus, F-38054 Grenoble, France
[2]Barcelona Supercomputing Center, Barcelona, Spain
[3]Eurotech SPA, Italy
[4]Foundation for Research and Technology Hellas, Greece
[5]STMicroelectronics, France
[6]Chalmers Tekniska Hoegskola AB, Sweden
[7]ARM, Cambridge, United Kingdom
[8]Technische Universitaet Dresden, Germany
[9]OnApp, United Kingdom

***Abstract*—EUROSERVER is a collaborative project that aims to dramatically improve data centre energy-efficiency, cost, and software efficiency. It is addressing these important challenges through the coordinated application of several key recent innovations: 64-bit ARM cores, 3D heterogeneous silicon-on-silicon integration, and fully-depleted silicon-on-insulator (FD-SOI) process technology, together with new software techniques for efficient resource management, including resource sharing and workload isolation. We are pioneering a system architecture approach that allows specialized silicon devices to be built even for low-volume markets where NRE costs are currently prohibitive. The EUROSERVER device will embed multiple silicon "chiplets" on an active silicon interposer. Its system architecture is being driven by requirements from three use cases: data centres and cloud computing, telecom infrastructures, and high-end embedded systems. We will build two fully integrated full-system prototypes, based on a common micro-server board, and targeting embedded servers and enterprise servers.**

Keywords—system architecture; virtualization; Three-dimensional (3D) integrated circuits; FD-SOI; ARM

I. INTRODUCTION

Data centres are crucial to modern society. There is a constant demand for improvements in online services, which means that the data centres that provide these services must continue to advance, both in capacity and throughput. Such improvements are, however, limited by the technical and financial characteristics of current IT equipment, including their thermal-limited density, high capital cost, and excessive energy consumption. As the interest of the computing industry moves from a tight focus on performance towards energy-efficiency and total cost of ownership (TCO), the basic components of future servers and their integration into a full system must be reconsidered from the ground up. The processor, memory hierarchy, I/O, system interconnects and system software all require fundamental changes to meet expectations for the application's performance, in less space and at drastically lower cost, while still maintaining support for existing applications and software frameworks.

EUROSERVER [5] is a collaborative project that is addressing all these challenges, through the coordinated application of several key recent innovations. Firstly the availability of **64-bit ARM cores** means that the world's leading energy-efficient processors are now fully capable of running data centre workloads. Over the last few years, there has been a growing interest in using ARM processors outside the mobile and embedded spaces, in data centres [4] and high-performance computing [9]. EUROSERVER builds on these previous projects, which, although limited by 32-bit virtual addressing, have done much of the important groundwork in demonstrating the feasibility of the approach and improving the maturity of the ARM software stack. The second key innovation is **3D silicon integration**, where multiple independent bare silicon die called "chiplets" are placed on a silicon interposer that either provides only passive connectivity, or in the case of EUROSERVER, is an active interposer that includes the on-chip interconnect and SoC interfaces. This approach improves manufacturing yield, reduces core-to-core and core-to-memory distances, and greatly reduces the cost of application specialization. The third key innovation is fully depleted silicon-on-insulator technology (**FD-SOI**), which brings market leading absolute performance and performance-per-watt.

The EUROSERVER micro-server system architecture is based on a large number of cores, each of them being much simpler than a conventional x86 processor. This builds on the findings of the **"scale out" architecture** investigated in the Eurocloud Project [4]. In addition, EUROSERVER will work to decrease the memory and I/O power dissipation. The EUROSERVER memory organization will enable resource **mutualization and data sharing**, in order to reduce unnecessary accesses. Furthermore, **I/O virtualization** will be facilitated through hardware and software support. Scaling up the number of cores and mutualizing the infrastructures are both made possible by 3D hardware integration, which enables groups of cores, memory modules, and interconnects to be integrated into a single package at a reasonable cost.

Category	Partners
IP provider	ARM, UK
Silicon vendor	STMicroelectronics, France
System integrator	Eurotech SPA, Italy
Semiconductor technology and project coordinator	Commissariat à l'Energie Atomique et aux énergies alternatives (CEA), France
Academia / Research	Barcelona Supercomputing Center (BSC), Spain; Foundation for Research and Technology Hellas (FORTH), Greece; Chalmers Tekniska Hoegskola AB, Sweden; Technische Universitaet Dresden, Germany
End user	OnApp, UK

Table 1: EUROSERVER project consortium

The final key innovation is a **new software architecture for efficient use of resources**. The system software will efficiently manage shared resources and processors, and dynamically assign workloads to the appropriate group of resources, reducing workload interference and achieving high resource utilization, without compromising performance.

The EUROSERVER system architecture will be driven by requirements gathered for three use cases: (a) data centres and cloud computing, (b) telecom infrastructures, and (c) high-end embedded systems. Typical workloads include web server hosting (LAMP/WAMP[1]), distributed databases (Hadoop), OLAP and OLTP workloads, relational databases (MySQL), network communications, vehicle on-board computers, and automatic vehicle location tracking.

We will build two fully integrated full-system prototypes, targeting embedded and enterprise servers, respectively. The prototypes will use a common micro-server packaged device and board integrating a single EUROSERVER device, containing multiple compute chiplets on an active silicon interposer. We will thereby demonstrate and evaluate the full EUROSERVER architecture, showing how the proposed approach can lead to a factor-of-ten improvement in data centre energy efficiency by 2020.

II. EUROSERVER PROJECT

A. Project objectives

EUROSERVER has the following key objectives:

1. **Reduced energy consumption** through: *(i)* 64-bit ARM cores, which are the world-leading low-power processors; *(ii)* novel silicon interposer packaging technology, which drastically reduces the core-to-memory distance; and *(iii)* improving on the energy proportionality.

[1] Linux/Windows, Apache, MySQL and PHP/Python

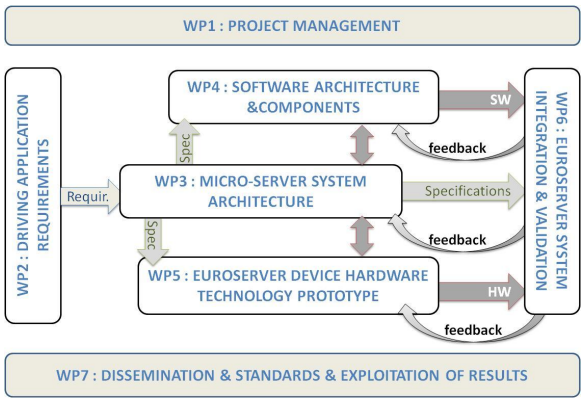

Figure 1: EUROSERVER project organisation

2. **Reduced cost** to build and operate each micro-server, owing to: *(i)* improved manufacturing yield by duplicating multiple small chiplets and placing them on a silicon interposer (3D); *(ii)* reduced physical volume of the packaged interposer module; and *(iii)* an energy efficient semiconductor process (FD-SOI).

3. **Better software efficiency**: Next generation system software that: *(i)* manages the resources in a server consisting of multiple coherence-islands; and *(ii)* isolates and protects the multiple workloads from each other when they use the shared server resources of I/O, storage, memory, and interconnects.

B. Project consortium

The EUROSERVER consortium has nine partners, who bring complementary expertise across the entire micro-server research and development value chain. As shown in Table 1, the project includes industry leaders in energy-efficient computing IP (ARM), chip and semiconductor integration (STMicroelectronics), and system integration at the board and machine levels (Eurotech). It also includes a key end user, OnApp, a leading IaaS cloud platform provider used primarily by the web-hosting industry. In addition, the consortium brings together leading academic and research partners in the areas of supercomputing (BSC), semiconductor integration (CEA), hardware–software integration (FORTH), architecture (Chalmers) and energy-efficient software for distributed and embedded applications (TU Dresden).

C. Project organisation

EUROSERVER is an FP7 collaborative project with a total budget of €12.9M, of which €8.6M is contributed by the European Commission. The project started in September 2013, and will run for three years.

The project follows a standard FP7 project methodology as shown in Figure 1. It is split into initial market analysis and requirements gathering for the three use cases (WP2), as summarized in Section III. These requirements will be translated into the next-generation micro-server system architecture (WP3), described in Section IV. The efficient software stack (WP4) is discussed in Section V. The discrete

component and micro-server packaged-device prototypes (WP5) are discussed in Section VI. The development efforts are then integrated for demonstration and validation (WP6), and the results are exploited and disseminated (WP7).

III. SYSTEM REQUIREMENTS

To ensure that the novel design proposed by EUROSERVER is relevant and will be considered for adoption by industry, an investigation was performed into a set of markets. The markets investigated were determined during the initial inception of the project based on the various domains of interest of the partners involved in the project.

A. Data centre and enterprise

Data centres and enterprise is a large target market for the innovative platform proposed in EUROSERVER, with over 500,000 data centres globally [10]. In 2005 the total data centre power consumption was calculated as 150 billion kilowatt hours per year, which equated to about 1% of the total electricity use for that year [8]. In Western and more economically developed countries, the total power demand as a proportion of the total electricity production was even higher with 1.5% in the U.S.

There are many reasons for this growth, not least of which is the rise of mobile, ubiquitous computing, social media sharing, and various other "Big Data" analytical systems that take an increasing number of available sensors and process the information. The amount of global, digital data is forecast to grow to 40 ZB in 2020 from 1.23 ZB in 2010 [6].

Generation and removal of the heat produced is one of the largest costs in running and operating a data centre. CPUs are the single most energy-demanding component within servers. As power efficiency is sought, the amount of useful computing is also being calculated such that new measures of performance per joule or per watt can be determined. There are not many benchmark systems that work fairly across multiple system architectures that can assist with this measurement.

Requirements for data centres depend on the workloads that are being run. Several popular workload scenarios have been investigated, including enterprise, web hosting, distributed databases, data-centric, high-performance computing (HPC), and scientific workloads.

The generic requirements are low power consumption, workload power efficiency, platform support, compatibility with existing software, same or lower cost of hardware, interfacing with existing hardware, dense processing, ability to mount in a rack, network stack support, reliability, error detection/correction, and heat tolerance.

B. Telecommunication applications

In telecommunications, the emphasis is on radio access networks (RAN) in mobile communications. The telecommunications market has traditionally had a number of purpose-built hardware systems that are customised for high performance and high reliability. Hardware is generally owned by a single operator, with the cables between points either owned or leased. Recently, new providers as well as those that are established are looking into using commodity hardware with virtualisation that can either be dedicated or eventually shared with other tenants to reduce costs and improve the manageability and adaptability of the infrastructure.

The Cloud-RAN (CRAN) [3] concept makes a paradigm shift from cell-centric wireless communications design to user-centric wireless communications design, i.e., by assigning virtual baseband stack algorithms to different clusters of users, which are then dynamically mapped to physical Base Station (BS) hardware resources.

This project will analyse feasibility of a CRAN concept using the EUROSERVER ARM 64-bit micro-server from the points of view of *(i)* integration of several BS on the same hardware, and *(ii)* hosting multiple vendors' algorithms on the same hardware (BS hotel concept).

A mandatory requirement is real-time response through a RTOS, which guarantees the following:

- real-time response requirements of PHY, MAC, RLC, and RRC layers.
- dynamic computing/storage resource allocation and scheduling;
- I/O channel allocation for routing signal to/from the compute cluster, with timing guarantees according to CPRI specification;
- management of the heterogeneity of processing nodes for protocol processing, including baseband signal processing and baseband FEC processing;
- routing of delay-constrained signals between the computing nodes.

Mixed speed and mixed type signals should be supported to allow radio frequency (RF) signals to not interfere with the rest of the components.

C. Embedded systems, through a transport use case

The resulting next-generation embedded server will enable new applications and solutions in many public and commercial vehicles; for rolling stock (trains, metros, trams), buses, public and commercial vehicles (e.g. Fleet Control/Management, Vehicle Control, Passenger services, Accurate Location and Tracking, Security/Surveillance).

Two scenarios were identified as potential candidates for EUROSERVER applications:

S1 - Multi-purpose rugged mobile computer: a compact, rugged, and low power server, highly configurable in terms of I/O, storage, and computation capabilities, designed to meet the transportation certification;

S2 - Multi-service gateway and edge controller: a compact specialised embedded server for in-vehicle and in-field multi-source data gathering and mining, with multi-connectivity wide area network routing capabilities.

The requirements include physical size, operating and storage temperature ranges, power input, remote configuration, I/O interfaces, vibration and dust tolerance,

lifetime, serviceability, and power-up and reset-time requirements.

Mobility presents additional challenges and constraints for the physical robustness of the prototypes. 3D hardware designs with stacked components will need to be able to cope with vibration, six-axis lateral and rotational forces, and ESR and other environmental conditions that are less significant in a fixed environment.

D. Security and reliability

As an important factor, orthogonal to the use cases presented, it was deemed important to analyse the security and reliability requirements of the proposed EUROSERVER design. These have an effect on the possible commercial success of the proposed platform. All of the use cases have important security requirements that dictate constraints on the exposure of resources. Each hardware resource and I/O channel can be potentially isolated or shared. The use cases indicated requirements for security between components as well as at a holistic system perspective. Well-defined access mechanisms to resources were indicated through the requirements for accountability, isolation of resources, and privilege escalation. Reliability of the components is important for all the use cases, as the units will typically be difficult to access. Another highlighted requirement was for indication of failure or impending failure.

The scenarios indicate upper and lower bounds for many technical requirements that must be met as a minimum for the use case to succeed. Beyond the minimal requirements, there are constraints and cost functions that indicate the upper bounds.

A number of non-technical requirements have also been taken into consideration to complete the requirement elicitation process. Governance, regulation and compliance factors, as well as commercial requirements have been investigated to indicate further restrictions as part of the requirement gathering.

IV. System Architecture and Implementation

The EUROSERVER system architecture is based on ARMv8-A Cortex 64-bit processors, 3D silicon-on-silicon integration, and FD-SOI semiconductor technology.

ARM's 32-bit processor cores are used in a significant portion of the mobile and embedded markets, with >95% market share in smartphones and tablets [1]. There has recently, however, been a strong interest in the use of ARM cores also in high-performance computing (HPC) [9] and data centres [4]. Previous and ongoing projects in this direction have demonstrated the feasibility of the approach and worked to greatly improve the maturity of the ARM software stack. Now that 64-bit ARM processors implementing ARMv8-A are becoming available, these processors have the potential to dramatically change the HPC and data centre landscape. The 64-bit ARMv8-A architecture, announced in 2011, increases the number of architectural registers to 31, with 64 bits each, introduces the cleaner and hence more power-efficient A64 instruction set, and extends the per-process virtual address space from 32 bits (4 GB) to 48 bits (256 TB) [7].

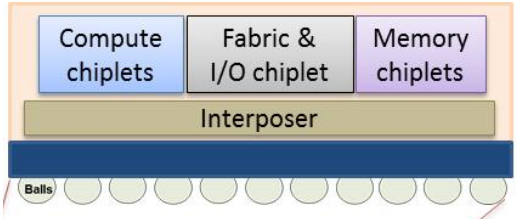

Figure 2: 2.5D integration using an interposer

EUROSERVER will also use the form of 3D packaging known as silicon on interposer. As illustrated in Figure 2, multiple small dies ("chiplets"), for example compute chiplets, I/O chiplets, or memory chiplets, are placed side-by-side in the same package on a small silicon board ("active interposer"). Compared to stacking chiplets on top of each other, the silicon-on-interposer approach eliminates the need for through-silicon vias (TSVs) in the chiplets and faces less difficulties with thermal hotspots. In common with all forms of 3D stacking, this approach **increases compute density** by integrating multiple chiplets, with each chiplet containing multiple cores, into a single package. It also **increases fabrication yield**: consider that a single large die in an advanced CMOS process node may have a yield of just 10% [2] or less, whereas multiple chiplets containing just the ARM compute, connected via an interposer, will deliver similar total aggregate performance and energy-efficiency, but with a fabrication yield in excess of 80%, comparable to that found in high volume mobile solutions. Alternatively, components that are currently located in separate chips can be placed inside a single package, reducing the amount of wiring and the number of contacts, improving reliability and energy-efficiency. For example, to drive today's DDR, one requires over 100 pJ of energy for each transferred bit; by contrast, an in-package DRAM die consumes closer to 1 pJ per bit.

The diversity of data centre applications means that a "one-fits-all" system-on-chip (SoC) solution does not exist. Different application domains benefit from different chip sizes and costs, ratios of compute to I/O and compute to memory, various I/O interfaces, as well as big or little cores, different cache sizes, and so on.[2] This is especially true if the server market is considered to also include HPC, which, for cost reasons, also uses commodity x86 server processors. In particular, more than 80% of the systems in the November 2013 TOP500 list use Xeon processors [13]. Similar devices

[2] Application domains include online transaction processing (strict transaction time boundaries, financial services and online sales security requirements); generic enterprise operations (business operations, finance, and messaging/communication); computationally intensive processing (logistics and production control, flight control, scientific research and modelling); multimedia content delivery (online entertainment and news, video conferencing, live meetings and content sharing); warehouse-scale computing (web-based email, web search and maps management).

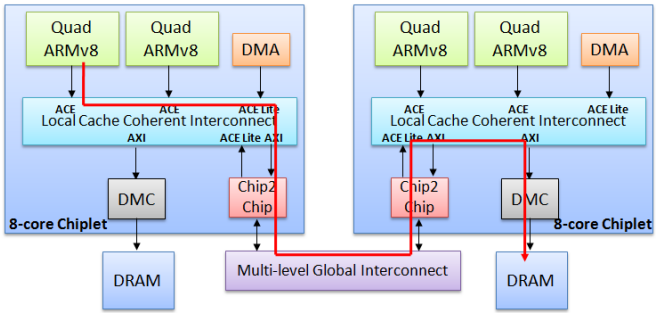

Figure 3: Remote memory accesses are routed from a chiplet to another and can be cached locally or remotely

are used across all these various requirements, and across the whole computing industry other than mobile, embedded, and telecoms—mostly based on the Intel architecture. There are essentially two micro-architectural implementations of the Intel architecture: Core/Xeon, and Atom.

In summary, there would be a significant improvement, in both performance and energy efficiency, if specialized devices were available. Unfortunately, however, limited sales volumes of a few tens of millions per specialized device cannot provide sufficient return on investment to justify the high Non-Recurring Engineering (NRE) costs.

The silicon-on-silicon integration approach is an important way to **reduce the cost of specialization**. Standard compute and memory chiplets can be fabricated, at advanced process nodes, and in very high volume. These compute chiplets can then be integrated into special-purpose devices, with the appropriate choices of compute chiplets, I/O interfaces, and I/O and memory bandwidths. The only component that needs to be customized for the application is the interposer, which is normally implemented in an older and less expensive technology, either directly integrating the I/O, or providing it through integration of an I/O chiplet, which is relatively small, dependent on commoditized IP blocks, and in an intermediate technology node. This approach is beginning to be used in the FPGA space: Xilinx have announced an interposer approach, based on a single FPGA chiplet design, with different FPGA densities supported by varying the number of chiplets in the package. Altera have also discussed a similar approach based on FPGA density, but targeting a fixed level of compute.

The silicon-on-silicon technique also allows chiplets manufactured in different technologies, including logic, analog, and RF, to be integrated into the same SoC package. The silicon interposer is now ready for implementation in a fashion that is widely accepted, similar to the old multi-chip modules. The above advantages mean that silicon-on-silicon integration will likely remain viable, even after full 3D is widely used.

In the EUROSERVER context, the flexibility in selection and combination of chiplets means that a specific configuration can be made according to the micro-server's intended use. For example, one micro-server may include a single compute chiplet, with DRAM and I/O located outside the package. Another micro-server, with the requirement for a high compute-to-I/O ratio, may include two or more compute chiplets in the interposer, with a shared I/O chiplet. A more memory-intensive micro-server, perhaps intended as a ramcache for a web server, may include one or more in-package memory chiplets. In all cases, the I/O can be attached directly to the processor bus, eliminating the PCIe interface abstraction, together with its performance and energy overheads. Today, where energy efficiency is critical, abstracting I/O through a general-purpose bus such as PCIe, which is the standard approach, is an unnecessary performance and energy penalty.

Each chiplet in the EUROSERVER prototype provides an 8-core ARM-based coherence "island". This is implemented by interconnecting two quad-core ARMv8 processors, a DMA engine, and other peripherals though the ARM Cache Coherent Interconnect (CCI-400), which provides full cache coherent accesses. The chiplets are interconnected through a multi-level global interconnect, which permits remote memory accesses, as depicted in Figure 3.

The EUROSERVER architecture can route a remote load/store from one chiplet to another chiplet. Each such remote memory access can be cached either locally (in the coherence island that initiated the access) or remotely (in the coherence island that the DRAM belongs to) by configuring appropriately the cache policies of the coherence island. In this way, we can have the following two useful scenarios:

- If coherence island A requires more DRAM, it can "borrow" memory from another coherence island B. This memory will be accessed and cached only in island A.
- Coherence island B can share a part of its local memory space with coherence island A. This memory space will be cached only in island B and accessed by both coherence islands.

Moreover, the chiplet architecture enables EUROSERVER to cost-effectively implement **cross-allocation of memory and I/O resources**, meaning that each compute chiplet can map any location in the global physical address space directly into its own memory space, to enable direct virtual address based sharing. This shared global address space will be also implemented outside the chiplets, and will allow sharing of memory and I/O resources across the entire system, an approach that relies on advances in the system software, as described in Section V.

The project also includes research into energy-saving techniques in all subsystems: chiplets, interposer, memory, processor cores, interconnect, and network energy proportionality,[3] as well as at the full systems level. It also includes the design of a next-generation system architecture, for implementation beyond the timescale of the project.

The flexibility in system architecture is seen as a unique feature of the EUROSERVER approach, allowing a market-

[3] Energy proportionality means that a component's energy consumption is proportional to its utilization.

specific configuration to be built, at significantly lower cost than re-fabricating the whole device. The approach has not as yet been disclosed or discussed by any vendor targeting the micro-server market.

V. SOFTWARE SUPPORT

The EUROSERVER system software will take advantage of the opportunities provided by its unique system architecture. It will allow efficient virtualization and sharing of resources across chiplets. It will also efficiently manage the shared resources and processors, dynamically assigning workloads, reducing workload interference. It will achieve high resource utilization with no compromise in performance.

The first aspect is **resource sharing**. Today's data centres are built as a cluster of fully independent nodes, each with a small number—typically two to sixteen—of "fat" cores, executing a manually partitioned part of the overall workload. These nodes have no shared resources, in that they are connected amongst themselves, to the outside world, and to external storage only via the high-speed network. The alternative to fat cores, followed by EUROSERVER, is a larger number of thinner cores, which promises significantly better energy efficiency.

In the Tilera approach [12], each chip has a large number of very thin cores, with full cache coherence. This supports a single OS or hypervisor per chip, in an extension of a conventional multi-core system. The main limitation, however, in the context of the data centre, is that there is no notion of separation of resources, resulting in unnecessary interference between multiple applications running on the same system. In addition, it is not possible to increase the number of cores managed by a single OS beyond the number of cores in a single chip. In contrast, in the "Calxeda" approach, each chip has a relatively small number of thin cores, again managed using one OS instance or hypervisor per chip. Since each unit contains a large number of these chips, the result is a "cluster in a box". In this approach, memory and I/O resources are permanently attached to the chip, meaning that excess memory or underutilized I/O resources cannot be reassigned to another OS instance running on another chip. Given that data centre workloads are highly diverse and dynamic, the Calxeda approach leads to an inefficient use of resources, both in terms of performance and overheads affecting the energy efficiency.

EUROSERVER uses an intermediate approach, combining the best aspects of the two extremes. The system architecture described in the previous section enables resource sharing between chiplets and between chips, so that cores and memory can be virtualized and made available seamlessly to the process or virtual machine. This provides the resource sharing of Tilera, together with the scalability, performance isolation, and hot-pluggability of Calxeda.

Several issues related to the **efficient use of resources** need to be addressed in the software stack, with a dual focus on performance and energy efficiency: virtualization support for CPU, memory and I/O resources, energy-aware virtual machine migration, use of multiple cores for scaling server I/O capabilities (especially high-speed connectivity and

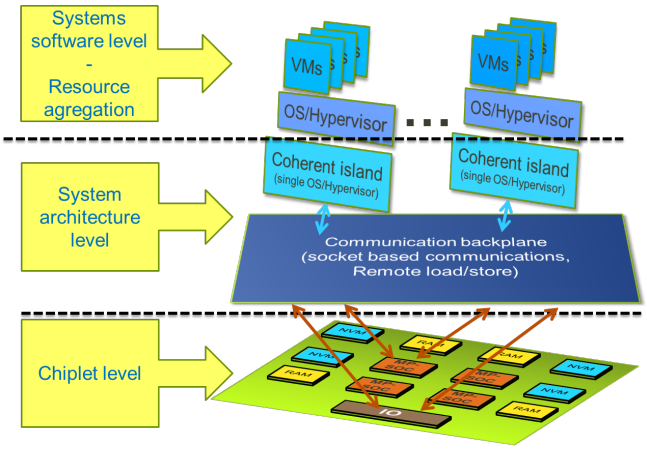

Figure 4: EUROSERVER Software architecture

networked storage), use of block-addressable NAND Flash and byte-addressable non-volatile memory (NVM) in the I/O path, resource sharing, including memory capacity, with and without coherence, and shared use of fast I/O devices. Also important is heterogeneous scheduling, which will be improved for the COMPSs programming model [11] and CloudRAN [3], as well as a reduction in performance interference between workloads and user-space tools and frameworks for managing, debugging and performance and/or energy-efficiency analysis of "thin" servers and appliances.

Future servers will include large numbers of cores, memory, and I/O resources, and will host large numbers of applications and services. Current approaches to managing resources in the OS kernel or the hypervisor require access to shared control and data paths resulting in interference and non-determinism, increased contention and high overheads, and ultimately limit server efficiency and scalability. Our goal is to **partition the resources** managed by the OS (or hypervisor) so that different applications do not contend or interfere for resources (storage, buffers, links, global policies, etc.) and allow access to resources from any core/application, while idle resources are shutdown.

We introduce the notion of "system slices", which will each consist of resource pools. For instance, two slices may include 16 cores, 32 GB of memory, 40 Gb/s network I/O, and 256 GB of NVM. Each "slice" may include resources from multiple chiplets and parts of the overall chip. In addition, slices will be isolated from each other. Cores that may belong to a specific slice at any point in time may be limited by architectural restrictions. Each slice will be cache-coherent, however, no cache coherence will be required across slices. We will design the abstraction based on which slices will be formed dynamically, according to application needs. Additionally, each slice will run applications without incurring any interference across slices at the device level, eliminating sharing overheads and reducing non-determinism in performance. We will provide OS and hypervisor support for slices in a transparent manner and as such provide direct portability for the higher levels of today's software frameworks and applications.

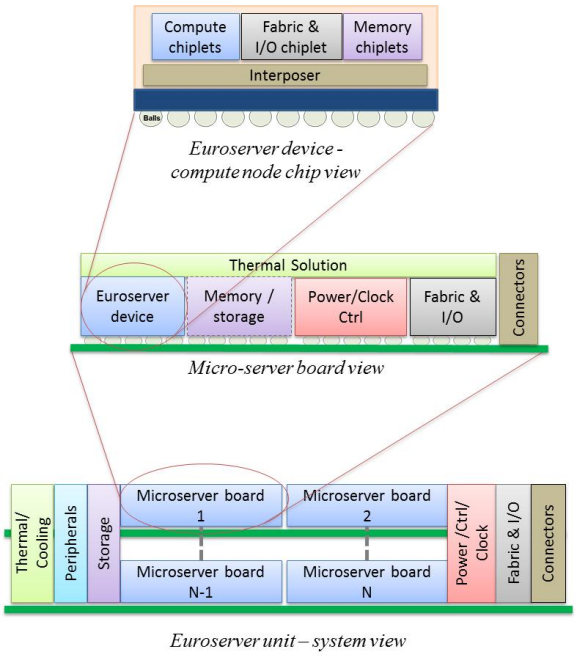

Figure 5: EUROSERVER prototype integration

VI. Expected Results

A. Project prototypes

We will develop two variants of the full-scale EUROSERVER prototype, targeting embedded servers and enterprise servers, respectively, in order to implement and evaluate the proposed hardware–software architecture. The system architecture, similar in the two prototypes, is illustrated in Figure 5.[4] At the top of the figure, the EUROSERVER packaged device will embed compute, I/O, and memory chiplets on a 3D interposer. A single such device will be integrated into a micro-server board, alongside memory, storage, clock, control, and I/O. The same micro-server board is used for the two prototype variants. The embedded server variant will hold one or two micro-server boards in a small, ruggedized, sealed form factor, with passive cooling. The enterprise server variant will integrate up to 64 micro-server boards, at very high density, together with networking, I/O, storage, and power supply, into a unit compatible with standard 42U server racks.

These full-scale prototypes are expected towards the end of the project. To enable system software development to proceed in parallel, a discrete component prototype was developed in the first year of the project. This platform uses FPGAs to emulate the most important characteristics of the full-scale prototype required by the system software stack. The discrete prototype will also be used to test some of the new features included in the full architecture.

[4] Since the prototype is still under development, the final details may not be exactly as described in this section.

B. Scientific Outcomes

The EUROSERVER project will advance the state-of-the-art in a coordinated fashion across many diverse aspects of micro-server technology. The system architecture will improve energy efficiency and total cost of ownership in all subsystems: low-power cores, chiplets, interposer, memory (DRAM and NVM), interconnect, network energy proportionality, as well as through the whole-system architecture. The use of 3D silicon-on-silicon integration reduces solution cost through improved device yield, compute density, improved core-to-core and core-to-memory energy and bandwidth, and enables application-targeted device specialization. Advances in system software will lead to more efficient use of resources across a diverse range of dynamic data centre workloads, through resource sharing (memory capacity and virtualized I/O), workload isolation, and energy-driven virtual machine (VM) placement algorithms. Changes in the application stack will improve heterogeneous scheduling for the COMPSs programming model and CloudRAN platform, as well as improving system monitoring, management and performance analysis.

VII. Conclusion

This paper presented EUROSERVER, a collaborative project aiming at an effective combination of micro-server architecture, silicon implementation, system integration and software development, to dramatically improve data centre energy-efficiency, total cost of ownership (TCO), and software efficiency. We are developing a next-generation micro-server architecture built on ARMv8-A 64-bit Cortex processors, 3D silicon-on-silicon integration, and FD-SOI process technology, together with an efficient system software stack for resource sharing and dynamic allocation, workload isolation, and intelligent VM placement. We argue in this paper that 64-bit ARM cores and 3D integration will bring a dramatic change in data centre architecture. In contrast to today's server industry, which is dominated by a small number of similar devices, 64-bit ARM-based chiplets will be integrated in application-specific configurations, bringing lower system cost and higher energy-efficiency.

Acknowledgment

This research project is supported by the European Commission under the 7th Framework Programme under the "Information and Communication Technologies" theme, with grant number 610456.

References

[1] ARM Annual Report, 2012.

[2] R. Chaware, K. Nagarajan, S. Ramal, "Assembly and Reliability Challenges in 3D Integration of 28nm FPGA Die on a Large High Density 65nm Passive Interposer", in Reliability Physics Symposium (IRPS), 15-19 April 2012, p. 279–283.

[3] China Mobile Research Institute, C-RAN, The Road Towards Green RAN, White Paper, 2010.

[4] Energy-conscious 3D Server-on-Chip for Green Cloud ("EuroCloud"). http://www.eurocloudserver.com.

[5] EUROSERVER project website: www.euroserver-project.eu.

[6] IDC - Digital Universe Study, sponsored by EMC, December 2012.

[7] J. Goodacre, Technology Preview: The ARMv8 Architecture, November 2011.

[8] J. Koomey for Analytics Press, Aug 2011.

[9] N. Rajovic et al. "Supercomputing with commodity CPUs: are mobile SoCs ready for HPC?." Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis. ACM, 2013.

[10] Talent Neuron Research and Analysis, 2013.

[11] E. Tejedor and R. M. Badia. COMP Superscalar: Bringing GRID superscalar and GCM Together. In 8th IEEE International Symposium on Cluster Computing and the Grid, May 2008.

[12] Tilera S2Q X5 Multi-node Series, http://www.tilera.com/sites/default/files/productbriefs/S2Q_server_overview.pdf

[13] TOP500, www.top500.org.