SEVENTH FRAMEWORK PROGRAMME

FP7-ICT-2013-10

DEEP-ER

DEEP Extended Reach

Grant Agreement Number: 610476

D1.8

Periodic progress report at month 42

Approved

Version: 2.0

Author(s): E.Suarez (JUELICH)

Contributor(s): S.Eisenreich (BADW-LRZ), H.Ch.Hoppe (Intel), K.Thust (JUELICH), V.Beltran (BSC), A.Zitz (JUELICH), I.Zacharov (Eurotech)

Date: 04.05.2017

Project and Deliverable Information Sheet

DEEP-ER Project	**Project Ref. №:** 610476	
	Project Title: DEEP Extended Reach	
	Project Web Site: http://www.deep-er.eu	
	Deliverable ID: D1.8	
	Deliverable Nature: Report	
	Deliverable Level: CO*	**Contractual Date of Delivery**: 31 / March / 2017
		Actual Date of Delivery: 31 / March / 2017
	EC Project Officer: Juan Pelegrín	

* - 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

Document	**Title:** Periodic progress report at month 42	
	ID: D1.8	
	Version: 2.0	**Status:** Approved
	Available at: Publishable part at: http://www.deep-er.eu	
	Software Tool: Microsoft Word	
	File(s):DEEP-ER_D1.8_Periodic_progress_report_M42_v2.0-PublishablePart	
Authorship	**Written by:**	E.Suarez (JUELICH)
	Contributors:	S.Eisenreich (BADW-LRZ), H.Ch.Hoppe (Intel), K.Thust (JUELICH), V.Beltran (BSC), A.Zitz (JUELICH), I.Zacharov (Eurotech)
	Reviewed by:	D.Tafani (BADW-LRZ), J.Kreutz (JUELICH)
	Approved by:	BoP/PMT

Document Status Sheet

Version	Date	Status	Comments
1.0	31/March/2017	Final	EC submission
2.0	04/May/2017	Approved	EC approved

Document Keywords

Keywords:	DEEP-ER, HPC, Exascale, progress report, month 42

Table of Contents

List of Figures

List of Tables

No table of figures entries found.

Executive Summary

The DEEP – Extended Reach (DEEP-ER) project started on 1st October 2013 and lasted for 42 months. The project addresses two significant Exascale challenges: the growing gap between I/O bandwidth and compute speed, and the need to significantly improve system resiliency. DEEP-ER extends the Cluster-Booster Architecture first realised in the DEEP project by a highly scalable I/O system. Additionally, an efficient mechanism to recover application tasks that fail due to hardware errors is implemented. The project builds a hardware prototype including new memory technologies to provide increased performance and power efficiency. As a result of employing these technologies, I/O parts of HPC codes run faster and scale up better. Furthermore, HPC applications are able to profit from checkpoint and task restart with reduced overhead. To demonstrate it a set of seven applications with high societal impact have been ported to the DEEP-ER architecture and make use of the I/O and resiliency capabilities available therein.

This report describes the objectives, work performed, resources used, and results achieved during **months 25 to 42** of the DEEP-ER project. The main achievements in the reporting period are enumerated below:

- Co-design effort stepping up: continuous discussions between hardware, software, and application developers to assure a coherent development that addresses all requirements.

- Update by partner Eurotech of the “Aurora Blade” architecture, with a new design of the KNL-based node board containing one KNL per board. Commitment by partner Eurotech to produce this design and complete the DEEP-ER prototype with in-kind resources.

- Completion of the installation of the Software Development Vehicle (SDV), a hardware platform for software and application developments, which will also become the Cluster part of the DEEP-ER prototype. Completion of SDV with eight pre-production KNL boards. Intense use of the SDV for application development and evaluation.

- DEEP-EST Booster construction: samples of the three key elements of the system (KNL blade, backplane and Root card) tested, production of the devices needed for the full DEEP-ER system completed. Room infrastructure prepared for the installation of the DEEP-ER Booster, which is scheduled for April 2017.

- I/O benchmarks and application mock-ups running on the Software Development Vehicle. Measurements show the I/O and resiliency performance achieved with the project’s developments.

- Two Network Attached Memory (NAM) devices installed in the SDV. Functionality and performance established and access library (libNAM) implemented. The libNAM integration with SIONlib and SCR for the realisation of the checkpointing functionality completed and verified with benchmarks.

- Development of the DEEP-ER I/O software stack – containing BeeGFS, SIONlib, and E10 – completed, taking into account the requirements from resiliency software and

applications. Verification with applications and benchmarks achieving impressive performance for checkpoint and local I/O on the NVM.

- Implementation of resiliency software layer completed. Tests with applications and benchmarks demonstrate the very low overhead of the applied techniques.
- Benchmarks integrated in JUBE environment and periodic test procedure established. GERShWIN, xPic, TurboRVB, FWI, and Chroma applications integrated in JUBE. Additionally, benchmarks and mini-apps for I/O and resiliency benchmarking have been included.
- Various application adaptations and improvements done: optimisation to take benefit from the second generation Intel® Xeon Phi™, code partition between Cluster and Booster parts of the DEEP/-ER architecture, integration with the I/O and resiliency tools developed in DEEP-ER. Porting to the SDV completed, benchmarking on SDV and other platforms show profit of the applied techniques.
- Dissemination of project goals and status in various workshops and conferences, amongst others the Supercomputing Conference (SC15, SC16) in the US and the International Supercomputing Conference (ISC 2016) in Germany.
- Coordination and co-organisation of joint dissemination activities with other European Exascale Projects, i.e. for the joint booth at SC15, SC16 and ISC 2016, as well as the European HPC Summit Week 2016.

1 Publishable summary

The DEEP-ER project tackles two important Exascale challenges. Firstly, the increasing gap in the growth rate of compute power versus the amount and performance of memory and storage available in HPC systems. Secondly, the high failure rates expected in Exascale systems as a consequence of the increased number of components and the need to take their performance and energy efficiency to the limits. To address these issues, DEEP-ER extended the heterogeneous Cluster-Booster Architecture implemented by the DEEP[1] project by additional I/O and resiliency functionalities.

DEEP-ER targets seamless integration of a high-performance I/O subsystem into the Cluster-Booster Architecture. The employment of novel memory technologies provides a multi-level I/O infrastructure capable of supporting data-intensive applications. Additionally, an efficient and user-friendly resiliency concept combining user-level checkpoints with transparent restart of application tasks restart has been developed, which enables applications to better cope with the higher failure rates expected in Exascale systems.

The DEEP-ER I/O and resiliency concepts have been evaluated using seven European HPC applications from fields that have proven their need for Exascale resources. These applications have been ported and optimised to demonstrate the usability, performance and resiliency of the DEEP-ER developments. Systems that leverage the DEEP-ER architecture will be able to run more applications at the same time, thereby increasing scientific results. This is due to improved computational efficiency, better and more scalable I/O performance and a substantial reduction in the loss of computational work caused by system failures.

1.1 Project objectives

The specific objectives of the DEEP-ER project and the results already achieved towards them are:

1. *Address two main Exascale challenges: I/O and resiliency. DEEP-ER will extend the DEEP Architecture by: i) a highly scalable, efficient and easy-to-use parallel I/O system; ii) providing a combination of low-overhead user-level checkpoint/restart and automatic task recovery.*

   → The implementation of the DEEP-ER I/O software stack– including BeeGFS, SIONlib, and E10 – has been completed taking the application and resiliency software requirements into account (see D4.1, D4.2, and D4.4).

   → The DEEP-ER I/O stack makes the local storage-class memory at the DEEP-ER nodes (see below) available in a transparent way and uses it internally to improve I/O throughput and scalability.

   → The resiliency software has been implemented (see D5.1, D5.2, and D5.3) taking the application requirements into account.

   → Integration of resiliency and I/O software layers with each other completed. Benchmarks and tests done and facilitated through user support.

2. *Develop a prototype system of the extended DEEP Architecture that leverages advances in hardware components (Intel's second generation Intel® Xeon Phi^{TM}*

[1] www.deep-project.eu

processors, high-speed interconnects and non-volatile memory devices) to further improve the performance and efficiency of the DEEP-ER Prototype and realise the novel I/O system and resiliency improvements. This prototype will allow proving the viability of the concept for 500 PFlop/s-class of supercomputers.

> → After initial exploratory studies of several architecture alternatives, in M18 it was decided to implement the DEEP-ER Booster using the Aurora Blade architecture from Eurotech. After a first design phase, the system design was completed for an Aurora blade using the second generation of Intel® Xeon Phi™ (code name "Knights Landing" or KNL), based on a half-width board from Intel. The node provides a full complement of 96 GBytes of DDR memeory, and is connected to its peripherals using PCIe generation 3.
>
> → Before deciding on the KNL node design, two additional design choices had been made: in light of the good progress with the EXTOLL TOURMALET implementation (based on an ASIC and performed outside the project), this interconnect was selected for the DEEP-ER Prototype. The exact network topology (a 3D torus open in one direction) was then settled through co-design discussions with the application developers. In addition, a NAND-Flash Intel SSD was selected as on-node fast storage device, attached by PCI Express and using the NVMe interface.
>
> → The Software Development Vehicle (SDV), a hardware platform for development of system (I/O and resiliency) and application software, was installed at JUELICH, The EXTOLL interconnect of the SDV has been upgraded by substituting the version A2 TOURMALET NICs by the final A3 version, which reaches a link bandwidth of over 100 Gbit/s. The SDV will serve as the Cluster side of the DEEP-ER prototype.
>
> → A further upgrade of the SDV was the integration of eight Intel Xeon Phi (KNL generation) nodes, which are used to port and optimise the parts of the applications that will later run on the DEEP-ER Booster. These eight KNL nodes have also been recently populated each with one 400 GB NVM device, providing all the functionality of a small 8-node DEEP-ER Booster. Together with the already existing 16 Intel® Xeon® nodes (of the "Haswell" generation), the SDV became the major software validation, application porting and benchmark platform of the project.
>
> → The room infrastructure has been prepared for the arrival of the DEEP-ER Booster, including electrical and water connections. The Booster rack was delivered and connected to the cooling loop. Installation of the Booster components is scheduled for April 2017.

3. *Explore the potential of new storage technologies (non-volatile and network attached memory) for use in HPC systems, with a focus on parallel I/O and system resiliency by integrating them with the DEEP-ER Prototype*.

> → NVM technology options for integration with the DEEP-ER prototype have been evaluated. A series of Intel SSD replacement devices based on advanced NAND-Flash was selected (see Deliverables D3.1 and D3.2). These devices implement the NVM Express (NVMe) interface and are connected via PCI Express to the compute nodes.

→ Extensive experiments were undertaken first with two samples of these devices at Juelich and later with the SDV and production NVM devices. A wide range of measurements with I/O benchmarks and application mock-ups are available (see D3.3, recently updated with the latest results). These clearly show substantial performance increases over best-of-breed SSDs, in particular for scenarios with many parallel I/O requests.

→ Two NAM prototypes have been integrated into the SDV and their full functionality has been verified and tested with benchmarks. The NAM uses UHEI's hybrid HMC controller implementation, which has been completed and validated and was made available as Open Source. A state-of-the-art Xilinx Virtex 7 FPGA implements the HMC controller, the NAM functional logic and two EXTOLL links compatible with the full TOURMALET EXTOLL fabric speed.

4. *Develop a highly scalable, efficient and user-friendly parallel I/O system tailored to HPC applications. The system will exploit innovative hardware features, optimise I/O routes to maximise data reuse, and expose a user friendly interface to applications. Its design will meet the requirements of traditional, simulation-based as well as emerging data-intensive HPC applications*.

→ The design of the DEEP-ER I/O system has been completed taking into account the outcome of the discussions with the experts from WP3 – to guarantee that the hardware provides the needed functionality – and with WP5 and WP6 to gather all their requirements on the I/O infrastructure.

→ The functionalities that each of the three I/O software packages – the BeeGFS file system of Fraunhofer, the parallel I/O library SIONlib, and the E10 software stack – must provide to the project, the interplay between them, and their interfaces have been described in deliverables D4.1 and D4.2.

→ In BeeGFS two new functions have been implemented: cache domain handling and user-level stripe-size definition. The cache domain is executed on the node-level NVM devices and can be executed synchronously or asynchronously. Both the synchronous and asynchronous versions have been implemented and verified. Recently, the implementation of native support for the EXTOLL network in BeeGFS started, activity that will be completed after the project's end, as it is not a mandatory feature for DEEP-ER now that the new EXTOLL stack has solved performance issues with IP previously identified.

→ SIONlib has been refactored and improved to eliminate code replication and increase the overall modularity and manageability of the library. The functionality required for buddy checkpointing has been implemented and tested. SIONlib also supports the checkpointing use case of the NAM, thanks to its integration with SCR and libNAM.

→ The E10 implementation is completed, and extensions have been developed and tested which use the caching functionalities of BeeGFS and other file systems. A new support library makes the integration of the new E10 functionality transparent to applications. Application measurements show improvements of over 50% on the performance of collective MPI-IO operations.

→ Access to the local fast NVM devices is possible for applications – either by using POSIX I/O with a specific directory path or by relying on the "BeeGFS on demand" (BeeOND) functionality, which provides the full BeeGFS interface for accessing local storage. Functionality and performance has been validated for both Intel Xeon and Intel Xeon Phi nodes.

→ The benchmarks to be used for the evaluation of the DEEP-ER I/O software have been identified (see D4.3) and they have been integrated into the JUBE benchmark environment. These benchmarks are used to regularly monitor the overall I/O performance and to measure the impact of the various system and application software development activities in the DEEP-ER project. This activity has allowed finding bugs and identifying the right configuration parameters in several software packages. Some applications have been integrated in JUBE and a user-guide has been provided to the developers to facilitate the integration of the remaining ones. The applications GERShWIN, xPic, TurboRVB, FWI, and Chroma are now fully integrated in JUBE.

5. *Develop a unified user-level system that significantly reduces the checkpointing overhead by exploiting multiple levels of storage and new memory technologies. Extend the DEEP programming model to combine automatic re-execution of failed tasks and recovery of long-running tasks from multi-level checkpoints, and introduce easy-to-use annotations to control checkpointing*.

→ In a co-design effort, the overall resiliency software stack has been defined (see D5.1) taking into account the requirements from the WP6 application developers, the WP3 hardware capabilities, and the I/O functionality required by WP4. In addition, the roles of the application-based and task-based resiliency functionalities and the interfaces between them have been defined.

→ The Scalable Checkpoint/Restart (SCR) library has been adapted to the needs of the DEEP-ER project including an API (abstraction layer) for the application users to apply SCR in their codes (see D5.2). The code has been adapted to reflect recent changes in the BeeGFS API, and the new SIONlib buddy-checkpointing functions have been implemented and tested.

→ OmpSs adaptations for task-based resiliency have been implemented, including the interaction with ParaStation MPI in order to extend the task-based implementation to support offloaded tasks.

→ Integration between the checkpoint/restart framework, OmpSs and ParaStation MPI has been completed.

→ In close collaboration with WP3, an event-based Monte Carlo failure model has been designed and implemented, in order to optimise policies that determine the frequency, redundancy level and storage-location of checkpoints for each application. A closed formula matching the simulation results has also developed and integrated with the SCR library, enabling the latter to provide indications on the required checkpointing frequency for each application to the user.

→ The checkpointing and recovery strategies developed in DEEP-ER have been tested with the applications, proving that higher resiliency levels can be achieved this way with a very low overhead.

→ The NAM has been used to perform checkpointing by storing parity data of global checkpoints. Current NAM prototypes have limited capacity but clearly demonstrate the potential of the technology.

6. *Analyse the requirements of HPC codes carefully selected to represent the needs of future Exascale applications with regards to I/O and resiliency, guide the design of the DEEP-ER hardware and software components, optimise these applications for the extended DEEP Architecture and use them to evaluate the DEEP-ER Prototype. Selected applications cover the fields of Health, Earthquake Physics, Radio Astronomy, Oil Exploration, Space Weather, Quantum Physics, and Superconductivity.*

→ Continuous co-design discussions have taken place to identify the application requirements – in terms of hardware capabilities, I/O and resiliency functionalities – and use them to define the hardware and software layers in the project. DDG teleconferences and face-to-face meetings have been used as the co-design platform, as applications evolve with time through code optimisations and implementation of new functionalities.

→ The structure of the applications has been analysed, and performance and scaling tests have taken place. In several cases mock-ups of the applications have been developed, to easily implement code changes and analyse their impact on the overall performance.

→ The results of these investigations had a significant impact in the overall improvement of the applications. Important topics in the code optimisations are: vectorisation, optimising communication strategies and/or numbering schemes, improving I/O, implementing checkpointing, etc.

→ BeeGFS, SIONlib, E10, SCR and OmpSs are now used by the DEEP-ER applications. In collaboration with WP3 and WP4, mock-ups for some applications (e.g. space weather and seismic imaging) have been used for I/O benchmarking.

→ The applications have been ported to the SDV (see D6.2 and D6.3). With this platform, the functionality of all project developments have been demonstrated, showing the advantages that they bring to real-world applications.

→ I/O and resiliency developments have been benchmarked using synthetic benchmarks and full applications. Results are described in D4.5, D5.4 and D6.3.

→ Platforms with similar hardware and software configuration to the DEEP-ER Booster have been employed for large-scale tests. Collaboration with the QPACE-3 project has been established for this purpose, a large KNL-system accessible to JUELICH. Some changes on the system configuration, done in agreement with the system administrators, allowed reproducing an environment as close as possible to the DEEP-ER Booster, providing very valuable results for our project.

7. *Demonstrate and validate the benefits of the extended DEEP Architecture and its first implementation (the DEEP-ER Prototype) with the DEEP-ER pilot applications and for applications that exploit generic multi-scale, adaptive grid and long-range force parallelisation models.*

   → First results, obtained by predicting the performance of three applications on the DEEP-ER Prototype with the Dimemas simulation tool by partner BSC and extrapolating the scaling characteristics have been obtained and documented in Deliverable D7.1. Further applications are being analysed and new modelling aspects, such as I/O performance, will be taken into account. Focus of work is now on I/O tracing and modelling.

   → The Extrae event tracing library was extended to record I/O traces, which contain events for relevant Posix and MPI-IO calls. Traces for two applications have been obtained on the DEEP-ER SDV, and analysis of the I/O data has produced insightful results. The existing performance model has been extended by taking into account I/O efficiency. Extrapolations of I/O performance were not possible due to the small scale of the SDV. First larger scale tests were run with the xPic code on QPACE3.

1.2 Work performed and main results

According to the amended DoW, four milestones had to be reached between **month 25 and month 36** of the DEEP-ER project:

- **MS8**: “Overall design of Aurora Blade prototype completed”: Deliverable D8.1, originally submitted in M24, has been updated to reflect the final design, which bases the KNL-boards for the Aurora Blade architecture on commercially available Intel KNL boards (S7200AP or “Adams Pass”). The document has been re-submitted in M29 (February 2016) and in its updated form specifies the prototype under construction.
- **MS9**: “Applications ported to the SDV”: Deliverable D6.2 submitted in M30, with the results already achieved by the applications running on the Software Development Vehicle.
- **MS10**: “BNC evaluator (i.e. KNL-blade) and rest of Aurora Blade components available”.
  - KNL blade: samples were available at Eurotech in M33, comprising a reference board by Intel (code name “Adam Pass”), and the three PCBs constituting the Blade Interface Board (BIB). Laboratory tests took place to verify its functionality. Thermal tests (with KNL blade attached to its cold plate) have been completed.
  - Backplane samples have been also tested and its functionality verified up to PCIe gen2 in combination with the KNL blade and Root Card.
  - Root card: design problems were found in the first samples, making a re-spin of the root card design necessary. This design has been tested to verify that

all previously identified issues have been solved,and it provides the intended functionality up toe PCIe gen2 speeds in combination with the Aurora KNL blades and the Backplane.

  - An in-depth investigation has identified a manufacturing problem that interferers with achieving the planned PCIe gen3 performance. Eurotech is working to correct this problem and produce new Backplanes or Root cards that will fully support PCI gen3 speeds.

- **MS11** ("First Aurora chassis delivered to Juelich") **and MS12 (**"DEEP-ER Aurora Blade Prototype installed and available to users"): due to difficulties during the test phase of the individual components, and the later-than-planned arrival of some of their parts to Eurotech, the delivery of the first chassis had to be re-scheduled to November 2016 (M38). Unfortunately, a leakage event occurred after 3 to 4 weeks of operation that prevented further usage of the chassis (see D8.3). Furthermor, problems found during the test phase of the Root Card required a re-design of the device, which delayed the overall system construction. The current Root Card and Backplane implementations provide stable PCIe gen2 connections. They will be used for the installation of the DEEP-ER Booster at Juelich, scheduled for April 2017.
- **MS13**: "I/O and resiliency software packages ready": this Milestone was achieved on time by M36. The software packages have been described in Deliverables D4.4 (I/O) and D5.3 (resiliency).

Management, legal and administrative tasks

A large part of the management activities in the present reporting period was dedicated to monitor the progress of the project with regards to the achievement of all technical goals specified in the Description of Work (DoW) and the fulfilment of all commitments to the European Commission, as well as for addressing the recommendations issued by the reviewers in M24 and the interim review at M32, the first of which included the preparation of the first DoW amendment. For this, internal discussions were concluded and a formal request for the first DoW amendment was completed, which extended the project by 6 months until end of March 2017. This request was granted in the reporting period.

The Project Management Team organised the agenda for the review meeting at month 24, which took place on December 9, 2015 in Brussels (Belgium), and the interim review at month 32, which took place on June 8, 2016 in Juelich (Germany). To fulfil the internal quality policies, rehearsal meetings took place one day before the reviews. As a result of both reviews, the project has been evaluated as having achieved "good progress". Additionally, all deliverables submitted between M13 and M32 have been approved. The comments from the reviewers in these reviews and their recommendations concerning future work are addressed in section 2.2.

The financial statements from all partners were submitted to the NEF server after the end of the second project year and the European Commission has approved the financial data.

Monthly teleconferences of the Team of Work Package leaders (ToW) have been organised to periodically discuss the progress in all Work Packages (WPs). Furthermore, bi-weekly teleconferences of the Design and Development Group (DDG) have been held to discuss the progress in the implementation of the different developments, and to drive co-design and cross-WP discussions.

Deliverables D1.6, D1.7, D1.8, D1.9, D2.4, D3.3, D3.4, D3.5, D3.6, D4.4, D4.5, D5.3, D5.4, D6.2, D6.3, D7.2, D8.1, D8.2, D8.3, and D8.4 have been submitted to the European Commission after having passed through the mandatory DEEP-ER internal review process. The already approved public deliverables (the publishable parts of D1.6, D3.3, and D3.4) have been uploaded to the project Website.

Dissemination, training and outreach

The DEEP-ER prototype breaks new ground in the combination of its principal components (KNL CPU, NVM devices, EXTOLL TOURMALET network), and the Aurora Blade architecture includes innovative ways to integrate, package and cool a highly efficient HPC system. In addition, the innovative DEEP-ER I/O and resiliency concepts did require the development of new techniques and tools never tested before. Access to the “know-how” achieved in this process shall not remain limited to the group of people directly involved in the project, but must be made available to a wider community to move the HPC field forward. For this reason, WP2 in DEEP-ER is entirely dedicated to the dissemination of the knowledge accumulated over the project duration, as well as to train the users on its application.

The centre of the dissemination activities of DEEP-ER is its Web site at www.deep-er.eu. The Web page is updated regularly and referred to in all other materials (articles, press releases, brochures, presentations, etc.). It is used to publish general information about the project, current activities, training opportunities, job vacancies, publications, tutorials, success stories, and achievements of the project.

Following previous recommendations, the content of the Website is being further developed. In particular, more focus is being put on the applications and on their presentation as a global and collaborative effort, as well as on highlighting the outcome of the strong co-design performed in the project.

Two social media platforms have been chosen to disseminate DEEP-ER news amongst the HPC world and the general public: LinkedIn and Twitter. The already existing DEEP LinkedIn group has been extended to host also DEEP-ER. The strong link existing between both projects justifies the use of a single group. The same applies for Twitter. Updates are being regularly posted (at least at bi-weekly basis) and frequently re-posted by other Twitter users in the HPC community. The most recent Twitter posts are visible also at the main page of DEEP-ER’s website. Continuous and steady increase in Twitter follower numbers has been observed. Re-tweets and interactions are in a solid state as well. The @DEEPprojects Twitter account established itself as a key player in the Twitter HPC community, providing impressions via re-tweets and mentions. LinkedIn is still slower, but postings are more frequent now and also more colleagues engage in the group. Although the total number of members is not too high, the number of project external members raises and a larger audience via likes and shares has been reached successfully.

Several high-profile dissemination activities have taken place in the reporting period. Partners from the DEEP-ER consortium presented the project concept and results in conferences and workshops, including two of the most important events in the HPC community, namely the Supercomputing Conference (SC), which took place in Austin (USA) in November 2015 and in Salt Lake City (USA) in November 2016, and the International Supercomputing Conference (ISC), which took place in Frankfurt (Germany) in June 2016.

At both SC'15 and ISC 2016, the DEEP-ER project co-organised a joint booth together with other European Exascale Projects (EEP) – Mont-Blanc (1 and 2), EPiGRAM, and EXA2CT. DEEP and DEEP-ER were jointly shown on a display panel describing the architecture and main goals of the projects. The HMC controller (the core component of DEEP-ER's NAM), was displayed as a demo at the booth. Additionally, a DEEP-ER flyer was prepared and distributed at the European Exascale Projects (EEP) booth and at the booths of other project partners. In addition, the DEEP-ER project was presented in the joint EEP BoF and at presentations and panel discussion at the Intel booth. In SC'16 the DEEP-ER presence was located at the booth of the project coordinator, JUELICH. There, a KNL-blade of the DEEP-ER Booster was displayed and project material was discussed and distributed.

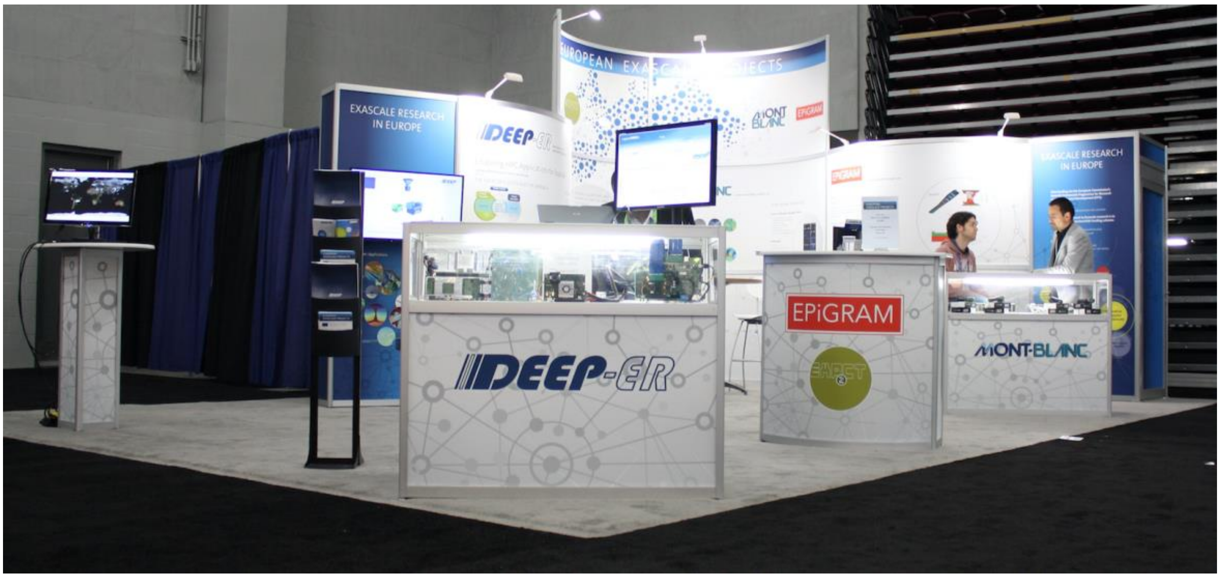

Figure 1: Joint EEP booth at SC15.

DEEP-ER was also presented at the European HPC Summit Week, which took place in Prage (Czech Republic) in May 2016, where FP7 and H2020 projects were presented. DEEP-ER members also participated in the ETP4HPC discussion in the Extreme Scale Demonstrators (EsDs) Workshop, which was organised at the last day of the summit week.

Additionally, several articles and publications on the project approach and results have been submitted. A list with all dissemination activities performed in the present reporting period is presented in Annex A.1 of this report.

With respect to business and industry relations, DEEP-ER focused on (a) increasing impact via industrial lobbying organisations (e.g. ETP4HPC or PROSPECT); (b) focus on communicating DEEP-ER software developments and the benefits of potential users; (c) support for partner marketing activities; (d) presence at industry-oriented conferences and events. To help the commercialisation of project results, the partners responsible for the different developments have made significant efforts to increase the productisation potential of their own and shared IP. Task 2.2 gathered information on all the actions and plans from the individual partners. With this overview, synergies for increasing project visibility were improved. Concerning point (b) content has been created targeting potential users of a DEEP-ER system as well as interested system developers, being leveraged via the Website and social media campaigns. Finally, concerning (d) the DEEP and DEEP-ER projects showcased their work to an industry-focused audience at CeBIT16 in March 2016 at Hannover (Germany).

The main goal of the training events in DEEP-ER is to teach the application developers participating in the project on how to use the software tools and programming environment that will run on the DEEP-ER Prototype and other intermediate hardware evaluators. A hands-on training event, jointly organised with Mont-Blanc, for the application developers of both projects took place in March 2016 in Barcelona (Spain). An additional training event was organised in the first days of September 2016 in Juelich, where Intel members instructed application developers on how to best prepare their codes to profit from the Intel Xeon Phi (KNL) architecture.

Technical Work

The technical work in DEEP-ER is grouped into three main topics: system architecture and hardware, system software (including I/O and resiliency software), and applications.

Overview

The DEEP-ER project designs and builds a second-generation prototype (see Figure 2) of the Cluster-Booster Architecture. In the DEEP-ER Prototype the second generation Intel Xeon Phi processors (KNL) provides the compute power for the Booster Nodes (BN), while Intel Xeon processors (Haswell generation) populate the Cluster Nodes (CN). A uniform high-speed interconnect runs across Cluster and Booster, and network-attached memory (NAM) devices connected to it provide high-speed shared memory access and checkpointing functions. The Cluster and Booster Nodes themselves also feature additional non-volatile memory (NVM) devices for efficiently buffering I/O and storing checkpoints.

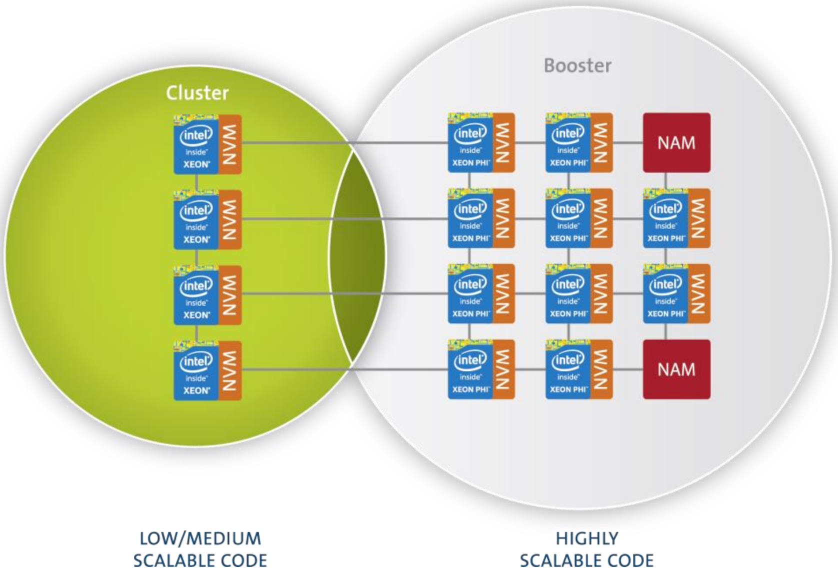

Figure 2: High-level view of the DEEP-ER Prototype. NVM=Non-Volatile Memory; NAM=Network Attached Memory

The DEEP-ER multi-level I/O infrastructure has been designed to support data-intensive applications and multi-level checkpointing/restart techniques. The project develops a scalable and efficient I/O software platform based on the BeeGFS parallel file system, the parallel I/O library SIONlib, and the Exascale10 (E10) I/O software package. It enables efficient and transparent use of the underlying hardware and to provide all functionality required by applications for standard I/O and checkpointing.

On top of this I/O infrastructure DEEP-ER has developed an efficient and user-friendly resiliency concept combining user-level checkpoints with transparent task-based application restart. OmpSs is used to identify individual tasks in an application and handle their interdependencies. The OmpSs runtime has been extended in DEEP-ER in order to automatically re-start tasks in case of transient hardware failures. In combination with a multi-level user-based checkpoint infrastructure to recover from non-transient hardware errors, applications are now able to cope with the higher failure rates expected in Exascale systems. DEEP-ER's I/O and resiliency concepts have been evaluated using seven HPC applications from fields that have proven their need for Exascale resources.

System Architecture and New Technologies

During the M18 interim review, it was decided to build the DEEP-ER Prototype based on the Aurora Blade architecture to the development of which Eurotech did commit. The initial implementation then foreseen comprised the design by Eurotech of a new KNL-board hosting two KNL chips, which would constitute two fully independent Booster nodes. Detailed analysis of the technical requirements and component placement led however to the conclusion that the complexity and risk of such development would be too high. Alternatively, in M24 Eurotech did present their decision to build a single-node KNL-blade for Aurora, which integrates a commercially available KNL board (Intel Server board S7200AP or "Adams Pass"). The density of the overall DEEP-ER Prototype is preserved by installing both the EXTOLL NICs and the NVM devices in the Root card. With this new approach, the design risk was significantly reduced. Eurotech's effort could be focused on the electrical and mechanical integration and cooling of all the Aurora Blade components. The Eurotech Aurora cooling technology has been enhanced to support memory DIMMs (of the ultra-low profile or ULPDIMM variety) in addition to soldered-down memory. The ability to adequately cool the ULP DIMMS that will populate the six memory slots in each KNL board has been validated.

During the reporting period, Eurotech has produced the KNL blades required for the DEEP-ER prototype Booster. The first samples of the Root Card were produced and their tests served to identify some design issues that have been corrected with a re-spin of the board. The current Root Card and Backplane implementations deliver PCIe gen2 speeds, yet exhibit severe problems in establishing PCIe gen3 connections. An in-depth investigation by Eurotech did identify the root cause in signal interference at the higher frequencies used by PCIe gen 3 caused by manufacturing problems in the Backplane to Root Card connector areas. To reach full PCIe gen3 performance, these problems have to be corrected, and this might require a re-spin of the Backplane and/or Root Card.

The DEEP-ER Booster to be installed at Juelich will include 72 compute nodes with one KNL CPU, 16 GBytes on-package memory and 96 GBytes of DDR4 memory on-board. Installation is scheduled for April 2017, and the system will at first only support PCIe gen2 speeds.

The EXTOLL TOURMALET ASIC-based NIC has progressed significantly in the reporting period. A new ASIC stepping (A3) has become available and was first tested by the project to integrate Intel Xeon Phi devices on the SDV. Later on, the full interconnect of the SDV was upgraded from A2 to the A3 version. The A3 EXTOLL TOURMALET delivers a stable 8.4

Gbit/s per lane bandwidth, achieving slightly over 100 Gbit/s per link and fully matching DEEP-ER requirements. These figures are for the raw bandwidth and do not include protocol overheads.

Further synthetic, application and tools benchmarks have been conducted with the NVM devices installed in the Xeon and KNL nodes of the SDV in Juelich, improving the initial crude application mock-ups and extending the scope of applications considered. It has also become possible to conduct multi-node runs and transition to use the full applications.

Further work has been done in close collaboration between UHEI and Micron on improving the HMC controller for current HMC silicon and the controller design has been put into the open source domain. The NAM architecture had been fixed earlier and the first NAM prototype became available, which uses a state-of-the-art Xilinx Virtex 7 FPGA. On it the HMC interface of 16 lanes and the NAM-specific logic responsible for RDMA operations and additional functionality to be provided by the NAM are implemented. Additionally, the FPGA also provides two EXTOLL links with twelve lanes each, compatible with the lane speeds achieved by the TOURMALET ASIC. The two NAM prototypes have been integrated in the SDV and their functionality is verified. The FPGA firmware implementation of the extended NAM functions, as well as the full software stack (combining libNAM, SIONlib and SCR) required for the checkpointing use case are completed. Benchmark tests show the potential of the NAM technology.

The installation of the Software Development Vehicle (SDV) has been completed: the system uses 16 dual-socket Intel Xeon E5 nodes (Haswell generation), has an Intel DC P3700 NVMe device (400 GByte capacity) attached to each node, and uses EXTOLL TOURMALET as the interconnect (now upgraded to the A3 version). Eight KNL nodes (using the stock S7200AP boards) have been integrated into the SDV, each accompanied with the same Intel DC P3700 devices and an EXTOLL TOURMALET NIC, which provides connection to the rest of the SDV. The Haswell part of the SDV constitutes the Cluster part of the DEEP-ER Prototype. The eight existing KNL nodes have been used as a small-size DEEP-ER Booster, allowing for full validation of the software stack and application improvements.

The SDV nodes have been procured together with the external storage. It contains a RAID system with 24 hard disks with a total capacity of 144 TByte. A meta-data server and two storage servers orchestrate the system. All three servers host an EXTOLL TOURMALET card and are connected via cables to the EXTOLL NICs of the compute nodes.

System Software

On the software side, the reporting period focused on the implementation of the various components involved in the I/O and resiliency software stacks. Regular discussions take place between the developers involved to guarantee a coherent and consistent global picture, where all the software components fit together. An overview of the DEEP-ER I/O and resiliency software layers has been presented in the DoW and is shown in Figure 3 and Figure 4, respectively.

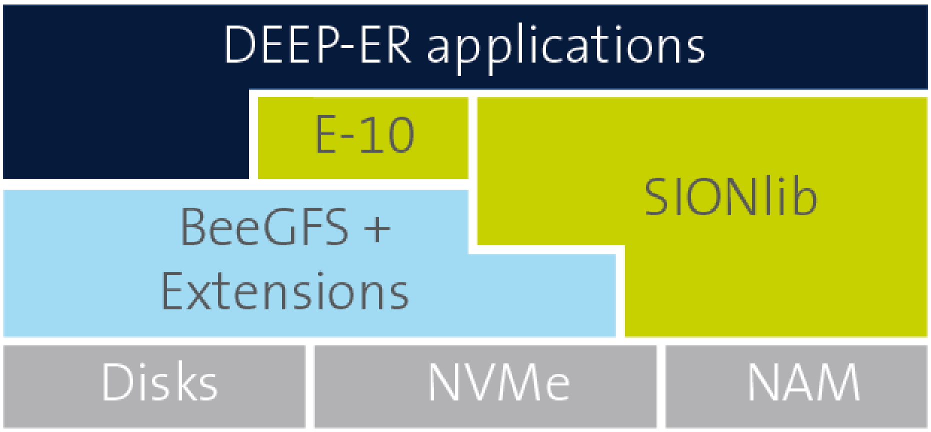

Figure 3: Sketch of DEEP-ER I/O software layers.

In particular, a close interrelation between BeeGFS, SIONlib, and the Scalable Checkpoint/Restart library (SCR) has been established. All three components cooperate to realise buddy checkpoints and the overall checkpointing mechanism in an efficient way. Functionality on the SIONlib and BeeGFS sides has been implemented and their integration with SCR is completed.

BeeGFS has completed the synchronous and asynchronous version of its implementation. Furthermore, in tight collaboration with UHEI (WP3), FHG-ITWM has started to develop native support of EXTOLL by BeeGFS. The goal of this activity was solving severe network performance issues formerly detected on the SDV. These have disappeared in the meantime thanks to an update of the EXTOLL software stack. The BeeGFS feature for native support of EXTOLL will be completed after the project's end.

The functionality required in SIONlib for the efficient implementation of buddy checkpointing is now available. The new features have been released to WP5 for their integration with SCR and verification of the overall usability and performance.

The implementation of the E10 API for I/O is completed and results of its evaluation have been published.

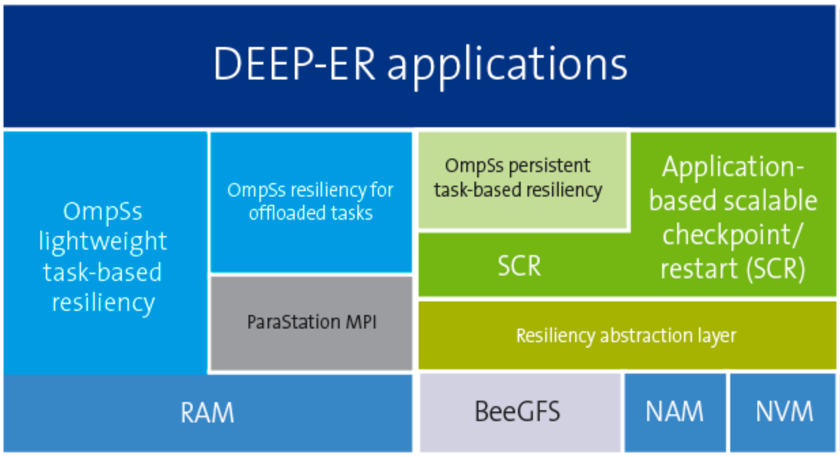
DEEP-ER applications
OmpSs lightweight task-based resiliency
OmpSs resiliency for offloaded tasks
OmpSs persistent task-based resiliency
Application-based scalable checkpoint/ restart (SCR)
SCR
ParaStation MPI
Resiliency abstraction layer
RAM
BeeGFS
NAM
NVM

Figure 4: Sketch of DEEP-ER resiliency layers

The DEEP-ER resilience architecture is based on user-level application checkpoint/restart techniques –which provide a high level of resiliency and are the most cost-effective in terms of I/O requirements– complemented with novel OmpSs task-based recovery techniques. With this combination, DEEP-ER develops new resiliency features to isolate partial failures of the system without requiring a full application restart, resulting in a more resilient, fine-grained and flexible architecture.

In addition to strong cooperation with the I/O software developers, the implementation of the failure recovery software packages themselves has been completed. The final version of the resiliency abstraction layer has been described in D5.2. The abstraction layer adds to SCR specific functions that allow efficiently exploiting DEEP-ER's I/O functionality for checkpoint/restart applications.

The SCR code makes use of the BeeGFS prefetch/flush functionality and already uses symbolic links to keep path structures synchronous. Additionally, the buddy checkpointing functions recently available in SIONlib have been implemented and tested. The use of NAM for checkpointing has been achieved by combining SIONlib with SCR and libNAM.

The task-based resiliency software is also complete. The required adaptions of OmpSs have been made and several use cases have been tested. In particular, the seismic imaging application from BSC is being used to test the task-based resiliency functionality.

Beyond that, the ParaStation management daemon has been extended to provide an interface for querying resiliency-related status information from the MPI layer and also from the OmpSs runtime environment. Tight collaboration between ParTec and BSC ensured the complete verification of the combined software, which by now has already been tested and validated.

Finally, the implementation of the failure model is complete and the closed formula derived from simulations has been integrated with the SCR library, enabling the latter to provide the user indications on the required checkpointing frequency for each application.

The software developments in WP4 and WP5 include benchmarking activities to document the progress in terms of performance and functionality. The Juelich Benchmarking Environment (JUBE) is used for this purpose. Benchmarks have been implemented in JUBE and run frequently on the DEEP Cluster and SDV to monitor the I/O performance, as the software was being developed and updates are installed. With this activity, some issues on the network performance on the SDV have been identified and finally solved. Benchmarking results of the I/O and resiliency performance achieved thanks to the DEEP-ER developments are reported in D4.5 and D5.4, respectively.

Applications

The application developers play a crucial role in the DEEP-ER project. Their work is two-folded: on the one hand they validate the work done by other technical work packages by porting their applications to the DEEP-ER Prototype; on the other hand, their input drives the development and future of both hardware and software architectures. Co-design discussions to gather more specific application requirements take place in the DDG and the consortium face-to-face meetings. Additionally, internal review meetings focused on the work done by WP6 take place during the consortium face-to-face meetings. In the two face-to-face meetings that took place in the reporting period, the developers have described their applications and presented the results that they had recently obtained, as well as the planned next steps. Other members of the consortium not involved in WP6 acted as internal reviewers and gave recommendations on the measures to be taken by each application team to achieve the results needed by the project. Additionally, issues on the specific requirements of each application were discussed, to continue with the co-design approach established in the DEEP-ER project.

In the reporting period, the application developers performed modifications to adapt the codes to the DEEP-ER hardware and software architecture, as well as general code optimisations. Additionally, the codes were ported and benchmarked on the SDV and other platforms available to the project partners. The results achieved are reported in detail in D6.2 and D6.3.

1.3 Expected final results

The project has designed and developed the DEEP-ER hardware and software prototype. The hardware comprises the new generation of Intel Xeon Phi processors, non-volatile memory in Cluster and Booster Nodes, as well as additional memory connected to the network.

The DEEP-ER Booster system is not yet installed at Juelich at the time of writing, caused by delays and technical problems in end-to-end validation of the Aurora Blade architecture and the cooling leak that occurred during tests of the first chassis. Status at the end of the project is that Eurotech has validated the functionality of the Aurora Blade architecture with PCIe gen2 speeds, and that installation of the Booster system will take place during April 2017.

A complete software stack has been created, fully tested and validated on the DEEP-ER SDV at Juelich. The programming environment based on ParaStation MPI and OmpSs has been complemented with the DEEP-ER I/O layers, which provide advanced parallel I/O functionality by combining the BeeGFS filesystem, the SIONlib I/O library and the MPI I/O optimisations from E10. Together with SCR and task-based resiliency, the above packages build an efficient infrastructure for failure recovery via easy-to-use application interfaces.

Porting and optimising applications on the DEEP-ER SDV has demonstrated the functionality and performance of the DEEP-ER software stack. Due to the small size of that system, scalability results are limited. To partly compensate, other KNL-based systems were used to collect benchmark data (such as CINECA’s Marconi A2 system and the QPACE3 system at Juelich). Full results of these activities are given in Deliverables D6.3 and D7.2.

Even with these limitations, the data gathered serves to demonstrate that systems using the DEEP-ER results will be able to run more applications in the same time, thus increasing scientific throughput, and that the loss of computational work through system failures will be substantially reduced. Once the DEEP-ER prototype configuration becomes available, extrapolation to larger scale system sizes will be feasible, demonstrating the capabilities of an Exascale-ready DEEP-ER system.

Annex A

A.1 Listing of dissemination activities

This list reflects the disseminations activities performed **between months 13 and 24** of the DEEP-ER project.

1.3.1.1 Conferences, workshops, and meetings:

- **LENS2015 International Workshop**, Akihabara, Japan, Oct 29 - 30, 2015
    - Eicker, N., "Taming Heterogeneity by Segregation – Taming Heterogeneity by Segregation -- The DEEP and DEEP-ER take on Heterogeneous Cluster Architectures" (presentation)
- **Supercomputing Conference SC15,** Austin, USA, November 16-19, 2015:
    - Joint booth of the European Exascale Projects (EEP). Booth #197. Participant projects: DEEP-ER, Mont-Blanc, EPiGRAM, and EXA2CT).
    - DEEP and DEEP-ER fliers distributed at the EEP and the partners' booths and on the attendees bag
    - DEEP+DEEP-ER video running at the booth of the European Exascale Projects
    - E.Suarez (JUELICH), presentation on DEEP-ER at the Intel Booth at a session called "An update on European HPC initiatives", November 19, 2015.
    - J. Schmidt (UHEI), "openHMC – Open Source Hybrid Memory Cube Controller" (presentation at the Emerging Technology Track)
    - S. Breuner (FHG-ITWM): BeeGFS presented at FHG-ITWM booth
    - W.Frings (JUELICH): SIONlib presented at JSC and DEEP-ER booths
- **The International Conference on RECONFIGurable Computing and FPGA**Mayan, Mexico, Dec 07-09, 2015
    - J.Schmidt (UHEI), "openHMC – Open Source Hybrid Memory Cube Controller" (poster)
- **AGU Fall Meeting**, San Francisco, USA, Dec 14, 2015.
    - J. Amaya (KULeuven), "First-principle modeling of planetary magnetospheres: Mercury and the Earth" (poster)
- **The VSC Users Day**, San Francisco, USA, Dec 14, 2015
    - J. Amaya (KULeuven), "Fully Kinetic 3D Simulations of the Interaction of the Solar Wind with Mercury" (poster)
- **HPC-LEAP Winter School 2016,** Juelich, Germany, January 15, 2016:
    - E.Suarez (JUELICH), "Implementing a new computing architecture paradigm" (presentation).
- **CeBIT 2016,** Hannover, Germany, March 14-18, 2016
    - DEEP + DEEP-ER topics presented at the booth of the North-Rhein Westfalia.
- **EGU General Assembly**, Vienna, Austria, April 17-22, 2016
    - J.Amaya (KULeuven), "Innovative HPC architectures for the study of planetary plasma environments" (accepted for presentation)
- **Parallel 2016 conference**, Heidelberg, Germany, April 07, 2016
    - C. Clauss & Th. Moschny (ParTec), "Verhalten von MPI Programmen im Fehlerfall (= Behaviors for MPI programmes when errors occur)" (presentation)
- **EGU conference**, Vienna, Austria, April 17-22, 2016
    - J.Amaya (KULeuven), "Innovative HPC Architectures for the Study of Planetary Plasma Environments" (poster presentation)
- **EASC 2016**, Stockholm, Sweden, May 10-14, 2016
    - J.Amaya (KULeuven), "Towards exascale simulations of space plasmas using the DEEP-ER architecture " (accepted for presentation)

- **European HPC Summit Week,** Prague, Czech Republic, May 10, 2016:
  - E.Suarez (JUELICH), “DEEP and DEEP-ER” (presentation).
- **BeeGFS User meeting 2016,** Kaiserslautern, Germany, May 18-19, 2016
  - C.Manzano (JUELICH), “BeeGFS in the DEEP/-ER Project” (presentation)
  - F.Kautz (FHG-ITWM), “BeeGFS User APIs” (presentation)
- **EMiT 2016,** Barcelona, Spain, June 3, 2016
  - E.Suarez (JUELICH), “Technology emerging from the DEEP & DEEP-ER projects” (presentation).
- **ISC 2016,** Frankfurt, Germany, June 20-23, 2016
  - J.Schmidt (UHEI), “Network Attached memory” (presentation at the PhD Forum)
  - N.Eicker (JUELICH), “Hardware Prototyping in DEEP and DEEP-ER” (presentation at Workshop ‘Developing Next-Gen HPC Architectures’)
  - R.Léger (Inria), “A feedback on approaching the DEEP-ER platform with a DGTD-based simulation software for Bioelectromagnetics applications” (presentation at Workshop ‘Form Follows Function’)
  - N.Eicker (JUELICH), (presentation at BoF ‘Exascale I/O: Challenges, Innovations & Solutions’)
  - J.Labarta (BSC), “The OmpSs Programming Model Vision” (presentation at BoF 11 ‘Programming Models for Exascale: Slow Transition or Complete Disruption’)
  - I.Zacharov (Eurotech), (presentation at BoF 15 ‘Monitoring Large-Scale HPC Systems: Data Analytics & Insights’)
  - I.Zacharov (Eurotech), “Aurora Tigon v4 with KNL, a system from research for research” (presentation at the Intel Collaboration Hub (booth #930))
  - E.Suarez (JUELICH), “System-level heterogeneity with Intel Xeon Phi processors” (presentation at Intel booth)
  - Video interview with insideHPC (to be published)
  - Written interview with Top500 blog (to be published)
- **JSC-KIT meeting,** Karlsruhe, Germany, June 30, July 1, 2016
  - E.Suarez (JUELICH), “DEEP/-ER cooling concept” (poster)
- **SAI Computing Conference**, July 15, 2016, London, UK
  - N. Eicker (JUELICH), “Taming Heterogeneity in HPC” (keynote presentation)
  - Th.Moschny (ParTec) and N.Eicker (JUELICH), presence at DEEP-ER booth
- **Kleine Nacht der Wissenschaft,** Juelich, Germany, September 2, 2016:
  - E.Suarez (JUELICH), “The future of supercomputing” (popular science colloquium).
- **Memsys 2016,** Washington D.C., USA, October 4, 2016.
  - J.Schmidt (UHEI), “Exploring Time and Energy for Complex Accesses to a Hybrid Memory Cube” (presentation and paper)
- **SC’16**, Salt Lake City, USA, November 14-18, 2016
  - N.Eicker (JUELICH), “Towards Modular Supercomputing” (presentation at Intel Community Hub, at Intel’s Booth)
  - N.Eicker (JUELICH), “ParaStation MPI” (Presentation at MPICH: A High-Performance Open-Source MPI Implementation BoF)
  - W.Frings (JUELICH), “The DEEP-ER take on I/O” (presentation at Workshop “Exascale I/O: Challenges, Innovations and Solutions”)
  - J.Schmidt (UHEI), “Network Attached Memory” (doctoral showcase with presentation + poster)
  - BeeGFS presented at Fraunhofer ITWM’s booth
  - Collaboration with BoF “European Exascale Projects and Their International Collaboration Potential”
  - Own space at Booth #2413, the booth of the Juelich Supercomputing Centre.
- **JLESC Workshop**, Kobe, Japan, December 1, 2016.

  - W.Frings (JUELICH), “HPC-Tools JUBE, LLview and SIONlib at JSC: Recent developments” (presentation containing DEEP-ER’s buddy checkpointing and NAM-XOR-checkpointing using SIONlib)
- **ISUM‘17**, International Supercomputing Conference in Mexico, Guadalajara, México, Feb 27 to March 4, 2017
  - E.Suarez (JUELICH), “Modular Supercomputing: the DEEP approach to hardware heterogeneity” (invited key-talk)
  - E.Suarez (JUELICH), „Woman in TICS: Women at Technology World“ (round table discussion on the role of women in technical and scientific fields)
- **ParFlow developers workshop**, Juelich, March 28, 2017.
  - E.Suarez (JUELICH), “The DEEP project(s) and the Cluster-Booster Architecture” (presentation)
- **Winter School of parallel computing,** Bologna, Italy, February 13-17
  - A.Emerson and F.Affinito (CINECA), description of DEEP-ER project (presentation)

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

- **JCP**, “Exactly Energy Conserving Implicit Moment Particle in Cell Formulation.”, G.Lapenta (KULeuven) et al. (submitted)
  - arXiv preprint arXiv:1602.06326
- **CeBIT 2016 press release** (JUELICH), 07/03/2016, “DEEP Project Presents Next-Generation of Supercomputers”
  - http://www.fz-juelich.de/SharedDocs/Pressemitteilungen/UK/DE/2016/16-03-07deep-cebit.html;jsessionid=A085C64D6693C3289B4ACAFDF4E3E9F5
- **Primeur Magazine**: “Exascale Project DEEP-ER to present at CeBIT”, 01/03/2016, http://primeurmagazine.com/flash/AE-PF-03-16-5.html
- **Science Node**
  - “Boosting Science with the next generation of Supercomputers”
  - https://sciencenode.org/feature/boosting-science-with-the-next-generation-of-supercomputers.php

- **insideHPC**:
  - Report about BeeGFS goes Open Source: http://insidehpc.com/2016/02/beegfs-parallel-file-system-now-open-source/
  - EXTOLL’s network chip enables network attached accelerators of any kind, 17/06/2016 http://insidehpc.com/2016/06/extolls-network-chip-enables-network-attached-accelerators-of-any-kind/
  - “DEEP-ER Project Moves Europe Closer to Exascale”, 04/07/2016 http://insidehpc.com/2016/07/deep-er-project/
  - “Taming Heterogeneity in HPC”, 17/08/2016, http://insidehpc.com/2016/08/taming-heterogeneity-in-hpc-the-deep-er-take/
  - “New Report Looks at European Exascale Projects”, 12/08/2016, http://insidehpc.com/2016/08/european-exascale-projects/
- **YouTube**
  - EMiT 2016: Interview to Estela Suarez, Juelich, 26/07/2016, https://www.youtube.com/watch?v=5KL0RMYW4A4
  - Taming Heterogeneity in HPC, 06/08/2016, https://www.youtube.com/watch?v=aM9AkgG5ud4&feature=youtu.be
- **Top500 Blog:**

  - "A Dive into DEEP-ER: Exascale Research with a distinctly European Flair ", 18/07/2016, https://www.top500.org/news/a-dive-into-deep-er-exascale-research-with-a-distinctly-european-flair/
  - "Extoll's network marches to the beat of a different drummer", 25/07/2016 https://www.top500.org/news/extolls-network-marches-to-the-beat-of-a-different-drummer/
- **DG Connect**, article on DEEP + DEEP-ER
- **Extoll**: "EXTOLL's network chip enables network attached accelerators of any kind", 16/06/2016, http://www.deep-er.eu/images/EXTOLL_Tourmalet_ISC_v1.0.pdf (press release)
- **DEEP-ER Status Update**: DEEP-ER on the right path, 17/06/2016, http://www.deep-er.eu/press-corner/news/184-deep-er-on-the-right-path.html (press release)
- Eurotech: "Eurotech introduces the Aurora Tigon v4", 21/06/2016, http://www.eurotech.com/en/press+room/news/?775 (press release)
- **European Exascale Projects**: A Lookback on 5 Years of European Exascale Research Collaboration, http://exascale-projects.eu/EuroExaFinalBrochure_v1.0.pdf 17/06/2016, (brochure)
- **Scientific Computing World**, Show preview (ISC 2016), 01/06/2016, print
- **INSIDE**, DEEPprojects at CeBIT 2016, 15/06/2016, http://inside.hlrs.de/#deep-project-at-cebit16
- **Europa.eu/digitalsinglemarket, "**Europe towards Exascale", 02/09/2016 https://ec.europa.eu/digital-single-market/en/news/europe-towards-exascale
- **IEEE Cluster Conference 2016**, September 13, 2016, Taipei, Taiwan
  - Congiu, G. (Seagate) "Improving Collective I/O Performance Using Non-Volatile Memory Devices" (paper and presentation)
  - https://ssl.linklings.net/conferences/ieeecluster/ieeecluster2016_program/views/at_a_glance.html
- **Intern Magazine**, Mitarbeitermagazine des Forschungszentrums Juelich, Germany, October, 2016
  - E.Suarez and N.Eicker (JUELICH), "Internationale Forschungskooperationen: Gemainsam Stark" (article and interview in german)
- **inside** – Innovatives Supercomputing Deutschland, published on October 1, 2016
  - Prototyping next-gen HPC architectures: ISC'16 workshop,
  - http://inside.hlrs.de/#prototyping-next-generation-supercomputing-architectures-isc16-workshop
- **Memsys 2016,** Washington D.C., USA, October 4, 2016.
  - J.Schmidt (UHEI), "Exploring Time and Energy for Complex Accesses to a Hybrid Memory Cube" (paper)
- **IEEE International Parallel and Distributed Processing Symposium (IPDPS'17)**, Orlando, FL, May 2017
  - B.Veenboer, M.Petschow, and J.W.Romein (ASTRON), "Image-Domain Gridding on Graphics Processors" (paper to appear)
- **HPCwire,** February 24, 2017
  - Advancing Modular Supercomputing with DEEP and DEEP-ER
  - https://www.hpcwire.com/2017/02/24/modular-supercomputing-deep-deep-er-architectures/
- **Intel**, February 24, 2017
  - DEEP-ER Modular Supercomputing
  - http://www.intel.com/content/www/us/en/high-performance-computing/julich-deeper-projects-video.html
- **Final Brochure,** March/April 2017
  - To be published in paper and online through the website
-

1.3.1.3 Industry and business cooperation:

- Desktop research:
  - Industrial application fields of the project technology: 2 applications identified
    - Enhancing Oil Exploration (OE)
    - High temperature superconductivity (HTS)
  - Products/services that leverage the project technology and market targets:
    - Present. OE: Seismic analysis/reservoir simulations in frontier domains (Oil&Gas market). HTS: MRI-NMR (medical market).
    - Future. HTS: Magnetic levitation devices, fusion reactors, motors and generators, fault current (transportation, electronics, energy markets).
  - Potential recipients for dissemination activities
    - CAPEX purchase of DEEP-ER system. After benchmarks and proof of concept, for big companies (i.e. oil companies, MagLev trains companies…)
    - Cloud services for SMEs.
- Market analysis structure and set-up.

List of Acronyms and Abbreviations

A

API: Application Programming Interface.

B

BADW-LRZ: Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften. Computing Centre, Garching, Germany

BeeGFS: **The** Fraunhofer Parallel Cluster File System (previously acronym FhGFS). A high-performance parallel file system to be adapted to the extended DEEP Architecture and optimised for the DEEP-ER Prototype.

BN: Booster Node (functional entity)

BNC: Booster Node Card is a physical instantiation of the BN

BoP: Board of Partners for the DEEP-ER 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

CINECA: Consorzio Interuniversitario, Bologna, Italy

CN: Cluster Node (functional entity)

Coordinator: The contractual partner of the European Commission (EC) in the project

CP/RS: Checkpoint / Restart

CPU: Central Processing Unit

CRB: Customer Reference Board. An early version of a KNL board developed by Intel.

CRESTA: Collaborative Research into Exascale Systemware Tools & Applications: EU-funded Exascale project.

D

DDG: Design and Developer Group of the DEEP-ER project

DEEP: Dynamical Exascale Entry Platform

DEEP-ER: DEEP Extended Reach: this project

DEEP-ER Network: high performance network connecting the DEEP-ER BN, CN and NAM; to be selected off the shelf at the start of DEEP-ER

DEEP-ER Prototype: Demonstrator system for the extended DEEP Architecture, based on second generation Intel$^{®}$ Xeon Phi^{TM} CPUs, connecting BN and CN via a single, uniform network and introducing NVM and NAM resources for parallel I/O and multi-level checkpointing

DEEP Architecture: Functional architecture of DEEP (e.g. concept of an integrated Cluster Booster Architecture), to be extended in the DEEP-ER project

DEEP System: The prototype machine based on the DEEP Architecture developed and installed by the DEEP project

E

E10: Exascale 10. Parallel I/O software developed by a consortium of partners around the EOFS community. Partner Xyratex is responsible for the development needed for the DEEP-ER project.

EC: European Commission

EC-GA: EC-Grant Agreement

EEP: European Exascale Projects

EESI: European Exascale Software Initiative (FP7)

EOFS: European Open File System.

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 1018 Floating point operations per second

EXTOLL: High speed interconnect technology for cluster computers developed by University of Heidelberg

ETP4HPC: European Technology Platform for High Performance Computing.

F

FhGFS: Acronym previously used to refer to BeeGFS.

FLOP: Floating point Operation

FP7: European Commission 7th Framework Programme.

FPGA: Field-Programmable Gate Array, Integrated circuit to be configured by the customer or designer after manufacturing

G

GRS: German Research School for Simulation Sciences GmbH, Aachen and Juelich, Germany

H

H5hut: Library implementing several data models for particle-based simulations that encapsulates the complexity of parallel HDF5.

HDF5: Hierarchical Data Format: A set of file formats and libraries designed to store and organise large amounts of numerical data

HMC: Hybrid Memory Cube

HPC: High Performance Computing

HW: Hardware

I

ICT: Information and Communication Technologies

IEEE: Institute of Electrical and Electronics Engineers

Intel: Intel Germany GmbH Feldkirchen,

IP: Intellectual Property

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

JUBE: Juelich Benchmarking Environment

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

JUELICH: Forschungszentrum Juelich GmbH, Juelich, Germany

K

KNC: Knights Corner, Code name of a processor based on the MIC architecture. Its commercial name is Intel$^{®}$ Xeon Phi^{TM}.

KNL: Knights Landing, second generation of Intel$^{®}$ Xeon Phi^{TM}

KULeuven: Katholieke Universiteit Leuven, Belgium

L

M

MIC: Intel Many Integrated Core architecture

Mont-Blanc: European scalable and power efficient HPC platform based on low-power embedded technology

Mont-Blanc 2: Follow-up project of Mont-Blanc

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

MTBF: Mean Time Between Failures.

N

NAM: Network Attached Memory, nodes connected by the DEEP-ER network to the DEEP-ER BN and CN providing shared memory buffers/caches, one of the extensions to the DEEP Architecture proposed by DEEP-ER

NASA: National Aeronautics and Space Administration, Washington, USA

NEF: Network of European Foundations: name of server where financial data is uploaded to provide it to the EC.

NetCDF: Network Common Data Form. A set of software libraries and data formats that support the creation, access, and sharing of array-oriented scientific data

NVM: Non-Volatile Memory

NVMe: NVM Express. Specification for accessing solid-state drives attached through the PCIe bus.

O

OEM: Original Equipment Manufacturer. Term used for a company that commercialises products out of components delivered by other companies.

OmpSs: BSC’s Superscalar (Ss) for OpenMP

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

Paraview: Open Source multiple-platform application for interactive, scientific visualisation

ParTec: ParTec Cluster Competence Center GmbH, Munich, Germany

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

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

PMT: Project Management Team of the DEEP-ER project

PRACE: Partnership for Advanced Computing in Europe (EU project, European HPC infrastructure)

PROSPECT: Promotion of Supercomputing Partnerships for European Competitiveness and Technology (registered association, Germany)

PTC: Persistent Task-based Checkpoint

Q

QCD: Quantum Chromodynamics

QPACE: QCD Parallel Computing Engine. Specialised supercomputer for QCD Parallel Computing

R

R&D: Research and Development

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

SCR: Scalable Checkpoint/Restart library

SDV: Software Development Vehicle: a HW system to develop software in the time frame where the DEEP-ER Prototype is not yet available.

SEO: Search Engine Optimisation: the process of improving the visibility of a website or a web page in a search engine's results.

SSD: Solid State Disk

SW: Software

T

TFlop/s: Teraflop, 10^{12} Floating point operations per second

ToW: Team of Work Package leaders within the DEEP-ER project

TP10: Third Party under special clause 10.

U

UHEI: University of Heidelberg, Germany

UREG: University of Regensburg, Germany

V

VI-HPS: Virtual Institute for High Productivity Supercomputing

VTune: Commercial application for software performance analysis

W

WP: Work Package

X

Y

Z