# SEVENTH FRAMEWORK PROGRAMME

FP7-ICT-2011-7

# DEEP

## Dynamical Exascale Entry Platform

## Grant Agreement Number: 287530

## D1.8
## Midterm management report at month 42

***Approved***


| | |
|---|---|
| Version: | 2.0 |
| Author(s): | E. Suarez, JUELICH |
| Contributions: | V.Beltran (BSC), D.Alvarez (JUELICH), A.Auweter (BADW-LRZ), N.Eicker (JUELICH), J.Kreutz (JUELICH), S.Eisenreich (BADW-LRZ), H.Ch.Hoppe (Intel) |
| Date: | 02.10.2015 |

## Project and Deliverable Information Sheet

<table>
<tr><td rowspan="7">DEEP Project</td><td colspan="2">Project Ref. №: 287530</td></tr>
<tr><td colspan="2">Project Title: Dynamical Exascale Entry Platform</td></tr>
<tr><td colspan="2">Project Web Site: http://www.deep-project.eu</td></tr>
<tr><td colspan="2">Deliverable ID: D1.8</td></tr>
<tr><td colspan="2">Deliverable Nature: Report</td></tr>
<tr><td rowspan="2">Deliverable Level:<br>CO*<br>(the present document contains only the public part of the deliverable)</td><td>Contractual Date of Delivery:<br>31 / May / 2015</td></tr>
<tr><td>Actual Date of Delivery:<br>31 / May / 2015</td></tr>
<tr><td colspan="2">EC Project Officer: Luis Carlos Busquets Pérez</td></tr>
</table>

* - The dissemination level are indicated as follows: PU – Public, PP – Restricted to other participants (including the Commission Services), RE – Restricted to a group specified by the consortium (including the Commission Services). CO – Confidential, only for members of the consortium (including the Commission Services).

## Document Control Sheet

<table>
<tr><td rowspan="6">Document</td><td colspan="2">Title: Midterm management report at month 42</td></tr>
<tr><td colspan="2">ID: D1.8</td></tr>
<tr><td>Version: 2.0</td><td>Status: Approved</td></tr>
<tr><td colspan="2">Available at: (Publishable part at) http://www.deep-project.eu</td></tr>
<tr><td colspan="2">Software Tool: Microsoft Word</td></tr>
<tr><td colspan="2">File(s): DEEP_D1.8_Midterm_management_report_at_M42_v2.0-ECapproved-PublishablePart</td></tr>
<tr><td rowspan="4">Authorship</td><td>Written by:</td><td>E. Suarez, JUELICH</td></tr>
<tr><td>Contributors:</td><td>V.Beltran (BSC), D.Alvarez (JUELICH), A.Auweter (BADW-LRZ), N.Eicker (JUELICH), J.Kreutz (JUELICH), S.Eisenreich (BADW-LRZ), H.Ch.Hoppe (Intel)</td></tr>
<tr><td>Reviewed by:</td><td>A.Auweter (BADW-LRZ), W.Gürich (JUELICH)</td></tr>
<tr><td>Approved by:</td><td>BoP/PMT</td></tr>
</table>

## Document Status Sheet

| Version | Date | Status | Comments |
|---|---|---|---|
| 1.0 | 31/May/2015 | Final | EC submission |
| 2.0 | 2/October/2015 | Approved | EC approved |

## Document Keywords

| | |
|---|---|
| Keywords: | DEEP, HPC, Exascale, status report, month 42 |

# Table of Contents

# List of Figures



# List of Tables

**No table of figures entries found.**

# Executive Summary

The Dynamical Exascale Entry Platform (DEEP) project started on 1st December 2011 and will last three years and 9 months[1]. The main goal of the project is to develop a prototype hardware and software supercomputing system, paving the way towards Exascale systems by the end of the decade. DEEP will optimise a set of grand-challenge applications with high societal impact and generic algorithmic structure for this platform. The key innovation of the DEEP project is its holistic Exascale-enabling concept integrating the architectural, system software and application level. The strategic goals of DEEP are *(i)* to contribute to an independent provision of general purpose Exascale performance supercomputers for the European HPC research infrastructure PRACE, *(ii)* to advance the growth of ICT and HPC hardware and software technology developed and produced in Europe, and *(iii)* to expand worldwide leadership and competitiveness of Europe's computational scientists and engineers.

This report describes the objectives, work performed, resources used, and results achieved during **months 36 to 42** of the DEEP project. The main achievements in the reporting period are enumerated below:

- Third review (at month 36) successfully passed.
- Proto-Booster with one pseudo-BIC (using an EXTOLL Galibier card) and one BNC containing two KNC cards installed and running at Mannheim, connected to the Super-BIC evaluator. Used by WP4 and WP5 for software development.
- Energy Efficiency Evaluator (EEE), a 16-Xeon Phi node system, installed at Garching. Bring-up of the system with debugging and testing is almost completed.
- ASIC Evaluator with 32 Xeon Phi nodes connected by EXTOLL-ASIC network in an immersion cooled design tested at UniHD. The ASIC-based Tourmalet NICs, system management and interaction with Xeon Phi nodes has been validated, and system bring-up is underway at UniHD.
- Two-half Booster chassis (32 KNC nodes) have been brought up at Jülich and were used for system software validation and optimization until scaling up of the machine started.
- Component production and assembly of the full Booster (384 KNC nodes) almost completed. Installation of the scaled-up Booster at Jülich in full swing at the time of writing.
- Excellent progress in the test and implementation of the RAS plane that measures the energy consumption of the final DEEP System.
- Tests and improvements in the implementation of the low-level Cluster-Booster protocol progressing on the DEEP Booster.
- Testing of ParaStation MPI and management software on the DEEP Booster ongoing.

[1] An amendment has been submitted to extend the project to a total duration of 45 months, finishing then in August 2015.

- Booster resource management with support for static and dynamic allocation available.
- Measurements of the OmpSs offloading functionality on different platforms done. Tests on DEEP Booster to start now. User support provided to the application developers and bugs fixed in the application layer.
- Standard procedure to evaluate the application's performance in a comparative way established. Benchmarking and measurements on various HPC platforms ongoing. Tests on DEEP Booster planned.
- Presentation of DEEP in publications, workshops and conferences.
- Co-organisation of a joint satellite event to the PRACE Days15 together with the European Exascale Projects (DEEP, DEEP-ER, CRESTA, Mont-Blanc 1 and 2, EPiGRAM, EXA2CT, and NUMEXAS) in Dublin, May 2015, focused on industrial users.
- Joint booth and dissemination activities at SC14, together with the other European Exascale Projects. Organisation of joint booth and workshop at ISC 2015.

# 1 Publishable summary

Numerous challenges must be overcome to reach Exascale by the end of the decade. When starting from current PetaFlop/s systems a factor of 1000 in performance increase is required. Investigating the roadmaps of standard CPUs reveals that it will not be sufficient to update and scale the concept of current cluster systems. In order to meet the requirements of energy-efficiency the use of accelerators becomes inevitable. However, today's solution of accelerated clusters –i.e. cluster with accelerators attached to each node– will not carry us to Exascale. On the one hand this is due to the inflexibility originating from the static assignment of standard CPUs and accelerators; on the other hand the competing use of the system bus by both accelerator and interconnect, combined with the lack of ability of the accelerator to use the interconnect autonomously, seriously limits the scalability of this solution.

Therefore, the DEEP Architecture proposes to detach the accelerators from the standard CPUs and to gather them into a separate cluster of accelerators that is called Booster. It is foreseen to run the highly scalable parts of the applications on this part of the DEEP System. The Booster is connected to a standard Cluster qualified to handle those parts of an application not suited for the Booster. The benefits are manifold: there is more flexibility on the ratio of standard CPUs and accelerators to be used by an application, the ability of the accelerators in the Booster to act autonomously allows for off-loading more complex and more parallel kernels, and an extended programming-model supports the application-developers in the identification of these offload-kernels and in porting their workload to the proposed architecture. Via its own interconnect the Booster Nodes can communicate between each other, independently of the use of the Cluster interconnect.

To implement the DEEP Architecture (see Figure 1), the project uses a standard Cluster, composed of multi-core processors and the high flexibility network InfiniBand®. Using special interface nodes, this cluster is attached to a Booster of Intel® Xeon Phi™ processors. The Booster nodes are connected via a highly scalable EXTOLL torus network.

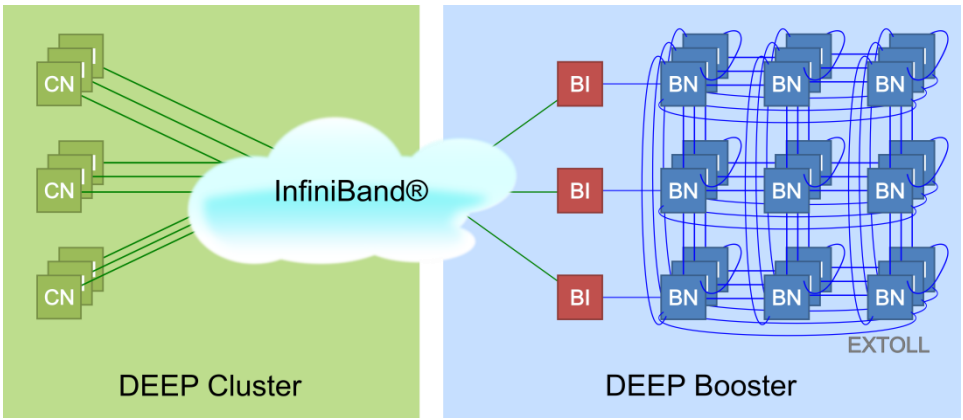


**Figure 1: Sketch of DEEP hardware architecture (CN: Cluster Node; BN: Booster Node; BI: Booster Interface)**

The DEEP software stack (see Figure 2) focuses on scalability and an efficient use of resources. The parts of the application with a complex communication pattern (low to medium scalability) will run on the DEEP Cluster, while the code parts with regular communication patterns (highly scalable) will run on the Booster. The DEEP programming environment provides a Global MPI layer below the OmpSs tasking model to help the

developers in decomposing their applications into tasks in order to efficiently overlap the computation done at Cluster and Booster sides of the system.

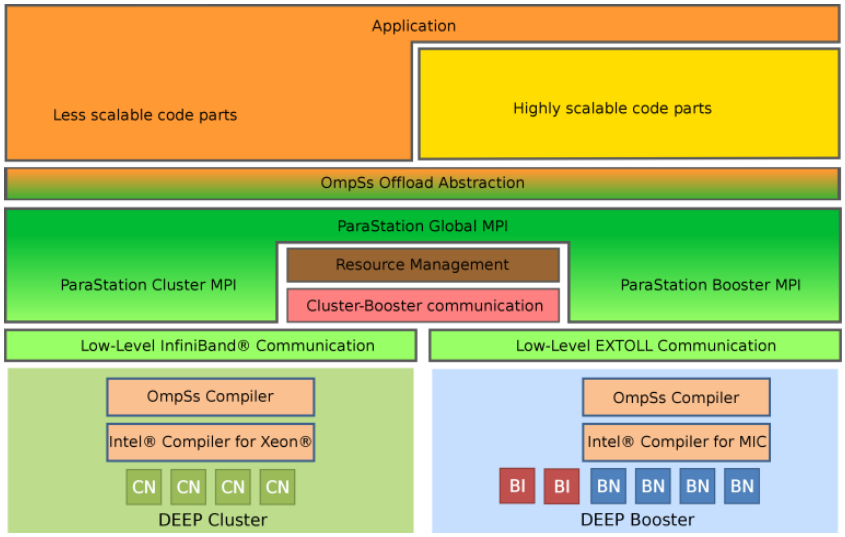


**Figure 2: Diagram of the updated DEEP software architecture**

This report describes the objectives, work performed, resources used, and results achieved by the DEEP project.

## 1.1 Project objectives

The objectives of the DEEP project as described in the DoW, and key results achieved in the project to reach its objectives are:

- Development of a prototype hardware platform consisting of a Cluster element based on multi-core chips, a Booster element based on many-core technology and a commensurate connectivity, following the components-off-the-shelf philosophy. This prototype of Cluster Booster Architecture will serve as proof-of-concept for a next-generation PRACE production system reaching up to 100 PFlop/s in the time-frame 2014/2015, with the potential to achieve Exascale between 2018 and 2020.

  → Cluster and Booster system integration improved with experience gathered after a hardware incident on the Cluster.

  → Proto-Booster (evaluator with one Booster Node Card containing 2 KNCs) used for system software development.

  → Two Booster chassis (each one holding a left half Backplane) with 32 Xeon Phi processors installed, brought up at Jülich, used for software tests and tuning.

  → Energy Efficiency Evaluator (small DEEP System with 4 Cluster Nodes and 16 Booster Nodes targeted at experiments related to hot water cooling) installed at BADW-LRZ.

  → New Booster Rack installed at Jülich for the scaled-up Booster, improving the original design. All 12 chassis integrated, eleven of them already populated with BNCs. The production and individual tests of the remaining BNCs is almost completed and their integration with the rest of the system is planned in the next days.

→ ASIC Evaluator with 32 Xeon Phi processors under tests at UniHD. Installation planned in May 2015, using the EXTOLL ASIC implementation (called Tourmalet, see below). Scale-up to 64 KNL nodes planned for June/July 2015 timeframe.

- Combination of innovative technologies for the Booster element: novel Intel® many-core processors; EXTOLL high speed interconnect; hot water cooling.

  → Booster Node Card (BNC) design completed. A BNC implements two Booster Nodes, each with an Intel Xeon Phi and an EXTOLL network controller implemented on an Altera Stratix V FPGA. Issues related to the Ethernet connection discovered on the early test platforms and solved with a hardware patch applied during production. Boards for the scaled-up Booster are already assembled into the Booster except for remaining minority of BNCs which in production or under test.

  → Booster Interface functionality has been implemented in the form of a combination of the Booster Interface Card (BIC) and a server board (Juno) in the EEE and the scaled-up Booster. The Juno-BIC solution provides the KNC remote boot functionality and Cluster-Booster connectivity required in DEEP. Bring-up and validation almost completed in the Energy Efficiency Evaluator. Bring-up ongoing on the Booster.

  → Backplane design (both left and right half planes) completed and production for the scaled-up Booster completed. Backplanes have been validated with full functionality.

  → A fully functional version of the EXTOLL ASIC implementation became available in early 2015. It has been fully qualified for a bandwidth of 5 Gbit/s per lane, and validation with 8,4 Gbit/s per lane is progressing. An EXTOLL chip offers 6 links with twelve lanes each. Additionally a 7$^{th}$ link is available.

- Development of a reliable, open source cluster operating system, interconnect and runtime software stack with high resilience while exploiting millions of cores.

  → Software environment and programming model for DEEP have been defined.

  → Cluster Booster Protocol plugin in ParaStation pscom implemented. Tests on Proto-Booster and first Booster chassis performed.

  → ParaStation component pscom supporting EXTOLL and its operation in combination with Xeon Phi.

  → Initial EXTOLL communication measurements with KNC showed less performance than expected. Modifications in the pscom's use of buffers have solved this problem. Recent measurements on the Proto-Booster and the first Booster chassis show good performance.

  → Resource management supporting static and dynamic allocation ready.

  → Software stack currently being installed and tested in the DEEP Booster and EEE.

- Development of programming models, scientific libraries and performance tools for standard x86-based many-core processors, in order to achieve high productivity and enabling unprecedented scalability.

→ OmpSs runtime (Nanos++) ported to Intel Xeon Phi.

→ OmpSs offloading mechanism supports now both C/C++ and Fortran codes, tested in several platforms.

→ Performance of the Intel MKL BLAS library routines needed by the DEEP applications measured and assessed.

→ Paraver/EXTRAE performance tool ported to MIC and supports the DEEP offload model.

- Improvement of current cluster energy efficiency by an order of magnitude exploiting novel many-core chip technologies and advanced software-aided cooling technologies with a power usage effectiveness approaching a value of 1.

  → Hot water cooling infrastructure prepared and in operation at Jülich.

  → Concepts for improving energy and cooling efficiency defined.

  → Required RAS plane functionality defined and implementation close to completion.

  → Firmware for the control and management of the DEEP Booster components (BNC and BIC) implemented.

  → With the newly introduced immersive liquid cooling approach, the ASIC Evaluator is expected to show very good energy efficiency and density.

- Optimisation of a set of application codes on the DEEP System that are representative for future Exascale computing and data handling requirements, chosen from the fields of Health and Biology, Climatology, Seismic Imaging, Computational Engineering, Space Weather, and Superconductivity and proving safe extrapolation to millions of cores as will be required with future Exascale systems.

  → Structure of the applications has been analysed and strategies to distribute the codes between Cluster and Booster have been identified.

  → Optimisations of the codes and porting to Intel Xeon Phi (KNC)well advanced.

  → "Taskification" of application codes with OmpSs completed in some applications, ongoing in others.

  → Offloading of application kernels with OmpSs ongoing.

  → Methodology for a uniform study and benchmarking of the applications established.

  → Benchmarking on several platforms ongoing, tests on the DEEP Booster about to start.

- Demonstration of scalability of the new hardware-software concept with respect to the generic multi-scale, adaptive grid and long-range force parallelisation models underlying the application codes.

  → Scalability of the DEEP offloading mechanism evaluated on JUQUEEN, Stampede and MareNostrum III supercomputers, amongst others.

  → Full demonstration awaiting availability of the scaled-up Booster and ASIC Evaluator.

- Dissemination of the innovations and results of the project to the public.

  → The DEEP project and its first results have been presented and discussed in many meetings, conferences, and workshops.

  → The project's website and social media channels are updated regularly with new content resulting in a steady increase in twitter followers.

## 1.2 Work performed and main results

The status of the milestones due between the **months 36 to 42** of the DEEP project is the following:

- MS9: DEEP Booster integrated into Chassis: shifted from M36 to M43 with the requested amendment.
- MS18: DEEP System installed: Installation of two Chassis, populated with one left half-Backplane and 8 BNCs each (16 BNCs in total) completed. Connection to the Cluster implemented by means of two Pseudo-BICs. System used for system software tests and optimization. Installation of scaled-up Booster ongoing at the time of writing.

*Management, legal and administrative tasks*

In the present reporting period the management activities focused on monitoring the progress of the project to guarantee the achievement of all technical goals specified in the Description of Work (DoW) and the fulfilment of all commitments to the European Commission, including the preparation and successful completion of the external review.

The Management team organised the agenda for the third review meeting (at month 36) which took place on the 16th January 2015 in Jülich (Germany). To fulfil the internal quality policies a rehearsal meeting one day before the review was conducted. As a result of the third review, the project has been evaluated as doing *"good progress".* Additionally, all deliverables submitted in the second year of the project were approved. All public approved deliverables have been uploaded to the project website. Minor corrections in the publishable part of D1.7 have been done to remove budget data, before it was uploaded to the project website.

One week after the review a report was submitted to the reviewers and the project officer to explain in detail the reasons for which a cost-neutral 3-month project extension is required, to make possible the scale-up of the existing hardware (Booster and ASIC Evaluator). With this information in hand, the review report recommended to request such a project extension Following this recommendation, the third amendment of the Description of Work was prepared and submitted to the European Commission. Additionally, the technical specifications of the Booster and ASIC Evaluator have been updated to describe their corresponding scale up and an amendment of the Side Agreement, in which the Consortium delegates the production of the Booster hardware to partner Eurotech, was also prepared. All the mentioned amendments (DoW, technical specifications, and Side Agreement) have been approved by the Board of Partners on 24.03.2015. The BoP voting process was executed via Email to speed up the process.

The eighth regular face-to-face meeting of the consortium (BoP) took place in Jülich (Germany) on the 28th – 29th May 2015. Here, the status of the project was discussed with particular emphasis on the progress of the hardware scale-up and software bring-up and tests.

Managing an integrated hardware-software co-design project like DEEP is by no means an easy task. Requirements of hardware, system software, and application teams have to be

addressed and various work plans have to be coordinated. Bi-weekly meetings of the Design and Development Group (DDG) have taken place to coordinate the access to the various hardware experimentations platforms. The Project Management Team (PMT) participates in the DDG as well as in the separate meetings of the work packages to track and prioritise all open issues and develop a suitable work plan, taking into account the available hardware and developer resources. Also monthly teleconferences of the Team of Work Package leaders (ToW) were organised to periodically discuss the progress in all Work Packages (WPs).

All due Deliverables were timely submitted to the European Commission after having passed through the mandatory DEEP internal review process.

### *Dissemination, training and outreach*

The centre of the dissemination activities of DEEP is its web site: www.deep-project.eu. The DEEP web page has been updated in the reporting period to keep track with the project results (i.e. with presentations, publications, and approved public deliverables) and to announce all upcoming events.

Partners from the DEEP consortium presented the project's concept in several conferences and workshops, such as the SC14 in New Orleans (USA), the Big Data and Extreme-Scale Computing (BDEC) workshop in Barcelona (Spain), and prepared for the ISC 2015 conference in Frankfurt (Germany).

Joint dissemination activities together with other European Exascale Projects – EEP (Mont-Blanc 1 and 2, CRESTA, DEEP-ER, EPiGRAM, EXA2CT and NUMEXAS) have been organised. A joint satellite event at the PRACEdays15 has taken place on 26th May in Dublin (Ireland), with the goal of attracting industrial users to use the concepts proposed in the European Exascale Projects. EEP-joint and project-specific activities took place at the SC14 conference. The organisation of the joint booth (#1039, see Figure 3), including floor-space reservation, booth layout, furniture renting, and graphic design, was led and taken over by the DEEP and DEEP-ER WP2 leader, who is also organising and chairing the regular EEP teleconferences. Joint flyers and give-aways were distributed at SC14. Additionally, DEEP-specific material (flyer, give-aways, etc.) was prepared, including a poster on the DEEP offload model, which was presented at the Emerging Technologies Track session of the conference.


**Figure 3: EEP booth at SC'14.**

For ISC 2015 several activities are being organised, including a joint EEP booth, a joint EEP workshop, and the preparation of DEEP-specific dissemination material.

Training the community on how to use the software and hardware developed in DEEP is an important part of the project. The main goal of the training events in DEEP is to teach the application developers participating in the project on how to use the software tools and programming environment running on the DEEP System and other intermediate prototypes. Since the project is arriving to its end, no training events have been performed in the reporting period. However, the support team continues, in these last months with redoubled effort, to assist the application developers in tuning and benchmarking their applications.

### *Technical Work*

Two major technical decisions taken during the reporting period were made:

- A commitment to a combined upscale of the existing hardware platforms in the form of a 384 node Booster and a 64 node ASIC evaluator was made.
- A mitigation plan addressing a problem with PCIe address space sizes encountered on the originally planned BIC design was enacted. The new BIC design features a server board (Juno) instead of an embedded COM Express module.

To ensure on-time delivery of the upscale hardware, the PMT has implemented thorough monitoring of the production milestones and supervises the detailed installation plan developed in WP6.

Regarding the work on the scientific applications, regular application review meetings foster the exchange between the application partners and the technical staff and ensures the continuity of the application partner's commitments.

The technical work in DEEP is grouped into the three main parts: system hardware, system software, and applications.

*System Hardware*

In the reporting period, the development, test and validation of the DEEP Booster HW components has progressed. Backplanes, BNC and BI designs have been completed. The BI includes now a Booster Interface Card (BIC) attached to a server (Juno), which provides sufficient MMIO space to remotely booth and control the 16 KNCs connected to each BI. Production of the components needed for the Energy Efficiency Evaluator (EEE) and the Booster, as well as their construction and installation at the corresponding sizes took place:

- The EEE was installed at Garching in February 2014 (M39). This system has constituted a crucial forerunner of the full-scale Booster, as it has been used to test the final configuration of the system, and verify the functionality of all its components and develop test-software and firmware needed later in the installation of the full Booster. Particularly important have been the validation of the Juno-BIC solution and confirmation of its ability to boot and orchestrate its corresponding 8 BNCs.
- The installation of the Booster has been done in two phases, starting with two-half chassis in December 2014 (M37) and continuing in May 2015 (M42) with the scale-up of the machine to a 384-KNC system. In the first phase "Pseudo-BICs" (based on EXTOLL Galibier cards connected to the backplane through a HDI6 adaptor and to external servers) have been used to implement the BI functionality, since the Juno-BIC solution was not yet ready at that time. In the full Booster this interim solution has been substituted by Juno-BICs. At the time of writing the full Booster rack with all its twelve chassis has been installed, populating it with 180 BNCs, and the bring up is ongoing. The remaining 12 BNCs are currently under production/test and will be inserted in the next days into the system.

The status of the ASIC development at EXTOLL GmbH, done outside DEEP, was closely watched by the project. A fully functional version of the EXTOLL ASIC implementation (named Tourmalet) was available in early 2015 and is currently in its test & validation phase. Full validation was completed for a lane speed of 5 Gbit/s (matching the PCIe lane speed of the Intel Xeon Phi). Validation at 8,4 Gbit/s is ongoing. With this technology, the first part of the "ASIC Evaluator" has been built and is being brought up at UniHD. This prototype contains 32 Intel Xeon Phi processors and 32 EXTOLL Tourmalet cards (at 5 Gbit/s/lane and PCIe gen2), cooled by an innovative immersion cooling technology. Potential issues with the immersive cooling of KNC processors with Novec have been discussed between EXTOLL GmbH and Intel, and solutions have been found. The existing ASIC Evaluator is planned to be installed at Jülich in June 2015.

*System Software*

On the software side, all the major components have been developed and the testing phase on the DEEP Booster has started. The Cluster-Booster Protocol has been tested on the upper half-chassis of the DEEP Booster, obtaining comparable performance to the one measured on the Proto-Booster. Tests on the EEE, where the interface between the Cluster and Booster parts of the machine is implemented with a Juno-BIC are pending. It is expected that this construction should provide an even better performance since the PLX included in the BIC should improve the communication speed between the EXTOLL NIC and InfiniBand NICs.

ParaStation and OmpSs are also already installed on the DEEP Booster and tests have started in the first Booster chassis and will continue on the scaled-up system.

*Applications*

In the reporting period important improvements have been achieved in different applications. The focus has been on achieving a working and usable Cluster/Booster division:

- EPFL has benchmarked its CoreNeuron application – where the Cluster/Booster division had been performed already – on MareNostrum III and Stampede. Preliminary analysis show a very significant improvement of the I/O phase thanks to the I/O offload. Even though I/O has a limited impact in CoreNeuron's total runtime, these results highlight the importance of avoiding direct I/O to NFS shares from Xeon Phi coprocessors. Besides this, the EPFL team successfully finished the memory layout transformation to structs of arrays (SoA), achieving also a very significant boost in performance.
- The code division of iPiC3D has been implemented using both OmpSs and MPI_Comm_spawn directly and benchmarking in alternative platforms will be finished soon.
- The team of CYI has been debugging EMAC during the last months, together with the support team. The detected issues have been solved very recently, and benchmarking is now on its way.
- The team at CERFACS has been working on improving performance on Xeon Phi, removing indirect accesses to improve vectorisation in certain kernels.
- The application developers of CINECA managed to have a working implementation of the OmpSs offload. However, the efforts to improve performance on Xeon Phi have not been successful due to the complexity of the code.
- Lastly, CGG made their RTM application ready to run on the DEEP System and currently wait for the Booster to become available.

## 1.3 Expected final results

At the end of the project, DEEP will have installed the full DEEP System in Jülich, composed of two parts, the Cluster and the Booster, running with a software stack that allows applications to distribute their code on both parts of the machine, dynamically assigning Cluster Nodes to Booster Nodes and vice versa. Additionally the ASIC evaluator platform in Jülich will enable performance tests of the interconnect based on EXTOLL ASIC. The Energy Efficiency Evaluator will have brought further inside into the feasibility and limitations of hot water cooling.

The experience gained by running six scientific applications on the DEEP System, with different code structures, requirements, and scientific goals, will demonstrate whether the proposed DEEP concept is suitable for the next generation Exascale supercomputers.

# 2 Annex A

## A.1 Listing of dissemination activities

This list reflects the disseminations activities performed between **months 36 and 42** of the DEEP project.

*Conferences, workshops, and meetings:*

- **JUELICH-JSC meeting (Visit C.Aubley),** Jülich, Germany, January 19, 2015:
  - E.Suarez (JUELICH), "DEEP and DEEP-ER" (presentation).

- **BDEC (Big Data and Extreme Scaling) Workshop,** Barcelona, Spain, January 28-30, 2015:
  - E.Suarez (JUELICH), "The DEEP (and DEEP-ER) projects" (presentation)

- **PARS Workshop, 26. GI/ITG Workshop Parallel -Algorithmen, -Rechnerstrukturen und –Systemsoftware,** Postdam, Germany, May 7-8, 2015. (http://www.cs.uni-potsdam.de/~schnor/potsdam/misc/workshops/2015/pars.html)
  - A.Jakobs (JUELICH), A. Zitz (JUELICH), N.Eicker (JUELICH), and G. Lapenta (KULeuven) "Particle-inCell algorithms on DEEP: The iPiC3D case study" (presentation)

- **ECL meeting,** Jülich, Germany, May 13, 2015:
  - E.Suarez (JUELICH), "The DEEP and DEEP-ER status" (presentation)

- **EASC Exascale Applications and Software Conference 2015,** Edinburgh, UK, April 21 – 23, 2015 (**http://www.easc2015.ed.ac.uk/home**)
  - P.Kumbhar (EPFL), "Leveraging a Cluster-Booster Architecture for Brain-Scale Simulations" (presentation)

- **JSC-LBL meeting (Visit S.Dosanjh),** Jülich, Germany, May 22, 2015:
  - N.Eicker (JUELICH), "DEEP and DEEP-ER" (presentation).

- **Enabling Exascale in Europe for Industry, Satellite event at the PRACE Days 2015,** Dublin, Ireland, May 26, 2015:
  - E.Suarez (JUELICH), M.Tchiboukdjian (CGGVS), G.Staffelbach (CERFACS), "DEEP and DEEP-ER: Innovative Exascale architectures in the light of user requirements" (presentation).

- **15th international Conference on Numerical Combustion, Towards exascale simulation of turbulent combustion mini-symposium**, Avignon, France, April 2015
  - G.Staffelbach (CERFACS), "Legacy codes and the jump to exascale: Fluid dynamics simulation with AVBP and HPC" (presentation)

*Publications, proceedings, press-releases, and newsletters:*

- **Advances in Space Research (2015)**

  - C.J.Schrijvera, et al, amonst others G.Lapenta (KULeuven), “Understanding space weather to shield society: A global road map for 2015-2025 commissioned by COSPAR and ILWS” (article)

- **Advances in Engineering Software,** to appear in 2015**.**
  - G. Lapenta (KULeuven), M.E. Innocenti (KULeuven), S. Markidis, J. Amaya (KULeuven),, A. Johnson (KULeuven),, J. Deca, V. Olshevsky, “Progress towards Physics-Based Space Weather Forecasting with Exascale Computing” (article)

- **Special Issue of Concurrency and Computation, Practice and Experience (CCPE), 2015**
  - N.E., Th. Lippert, Th. Moschny, and E. Suarez for the DEEP project. “The DEEP Project – An alternative approach to heterogeneous cluster-computing in the many-core era” (HUCAA-Paper from 2013 accepted as invited paper)

- **HPCwire: Weekly Twitter Roundup (Jan, 22)**
  http://www.hpcwire.com/2015/01/22/weekly-twitter-roundup-27/

- **21st IEEE Symp. on High Performance Computer Architecture (HPCA),** February 7-11, 2015, San Francisco Bay Area, California, USA
  - S.Neuwirth, D.Frey, M.Nuessle, U.Bruening; “Scalable Communication Architecture for Network-Attached Accelerators”.

*Media relations:*

- DEEP image video will be featured in the February issue of online journal iSGTW; on top: opinion piece planned with this magazine towards end of DEEP
- Interview request (via Twitter) by Robert Roe from Scientific Computing; appointment will be fixed after journalist’s editorial deadline; focus: Extoll ASIC
-

# List of Acronyms and Abbreviations

## *A*

**ADI3 layer:** MPICH Abstract Device Interface Version 3
**AoS:** Array of Structs
**AoSoA:** Array of Structs of Arrays
**API:** Application Programming Interface
**ASIC:** Application Specific Integrated Circuit: Integrated circuit customised for a particular use
**Aurora:** The name of Eurotech‘s cluster systems
**AVBP:** A parallel CFD code for reactive unsteady flow simulations on hybrid grids developed by partner CERFACS

## *B*

**BADW-LRZ:** Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften.Computing Centre, Garching, Germany
**BI:** Booster Interface (functional entity)
**BIC:** Booster Interface Card: Interface card to connect the Booster to the Cluster InfiniBand® network
**BIC evaluator:** A platform consisting of three x86-based nodes equipped with (i) an EXTOLL NIC, (ii) an InfiniBand® HCA, (iii) both, EXTOLL NIC and InfiniBand® HCA, developed and used only in the DEEP project
**BLAS:** Basic Linear Algebra Subprograms: Standard application programming interface to publish basic linear algebra libraries
**BlueGene/Q:** Supercomputing architecture developed by IBM, well known for its energy efficiency, massive parallelism, 5D torus network and wide vector units.
**BMC:** Baseboard Management Controller
**BN:** Booster Node (functional entity)
**BNC:** Booster Node Card: A physical instantiation of the BN
**BNC evaluator:** Same as EXTOLL evaluator
**BoF:** Birds of a Feather Session: Informal meeting during a Conference where people can discuss about the topic of their common interest
**Booster System**: Hardware subsystem of DEEP comprising of BNC, BIC and Intra-Booster network
**BoP**: Board of Partners for the DEEP project
**BSC:** Barcelona Supercomputing Centre, Spain
**BSCW:** Basic Support for Cooperative Work: Software package developed by the Fraunhofer Society, used to create a collaborative workspace for collaboration over the web

## *C*

**CBP:** Cluster-Booster protocol
**CERFACS**: Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, Toulouse, France
**CFD:** Computational Fluid Dynamics

**CG:** Conjugate Gradient
**CGGVS:** CGGVeritas Services SA, Paris, France
**CINECA:** Consorzio Interuniversitario, Bologna, Italy
**CN:** Cluster Node (functional entity)
**COMe:** Computer-on-module (COM) form factor: highly integrated and compact PC that can be used like an integrated circuit component.
**CooLMUC:** Prototype at BADW-LRZ with direct warm water cooling
**Coordinator:** The contractual partner of the European Commission (EC) in the project
**CPU:** Central Processing Unit
**CRESTA:** Collaborative Research into Exascale Systemware, Tools & Applications: EU-FP7 Exascale Project led by the University of Edinburgh.
**CUDA:** Compute Unified Device Architecture: Parallel computing architecture developed by NVIDIA
**CYI:** Cyprus Institute, Nicosia, Cyprus

## *D*

**DC:** Direct Current (electricity)
**DDG:** Design and Developer Group of the DEEP project
**DEEP:** Dynamical Exascale Entry Platform: EU-FP7 Exascale Project led by Forschungszentrum Jülich
**DEEP Architecture:** Functional architecture of DEEP (e.g. concept of an integrated Cluster Booster Architecture)
**DEEP Booster:** Booster part of the DEEP System
**DEEP Supercomputer:** A future Exascale supercomputer based on the DEEP Architecture
**DEEP System:** The production machine based on the DEEP Architecture developed and installed by the DEEP project
**DFF:** Dense Form Factor
**DGEMM:** Double precision General Matrix Matrix multiplication
**DGEMV:** Matrix-vector multiplication
**Dimemas:** A performance analysis tool for message-passing programs developed at BSC
**DMA:** Direct Memory Access
**DoW:** Description of Work: Annex I of the Grant Agreement
**DSL:** Domain Specific Language

## *E*

**EC:** European Commission
**EESI:** European Exascale Software Initiative (FP7)
**EMAC:** ECHAM/MESSy (Application coupling together the ECHAM model with the MESSy framework)
**EMEA:** Europe, the Middle East and Africa: Regional designation used for government, marketing and business purposes
**Energy Efficiency evaluator:** Platform used for the investigations of the energy-aware functionality of DEEP, used only in the DEEP project
**EPFL:** École Polytechnique Fédérale de Lausanne, Switzerland
**ETP4HPC:** European Technology Platform for High Performance Computing
**EU:** European Union

**Eurotech:** Eurotech S.p.A., Amaro, Italy
**Exaflop**: $10^{18}$ floating point operations per second
**Exascale**: Computer systems or applications, which are able to run with a performance above $10^{18}$ floating point operations per second
**EXTOLL:** High speed interconnect technology for cluster computers developed by University of Heidelberg
**EXTOLL evaluator:** Platform for evaluation of EXTOLL technology, developed and used in the DEEP project

## *F*

**FLOP:** Floating point Operation
**FPGA:** Field-Programmable Gate Array: Integrated circuit to be configured by the customer or designer after manufacturing

## *G*

**Global MPI:** MPI allowing communication between the Booster and Cluster part of the DEEP System. Based on the ParaStation process-management and the Cluster-Booster protocol acting as a plug-in for the pscom library. Provides the MPI_Comm_spawn() call used by application processes running on the CNs to start additional processes on the BNs
**GPU:** Graphics Processing Unit
**GRS:** German Research School for Simulation Sciences GmbH, Aachen and Jülich, Germany

## *H*

**H4H:** Hybrid programming For Heterogeneous architectures (EU project)
**HCA:** Host Channel Adapter
**HOPSA:** HOlistic Performance System Analysis (EU-Russia FP7 project)
**HPC:** High Performance Computing
**HW:** Hardware

## *I*

**IB:** InfiniBand®
**ICPP:** International Conference on Parallel Processing: Yearly conference on parallel and distributed computing
**ICT:** Information and Communication Technologies
**IEEE:** Institute of Electrical and Electronics Engineers
**INFSO:** Information Society
**Intel:** Intel GmbH, Feldkirchen, Germany
**Intel Xeon® Phi™:** official product name of the Intel Many Core (MIC) architecture processors. The first available Intel Xeon® Phi™ product is code-named Knights Corner (KNC).
**Interconnect evaluator:** Hardware for interconnect studies on physical and mechanical layer, developed and used in the DEEP project

**I/O:** Input/Output
**IP:** Intellectual Property or Internet Protocol (depending on the context)
**iPIC3D:** Programming code developed by the University of Leuven to simulate space weather
**ISC:** International Supercomputing Conference: Yearly conference on supercomputing which has been held in Europe since 1986

## *J*

**JSC:** Jülich Supercomputing Center
**JUDGE:** Jülich Dedicated GPU Environment: A cluster at the Jülich Supercomputing Centre
**JUELICH:** Forschungszentrum Jülich GmbH, Jülich, Germany
**JUQUEEN:** Jülich's BlueGene/Q machine: A supercomputer installed at the Jülich Supercomputing Centre

## *K*

**KNC:** Knights Corner: Code name of a processor based on the MIC architecture. The commercial name of this product is Intel Xeon® Phi™.
**KNF:** Knights Ferry: Intel first available processor based on the MIC
**KULeuven**: Katholieke Universiteit Leuven, Belgium

## *L*

**LINPACK:** Software library to perform numerical linear algebra calculations used as benchmark
**LINUX:** A Unix-like computer operating system assembled under the model of free and open source software development and distribution

## *M*

**MareNostrum:** Supercomputer system hosted by BSC
**Maui:** Job scheduler for use on clusters and supercomputers
**MB:** Mega Byte or Mother Board (depending on the context)
**MC:** Monte Carlo
**MECCA:** Module Efficiently Calculating the Chemistry of the Atmosphere
**Mercurium compiler:** OmpSs' source-to-source compiler
**MLNX:** Mellanox Technologies, Ltd., Sunnyvale, California and Yokneam, Israel
**MIC:** Intel Many Integrated Core architecture
**MIC evaluator:** Platform for evaluation of the MIC architectural concept, used only in the DEEP project
**MIC-OS:** Operating System of the MIC architecture
**Mini Booster prototype:** Minimal instantiation of a DEEP Booster used for analysis of the energy-aware functionality, developed and used in the DEEP project
**Mini DEEP System:** A fully featured DEEP System of minimal size comprising the Mini Booster

**MKL:** Intel® Math Kernel Library

**Mont-Blanc:** European scalable and power efficient HPC platform based on low-power embedded technology: EU-FP7 Exascale Project led by the Barcelona Supercomputing Centre

**MPI:** Message Passing Interface: API specification typically used in parallel programs that allows processes to communicate with one another by sending and receiving messages

**MPICH:** Freely available, portable implementation of MPI

**MPSS:** Intel many-core platform software stack. Software bundle to operate Xeon® Phi™ devices

**MQTT protocol:** Message Queue Telemetry Transport. Open message protocol for machine to machine communications. It enables the transfer of telemetry-style data in the form of messages from pervasive devices, along high latency or constrained networks, to a server or small message broker.

## *N*

**NIC:** Network Interface Card: Hardware component that connects a computer to a computer network

## *O*

**OmpSs:** BSC‘s Superscalar (Ss) for OpenMP

**OpenCL:** Open Computing Language to program GPUs

**OpenMP:** Open Multi-Processing: Application programming interface that support multiplatform shared memory multiprocessing

**OS:** Operating System

## *P*

**ParaStation Consortium:** Involved in research and development of solutions for high performance computing, especially for cluster computing

**ParaStationMPI:** Software for cluster management and control developed by ParTec

**Paraver:** Performance analysis tool developed by BSC

**ParTec:** ParTec Cluster Competence Center GmbH, Munich, Germany

**PC:** Normally Personal Computer, but in the context of the proposal also Project Coordinator

**PCI:** Peripheral Component Interconnect: Computer bus for attaching hardware devices in a computer

**PCIe:** PCI Express: Standard for peripheral interconnect, developed to replace the old standards PCI, improving their performance

**PFlop/s:** Petaflop, $10^{15}$ floating point operations per second

**PIC:** Particle In Cell

**PLX switch**: switch to connect a single PCI Express port to multiple end-points, produced by the company PLX Technology

**PM:** Person Month or Project Manager of the DEEP project (depending on the context)

**PMT:** Project Management Team of the DEEP project

**PR:** Public Relations

**PRACE:** Partnership for Advanced Computing in Europe (EU project, European HPC infrastructure)
**PRACE-1IP:** PRACE First Implementation Phase (EU project)
**Project Coordinator:** Leading scientist coordinating and representing the DEEP project
**Proto-Booster:** Minimal instantiation of a DEEP Booster based on early access techonologies (EXTOLL FPGA and KNC in PCIe form factor). Developed and used in the DEEP project for software development
**PROSPECT:** Promotion of Supercomputing Partnerships for European Competitiveness and Technology (registered association, Germany)
**PUE:** Power Usage Effectiveness

## *Q*

**QDR**: Quad Data Rate: Communication signalling technique of InfiniBand®

## *R*

**RAS:** Reliability, Availability and Serviceability
**RDMA:** Remote Direct Memory Access
**RML:** Risk management list used in the DEEP project
**RTD:** Research and Technological Development
**RTM:** Reverse Time Migration

## *S*

**SC:** International Conference for High Performance Computing, Networking, Storage, and Analysis, organised in the USA by the Association for Computing Machinery (ACM) and the IEEE Computer Society
**SCIF:** Symmetric Communication Interface from Intel
**Scalasca:** Performance analysis tool developed by JUELICH and GRS
**SIMD:** Single Instruction, Multiple Data. Describes computers with multiple processing elements that perform the same operation on multiple data points simultaneously
**SMFU:** Shared Memory Functional Unit
**SoA:** Struct of Arrays
**SRMIP:** Soubaras-Remez Migration Parallel. Simulation code for seismic imaging used at partner CGGVS
**StarSs:** Generic programming environment developed by BSC
**Stampede:** Supercomputer (Dell PowerEdge C8220 Cluster with Intel Xeon® Phi™ coprocessors) installed at Texas Advanced Computing Center from Univ. of Texas, in the USA. It is nr. 7 in the TOP500 list today
**STRATOS:** PRACE advisory group to foster development of HPC technologies in Europe
**SW:** Software

## *T*

**TCO:** Total Cost of Ownership
**TFlop/s:** Teraflop, $10^{12}$ floating point operations per second

**Tier-0, Tier-1, …:** Different classes of supercomputers ordered by their performance
**TIM:** Thermal Interface Material
**TK:** Task, followed by a number: Term to designate a task inside a work package of the DEEP project
**Torque:** Distributed resource manager providing control over batch jobs and distributed compute nodes
**ToW:** Team of Work Package leaders within the DEEP project
**TP10:** Third Party under Clause 10
**TurboRVB:** Quantum Monte Carlo Software for electronic structure calculations, developed by SISSA

## *U*

**UniHD:** University of Heidelberg, Germany
**UniReg:** University of Regensburg, Germany
**USP:** Unique Sell Point

## *V*

**VELO:** Virtualised Engine for Low Overhead: An EXTOLL communications channel

## *W*

**WP:** Work Package

## *X*

**x86**: Family of instruction set architectures based on the Intel 8086 CPU

## *Y*

## *Z*