SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

FP7-ICT-2011-7

DEEP

Dynamical Exascale Entry Platform

Grant Agreement Number: 287530

D1.5
Periodic progress report at month 24

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), L.Palm (BADW-LRZ), H.Ch.Hoppe (Intel)
Date:	24.02.2014

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: D1.5</td></tr>
<tr><td colspan="2">Deliverable Nature: Report</td></tr>
<tr><td rowspan="2">Deliverable Level:
CO *</td><td>Contractual Date of Delivery:
30 / November / 2013</td></tr>
<tr><td>Actual Date of Delivery:
30 / November / 2013</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: Periodic progress report at month 24</td></tr>
<tr><td colspan="2">ID: D1.5</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_D1.5_Periodic_progress_report_at_M24_v2.0-ECapproved-PublishablePart.docx</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), L.Palm (BADW-LRZ), H.Ch.Hoppe (Intel)</td></tr>
<tr><td>Reviewed by:</td><td>S.Höfler-Thierfeldt (JUELICH), W.Guerich (JUELICH), I.Schmitz (ParTec)</td></tr>
<tr><td>Approved by:</td><td>BoP/PMT</td></tr>
</table>

Document Status Sheet

Version	Date	Status	Comments
1.0	30/November/2013	Final	Ready for EC submission
2.0	24/February/2014	Final	Approved by EC

Document Keywords

Keywords:	DEEP, HPC, Exascale, status report, month 24

Table of Contents

List of Figures

Executive Summary

The Dynamical Exascale Entry Platform (DEEP) project started on 1st December 2011 and will last three years. The main goal of the project is to develop a prototype hardware and software supercomputing system with the potential to reach a peak performance of 100 PFlop/s in 2014, 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 13 to 24** of the DEEP project. The main achievements in the reporting period are enumerated below:

- Second review (at month 12) successfully passed.
- Submission of an amendment of the DoW to include two Third Parties in the consortium, adapt the working schedule, and prolong the project in cost-neutral basis.
- Installation of the DEEP Cluster at Juelich completed with the connection of its InfiniBand network.
- Progress in the definition of the functionality to be implemented in the RAS plane to measure the energy consumption of the final DEEP System.
- Progress in the DEEP Booster design and construction: BIC (based on EXTOLL's FPGA version) production finished; bring-up of BNC board (based on Intel Xeon® Phi™ in PCIe form factor and on EXTOLL FPGA) on-going.
- Implementation of the low-level Cluster-Booster protocol completed (all functionality available; performance improvements expected in final hardware).
- Extension of the ParaStation component *pscom* to support EXTOLL completed.
- OmpSs tool-chain ported and evaluated on the Intel® Xeon Phi™ processor (KNC).
- OmpSs extended with the MPI-offloading directives. Fully functional implementation (on top of the Intel MPI library) that can already be used by the DEEP applications on the DEEP Cluster.
- Extrae instrumentation library extended to support offloaded applications and provide a unified view of Cluster and Booster sides.
- Intel MKL evaluated on the Intel® Xeon Phi™.
- Application improvements in performance and porting to the MIC evaluator (Intel Xeon® Phi™ cards). Codes tested on an Intel cluster with 32 Xeon® Phi™ cards.
- Presentation of DEEP in publications, workshops and conferences.
- Organisation of the "European Exascale Projects" booth at SC13, joint presentation of the DEEP/Mont-Blanc/CRESTA projects.

- Organisation of a joint CRESTA/DEEP/Mont-Blanc workshop for technical discussions on common topics.

1 Publishable summary

Numerous challenges must be overcome to reach Exascale by the end of the decade. When starting from current PFlop 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 multi core 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 will support the application-developers in the identification of these offload-kernels and in porting their workload to the proposed architecture.

1.1 Project objectives

The specific objectives of the DEEP project and first results 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.
  → Installation of Cluster part in Jülich is completed.
  → Prototypes of several Booster components have been built and are used for software development.
  → Ongoing: Bring-up phase of the first prototype representative of the DEEP Booster, called “Proto-Booster”. The “Proto-Booster” contains two or more Booster Node Cards (BNC) and one Booster Interface Card (BIC).
- Combination of innovative technologies for the Booster element: novel Intel many-core processors; EXTOLL high speed interconnect; hot water cooling.
  → First prototypes of Booster Node Card (BNC) are in bring-up at Eurotech. One BNC contains two nodes, each with an Intel Xeon® Phi™ and an EXTOLL FPGA to implement the network.
  → Prototypes of Booster Interface Card (BIC) available at Eurotech.

→ EXTOLL protocol ported to the FPGA technology from Altera (Stratix V).

- 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.
  → KNC remote boot successfully tested
  → Cluster Booster communication layer implemented and fully functional (performance improvements targeted when final hardware is available).
  → Cluster Booster Protocol plugin in ParaStation *pscom* implemented.
  → ParaStation component *pscom* supporting EXTOLL (D4.3).
  → OmpSs runtime (Nanos++) ported to Intel Xeon® Phi™.
  → OmpSs extended with the MPI-offloading directives and fully functional, supporting both C and Fortran codes.
  → First version of the Booster resource management with static allocation in Torque and Maui already available

- 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.
  → Libraries and mathematical functions needed by the application codes and the parameters they used have been identified.
  → Performance measurements of Intel MKL BLAS library done for the most important routines used by the applications. Small issues reported to Intel for future improvements.→ Paraver/Extrae performance tool ported to MIC.
  → Paraver/Extrae instrumentation library extended to support offloaded applications and provide a unified view of Cluster and Booster sides.
  → Scalasca performance tool ported to MIC.

- 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 operating in Juelich.
  → Concepts for improving energy and cooling efficiency defined.
  → Functionality needed on the RAS plane defined. Implementation on-going.

- 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 initial strategies to distribute the codes between Cluster and Booster has been identified.
  → Five of the six applications ported to MIC.
  → Optimisations of the codes on-going.
  → Measurements performed on large Xeon Phi™-based systems.

- 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.
  → Demonstration will be first possible when the complete system is available.

- Dissemination of the innovations and results of the project to the public.
  → The DEEP project and its first results have been presented in multiple meeting, conferences, and workshops. Details are given in the list of dissemination activities provided in Annex A.1 of this report.
  → The DEEP project collaborates closely with the projects CRESTA and Mont-Blanc. The three projects jointly organised BoFs at ISC'13 and SC13, a mutual booth at the SC13, and a Workshop in Barcelona. This kind of joint technical and dissemination activities will continue (floor plan already reserved for ISC'14 and SC14).

1.2 Work performed and main results

During the months 13 to 24 the DEEP project has achieved the following milestones:

- MS3: successful completion of the second project review in January 18th 2013.
- MS8 partially achieved: software development prototype (Proto-Booster) available and under tests, but not yet ready for software bring-up
- MS10: MIC-OS has been adapted to the DEEP Architecture. Results are shown in deliverable D4.4.
- MS11 partially achieved: Cluster-Booster protocol ported to the Super-BIC evaluator, but not yet to final hardware, since the DEEP System will be available in month 36. However, all functionalities are already implemented, so that the only difference expected in final hardware is a better performance.
- MS15: OmpSs and BLAS ready for use on DEEP System
- MS17: installation of the Cluster Subsystem completed with the installation of its InfiniBand network at the beginning of 2013.
- MS20: partially achieved: all applications, except NEURON, have been ported to the MIC architecture. Application/workload division between Cluster and Booster as well as the concrete offloading strategies used (MPI vs. OmpSs) often remain unclear.

Management, legal and administrative tasks

In the present reporting period (month 13 to 24) 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. Additionally the preparation of an amendment of the DoW constituted a large part of the Project Management Team's work during the period. Finally, the management team was also responsible for the preparation of the second external review and for the upcoming third external review.

The Management Team organised the agenda for the second review meeting (at month 12) which took place on the 18th January 2013 in Mannheim (Germany). To fulfil the internal quality policies a rehearsal meeting one day before the review was conducted. As a result of the second review, the project has been evaluated as doing *"good progress".* All deliverables submitted in the first year of the project were approved.

Following the recommendations from the review, the DEEP project has prepared an amendment of the Description of Work (DoW) to adapt the work-plan to the hardware availability and include two Third Parties in the consortium. The proposed new DoW was sent to the project's officer beginning of August 2013. Following that, the DoW was updated to the NEF server and an official request letter was sent to the European Commission in September 2013.

Monthly teleconferences of the Team of Work Package leaders (ToW) were organised to periodically discuss the progress in all Work Packages (WPs). The fourth regular face-to-face meeting of the consortium (BoP) took place in Bologna (Italy) on the 15th – 16th May 2013, and the fifth one in Toulouse (France) on the 15th – 16th October 2013. There, the status of the project was discussed. All 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 2nd IS-ENES Workshop on High performance computing for climate models in Toulouse (France), the Exascale Applications and Software Conference (EASC) in Edinburgh (UK), or the Workshop Series on Heterogeneous and Unconventional Cluster Architectures and Applications (HUCAA) in Lyon (France). Three papers showing DEEP results were accepted at the EASC conference and two further ones at the HUCAA workshop. A joint BoF session together with Mont-Blanc and CRESTA was organised at the ISC'13. Two additional talks containing DEEP topics were presented at another BoF session at ISC'13, entitled "Towards Exascale Efficiency". At the Supercomputing Conference (SC13), DEEP has co-organised a joint BoF and a joint booth (see Figure 1) also together with CRESTA and Mont-Blanc. Floor space for joint booths at ISC'14 and SC14 have been already reserved. The DEEP flyer has been regularly updated and distributed both at the ISC'13 and at SC13.

Figure 1: DEEP, CRESTA, and Mont-Blanc people in front of the "European Exascale Projects" booth at SC13.

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. The third Mont-Blanc/DEEP Training Workshop took place from the 11th to the 13th March 2013 in Barcelona (Spain). An intensive course including lectures and numerous hands-on sessions on OpenCL and OmpSs programming with special focus on its support for Fortran, was offered to all members of the DEEP and Mont-Blanc projects. On 10th-12th July 2013, a training on Intel's MIC was organised specifically for the DEEP application developers in Barcelona.

Coupled with the previously mentioned training event a workshop to strengthen the collaboration between the EU Exascale Projects DEEP and Mont-Blanc was held in Barcelona on 14th March 2013. This one-day meeting allowed for an open technical discussion on various aspects of the two projects, including their hardware architectures, software environment, and applications work. Common topics were discussed and a rich interchange of experiences took place. The discussion was highly appreciated by members of both projects.

Following the success of the event, a new workshop, this time with the three Exascale Projects CRESTA, DEEP, and Mont-Blanc took place on 10th-11th June 2013 in Barcelona, focusing on software environment, performance analysis tools, and applications. This workshop showed several common interests between the projects identified and strengthened the ties between them. The success of this event was such that the next one is already planned to take place in Edinburgh in March 2014.

All dissemination and training activities performed in the present reporting period are enumerated in Annex A.1 of this report.

Technical Work

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

Overview

In the hardware concept of DEEP (see Figure 2) a standard Cluster, composed of multi-core processors and the high-flexibility network InfiniBand®, is attached to a Booster of Intel Xeon® Phi™ processors connected via a highly scalable EXTOLL torus network.

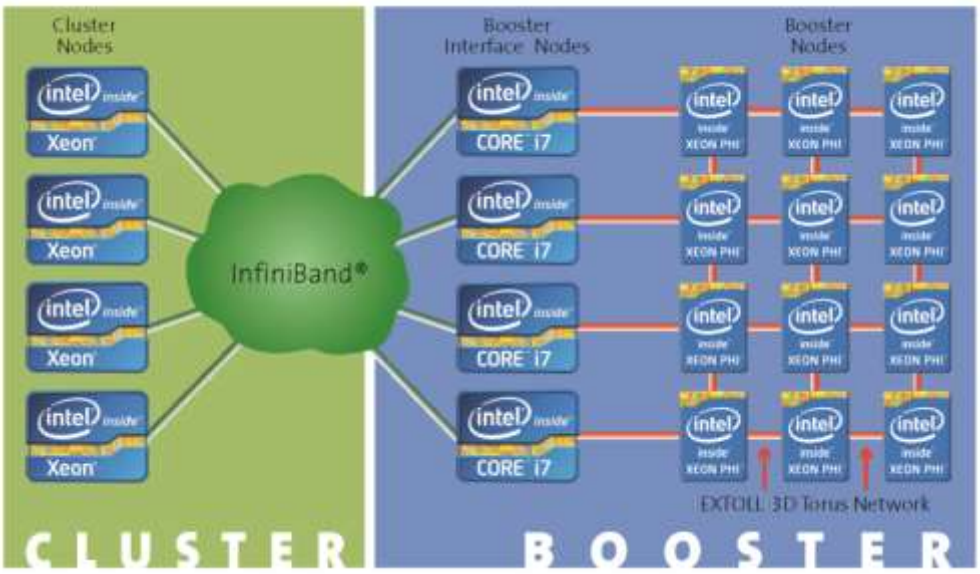

Figure 2: Sketch of DEEP hardware architecture as realised in the DEEP System.

The DEEP software stack (see Figure 3) 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 (BSC) tasking model to help the developers in decomposing their applications into tasks in order to efficiently overlap the use of the Cluster and the Booster sides.

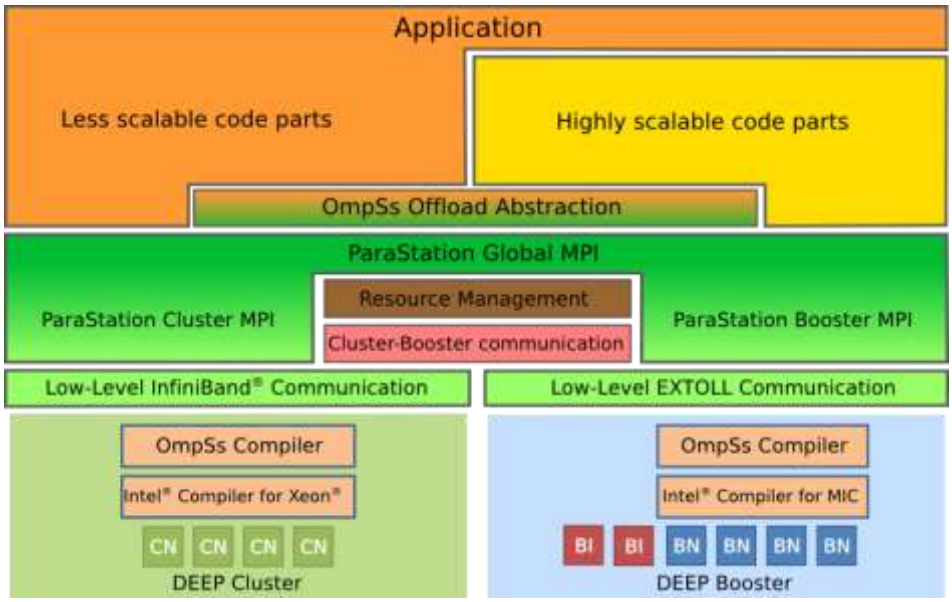

Figure 3: Diagram of the DEEP software architecture

System Hardware

On the hardware side, the work within the reporting period did focus on the construction and bring-up of the software development prototype, referenced to as the “Proto-Booster”. The

Proto-Booster uses the FPGA-version of the Booster Node Card (BNC) with the EXTOLL NIC implemented in an Altera Stratix V FPGA and standard PCIe form factor Intel Xeon® Phi™ cards. A "Mini-Backplane" can take up to four BNCs, and it provides additional connectors for test and debugging.

Bring up of this Proto-Booster (see Figure 4), using a standard server with an EXTOLL Galibier card as "Pseudo-BIC", has made progress, with several hardware and software issues being identified, analysed and corrected. First of all the flow of electrical signals through the BNC and Mini-Backplane was checked. Signal integrity tests were used to verify the connection between FPGA and Mini-backplane, first for the 7th link to the BIC. Further signal integrity measurements are on-going to test and validate all connections between BNCs on the Mini-Backplane, as well as those to connectors that will go to other chassis in a larger system.

The EXTOLL implementation for Altera Stratix V was ported to the BNC, with modificatons required to enable reliable communication. Remote boot of the Intel® Xeon Phi™ is working, which in turn enables validation of the cooling system and integration of the Proto-Booster in its self-contained 19" rack, which is already available at Eurotech.

Figure 4: *Left:* Photo of a BNC Evaluator Board (bottom) connected to a mini-backplane. *Right*: BIC prototype board, built for the Proto-Booster.

In addition, the FPGA-based BIC cards have become available and will be integrated with the Proto-Booster after they have been tested and validated.

In the present reporting period the installation of the DEEP Cluster was completed with the integration of its InfiniBand network. Operating system, file system and a general software environment were previously available and have been frequently updated. Several stability and stress tests have been used to verify the correct functioning of all hardware and software elements of the system. A parallel optimised LINPACK run on the Cluster nodes showed about 87% of the theoretical peak performance. The DEEP Cluster, its environment and the login and file system servers are part of the in-kind contribution to the project by JUELICH.

The MIC evaluator has been upgraded substituting its KNF by KNC[1] cards. Currently it is constituted by two workstations, one with two and the other with four KNC cards. Its configuration has been updated to allow for running MPI programs on all cards and nodes. The MIC evaluator is being used for software development and application porting.

System Software

On the software side, the implementation and porting of the various elements involved in the DEEP software stack have progressed in the present reporting period. The MIC-OS has been adapted to allow for remote boot of the KNC card. Extensions of MIC-OS have been implemented and tested in order to add support for the EXTOLL interconnect on KNC. The Cluster Booster Protocol has been further tested, debugged and refined and based on it, a plugin for ParaStation *pscom* has been implemented. The first version of the Booster resource management with static allocation in Torque and Maui is already available. ParaStation MPI has been extended to support EXTOLL (documented in D4.3). The realisation of Global MPI has progressed with the implementation of *MPI_Comm_connect* and the parts of *MPI_Comm_spawn* responsible for communication and process management.

The work on the OmpSs programming environment has also progressed: the Nanos++ execution runtime and the Extrae instrumentation library have been ported to the Intel Xeon® Phi™. The Scalasca measurement system has also been ported to the Intel Xeon® Phi™. Nanos++ and the Mercurium compiler have been extended to fully support the MPI offloading directives proposed in the context of the DEEP project, for both C and Fortran codes. WP8 applications written in C and Fortran have been used to validate the offloading concept.

Applications

The application developers, with the help of the support team from Task 8.7 through regular teleconference meetings, have progressed in porting their applications to Intel® Xeon Phi™. Some of the developers have started to hybridise their codes including OpenMP or OmpSs directives. Application profiling, debugging and optimisation have continued. Following the recommendations from last review, the project has made efforts to obtain access to larger Intel Xeon® Phi™ based systems. Accounts on Stampede (University of Texas, USA) have been setup for all members of the support team and for the application developers that requested it. Additionally, access to an Intel Xeon® Phi™ cluster owned by Intel has been granted to the DEEP application developers. The use of this system is on a reservation basis. Two weeks were reserved in October 2013. The first week was used for preparation of the codes and first tests. In the second week application developers and the support team performed measurements and analysed the results. Conclusions can be found in deliverable D8.2.

1.3 Expected final results

The DEEP project will have installed the DEEP System in Juelich (Germany), 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. The experience gained running scientific applications on the

[1] To ease the reading, along the rest of the document the names KNC, MIC, or Xeon Phi are used to designate Intel Xeon® Phi™.

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.

A.1 Listing of dissemination activities

This list reflects the disseminations activities performed between months 13 and 18 of the DEEP project.

Conferences, workshops, and meetings:

- **"Advanced Computer Architecture"** at Technical University Munich, Germany, January 30, 2013:
  - M.Ott (BADW-LRZ): "DEEP: Dynamical Exascale Entry Platform" (presentation)

- **2nd IS-ENES Workshop on "High performance computing for climate models"**, Fondation Bembert, Toulouse, France, January 30 to February 1, 2013:
  - E.Suarez (JUELICH): "DEEP: Dynamical Exascale Entry Platform" (presentation)

- **Joint DEEP/Mont-Blanc internal Workshop,** Barcelona, Spain, March 14, 2013:
  - E.Suarez (JUELICH): " DEEP: Overview" (presentation)
  - H.Ch.Hoppe (Intel): "DEEP Architecture" (presentation)
  - Th.Moschny (ParTec): "DEEP Programming Model" (presentation)
  - J.Labarta (BSC): "OmpSs in DEEP and Mont-Blanc (presentation)
  - D.Alvarez (JUELICH): "DEEP Applications" (presentation)

- **HPC Advisory Council Switzerland Conference 2013**, Lugano, Switzerland, March 13-15, 2013:
  - N.Eicker (JUELICH): "The DEEP project" (presentation). Available online in: http://insidehpc.com/2013/03/19/the-deep-project-developing-a-novel-exascale-enabling-supercomputing-platform/

- **Workshop on Strategic Development of High Performance Computers**, Tsukuba, Japan, March 18-19, 2013
  - N.Eicker (JUELICH): "The DEEP project" (presentation)

- **6th MEGWARE HPC User Meeting**, Chemnitz, Germany, April 25, 2013
  - N.Eicker (JUELICH): "The DEEP project" (presentation)

- **Exascale Applications and Software Conference,** Edinburgh, Scotland, UK. 9-11 April 2013:
  - F.Delalondre (EPFL), M.Lee Hines, M.Knobloch (JUELICH), J.Gimenez (BSC), J.Labarta (BSC), D.Alvarez (JUELICH) and F.Schürmann (EPFL). "Benchmark and Profiling of NEURON in Strong Scaling Configuration" (presentation)
  - G. Lapenta (KULeuven), et al. "Space Weather Applications for Exascale Computing" (presentation)
  - D.Alvarez (JUELICH), N.Eicker, E.Suarez and W.Guerich (JUELICH). "Programming model and application porting to the Dynamical Exascale Entry Platform (DEEP)" (presentation)

- **Programming models for heterogeneous hardware seminar,** Bucaramanga, Colombia, June 1-5, 2013.
  - V.Beltran (BSC), "The DEEP programming model and analysis tools" (presentation)

- **Intel Innovation Day,** Gravenbruch, Germany, June 6, 2013.
  - H.-Ch.Hoppe (Intel), “The DEEP architecture” (presentation)

- **First joint DEEP/CRESTA/Mont-Blanc Workshop,** Barcelona, June 10-11, 2013.
  - E.Suarez (JUELICH), “DEEP Introduction and Status” (presentation)
  - Th.Moschny (ParTec) , J.Labarta (BSC), J.Gimenez (BSC), M.Knobloch (JUELICH) “The DEEP Programming model and analysis tools” (presentations)
  - D.Alvarez (JUELICH), “DEEP applications” (presentation)

- **ISC’13,** Leipzig, Germany, June 17-20, 2013:
  - P.Arts (Eurotech), “The DEEP System” (presentation at BoF session: “Exascale Research – The European Approach”, June 18, 2013)
  - Th.Moschny (ParTec) DEEP content within “ParTec presentation” (presentation at the Vendor session, June 17, 2013)
  - A.Auweter (BADW-LRZ), “Meet a Beast – and keep cool” (presentation at BoF session “Towards Exascale Efficiency: Research Results & Steps Ahead in Cooling, Adaptive Power Management & Power-Aware Applications”, June 19, 2013)
  - M.Knobloch (JUELICH), “Energy-Efficient HPC - A Tools Perspective” (presentation at BoF session “Towards Exascale Efficiency: Research Results & Steps Ahead in Cooling, Adaptive Power Management & Power-Aware Applications”)
  - H.Ch.-Hoppe (Intel) and G.Lapenta (KULeuven), “Space Weather application running on Intel Xeon Phi” (demonstration of the DEEP application at Intel booth)
  - G. Staffelbach (CERFACS), "The HPC Revolution: Innovation on the Field of Combustion Applications” (presentation at Bof “PRACE & the HPC Services for Industry”)
  - G. Staffelbach (CERFACS) "Progress in the understanding of combustion: High performance computing and the AVBP code” (presentation at the PRACE booth)
  - DEEP flyer distributed at the partners’ booths and on the attendees bag
  - DEEP video running at the booths from the Gauss Center for Supercomputing and Eurotech

- **ScalPerf’13,** Bertinoro (Forli-Cesena), Italy, September 22-27, 2013.
  - B.Mohr (JUELICH), “Jülich – still – on the way to Exascale” (presentation)

- **HUCAA 2013 (Workshop on Heterogeneous and Unconventional Cluster Architectures and Applications),** 4th International Conference on Parallel Processing Workshop (ICPPW)**,** Lyon, France, October, 1, 2013**:**
  - N.Eicker (JUELICH), Th.Lippert (JUELICH), Th.Moschny (ParTec), and E.Suarez (JUELICH): “The DEEP project: Pursuing cluster-computing in the many-core era (presentation)
  - S.Rinke (GRS), S.Prabhakaran (GRS), F.Wolf (GRS): “Efficient Offloading of Parallel Kernels Using MPI_Comm_Spawn” (presentation)

- **European Research & Innovation Conference 2013" (ERIC 2103),** Nice, France, October 23, 2013:
  - N.Eicker (JUELICH): “The DEEP-ER Project - Extending the reach of the Cluster-Booster Architecture“ (presentation)

- **HBP Summit,** Lausanne, Switzerland, October, 9 2013
  - N.Eicker (JUELICH): “DEEP and DEEP-ER – Booster for HPC“ (presentation)

- **SC13,** Denver, USA, November 17-22, 2013:
  - N.Eicker (JUELICH), "DEEP and DEEP-ER: Innovative Cluster architecture for Intel Xeon Phi" (Intel Theatre presentation at the Intel booth, November 18, 2013)
  - N.Eicker (JUELICH), "The DEEP Project" (presentation at BoF session: "Building on the European Exascale Approach", November 19, 2013)
  - H.Ch.-Hoppe (Intel), E.Suarez (JUELICH), "DEEP and DEEP-ER – Innovative Cluster Architectures for Intel® Xeon Phi™" (repeated presentations at the Intel booth, November 18-21, 2013)
  - J.Labarta (BSC), "DEEP Programming Environment" (presentation at the Intel booth, November 20, 2013)
  - H.Ch.-Hoppe (Intel) and D.Alvarez (JUELICH), "DEEP applications" (repeated presentation at the Intel booth, November 18-21, 2013)
  - E.Suarez (JUELICH), DEEP presentation and discussion at the Panel "Emerging Technologies and Big Data (Euro-Centric)", November 21, 2013
  - Joint booth of the "European Exascale Projects", joining DEEP, CRESTA, and Mont-Blanc
  - DEEP flyer distributed at the partners' booths and on the attendees bag
  - DEEP video running at the booths the European Exascale Projects, JSC, Intel, and Eurotech
  - N.Eicker (JUELICH) interview to Computer World
  - E.Suarez (JUELICH) video interview on DEEP and DEEP-ER, as part of the "Discover Your Parallel Universe video project" series, November 21, 2013.
  - Presentation of DEEP Proto-Booster components (FPGA-based BNC boards, Mini-Backplane and FPGA-based BIC boards) at the Eurotech and "European Exascale Projects" booths, November 18-21, 2013.
  - Presentation of packaged A0 silicon for the EXTOLL ASIC at the EXTOLL booth, November 18-21, 2013.

- **ExaMPI13 - Workshop on Exascale MPI at SC13,** Denver, USA, November 22, 2013
  - N.Eicker (JUELICH), "Message-Passing Challenges in DEEP's Heterogeneous Cluster-Booster Architecture" (presentation)

- **Vlasovia Workshop 2013,** Nancy, France, November, 25-28 2013
  - E.A.Johnson (KULeuven), S.Markidis, G.Lapenta (KULeuven): "The kinetic implicit fluid method for collisionless plasma" (poster)

Publications, proceedings, press-releases, and newsletters:

- **Journal of Concurrency and Computation**
  - A. E. Vapirev, G. Lapenta (KULeuven), J.Deca, S. Markidis, I. Hur, "Initial Results on Computational Performance of Intel MIC, Sandy Bridge, and GPU Architectures: Implementation of a 1D C++/OpenMP Electrostatic Particle-In-Cell Code" (submitted)

- **Deutschlandfunk**, Radio interview by M.Schönherr, October 6, 2013:
  - E.Suarez (JUELICH) and U.Brüning (UniHD), interview within the emission "Schneller, lauter, heißer. Exascale – Die nächste Generation von Supercomputern"

- **Proceedings of HUCAA 2013 (Workshop on Heterogeneous and Unconventional Cluster Architectures and Applications),** 4th International

Conference on Parallel Processing Workshop (ICPPW)**,** Lyon, France, October, 1, 2013**:**

- N.Eicker (JUELICH), Th.Lippert (JUELICH), Th.Moschny (ParTec), and E.Suarez (JUELICH): “The DEEP project: Pursuing cluster-computing in the many-core era” (proceedings)
- S.Rinke (GRS), S.Prabhakaran (GRS), F.Wolf (GRS): “Efficient Offloading of Parallel Kernels Using MPI_Comm_Spawn” (proceedings)

- **Proceedings of the Exascale Applications and Software Conference,** Edinburgh, Scotland, UK. 9-11 April 2013:
  - G. Lapenta (KULeuven), et al. “Space Weather Applications for Exascale Computing” (proceedings)
  - D.Alvarez (JUELICH), N.Eicker, E.Suarez and W.Guerich (JUELICH). “Programming model and application porting to the Dynamical Exascale Entry Platform (DEEP)” (proceedings)

- **Intel Exascale Labs Yearly Report 2012**. Revised version of the article published at International Conference on Supercomputing 2012 (ICS 2012).
  - D. Alvarez Mallon (JUELICH), N. Eicker (JUELICH), M.E. Innocenti (KULeuven), G. Lapenta(KULeuven), Th. Lippert (JUELICH), and E. Suarez (JUELICH): “On the Scalability of the Cluster-Booster concept”

List of Acronyms and Abbreviations

A

ADI3 layer: MPICH Abstract Device Interface Version 3
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

BACNet: Building Automation and Control Network; a protocol used in building automation and monitoring. In DEEP, BACNet is used to manage the cooling and power infrastructure surrounding the DEEP System.
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.
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

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)
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 Juelich
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
DGETRF: LU factorisation of a ream *m* by *n* matrix.
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: The 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

FIFO: First In, First Out. Method for organizing and manipulating a data buffer, or data stack, where the oldest entry, or 'bottom' of the stack, is processed first.

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 Juelich, Germany

H

HCA: Host Channel Adapter

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). Along this document this product is indistinctively called KNC, MIC, or Intel Xeon Phi.

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: Juelich Supercomputing Centre

JUDGE: Juelich Dedicated GPU Environment: A cluster at the Juelich Supercomputing Centre

JUELICH: Forschungszentrum Juelich GmbH, Juelich, Germany

JUQUEEN: Juelich's BlueGene/Q machine: A supercomputer installed at the Juelich 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

LU decomposition: factors a matrix as the product of a lower triangular matrix and an upper triangular matrix.

LVDS: Low Voltage Differential Signalling

M

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

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

OpenMP ARB: Open MP Architecture Review Board: a group of leading hardware and software vendors and research organizations that create the OpenMP standard.

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 standard PCIs, improving their performance

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

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

PMT: Project Management Team of the DEEP project

POD: Plain Old Data

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 technologies (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
Scalasca: Performance analysis tool developed by JUELICH and GRS
SCIF: Symmetric Communication Interface from Intel
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
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
SysFs: Virtual file system provided by the Linux kernel to export information from devices and drivers to the user.

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
UPC: Universitat Politècnica de Catalunya, Barcelona, Spain

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